

# State of the Art in Data Stream Mining

**Antoine Cornuéjols**

Computer Science Department  
UniAgroParisTech, Paris

KDubiQ Symposium. September 16th 2007

# Plan

- 1 Une approche de l'apprentissage inductif
  - L'induction supervisée
  - Analyse statistique 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**

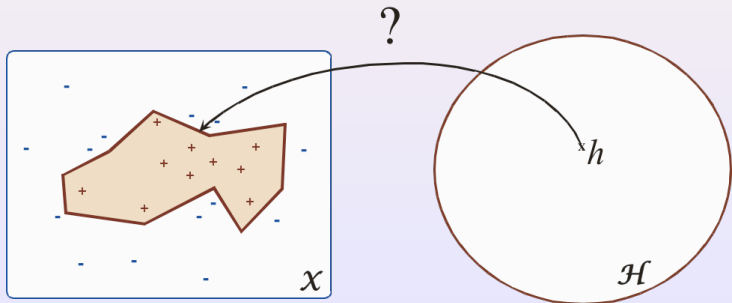
- Des processus se déployant dans le temps
- Des connaissances complexes

# L'essence du problème

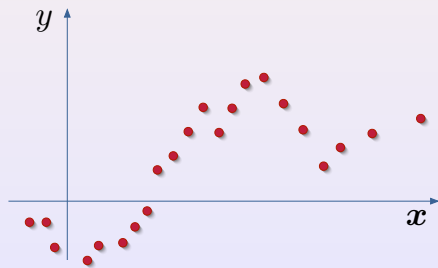
Définition [Mitchell, 82]

**Apprentissage = Recherche dans un espace d'hypothèses**

- Sous la **contrainte** des exemples d'apprentissage



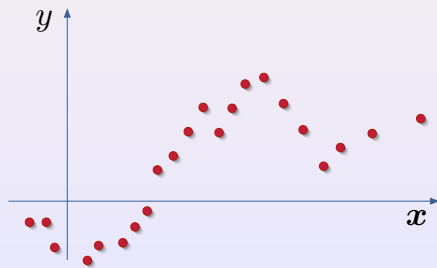
# Illustration : la régression



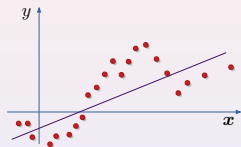
# Illustration : la régression

## Le problème

- Trouver une **loi rendant compte des observations**
- ... et permettant de **faire des prédictions**.

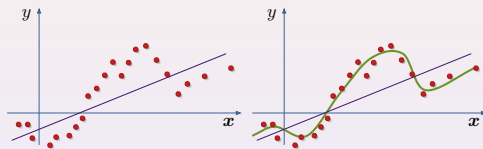


# Illustration : la régression

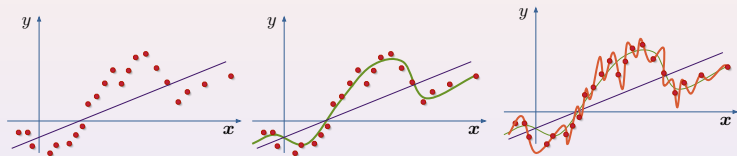




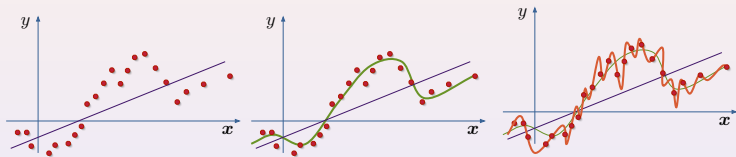
# Illustration : la régression



# Illustration : la régression



# Illustration : la régression

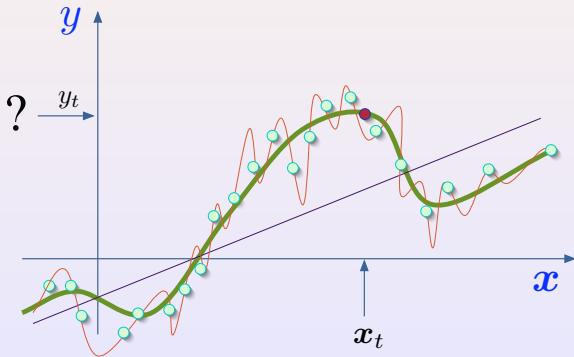


Quelle loi choisir ?

# Illustration : la régression

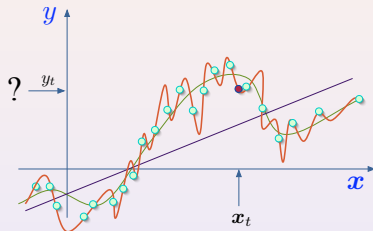
Le critère inductif

Le but : prédire



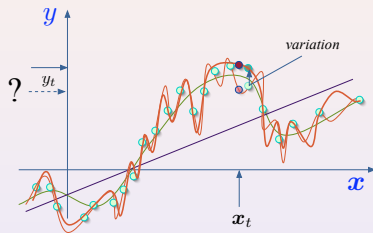
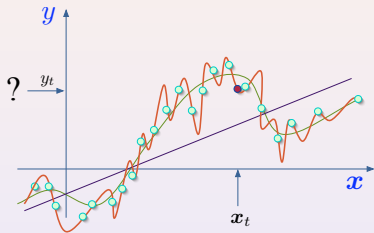
# Illustration : la régression

Le critère inductif



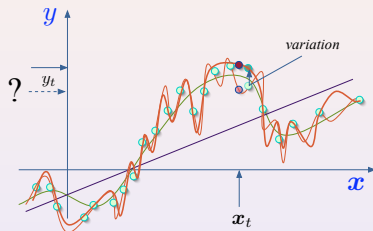
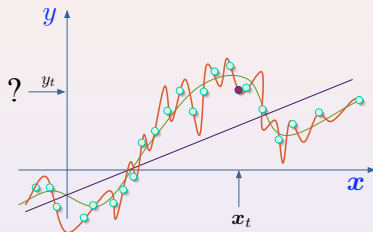
# Illustration : la régression

Le critère inductif



# Illustration : la régression

Le critère inductif

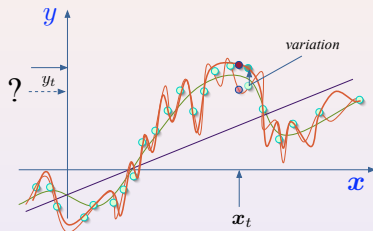
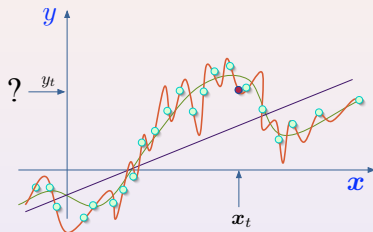


## Critère inductif

- **Fidélité** aux données d'apprentissage

# Illustration : la régression

Le critère inductif



## Critère inductif

- **Fidélité** aux données d'apprentissage
- ... en contrôlant la **capacité** de l'espace des hypothèses



# Théorie de l'apprentissage

Le cadre actuel

## Questions jugées essentielles

- 1 ***Quelle classe de fonctions (hypothèses) adéquate ?***
- 2 Choix d'une hypothèse dans cette classe : ***quel critère inductif ?***

# 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 auxiliaires

- 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

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

# Le paradigme

## Apprentissage = Problème inverse mal posé

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

## Proximité : Espérance de risque

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

## MRE

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

$$R_{\text{Emp}}(h) = \frac{1}{m} \sum_{(\mathbf{x}_i, u_i) \in S} \ell(h(\mathbf{x}_i), u_i)$$

# 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

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



# La théorie statistique de l'apprentissage : actualités

Estimations plus fines de la borne :  $\sup_{h \in \mathcal{H}} [R_{\text{Réel}}(h) - R_{\text{Emp}}(h)]$

## Directions actuelles

- 1 Nouvelles mesures de diversité
  - Nombre de couverture
  - Prise en compte de moments du 2ème ordre sur  $\mathcal{H}$
  - Inégalités de concentration
- 2 Nouvelles mesures empiriques de diversité
  - Moyennes de Rademacher
- 3 Restriction au sous-espace de  $\mathcal{H}$  exploré ( $R_{\text{Emp}}$  faible)
- 4 Apprentissage actif : modification de la distribution en apprentissage

# La théorie statistique de l'apprentissage : actualités

Estimations plus fines de la borne :  $\sup_{h \in \mathcal{H}} [R_{\text{Réel}}(h) - R_{\text{Emp}}(h)]$

## Directions actuelles

- 1 Nouvelles mesures de diversité
  - Nombre de couverture
  - Prise en compte de moments du 2ème ordre sur  $\mathcal{H}$
  - **Inégalités de concentration**
- 2 Nouvelles mesures empiriques de diversité
  - Moyennes de Rademacher
- 3 Restriction au sous-espace de  $\mathcal{H}$  exploré ( $R_{\text{Emp}}$  faible)
- 4 Apprentissage actif : modification de la distribution en apprentissage

# La théorie statistique de l'apprentissage : actualités

Estimations plus fines de la borne :  $\sup_{h \in \mathcal{H}} [R_{\text{Réel}}(h) - R_{\text{Emp}}(h)]$

## Directions actuelles

- 1 Nouvelles mesures de diversité
  - Nombre de couverture
  - Prise en compte de moments du 2ème ordre sur  $\mathcal{H}$
  - **Inégalités de concentration**
- 2 Nouvelles mesures empiriques de diversité
  - **Moyennes de Rademacher**
- 3 Restriction au sous-espace de  $\mathcal{H}$  exploré ( $R_{\text{Emp}}$  faible)
- 4 Apprentissage actif : modification de la distribution en apprentissage

# La théorie statistique de l'apprentissage : actualités

Estimations plus fines de la borne :  $\sup_{h \in \mathcal{H}} [R_{\text{Réel}}(h) - R_{\text{Emp}}(h)]$

## Directions actuelles

- 1 Nouvelles mesures de diversité
  - Nombre de couverture
  - Prise en compte de moments du 2ème ordre sur  $\mathcal{H}$
  - **Inégalités de concentration**
- 2 Nouvelles mesures empiriques de diversité
  - **Moyennes de Rademacher**
- 3 **Restriction au sous-espace de  $\mathcal{H}$  exploré** ( $R_{\text{Emp}}$  faible)
- 4 **Apprentissage actif** : modification de la distribution en apprentissage

# La théorie statistique de l'apprentissage : actualités

Estimations plus fines de la borne :  $\sup_{h \in \mathcal{H}} [R_{\text{Réel}}(h) - R_{\text{Emp}}(h)]$

## Directions actuelles

- 1 Nouvelles mesures de diversité
  - Nombre de couverture
  - Prise en compte de moments du 2ème ordre sur  $\mathcal{H}$
  - **Inégalités de concentration**
- 2 Nouvelles mesures empiriques de diversité
  - **Moyennes de Rademacher**
- 3 **Restriction au sous-espace de  $\mathcal{H}$  exploré** ( $R_{\text{Emp}}$  faible)
- 4 **Apprentissage actif** : modification de la distribution en apprentissage

[BBL05] [Bousquet, O., S. Boucheron, and G. Lugosi](#)  
*Theory of classification : A survey of recent advances.*  
ESAIM : Probability and Statistics.

[Mar05] [Jérémie Mary](#)  
*Étude de l'apprentissage actif. Applications à la conduite d'expériences*  
Thèse, LRI, Orsay, Déc. 2005.

# L'approche statistique de l'apprentissage : algorithmes

## ... de nouvelles méthodes

- Minimisation d'un  $\phi$ -risque empirique : *Un seul optimum*
  - Méthodes à noyaux (**SVM**, ...)
- Combinaison linéaire d'experts
  - Méthodes de votes (**bagging**, **boosting**, ...)

---

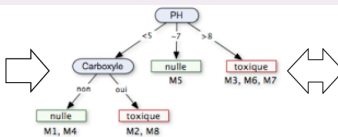
[MS03] Mendelson, S. and A. Smola, (Eds.)  
*Advanced Lectures on Machine Learning (Machine Learning Summer School 2002).*  
Springer-Verlag. LNAI-2600

# Le paradigme est adapté à ...

... l'analyse de données

**BD**  $\implies$  *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
  - Pas de relations de généralité
  - Pas de 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)



Prise en compte de :

- 1 **la structure de  $\mathcal{H}$**
- 2 **l'information apportée par les exemples**
- 3 **des stratégies d'exploration de  $\mathcal{H}$**

# Plan

- 1 Une approche de l'apprentissage inductif
  - L'induction supervisée
  - Analyse statistique 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

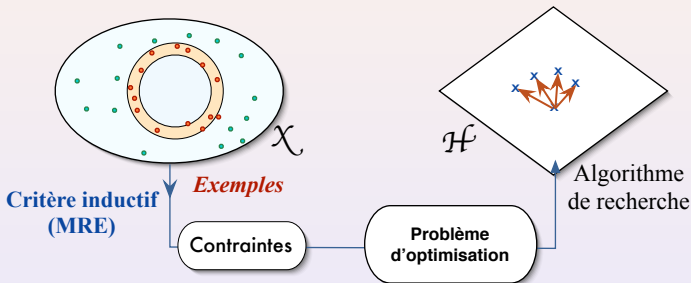
# Transition de phase en induction

## Collaborations

- Mario Botta
- Attilio Giordana
- Lorenza Saitta
  
- Nicolas Baskiotis (*Doctorant*)
- Jérôme Maloberti (*Thèse, 2005*)
- Nicolas Pernet (*stage DEA, 2004*)
- Sandra Pinto (*stage DEA, 2002*)
- Raymond Ros (*Doctorant*)
- Michèle Sebag (DR, CNRS)

# Sous quelles conditions l'induction est-elle possible ?

Des conditions sur le gain d'information

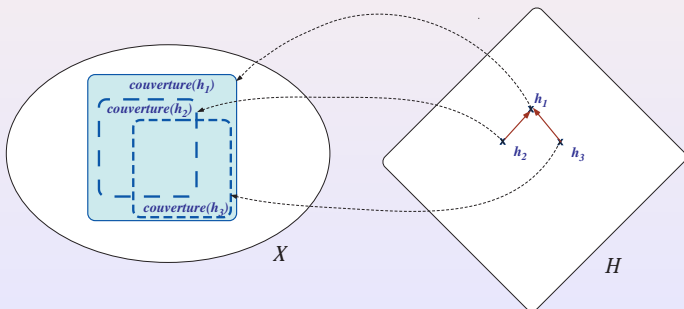


*Les entrées doivent **se traduire en « différences »** sur les hypothèses*

# 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)

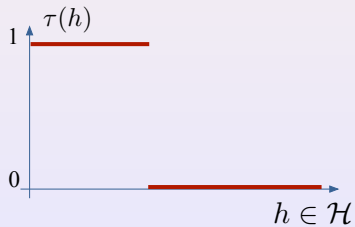
$$\tau(h) = P_{\mathcal{D}_X}(h)$$

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



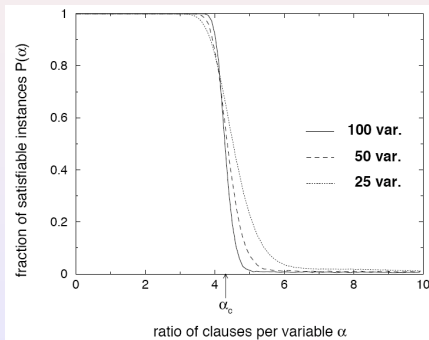
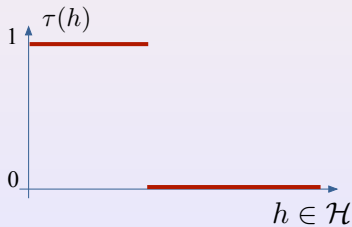
# Des conditions sur le gain d'information

Un cas limite ... mais ...



# Des conditions sur le gain d'information

Un cas limite ... mais ...



# Des conditions sur le gain d'information

Un CSP : le 3-SAT

$$\underbrace{(x \vee \bar{y} \vee z)}_{\text{Une clause}} \wedge (\bar{x} \vee v \vee w) \wedge \dots$$

$\underbrace{\hspace{15em}}_{N \text{ clauses}}$

- $N$  clauses
- $n$  variables
- $\alpha = N/n$  : taux de contraintes

# Des conditions sur le gain d'information

Un CSP : le 3-SAT

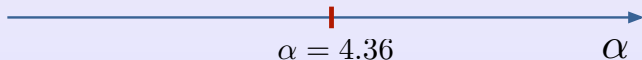
$$\underbrace{(x \vee \bar{y} \vee z)}_{\text{Une clause}} \wedge (\bar{x} \vee v \vee w) \wedge \dots$$

$\underbrace{\hspace{15em}}_{N \text{ clauses}}$

- $N$  clauses
- $n$  variables
- $\alpha = N/n$  : taux de contraintes

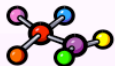
*Beaucoup  
de solutions*

*Pas  
de solution*

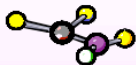


# L'induction de connaissances structurales

## Programmation Logique Inductive (ILP)



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

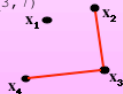


### active (d1)

```
lumo (d1, -1.246)
logp (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)
```

### Nonactive (d167)

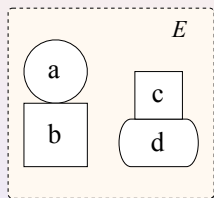
```
lumo (d167, -1.246)
logp (d167, 4.23)
atm (d167, d167_1, n, 22, -0.117)
atm (d167, d167_2, c, 22, -0.117)
atm (d167, d167_3, n, 22, -0.117)
atm (d167, d167_4, c, 195, -0.087)
atm (d167, d167_5, n, 195, 0.013)
bond (d167, d167_1, d167_2, 7)
bond (d167, d167_2, d167_3, 7)
....
```



**Relational Learner**

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

# Programmation Logique Inductive (ILP)



on(X, Y)

X	Y
a	b
c	d

left(X, Y)

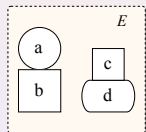
X	Y
a	c
a	d
b	c
b	d



$h(X, Y, Z) : - \text{left}(X, Y), \text{on}(Y, Z)$

X	Y	Z
a	c	d
b	c	d

# Programmation Logique Inductive (ILP)



$\text{on}(X, Y)$

X	Y
a	b
c	d

$\text{left}(X, Y)$

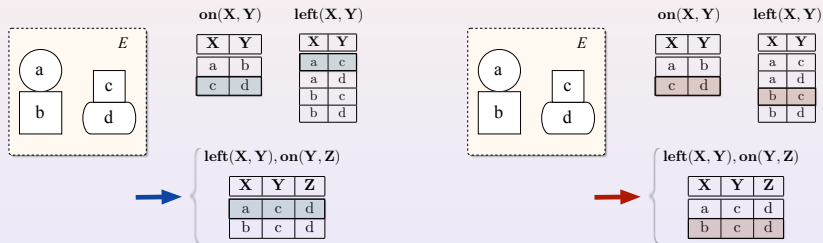
X	Y
a	c
a	d
b	c
b	d



$\left\{ \text{left}(X, Y), \text{on}(Y, Z) \right.$

X	Y	Z
a	c	d
b	c	d

# Programmation Logique Inductive (ILP)

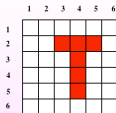




# L'induction de connaissances structurales

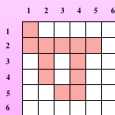
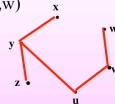
## Programmation Logique Inductive (ILP)

### Matching Problem



**Formula  $\varphi$**

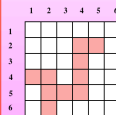
$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \\ \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \\ \text{adj}(y,u) \wedge \text{north}(u,v)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



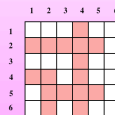
1 solution

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

Universe  $U$

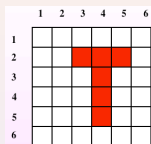


No solution



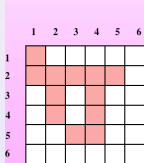
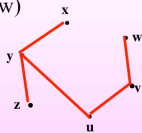
4 solutions

# ILP et satisfaction de contraintes



## Formula $\varphi$

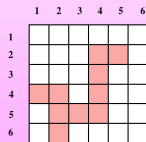
$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \\ \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \\ \text{adj}(y,u) \wedge \text{north}(u,v)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



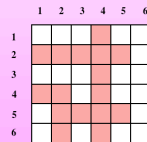
1 solution

west(x,y)		north(x,y)		adj(x,y)	
x	y	x	y	x	y
a <sub>21</sub>	a <sub>22</sub>	a <sub>11</sub>	a <sub>21</sub>	a <sub>11</sub>	a <sub>21</sub>
a <sub>21</sub>	a <sub>23</sub>	a <sub>22</sub>	a <sub>32</sub>	a <sub>21</sub>	a <sub>11</sub>
a <sub>21</sub>	a <sub>24</sub>	a <sub>22</sub>	a <sub>42</sub>	a <sub>21</sub>	a <sub>22</sub>
...	...	a <sub>22</sub>	a <sub>32</sub>	a <sub>22</sub>	a <sub>21</sub>
a <sub>21</sub>	a <sub>32</sub>	a <sub>32</sub>	a <sub>42</sub>	a <sub>22</sub>	a <sub>23</sub>
...	...	...	...	a <sub>22</sub>	a <sub>32</sub>
...	...	...	...	...	...
a <sub>54</sub>	a <sub>53</sub>	a <sub>44</sub>	a <sub>54</sub>	a <sub>53</sub>	a <sub>54</sub>

Universe  $U$

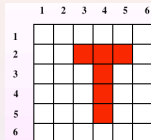


No solution



4 solutions

# ILP et satisfaction de contraintes

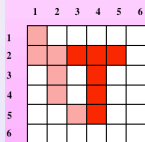
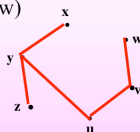


## Formula $\varphi$

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

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

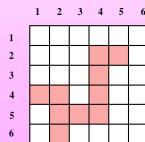
$$\text{adj}(y,u) \wedge \text{north}(u,v) \wedge \text{adj}(u,v) \wedge \text{north}(v,w) \wedge \text{adj}(v,w)$$



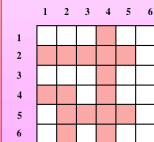
1 solution

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

Universe U

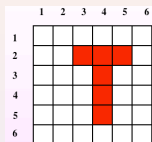


No solution



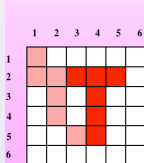
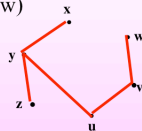
4 solutions

# ILP et satisfaction de contraintes



**Formula  $\varphi$**

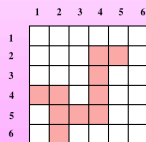
$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \text{adj}(y,u) \wedge \text{north}(u,v)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



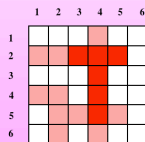
1 solution

west(x,y)		north(x,y)		adj(x,y)	
x	y	x	y	x	y
a <sub>21</sub>	a <sub>22</sub>	a <sub>11</sub>	a <sub>21</sub>	a <sub>11</sub>	a <sub>21</sub>
a <sub>21</sub>	a <sub>23</sub>	a <sub>22</sub>	a <sub>32</sub>	a <sub>21</sub>	a <sub>11</sub>
a <sub>21</sub>	a <sub>24</sub>	a <sub>22</sub>	a <sub>42</sub>	a <sub>21</sub>	a <sub>22</sub>
...	...	a <sub>22</sub>	a <sub>32</sub>	a <sub>22</sub>	a <sub>21</sub>
a <sub>21</sub>	a <sub>32</sub>	a <sub>32</sub>	a <sub>42</sub>	a <sub>22</sub>	a <sub>23</sub>
...	...	a <sub>32</sub>	a <sub>42</sub>	a <sub>22</sub>	a <sub>32</sub>
...	...	...	...	...	...
a <sub>54</sub>	a <sub>53</sub>	a <sub>44</sub>	a <sub>54</sub>	a <sub>53</sub>	a <sub>54</sub>

**Universe U**

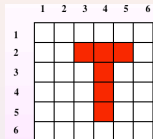


No solution



4 solutions

# ILP et satisfaction de contraintes

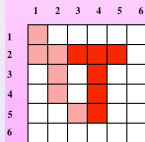
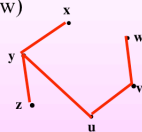


**Formula  $\varphi$**

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

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

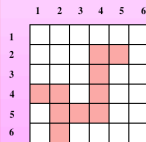
$$\text{adj}(y,u) \wedge \text{north}(u,v) \wedge \text{adj}(u,v) \wedge \text{north}(v,w) \wedge \text{adj}(v,w)$$



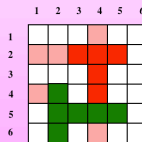
1 solution

west(x,y)		north(x,y)		adj(x,y)	
x	y	x	y	x	y
a <sub>21</sub>	a <sub>22</sub>	a <sub>11</sub>	a <sub>21</sub>	a <sub>11</sub>	a <sub>21</sub>
a <sub>21</sub>	a <sub>23</sub>	a <sub>22</sub>	a <sub>32</sub>	a <sub>21</sub>	a <sub>11</sub>
a <sub>21</sub>	a <sub>24</sub>	a <sub>22</sub>	a <sub>42</sub>	a <sub>21</sub>	a <sub>22</sub>
...	...	a <sub>22</sub>	a <sub>32</sub>	a <sub>22</sub>	a <sub>21</sub>
a <sub>21</sub>	a <sub>32</sub>	a <sub>32</sub>	a <sub>42</sub>	a <sub>22</sub>	a <sub>23</sub>
...	...	...	...	a <sub>22</sub>	a <sub>32</sub>
a <sub>54</sub>	a <sub>53</sub>	a <sub>44</sub>	a <sub>54</sub>	...	...
		a <sub>44</sub>	a <sub>54</sub>	a <sub>53</sub>	a <sub>54</sub>

**Universe U**

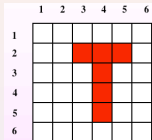


No solution



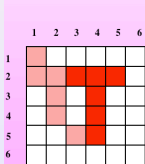
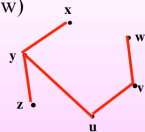
4 solutions

# ILP et satisfaction de contraintes



## Formula $\varphi$

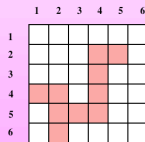
$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \text{adj}(y,u) \wedge \text{north}(u,v)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



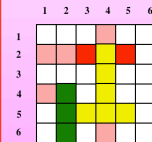
1 solution

west(x,y)		north(x,y)		adj(x,y)	
x	y	x	y	x	y
a <sub>21</sub>	a <sub>22</sub>	a <sub>11</sub>	a <sub>21</sub>	a <sub>11</sub>	a <sub>21</sub>
a <sub>21</sub>	a <sub>23</sub>	a <sub>22</sub>	a <sub>32</sub>	a <sub>21</sub>	a <sub>11</sub>
a <sub>21</sub>	a <sub>24</sub>	a <sub>22</sub>	a <sub>42</sub>	a <sub>21</sub>	a <sub>22</sub>
...	...	a <sub>22</sub>	a <sub>32</sub>	a <sub>22</sub>	a <sub>21</sub>
a <sub>21</sub>	a <sub>32</sub>	a <sub>32</sub>	a <sub>42</sub>	a <sub>22</sub>	a <sub>23</sub>
...	...	...	...	a <sub>22</sub>	a <sub>32</sub>
a <sub>54</sub>	a <sub>53</sub>	a <sub>44</sub>	a <sub>54</sub>	...	...
		a <sub>44</sub>	a <sub>54</sub>	a <sub>53</sub>	a <sub>54</sub>

Universe  $U$

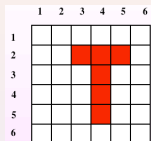


No solution



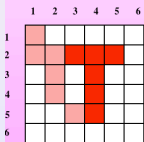
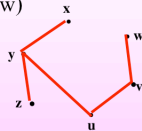
4 solutions

# ILP et satisfaction de contraintes



**Formula  $\varphi$**

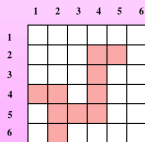
$$\varphi(x,y,z,u,v,w) = \text{red}(x) \wedge \text{red}(y) \wedge \text{red}(z) \wedge \text{red}(u) \wedge \text{red}(v) \wedge \\ \text{red}(w) \wedge \text{west}(x,y) \wedge \text{adj}(x,y) \wedge \text{west}(y,z) \wedge \text{adj}(y,z) \wedge \text{north}(y,u) \wedge \\ \text{adj}(y,u) \wedge \text{north}(u,v)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



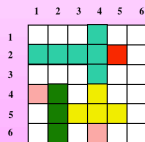
1 solution

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

**Universe U**



No solution



4 solutions

# Principe des expériences

## Étude en cas moyen

### *Génération aléatoire de problèmes*

#### Hypothèses

- $n$  : nombre de *variables* dans l'hypothèse  $h$  testée,
- $m$  : nombre de *symboles de prédicats* dans  $h$ ,

#### Exemples

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

→ *Étude dans le cas moyen*

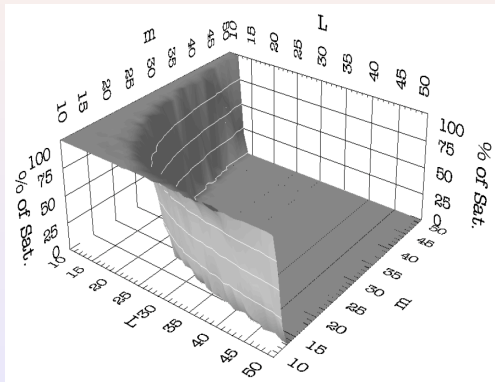


# Des conditions sur le gain d'information

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 prédicats* dans  $h$ ,
- $L$  : nombre total de *constantes* dans l'exemple  $e$ ,
- $N$  : nombre de *littéraux* construits sur chaque symbole de prédicat dans  $e$ .

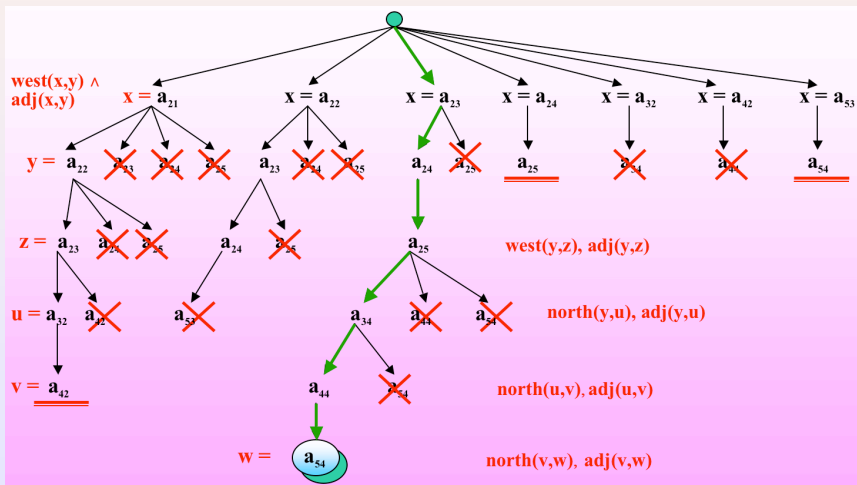


[bot03]

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.

# ILP et satisfaction de contraintes

Complexité de la recherche

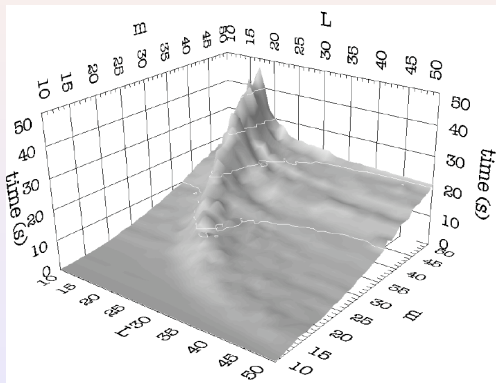


# Coût du test de couverture

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 prédicats* dans  $h$ ,
- $L$  : nombre total de *constantes* dans l'exemple  $e$ ,
- $N$  : nombre de *littéraux* construits sur chaque symbole de prédicat dans  $e$ .



[mal04]

Maloberti, J. and M. Sebag

Fast Theta-Subsumption with Constraint Satisfaction Algorithms.

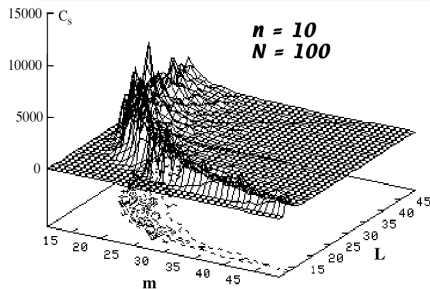
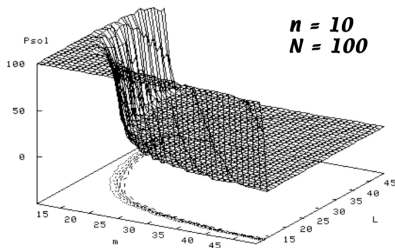
Machine Learning Journal, 55, 137-174, 2004.

# 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



100 problems  
for each pair  $(m, L)$

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

Set of 900,000  
matching problems

# Conséquences sur l'apprentissage de concept

## Paramètre :

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

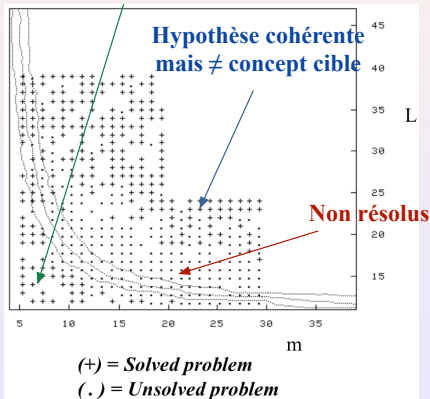
## Contours :

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

## Succès :

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

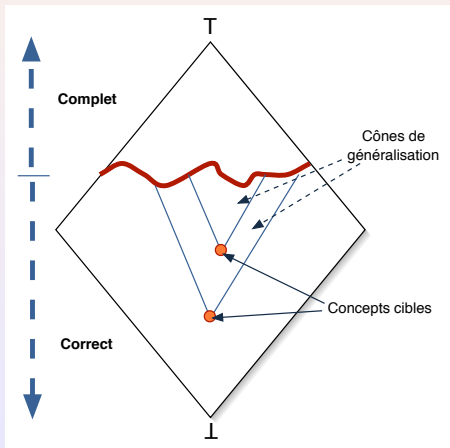
## Concept cible ~ identifié



# Conséquences sur l'apprentissage de concept

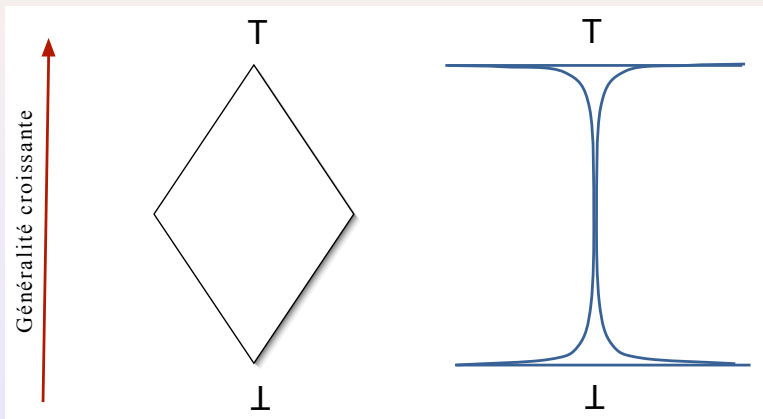
	Probl.	#Nof Clauses	Complexity	Class.Rate (LS) [%]	Class. Rate (TS) [%]	CPU Time [sec]	Avg. #N of Models
NO	LP <sub>1</sub> *	10	<7-13> 8.9	88	50	398.2	1.7
	LP <sub>7</sub> *	11	<6-11> 8.6	92	53	624.7	2.0
	LP <sub>3</sub> *	15	<7-11> 8.9	98.5	52	513.9	4.9
	LP <sub>4</sub> *	1	6	100	100	43.3	2.1
	LP 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

# Analyse



# Des conditions sur le gain d'information

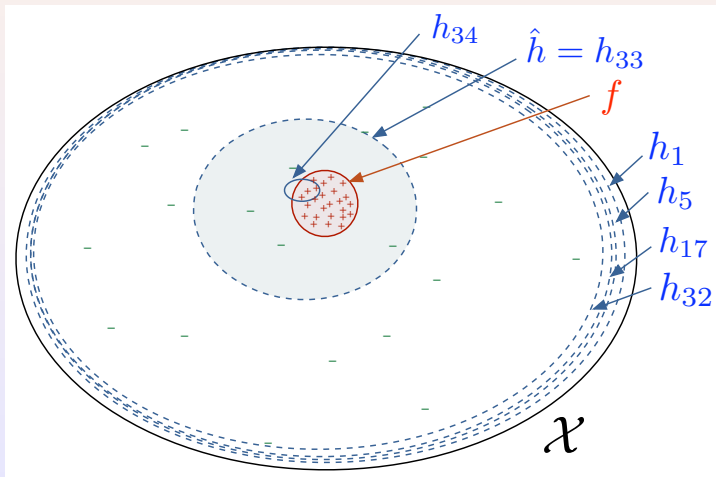
Transition de phase et espace des versions





# 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 :

- 1  $\mathcal{L}_{\mathcal{H}}$  : langage des hypothèses
- 2  $\mathcal{L}_{\mathcal{X}}$  : langage des exemples

# Des conditions sur le gain d'information

## Questions ouvertes

1. *Quels sont les langages affectés ?*

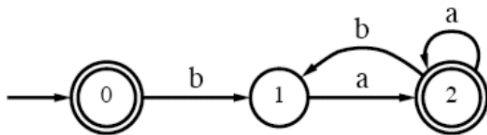
2. *Peut-on contourner le problème ?*

## Inférence grammaticale : rappels

- **Entrée** : *chaînes* sur un alphabet  $\Sigma$ , de longueur  $\ell$
- **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  $\ell$  des exemples testés.

---

[Pin01]

S. Pinto.

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

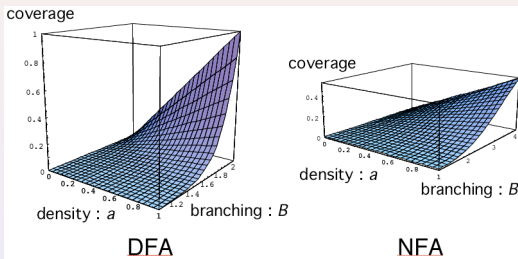
Rapport de DEA (LRI, Univ. Paris-Sud, Orsay), 2001.

# 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  $a \in [0, 1]$  d'états acceptants
- Taille  $|\Sigma|$  de l'alphabet
- Longueur  $\ell$  des exemples testés.



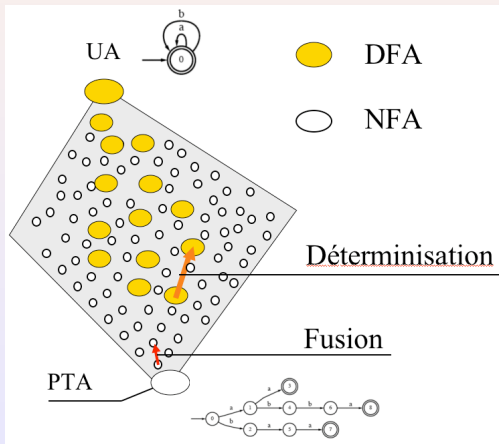
$$P(\text{accept}) = \begin{cases} a \cdot \left(\frac{B \cdot L}{|\Sigma|}\right)^\ell & \text{pour un DFA} \\ a \cdot \left[1 - \left(1 - \frac{L}{|\Sigma|}\right)^B\right]^\ell & \text{pour un NFA} \end{cases}$$

[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

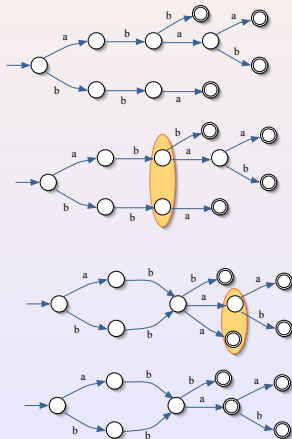
Principe des algorithmes d'apprentissage



# Gain d'information en inférence grammaticale

Distribution uniforme avec paramètres de contrôle

$$S^+ = \{ \langle a, b, b \rangle, \langle a, b, a, a \rangle, \langle a, b, a, b \rangle, \langle b, b, a \rangle \}$$





# Gain d'information en inférence grammaticale

Étude sur l'espace d'hypothèses **effectivement** exploré

## Protocole expérimental

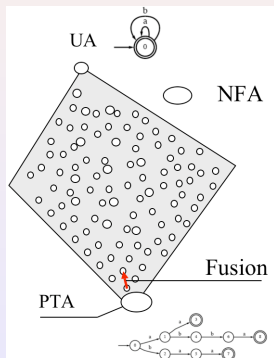
- 1 Génération aléatoire d'un **échantillon d'apprentissage** :  $|S^+|$  (= 200) chaînes de taille  $\ell$
- 2 Construction du **PTA** pour chaque échantillon  $S^+$
- 3 **Calcul de chemins de généralisation partant du PTA** :
  - **Fusions aléatoires**
  - **Couverture** calculée pour chaque automate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

## Expériences

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

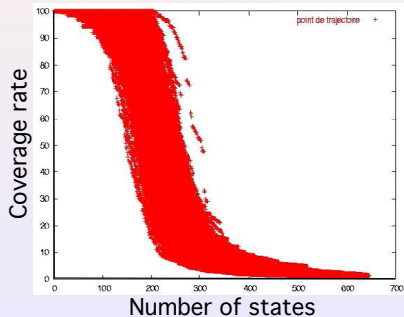
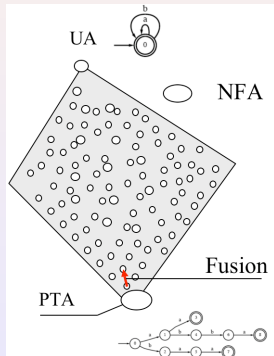
# Gain d'information en inférence grammaticale

Cas non-déterministe : NFA



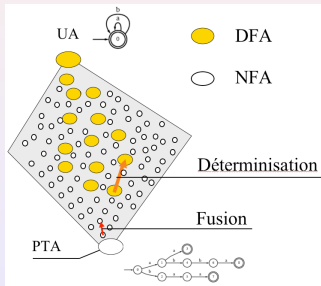
# Gain d'information en inférence grammaticale

Cas non-déterministe : NFA



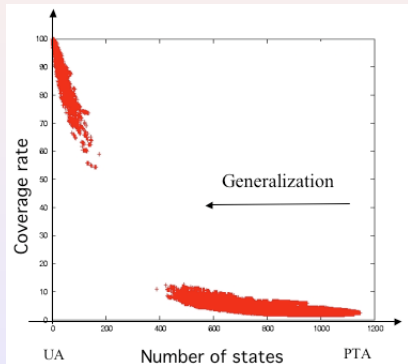
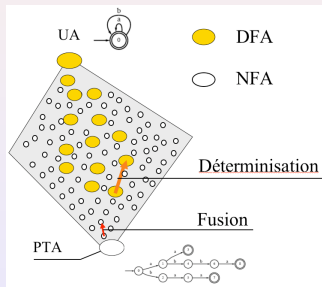
# Gain d'information en inférence grammaticale

Cas déterministe : DFA



# Gain d'information en inférence grammaticale

Cas déterministe : DFA



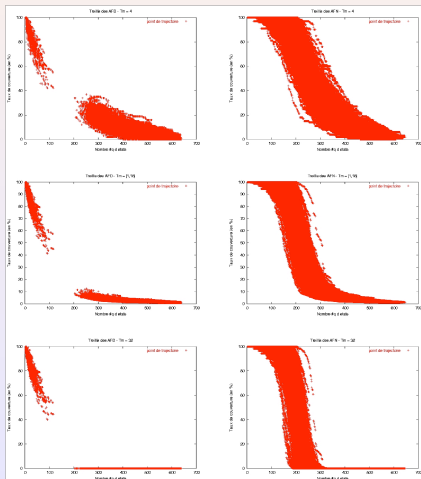
$$|\Sigma| = 8; \ell = 8$$

# Gain d'information en inférence grammaticale

Variété de situations

$$\begin{aligned} |\Sigma| &= 4 \\ \ell &= 16 \\ |S^+| &= 100 \end{aligned}$$

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



# Gain d'information en inférence grammaticale

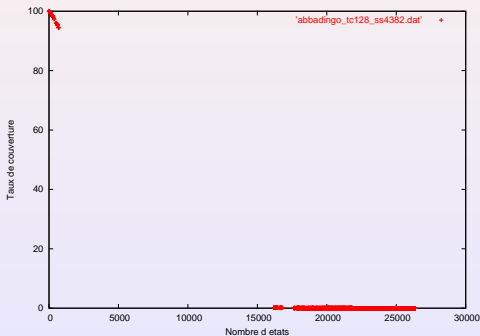
Le défi *Abbadingo*

$$|\Sigma| = 2$$

$$l = 17$$

$$|S^+| = 4382$$

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*

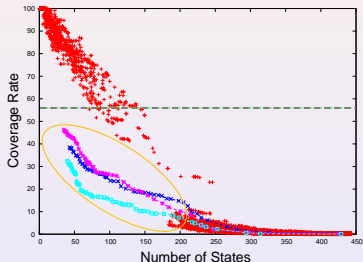
## EDSM [Lan et al.98][Lan98]

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

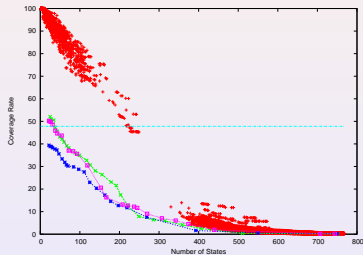


# Analyse pour les DFA

Algorithmes avec stratégies de recherche : **résultats**



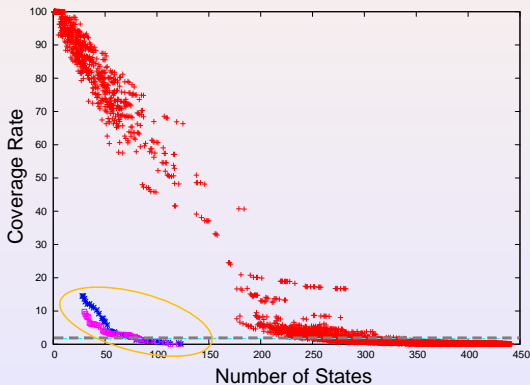
RPNI



EDSM

# Analyse pour les DFA

Algorithmes avec stratégies de recherche : **résultats**



# Analyse pour les DFA

Algorithmes avec stratégies de recherche : **résultats**

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 Une approche de l'apprentissage inductif
  - L'induction supervisée
  - Analyse statistique 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

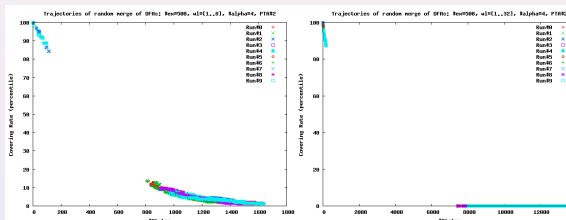
## Un saut de couverture spectaculaire

$$|\Sigma| = 4$$

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

$$|S^+| = 500$$

8 trajectoires aléatoires



# Analyse pour les DFA

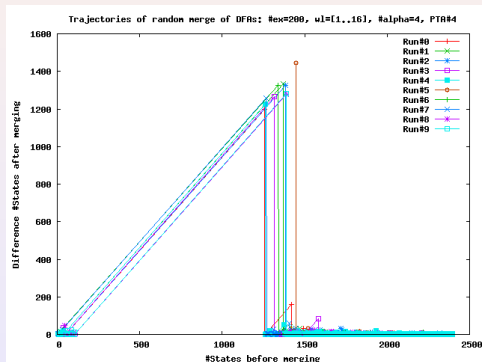
L'explication : avalanches de fusions pour déterminisation

$$|\Sigma| = 4$$

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

$$|S^+| = 200$$

8 trajectoires aléatoires



# Analyse pour les DFA

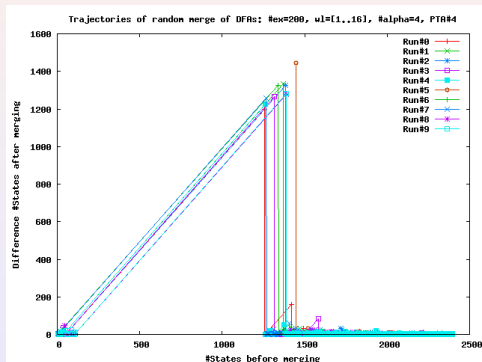
L'explication : avalanches de fusions pour déterminisation

$$|\Sigma| = 4$$

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

$$|S^+| = 200$$

8 trajectoires aléatoires



**Phénomène d'avalanche** (ou de réactions en chaînes)

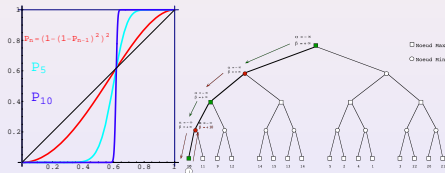
# Analyse pour les NFA

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**

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





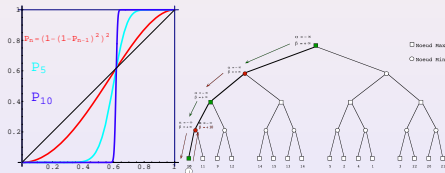
# Analyse pour les NFA

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**

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

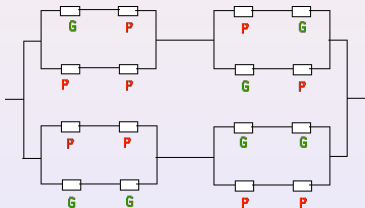


En informatique :

**La récurrence peut conduire à un phénomène de TP**

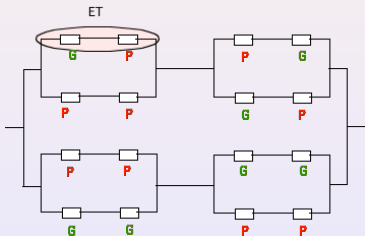
# Analyse pour les NFA

## Structures auto-similaires



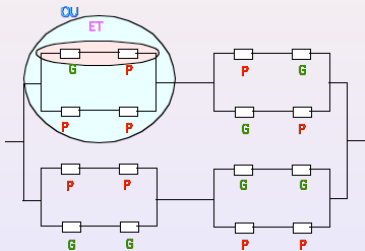
# Analyse pour les NFA

## Structures auto-similaires



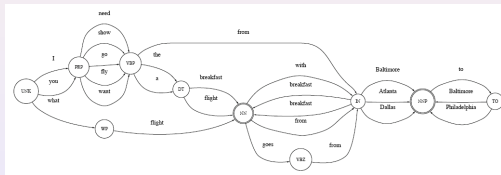
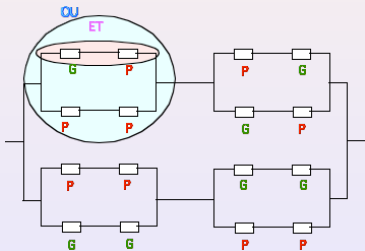
# Analyse pour les NFA

## Structures auto-similaires



# Analyse pour les NFA

## Structures auto-similaires



## Remèdes ?

### Changement de langage de représentation

- *E.g. par abstraction* [Saitta & Zucker, 02]
- *Autres opérateurs de spécialisation / généralisation*

### Nouvelles mesures de gain d'information

- *Mesure d'appariement souple*
- *Exploration aléatoire en avant*
- ...

### Autres types d'apprentissage : choix des exemples

- *Apprentissage actif*
- *Apprentissage guidé*

# Analyse

## *Alors ... pour quels langages ?*

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

# 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 récurrent

*Encore du domaine de la recherche*



# Gain d'information et induction

## Analyse statistique de l'induction

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

# Gain d'information et induction

## Analyse statistique de l'induction

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

## Étude des variations du taux de couverture

- **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

## Analyse statistique de l'induction

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

## Étude des variations du taux de couverture

- **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 !***