Apprentissage Supervisé Relationnel par Algorithmes d'Évolution

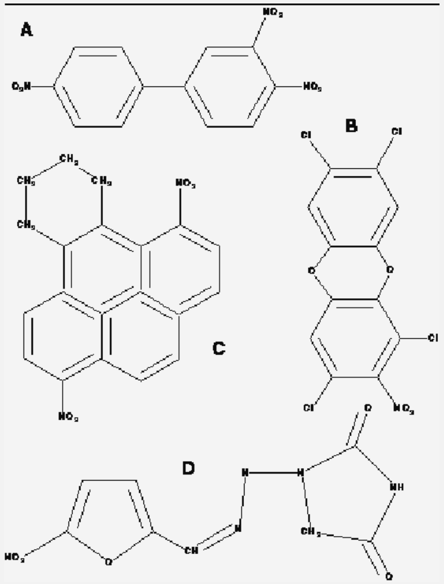
Sébastien Augier

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

19/12/2000

L'Apprentissage Supervisé Relationnel

Le problème de PLI mutagénèse : [Srinivasan et al., 1994]



$$\longrightarrow \quad atm(X, Z, *) \wedge atm(Y, Z, *) \wedge$$
$$atm(*, Z, [0.106, \infty]) \wedge lien(X, Y, *)$$

Soit :
- $\mathcal{B}$ une base d'apprentissage
- $\mathcal{L}_h$ un ensemble de formules (hypothèses)
- $f_{\mathcal{B}} : \mathcal{L}_h \longrightarrow \Re$

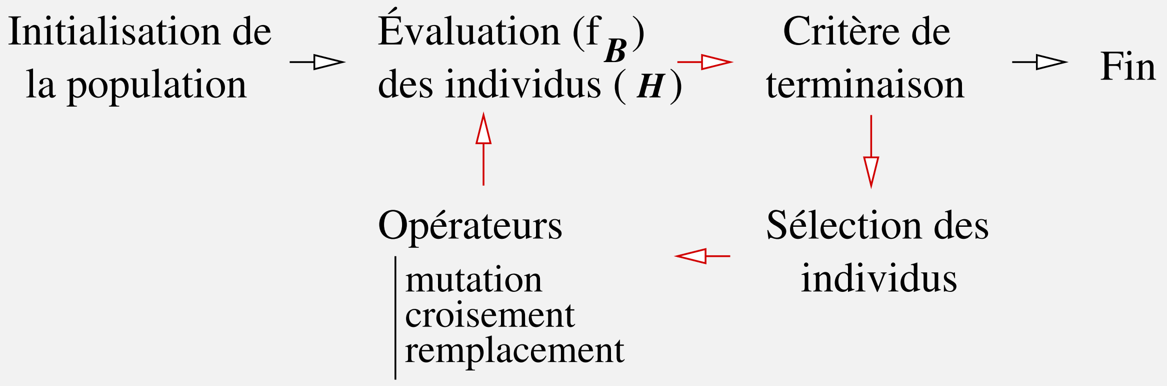
Trouver une hypothèse $\mathcal{H}$ telle que $f_{\mathcal{B}}(\mathcal{H})$ soit maximale.

Les Algorithmes d'Évolution (AE)

– Fondés sur les mécanismes de l'évolution biologique darwinienne.

[Fogel, 1966] [Schwefel, 1968] [Holland, 1975] [Goldberg, 1989] [Koza, 1994]

– Boucle générique :

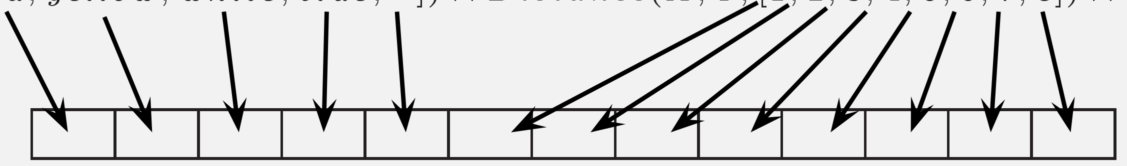


Application des AE à l'Apprentissage Relationnel

1. Algorithmes génétiques : Chaînes binaires de taille fixe

[Giordana et al., 1997] [Hekanaho, 1998]

$Color(X, [red, yellow, white, blue, \neg]) \wedge Distance(X, Y, [1, 2, 3, 4, 5, 6, 7, 8]) \wedge Follow(X, Y)$

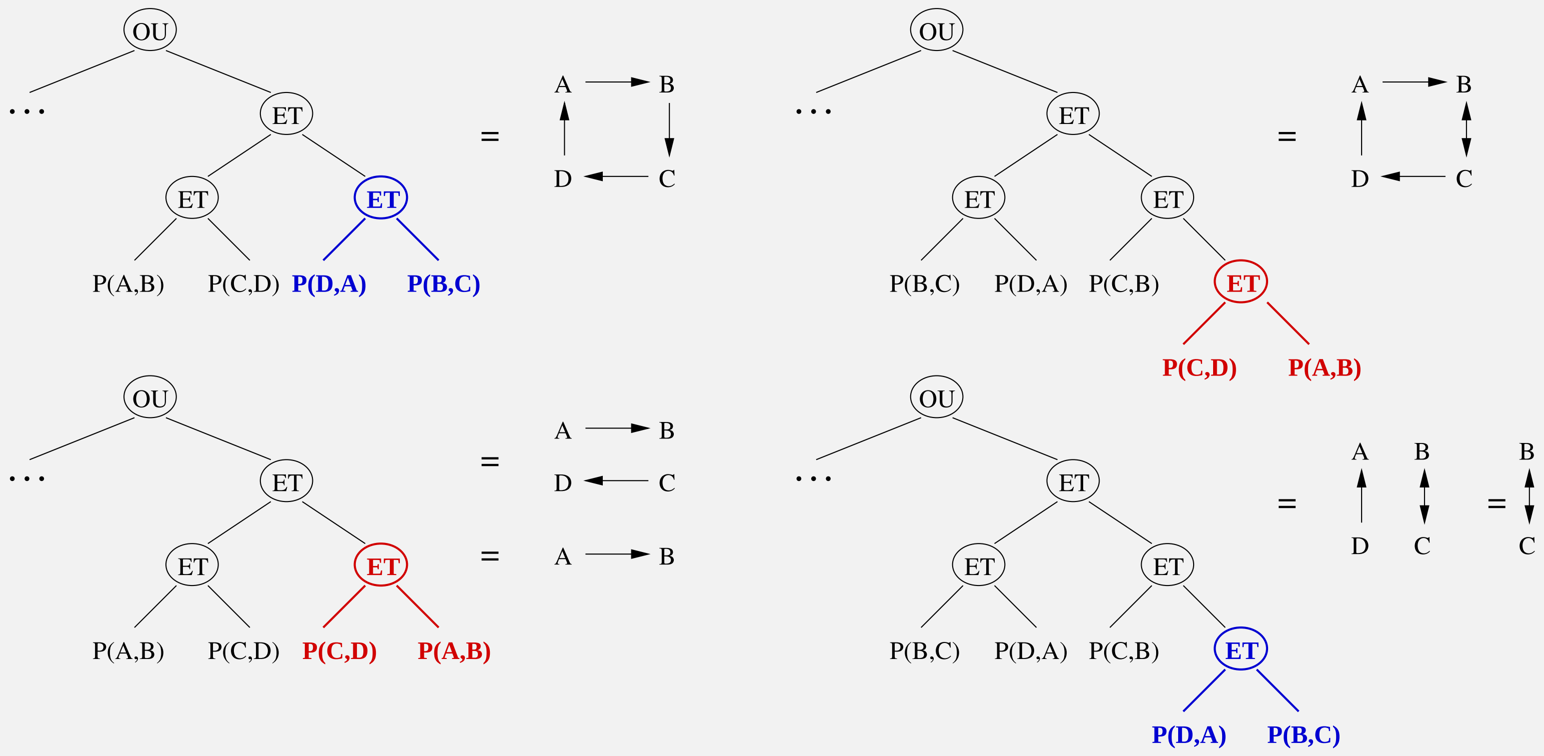


Parent 1 :	1	1	1	0	0	1	0	0	0	1	0	1	0
Parent 2 :	1	1	0	0	0	0	0	0	0	0	1	1	1
Enfant 1 :	1	1	1	0	0	1	0	0	0	0	1	1	1
Enfant 2 :	1	1	0	0	0	0	0	0	0	1	0	1	0

Application des AE à l'Apprentissage Relationnel

2. Programmation génétique : Arbres ET/OU

[Wong & Leung, 1995]



Définition du langage $\mathcal{L}_h$

Fondé sur DATALOG

- Les formules sont du type : $classe \leftarrow a_1 \wedge a_2 \wedge ... \wedge a_n.$
- Théorie du domaine : hiérarchies sur les constantes et les prédicats.

Extensions

- Disjonctions internes et intervalles numériques :

$$atm(X, \{h \ \vee \ n\}, [0,143..0,812])$$

- Symboles spéciaux typés `*` et `?` :

classe	identifiant	nb_roues	couleur	contenance_remorque	marque
voiture	$283LZZ75$	4	$*$	$\#$	?

$$classe(X, voiture) \leftarrow id(X, 283LZZ75) \wedge nb\_roues(X,4) \wedge couleur(X,*) \wedge marque(X,?).$$

Le typage

- Le typage restreint les généralisations :

$$Objet(X, rectangle, bleu) \not\rightarrow Objet(X, rectangle, \{bleu \vee cercle\})$$

$$Objet(X, rectangle, bleu) \rightarrow Objet(X, rectangle, \{bleu \vee rouge\})$$

- Il est enrichi par deux modalités : [Sebag & Rouveirol 2000]

  - types fonctionnels (valeur) $\rightarrow$ extensions de domaines
  - types relationnels (structure) $\rightarrow$ variabilisations

Apprentissage par interprétations

Molécules

Id_m	Classe
$m1$	$classe1$
$m2$	$classe2$
$\ldots$	

Atomes

Id_m	Id_a	Élément
$m1$	$a1$	o
$m1$	$a2$	s
$m1$	$a3$	o
$m2$	$a4$	c
$\ldots$		

Liaisons

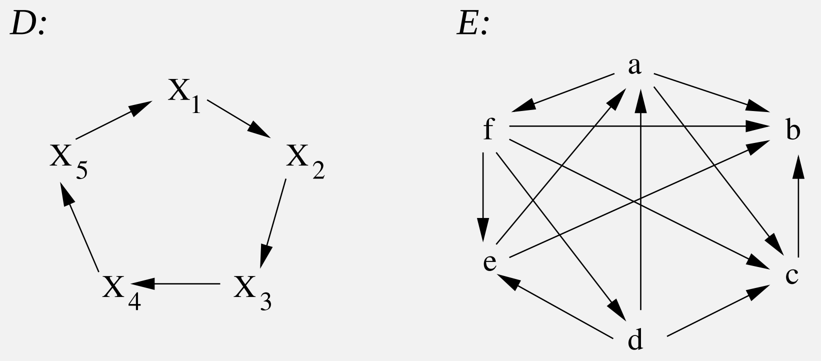
Id_m	Id_a	Id_a	Liaison
$m1$	$a1$	$a2$	$dative\_2e$
$m1$	$a2$	$a3$	$covalente\_4e$
$\ldots$			

$$\begin{aligned} m1 &: classe1 \leftarrow atome(a1, o) \ \wedge \ atome(a2, s) \ \wedge \ atome(a3, o) \ \wedge \\ &\qquad liaison(a1, a2, dative\_2e) \wedge liaison(a2, a3, covalente\_4e). \\ m2 &: classe2 \leftarrow atome(a4, c) \ \wedge \ \ldots \end{aligned}$$

Le test de couverture $\vdash$

– Défini en étendant la θ-subsomption utilisée en PLI à $\mathcal{L}_h$

– Comme la θ-subsomption, il est décidable mais NP-difficile



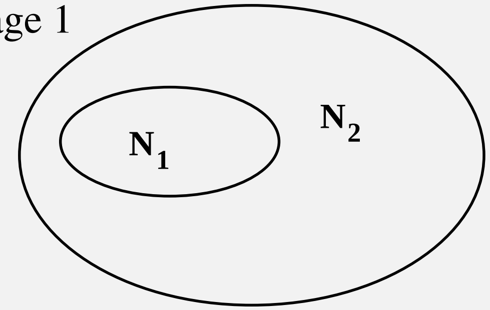
– Il est utilisé un très grand nombre de fois ($\propto I.|\mathcal{B}|$)

– L'optimisation de ce test est cruciale

Stratégie de recherche de $\mathcal{SIAO}1$

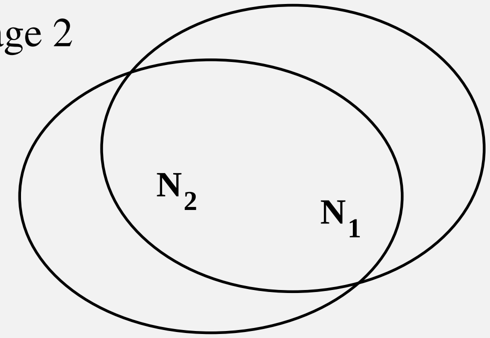
- Système SIA [Venturini,1994]
- Direction de recherche ascendante
- Développement simultané de plusieurs noyaux
- Processus d'élagage tous les $N_{cycle} = 100$ générations

Élagage 1



Couverture ($\mathbf{N_1}$) > Couverture ($\mathbf{N_2}$)
Cohérence ($\mathbf{N_1}$) $\geq$ Cohérence ($\mathbf{N_2}$)

Élagage 2

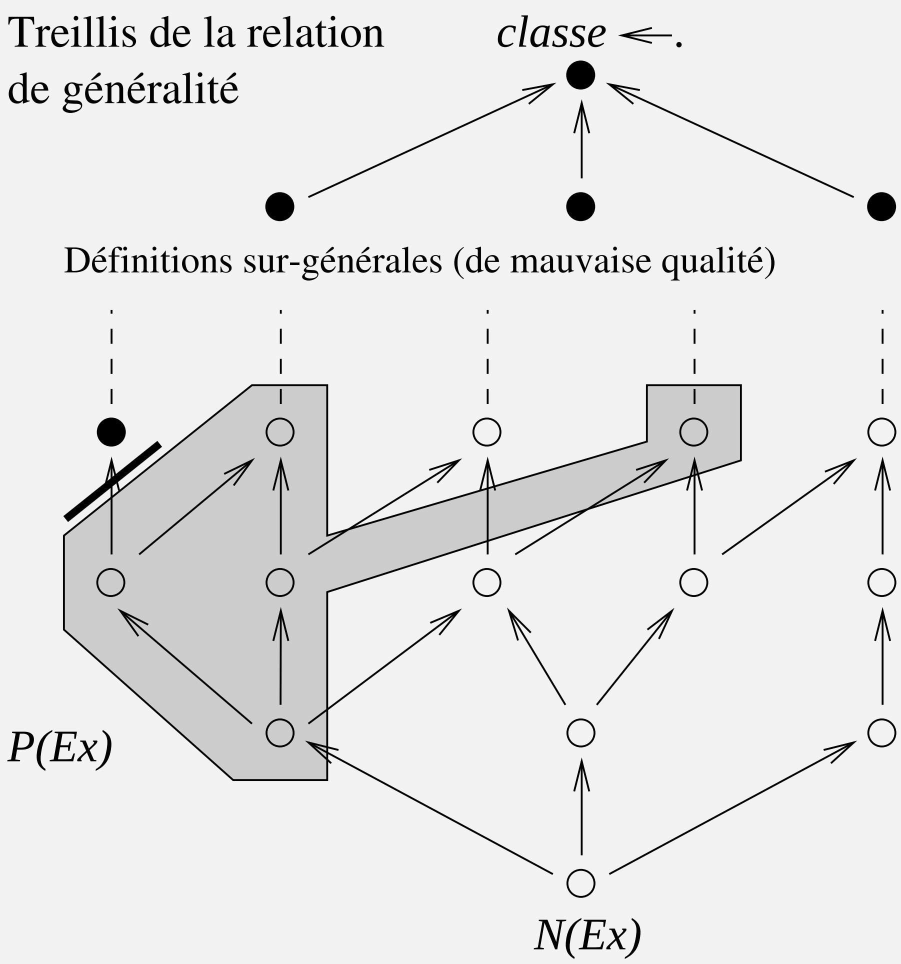


Cohérence ($\mathbf{N_1}$) $\geq$ Cohérence ($\mathbf{N_2}$)

Croissance d'un noyau

Propriétés :

- Recherche ascendante
- Terminaison de l'algorithme
- Incrémentalité *steady-state*
- Élitisme
- Diversité maximale
- Auto-adaptation des taux des opérateurs



Calcul de la qualité

$$Q(H) = \underbrace{\frac{p+1}{p+n+k}}_{\textit{estimateur de Laplace}} + \alpha \; . \; \underbrace{BS(H)}_{\textit{biais syntaxique}} \qquad (0 \leq \alpha \ll 1)$$

- Fondé sur la cohérence
- Sensible aux effectifs couverts
- Robuste au bruit

$$P(*) > P(X) > P(a)$$

La mutation

[APC, Michalski]

$\mathcal{R}_1$ Voiture(X) $\xrightarrow{\mu}$ Véhicule(X) ; Véhicule(X) $\xrightarrow{\mu}$ Vrai

$\mathcal{R}_2$ Élément(X,rectangle,rouge) $\xrightarrow{\mu}$ Élément(X,polygone,rouge)

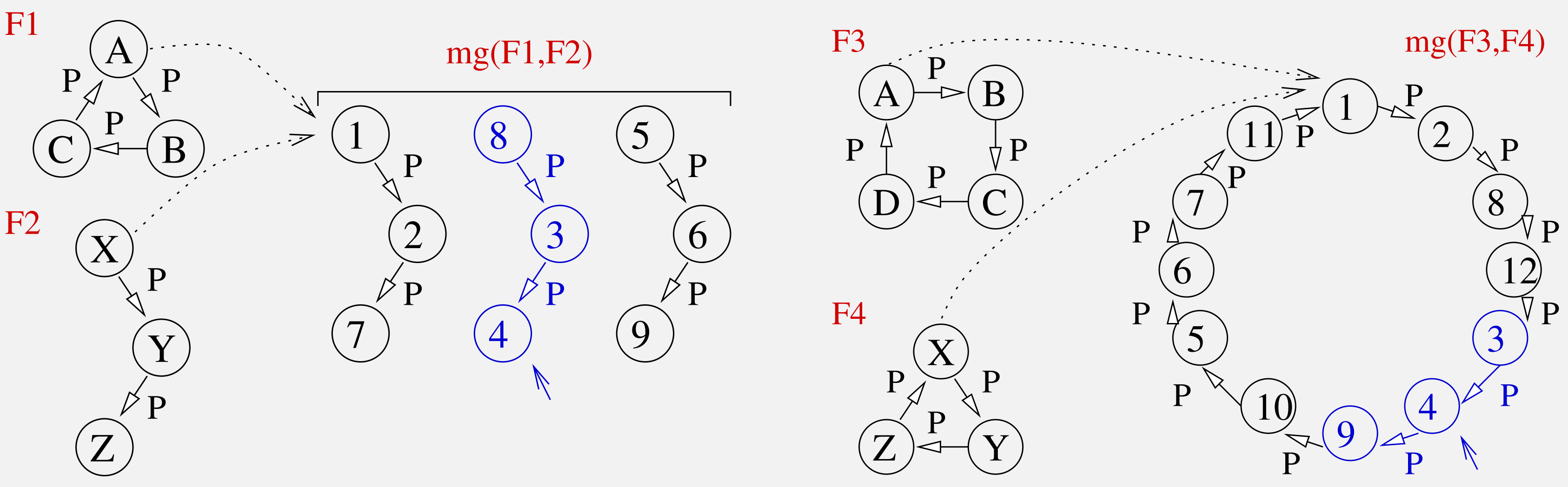
$\mathcal{R}_3$ Père(pierre,jean) $\wedge$ Père(jean,christophe) $\wedge$ Mère(jean,claire)
$\xrightarrow{\mu}$ Père(pierre,X) $\wedge$ Père(X,christophe) $\wedge$ Mère(X,claire)

$\mathcal{R}_4$ Température(X,97,3) $\xrightarrow{\mu}$ Température(X,[97,3 .. 98,5])

$\mathcal{R}_5$ Élément(X,{ rectangle $\vee$ cercle },rouge)
$\xrightarrow{\mu}$ Élément(X,{ rectangle $\vee$ cercle $\vee$ triangle},rouge)

Le croisement

Fondé sur le moindre généralisé (mg) ... [Plotkin, 1971]



1 = A+X, 2 = B+Y, ...

... la taille de l'individu générée étant bornée.

Évaluation expérimentale 1/2

$\mathcal{SIAO}1$ est polyvalent et sollicite peu l'utilisateur :

Iris

Système	Taux d'erreur (%)
C4.5	5,7
$\mathcal{SIAO}1$	6,0
RBN	6,3
CART	7,0

1000-Trains

Système	Taux d'erreur (%)
G-Net	0,0
$\mathcal{SIAO}1$	2,0
REGAL	2,6

Mushrooms

Système	Taux d'erreur (%)
$\mathcal{SIAO}1$	0,0
C4.5	0,0
G-NET	0,0
REGAL	0,0
RBN	0,3
STAGGER	5,0
HILLARY	5,0

Mutagenesis

Système	BK	Taux d'erreur (%)
PROGOL	B_2	19
TILDE	B_2	21
$\mathcal{SIAO}1$	B_2	22
FOIL	B_2	39

Évaluation expérimentale 2/2

Census-Income
(300.000 exemples)

Système	Taux d'erreur (%)
C4.5 + boosting	4,8
$\mathcal{SIAO}$1	4,9
C4.5	6,2
RBN	23,2

Adult
(30.000 exemples)

Système	Taux d'erreur (%)
NBTree	14,10
C4.5	15,54
CN2	16,00
RBN	16,12
$\mathcal{SIAO}$1	16,38
ID3 (vote)	16,47
T2	16,84
1R	19,54
PPV (3)	20,35
PPV (1)	21,42
Pebls	Crashed

Le changement d'échelle

– Un axe important de l'apprentissage automatique

[Dietterich, 1997]

– Changement d'échelle en espace

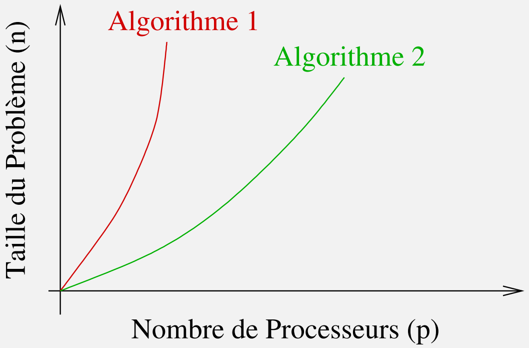
– Changement d'échelle en temps

$\Rightarrow$ formalisation dans le cadre du parallélisme

Parallélisme (Objectifs)

Contexte d'évaluation

- p processeurs identiques ; on mesure $T_p(n)$
- Accélération $A(n,p) = \frac{T_1(n)}{T_p(n)}$ *Gain pour l'utilisateur*
- Efficacité $E(n,p) = \frac{A(n,p)}{p}$ *Taux de rendement*
- Iso-efficacité : changement d'échelle



$\rightarrow$ Nécessité d'un modèle

Architectures proposées

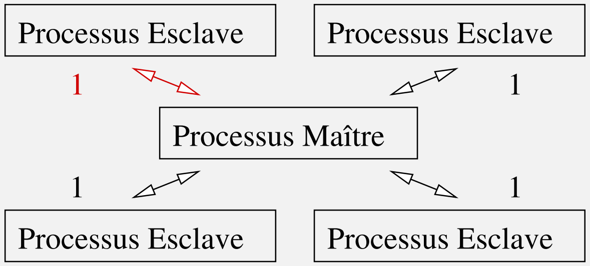
Observations

– $n \propto$ nombre d'exemples

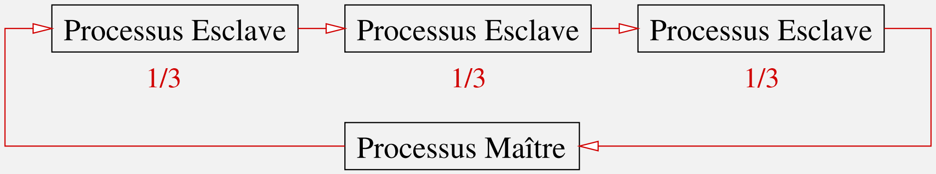
– 97% du temps passé en appariements

Architectures

Maître-esclave

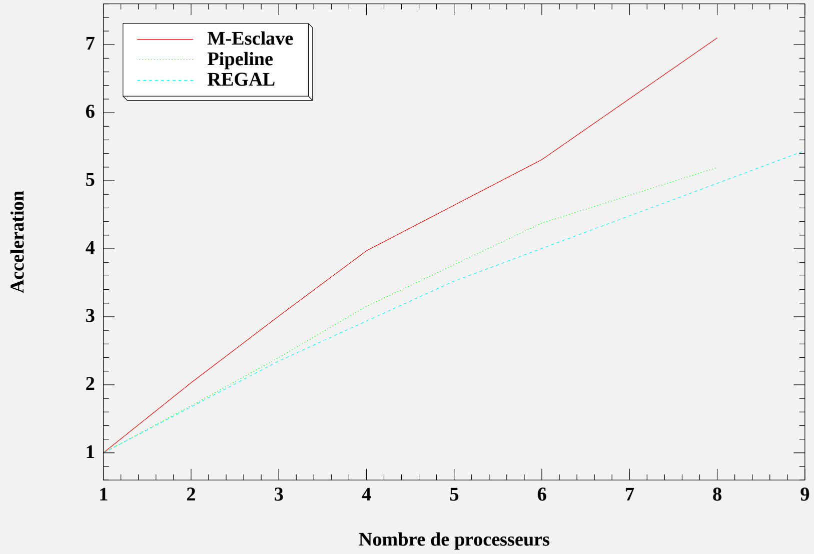


Pipeline

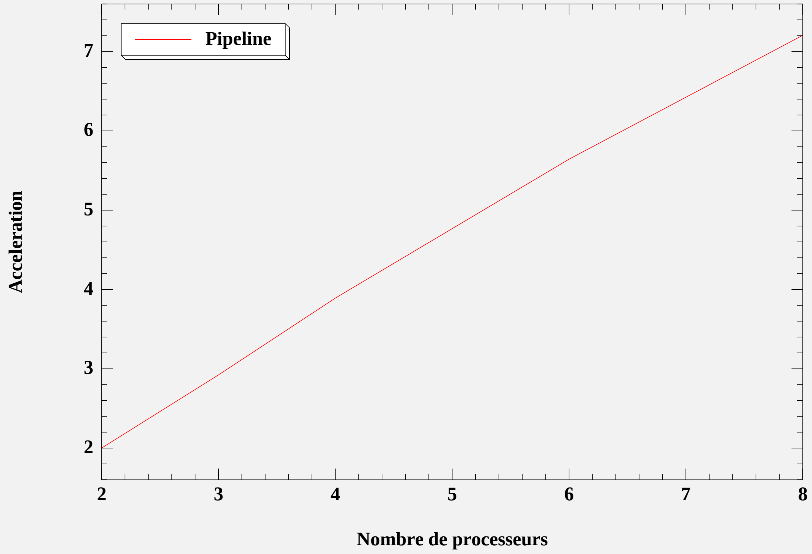


Premiers résultats

Mushrooms

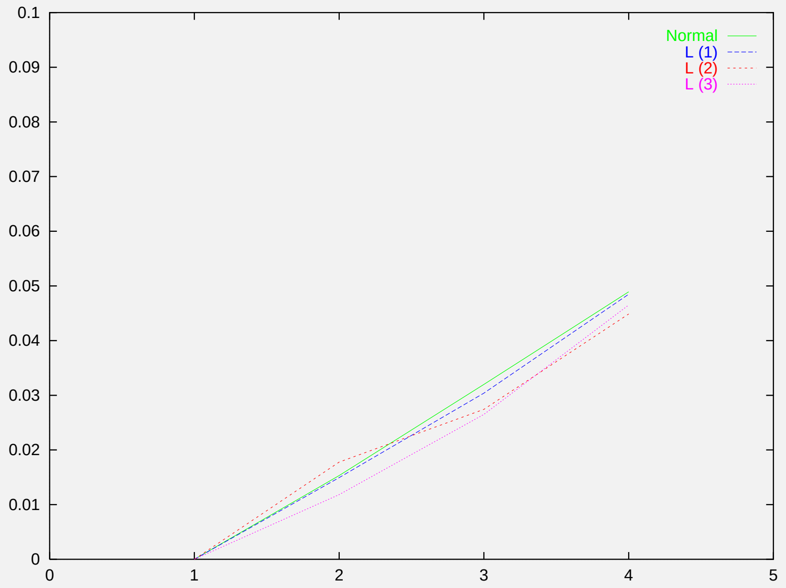


Census – income

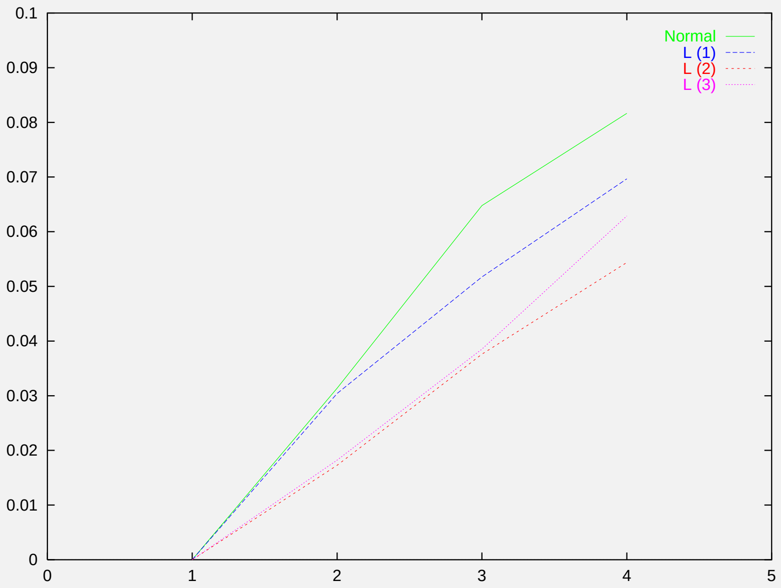


$T_{Com}(n,p)$ sur *mushrooms*

Architecture maître-esclave



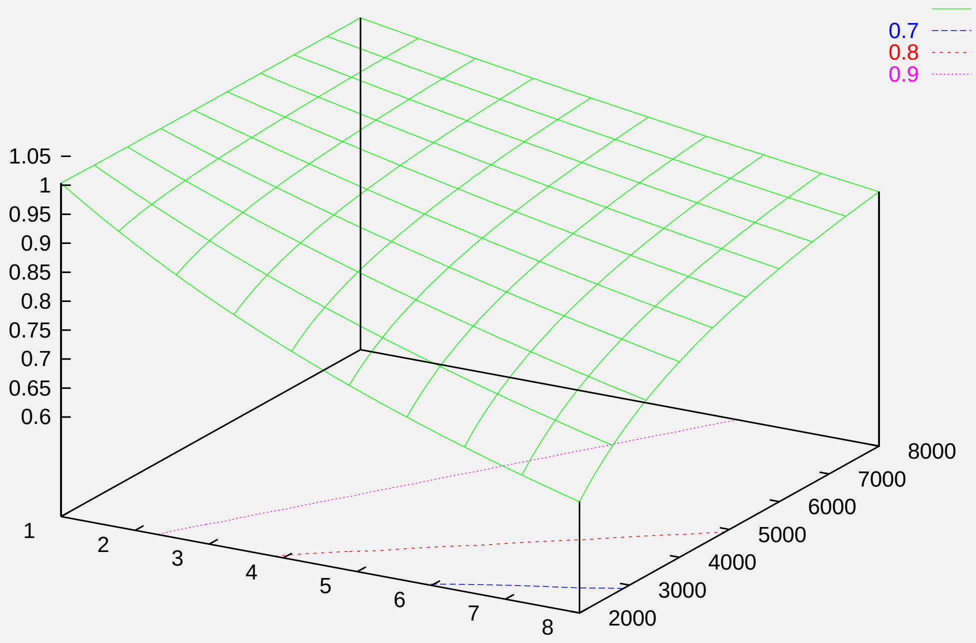
Architecture pipeline



$\Rightarrow$ Le modèle obtenu est précis :

Test	Sur les données	Processeurs $\uparrow$	Exemples $\updownarrow$
Erreur max	$\leq 1\%$	$\leq 2\%$	$\leq 5\%$

Changement d'échelle et Loi d'Amdahl



Courbes d'iso-efficacité obtenues grâce au modèle.

Problème : *mushrooms* Architecture : maître-esclave Taille de la file : 1 individu

Conclusions

1. $\mathcal{SIAO}1$

   – Compromis expressivité/complexité

   – L'apprentissage est un problème d'optimisation particulier :

     - Recherche dirigée par la relation d'ordre
     - Opérateurs dirigés par les données

2. Parallélisme

   – Maître-esclave : changement d'échelle en temps

   – Pipeline : changement d'échelle en espace

   – Modèle prédictif des performances

Perspectives

- Ne plus demander à l'utilisateur le nombre maximum de définitions : hybridation coevolution/incrémentalité des noyaux
- Coopération avec l'utilisateur afin de gérer le cycle entier de découverte des connaissances
- Analyse Discriminante **en cours** d'évolution

Durées d'apprentissage de $\mathcal{SIAO}1$ et G-Net

	$\mathcal{SIAO}1$ [a]	G-Net [b]
$Mushrooms$	$\simeq 300\ s.$	$\simeq 5900\ s.$
$1000 - Trains$	$\simeq 90\ mn.$	$\geq 300\ mn.$

Résultats obtenus sur une même machine uniprocesseur (PC).

[a] Temps total de l'apprentissage (découverte de la solution + convergence).

[b] Temps nécessaire à la découverte de la solution

La relation de généralité dans $\mathcal{L}_h$

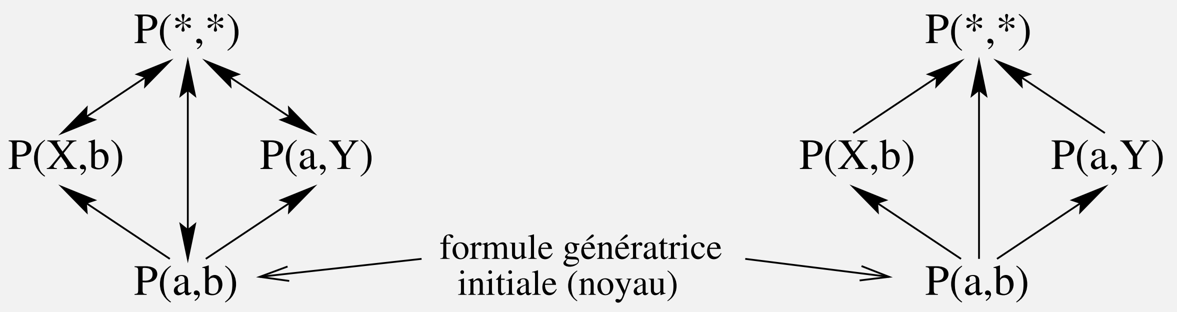
– Cas particulier des symboles `*` et `?`

$$Prison \leftarrow Coupable(X, Meurtre, Londres) \overset{?}{\vdash} Prison \leftarrow Coupable(Jack, Meurtre, *)$$

$$Frère \leftarrow Père(Pierre, X) \wedge Père(Paul, X) \overset{?}{\vdash} Frère \leftarrow$$
$$Père(Pierre, ?) \wedge Père(Paul, ?)$$

– Relation de couverture ($\vdash$) et relation d'ordre ($\vdash_2$)

Graphe de la relation $\vdash$ Treillis défini par $\vdash_2$



Implantation du test $\vdash$

1. Reformulation sous forme de PSC

$$\mathcal{L} = \left[\begin{array}{lllll} P(W) : & [e,*,*,*] & [f,*,*,*] & [g,*,*,*] & \\ \hline\hline Q(Y,Z) : & \color{red}{[*,*,a,b]} & [*,*,*,b] & \color{blue}{[*,*,c,c]} & [*,*,d,b] \\ R(Y,Z) : & [*,*,a,b] & [*,*,b,c] & [*,*,c,d] & [*,*,d,b] \\ R(Z,X) : & \color{blue}{[*,b,*,a]} & [*,c,*,b] & [*,d,*,c] & [*,b,*,d] \end{array}\right.$$

2. Réduction du tableau
   - Composantes indépendantes
   - Cohérence par nœud et par arc
   - Cohérence par chemin (heuristique)

$$\left.\begin{array}{l} [*,*,a,b][*,*,b,c][*,*,c,d][*,*,d,b] \\ [*,c,*,b][*,d,*,c][*,b,*,d] \end{array}\right\} \rightarrow [*,c,a,b][*,d,b,c][*,b,c,d][*,c,d,b]$$

3. Résolution avec propagation des contraintes de type *forward checking*

Bonjour,

Je vous convie à assister à ma soutenance de Thèse intitulée :

Apprentissage Supervisé Relationnel par Algorithmes d'Évolution

le mardi 19 décembre en salle 79 à 14h30

... ainsi qu'au pot qui suivra.

Sébastien Augier