

Apprentissage et circulation de l'information

Antoine Cornuéjols

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

Soutenance d'Habilitation à diriger des Recherches

Plan

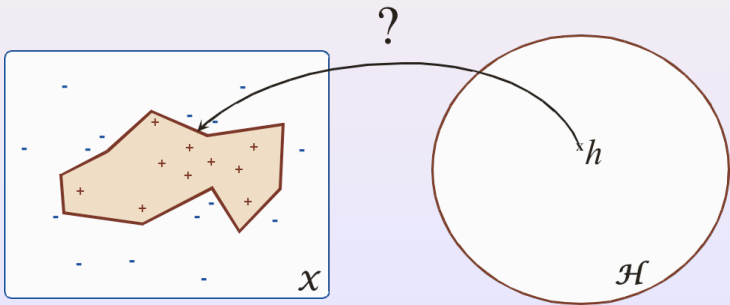
- 1 Un point de vue sur l'apprentissage
 - Etat de l'art et limites
 - Des applications
 - Gain d'information et transition de phase
- 2 Pour une science de la dynamique de l'apprentissage
 - Le cadre i.i.d.
 - Les effets de séquences
 - Contributions
- 3 Bilan et perspectives
 - Bilan
 - Perspectives

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

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

Le paradigme

Apprentissage = Problème inverse mal posé

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

Proximité : Espérance de risque

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

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

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

La théorie statistique de l'apprentissage : actualités

Théorie

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

[Mar05]

Jérémie Mary

Étude de l'apprentissage actif. Applications à la conduite d'expériences

Thèse, LRI, Orsay, Déc. 2005.

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

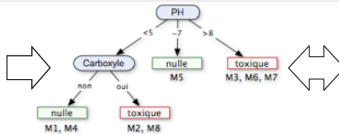
- Minimisation du ϕ -risque empirique
 - Méthodes de votes (**bagging**, **boosting**, ...)
 - Méthodes à noyaux (**SVM**, ...)

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)

Des problèmes difficiles

Évolution (dépendances) dans le temps

«Nouveaux » apprentissages

- **Dérive** de la dépendance cible
- **agents autonomes**
- par «démonstration »/ guidé par **un professeur**
- à **long terme** (*long-life learning*)
 - Articulation
 - Transfert

Applications

Projet INDANA (2001-2004)

Prédiction du risque cardio-vasculaire

- Divers types de bruit : description et classes
- Une classe « sûre », l'autre mal déterminée
- Déséquilibre entre classes
- Classification et régression

Contributions

- 1 Application de EM (apprentissage non supervisé)
- 2 Extension de EM à des étiquettes réelles

[Col02]

I. Colombet.

Aspects méthodologiques de la prédiction du risque cardiovasculaire : apports de l'apprentissage automatique.

Thèse de doctorat, Paris-5, Juin 2002.

Applications

FISICA (2001-2003)

Reconnaissances de scènes naturelles

- Données en très grandes dimensions ($> 10^4$)
- Recherche de bons descripteurs

Contributions

- 1 Nouveau système de **codage clairsemé**
- 2 à partir de **motifs fréquents**

[JCSTL03] S. Jouteau, A. Cornuéjols, M. Sebag, P. Tarroux and J.-S. Liénard.
Nouveaux résultats en classification à l'aide d'un codage par motifs fréquents.
Revue d'Intelligence Artificielle (Proc. of the EGC-03 Conf.), vol. 17, No.1-3, 521-532, 2003.

[ASM04] A. Cornuéjols, M. Sebag and J. Mary.
Classification d'images à l'aide d'un codage par motifs fréquents.
RFIA-04. (Workshop sur la fouille d'images), Toulouse, France, 2004.

Applications

Analyse du transcriptome (2002- ...)

- Données en très grandes dimensions ($> 10^3$)
- Très peu de données
- Très bruitées

Contributions

- 1 Développement d'une méthode d'évaluation d'attributs
- 2 Combinaison de méthodes

[NAR-04] G. Mercier, N. Berthault, J. Mary, A. Antoniadis, J-P. Comet, A. Cornuéjols, Ch. Froidevaux and M. Dutreix.

Comparing and combining feature estimation methods for the analysis of microarray data.
Nucleic Acids Research (NAR), vol.32, No.1, 1-8 (2004).

[CFM05] A. Cornuéjols, Ch. Froidevaux and J. Mary.

Biological detection of low radiation by combining results of two analysis methods.
JOBIM-05 (Poster), Lyon, France, 2005.

[Book-chapter,03], [CTDDR'04], [Book-chapter,04], [PKDD'04]

Applications

ACCAMBA (2004-2007)

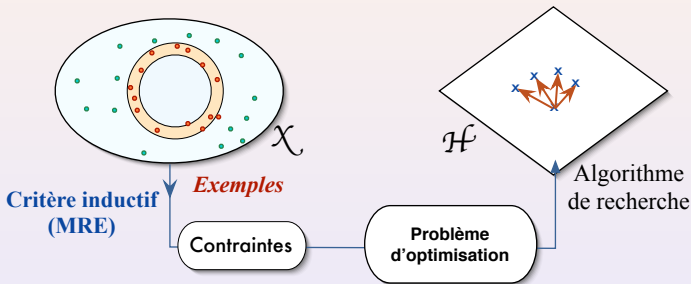
Prédiction de bio-activité de molécules

- Données structurées
- Recherche de bons descripteurs

[MRL05] Maréchal E., Roy S., Lafanechère L.(éditeurs).
Le criblage pharmacologique automatisé : du haut-débit au haut-contenu d'information. Une introduction à l'usage des biologistes, des chimistes et des informaticiens.
Ouvrage collectif, en cours d'édition (PUG).

Sous quelles conditions l'induction est-elle possible ?

Des conditions sur le gain d'information

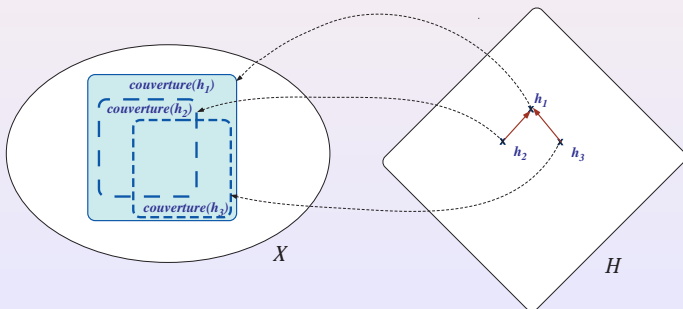


Les entrées doivent **se traduire en « différences »** sur les hypothèses

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 ?

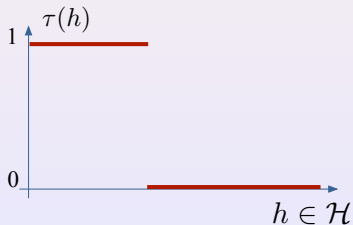
Definition (Taux de couverture)

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

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

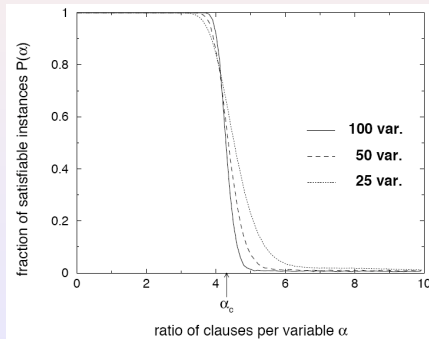
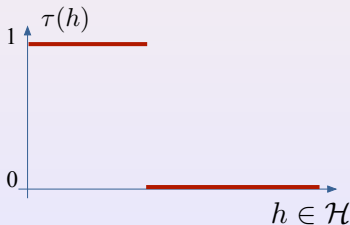
Des conditions sur le gain d'information

Un cas limite ... mais ...



Des conditions sur le gain d'information

Un cas limite ... mais ...

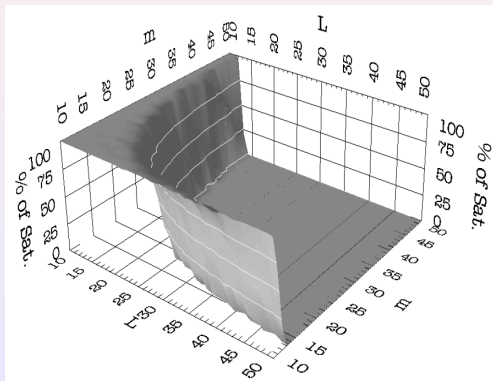


Des conditions sur le gain d'information

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

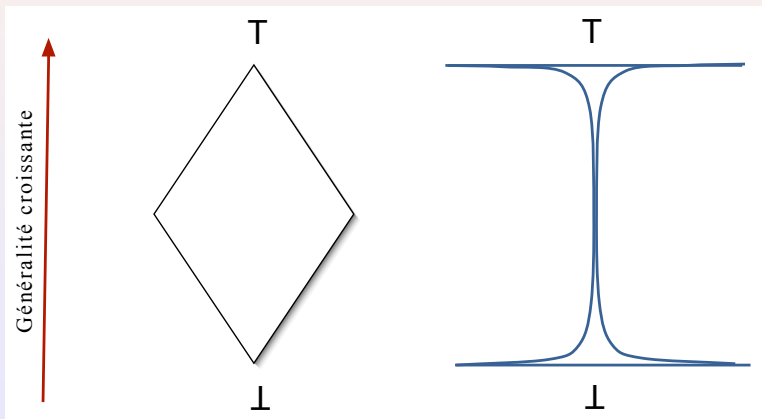
Paramètres de contrôle :

- n : nombre de *variables* dans l'hypothèse \underline{h} testée,
- m : nombre de *symboles de prédicats* dans \underline{h} ,
- L : nombre total de *constantes* dans l'exemple \underline{e} ,
- N : nombre de *littéraux* construits sur chaque symbole de prédicat dans \underline{e} .



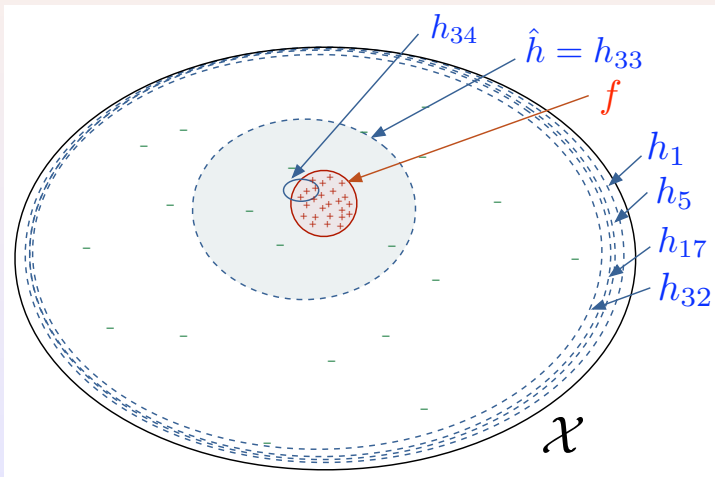
Des conditions sur le gain d'information

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 ?*

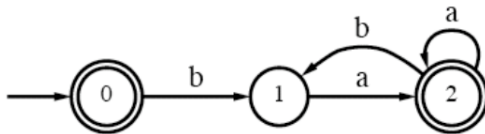
2. *Peut-on contourner le problème ?*

Inférence grammaticale : rappels

- **Entrée** : *chaînes* sur un alphabet 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.

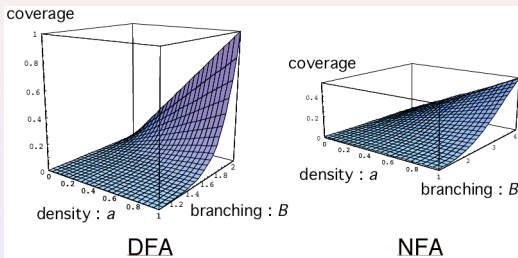
Rapport de DEA (LRI, Univ. Paris-Sud, Orsay), 2001.

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



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

[Pin01]

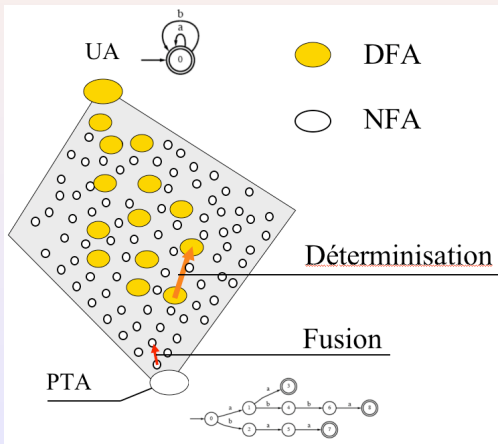
S. Pinto.

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

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

Protocole expérimental

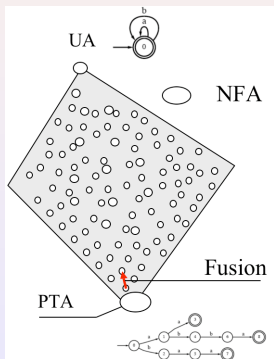
- 1 Génération aléatoire d'un **échantillon d'apprentissage** : $|S^+|$ (= 200) chaînes de taille ℓ
- 2 Construction du **PTA** pour chaque échantillon S^+
- 3 **Calcul de chemins de généralisation partant du PTA** :
 - **Fusions aléatoires**
 - **Couverture** calculée **pour chaque automate engendré**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

Expériences

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

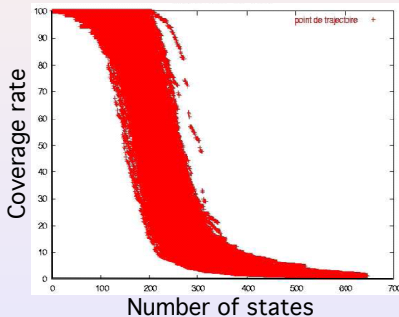
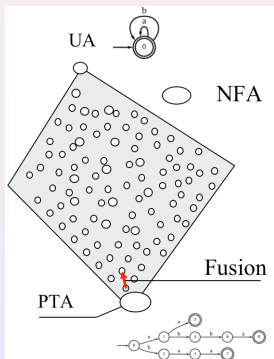
Gain d'information en inférence grammaticale

Cas non-déterministe : **NFA**



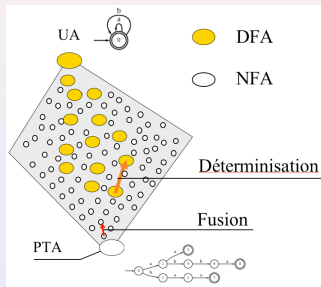
Gain d'information en inférence grammaticale

Cas non-déterministe : NFA



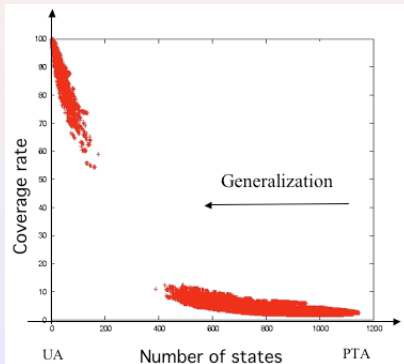
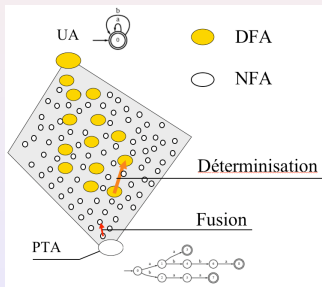
Gain d'information en inférence grammaticale

Cas déterministe : **DFA**



Gain d'information en inférence grammaticale

Cas déterministe : **DFA**



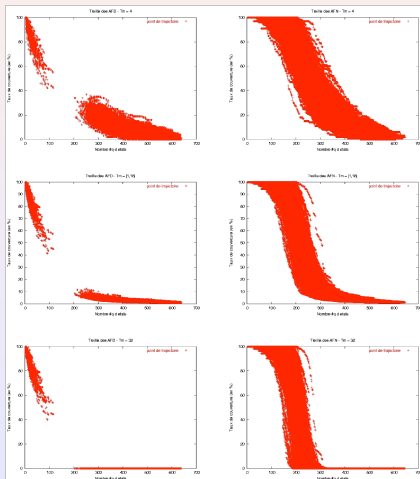
$$|\Sigma| = 8; \ell = 8$$

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

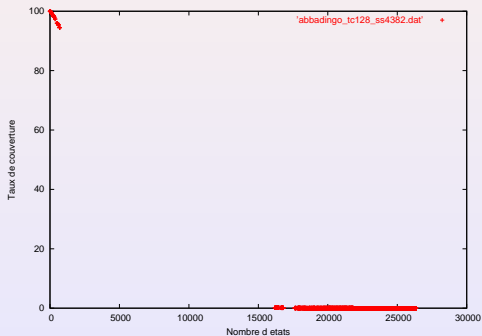
Le défi **Abbadingo**

$$|\Sigma| = 2$$

$$\ell = 17$$

$$|S^+| = 4382$$

Test sur 1000 chaînes
de taille : 17



Gain d'information et induction

Conclusions pour l'inférence grammaticale

Bilan

- 1 Importance de l'étude des **variations du taux de couverture**
- 2 ... par rapport à l'**espace de recherche effectif**

- Permet d'expliquer des comportements
- Orienter les recherches

-
- [PCS05] N. Pernot, A. Cornuéjols and M. Sebag.
Phase transition within grammatical inference.
Int. Joint Conf. on Artificial Intelligence (IJCAI-05), Edinburgh, UK, 2005, (Ed. L. P. Kaelbling), pp.811-816.
- [CS05] A. Cornuéjols and M. Sebag.
Phase transition, induction and grammatical inference.
Second Franco-Japanese Workshop on Information Search, Integration and Personalization (ISIP-05), Lyon, France, 2005.

Gain d'information et induction

Conclusions pour l'inférence grammaticale

Perspectives

- Autres ***opérateurs ou heuristiques*** pour modifier le paysage apparent
- Possibilités d'***apprentissage guidé ou actif***
 - Modification de l'espace des exemples
 - longueur des chaînes (e.g. les longues avant les courtes !)
 - ignorer des lettres de l'alphabet (?)
 - dynamiquement

Gain d'information et induction

Conclusions générales

- Étude plus fine de l'induction
- prenant en compte le gain d'information ($\frac{\partial \tau_{\mathbf{s}}(h)}{\partial h}$)
- et l'espace explorable

Gain d'information et induction

Conclusions générales

- Étude plus fine de l'induction
- prenant en compte le gain d'information ($\frac{\partial \tau_{\mathbf{s}}(h)}{\partial h}$)
- et l'espace explorable

Mais reste dans le cadre i.i.d. !!

Plan

- 1 Un point de vue sur l'apprentissage
 - Etat de l'art et limites
 - Des applications
 - Gain d'information et transition de phase
- 2 Pour une science de la dynamique de l'apprentissage
 - Le cadre i.i.d.
 - Les effets de séquences
 - Contributions
- 3 Bilan et perspectives
 - Bilan
 - Perspectives

Le cadre i.i.d.

Definition (Le cadre i.i.d.)

Pour faire une prédiction sur la prochaine entrée, il suffit de connaître la distribution génératrice sous-jacente.

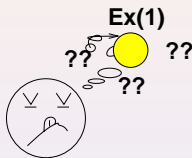
Corollaire

La connaissance d'autres entrées n'apporte aucune information supplémentaire.

⇒ *Efface la notion d'histoire*

Les effets de séquences

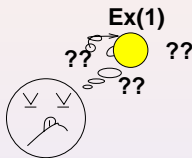
1er exemple (du à Laurent Chaudron [HDR,2005])



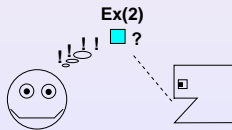
Exercice(1) Soit $\left(\frac{p_n}{q_n}\right)_{n \in \mathbb{N}}$
une séquence de nombres
rationnels convergeant vers
 x irrationnel. Prouver que
 (p_n) et (q_n) convergent tous
les deux vers l'infini.

Les effets de séquences

1er exemple (du à Laurent Chaudron [HDR,2005])



Exercice(1) Soit $\left(\frac{p_n}{q_n}\right)_{n \in \mathbb{N}}$ une séquence de nombres rationnels convergeant vers x irrationnel. Prouver que (p_n) et (q_n) convergent tous les deux vers l'infini.



Exercice(2) Prouver que l'image de n'importe quelle séquence finie de nombres naturels est un ensemble fini.

Les effets de séquences

2ème exemple

[Sur 24 étudiants de DEA, 1996]

a b c



a b d



a a b a b c



?

- Long et difficile
- Grande variété de réponses

Les effets de séquences

2ème exemple

[Sur 24 étudiants de DEA, 1996]

a b c



a b d

a a b a b c



- Long et difficile
- Grande variété de réponses

a b c



a b d

a a b a b c



- Beaucoup plus rapide
- Spectre de réponses beaucoup plus serré

i j j k k k



Les effets de séquences

2ème exemple

[Sur 24 étudiants de DEA, 1996]

a b c



a b d

a a b a b c



- Long et difficile
- Grande variété de réponses

1 2 3



1 2 4

a b c



a a b a b c



- Chemin rouge : plus difficile et réponses plus confuses

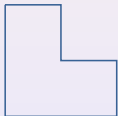
Les effets de séquences

3ème exemple

[Sur quelques étudiants de Polytechnique, 1994]

Consigne : découper la figure suivante en n parties superposables.

En **2** :



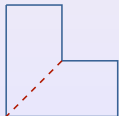
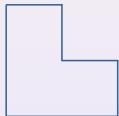
Les effets de séquences

3ème exemple

[Sur quelques étudiants de Polytechnique, 1994]

Consigne : découper la figure suivante en n parties superposables.

En **2** :



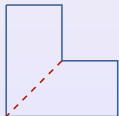
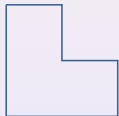
Les effets de séquences

3ème exemple

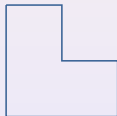
[Sur quelques étudiants de Polytechnique, 1994]

Consigne : découper la figure suivante en n parties superposables.

En **2** :



En **3** :



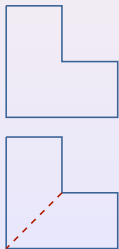
Les effets de séquences

3ème exemple

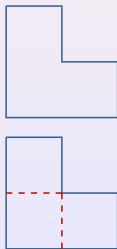
[Sur quelques étudiants de Polytechnique, 1994]

Consigne : découper la figure suivante en n parties superposables.

En **2** :



En **3** :



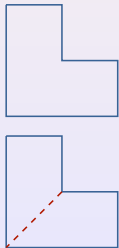
Les effets de séquences

3ème exemple

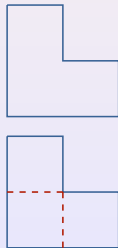
[Sur quelques étudiants de Polytechnique, 1994]

Consigne : découper la figure suivante en n parties superposables.

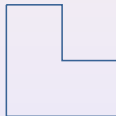
En **2** :



En **3** :



En **4** :



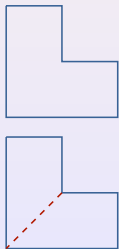
Les effets de séquences

3ème exemple

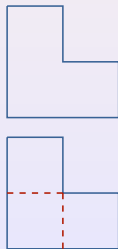
[Sur quelques étudiants de Polytechnique, 1994]

Consigne : découper la figure suivante en n parties superposables.

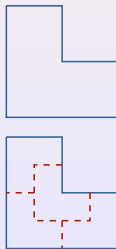
En **2** :



En **3** :



En **4** :



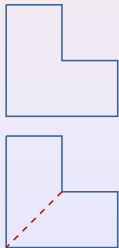
Les effets de séquences

3ème exemple

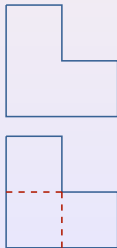
[Sur quelques étudiants de Polytechnique, 1994]

Consigne : découper la figure suivante en n parties superposables.

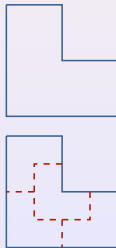
En **2** :



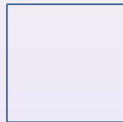
En **3** :



En **4** :



En **5** :



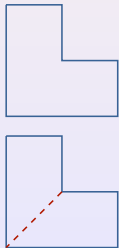
Les effets de séquences

3ème exemple

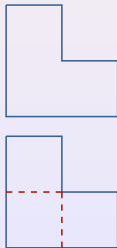
[Sur quelques étudiants de Polytechnique, 1994]

Consigne : découper la figure suivante en n parties superposables.

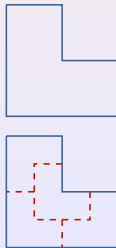
En **2** :



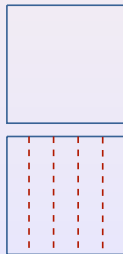
En **3** :



En **4** :



En **5** :



Effets de séquences : constatations

Dépendance sur l'histoire

- De la vitesse de résolution
- Du résultat

Effets de séquences : constatations

Dépendance sur l'histoire

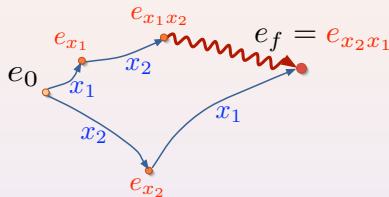
- De la vitesse de résolution
- Du résultat

Courbure de l'espace des états

→ ce n'est plus un espace euclidien

Mesure de courbure

Le « **crochet de Lie** »
 $[x_1, x_2]_{\text{état}}$



Remarque

Le **crochet de Lie** (la courbure) **s'annule si** :

- 1 Ressources de **calcul** suffisantes
- 2 Ressources **mémoire** suffisantes

Hors-ligne vs. **en-ligne**

Pour sortir du cadre i.i.d. ...

Il faut **aborder de front** *les effets de séquences*

- 1 Quelles sont les **entrées utiles** (les plus utiles) ?
- 2 Quelles sont les entrées nuisibles ?

Nouvelles questions

- 1 *Ordres de présentation les plus favorables ?*
- 2 Quels sont les *systèmes sensibles aux effets de séquence* ?

Pour sortir du cadre i.i.d. ...

... Il faut de nouveaux outils

Outils nécessaires

- ① Une **métrique** (entre programmes)
 - pour mesurer la *distance entre états*
 - pour mesurer la *corrélacion entre des entrées*
- ② Une **mesure de courbure** de l'espace

Pour sortir du cadre i.i.d. ...

... Il faut de nouveaux outils

Outils nécessaires

- ① Une **métrique** (entre programmes)
 - pour mesurer la *distance entre états*
 - pour mesurer la *corrélacion entre des entrées*
- ② Une **mesure de courbure** de l'espace

Outils existants

- Entropie relative / information mutuelle / complexité algorithmique

Pour sortir du cadre i.i.d. ...

... Il faut de nouveaux outils

Outils nécessaires

- 1 Une **métrique** (entre programmes)
 - pour mesurer la *distance entre états*
 - pour mesurer la *corrélation entre des entrées*
- 2 Une **mesure de courbure** de l'espace

Outils existants

- Entropie relative / information mutuelle / complexité algorithmique

Limites

- 1 Ne permettent pas de rendre compte de corrélations négatives
- 2 Inadaptés à espaces courbes

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

Le *problème* : **évaluation (tri) d'attributs**
(analyse du transcriptome)

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

Le *problème* : **évaluation (tri) d'attributs**
(analyse du transcriptome)

Inférence très précaire

- 1 Beaucoup plus d'attributs que de dimensions
- 2 Nombreuses sources de « bruit »

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

Le *problème* : **évaluation (tri) d'attributs**
(analyse du transcriptome)

Inférence très précaire

- 1 Beaucoup plus d'attributs que de dimensions
- 2 Nombreuses sources de « bruit »

Comment évaluer le résultat ?

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

Le *problème* : **évaluation (tri) d'attributs**
(analyse du transcriptome)

Inférence très précaire

- 1 Beaucoup plus d'attributs que de dimensions
- 2 Nombreuses sources de « bruit »

Comment évaluer le résultat ?

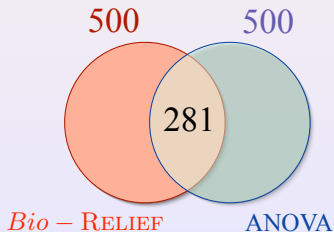
Deux méthodes (non supervisées)

valent-elles mieux qu'une ?

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

Illustration

- 1 6135 gènes ; 18 exemples (6+, 12-)
- 2 Deux méthodes d'évaluation : ANOVA et *Bio*-RELIEF

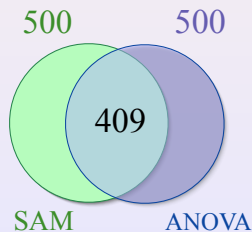
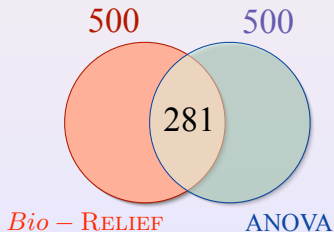


Comment juger ces 281 gènes en commun ?

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

Illustration

- 1 6135 gènes ; 18 exemples (6+, 12-)
- 2 Deux méthodes d'évaluation : ANOVA et *Bio*-RELIEF



Comment juger ces 281 gènes en commun ?

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

L'intersection est due :

- 1 au **hasard** (k) : $H(d, n, k) = \frac{\binom{n}{k} \cdot \binom{d-n}{n-k}}{\binom{d}{n}}$
- 2 à la **corrélation des méthodes *a priori***
- 3 aux **régularités dans les données**

Sorte d'hypothèse nulle :

$$\langle M_1, M_2 \rangle = E_D(|\text{top}_n(M_1) \cap \text{top}_n(M_2)|)$$

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

$$\langle M_1, M_2 \rangle = E_{\mathcal{D}}(|\text{top}_n(M_1) \cap \text{top}_n(M_2)|)$$

0 :?

40 :?

281 :?

500 :?

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

$$\langle M_1, M_2 \rangle = E_{\mathcal{D}}(|\text{top}_n(M_1) \cap \text{top}_n(M_2)|)$$

0 :?

Anticorrélés

40 :?

281 :?

500 :?

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

$$\langle M_1, M_2 \rangle = E_{\mathcal{D}}(|\text{top}_n(M_1) \cap \text{top}_n(M_2)|)$$

0 :? Anticorrélés

40 :? Décorrélés

281 :?

500 :?

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

$$\langle M_1, M_2 \rangle = E_{\mathcal{D}}(|\text{top}_n(M_1) \cap \text{top}_n(M_2)|)$$

0 :?

Anticorrélés

40 :?

Décorrélés

281 :?

Pas de sur-représentation des régularités

500 :?

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

$$\langle M_1, M_2 \rangle = E_{\mathcal{D}}(|\text{top}_n(M_1) \cap \text{top}_n(M_2)|)$$

0	:?	Anticorrélés
40	:?	Décorrélés
281	:?	Pas de sur-représentation des régularités
500	:?	Méthodes totale ment corrélées

Une nouvelle mesure de corrélation entre programmes qui peut prendre une valeur négative

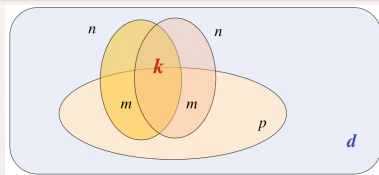
$$\langle M_1, M_2 \rangle = E_{\mathcal{D}}(|\text{top}_n(M_1) \cap \text{top}_n(M_2)|)$$

0	:?	Anticorrélés
40	:?	Décorrélés
281	:?	Pas de sur-représentation des régularités
500	:?	Méthodes totale ment corrélées

Ici : **170** \pm 40

Une nouvelle mesure de corrélation entre programmes

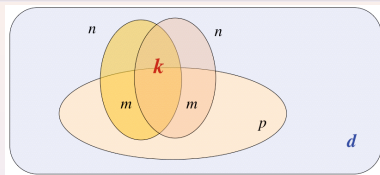
Application



$$p(n = k | d, p, n, m, \mu_{\mathcal{H}_0}) = \frac{\binom{p}{m} \binom{d-p}{n-m} \sum_{k^+=2m-p}^m \binom{m}{k^+} \binom{p-m}{m-k^+} \binom{n-m}{k-k^+} \binom{d-n-(p-m)}{n-m-(k-k^+)}}{\binom{d}{n} \cdot \binom{d}{n}} / C(\mu_{\mathcal{H}_0})$$

Une nouvelle mesure de corrélation entre programmes

Application



$$p(n = k | d, p, n, m, \mu_{\mathcal{H}_0}) = \frac{\binom{p}{m} \binom{d-p}{n-m} \sum_{k^+=2m-p}^m \binom{m}{k^+} \binom{p-m}{m-k^+} \binom{n-m}{k-k^+} \binom{d-n-(p-m)}{n-m-(k-k^+)}}{\binom{d}{n} \cdot \binom{d}{n}} / C(\mu_{\mathcal{H}_0})$$

Résultats

- $p = 420 \pm 20$
- $m = 340 \pm 20$
- ≈ 265 des 281 sont pertinents ! (*précision* = 0.94)

Une nouvelle mesure de corrélation entre programmes

Portée

- Permet de rendre compte de **corrélations négatives**
- Applicable aussi à des **algorithmes d'apprentissage supervisé**

[CFM05]

A. Cornuéjols, Ch. Froidevaux and J. Mary.

Comparing and combining feature estimation methods for the analysis of microarray data.

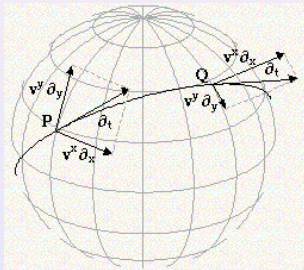
JOBIM-05 : Journées Ouvertes Biologie Informatique Mathématiques (poster), Lyon, France, 2005.

Changement de référentiel

Espace courbe

→ Pour comparer deux états du système
en deux situations-problèmes différents

→ Notion de **transport parallèle**



Propriétés des trajectoires d'apprentissage

Caractéristiques d'une trajectoire d'apprentissage fonction de :

- Propriétés de l'apprenant
- Caractéristiques de la séquence d'entrées

Quelle séquence idéale pour passer d'un état à un autre ?

Propriétés des trajectoires d'apprentissage

Systèmes insensibles à l'ordre



[Cor93b]

A. Cornuéjols.

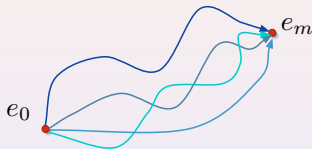
Getting Order Independence in Incremental Learning.

European Conference on Machine Learning (ECML-93), (Ed. P. Brazdil), Vienna, Austria (1993).

Springer-Verlag, LNAI-667, pp. 196-212.

Propriétés des trajectoires d'apprentissage

Systèmes insensibles à l'ordre



Espace mémoire :

- Calcul du **Max** :
- Calcul de la **Moyenne** :
- **ID5R** :
- **Espace des versions** :

[Cor93b]

A. Cornuéjols.

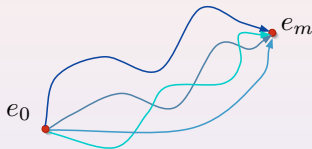
Getting Order Independence in Incremental Learning.

European Conference on Machine Learning (ECML-93), (Ed. P. Brazdil), Vienna, Austria (1993).

Springer-Verlag, LNAI-667, pp. 196-212.

Propriétés des trajectoires d'apprentissage

Systèmes insensibles à l'ordre



Espace mémoire :

- Calcul du **Max** : $\max(t)$
- Calcul de la **Moyenne** :
- **ID5R** :
- **Espace des versions** :

[Cor93b]

A. Cornuéjols.

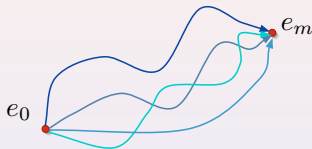
Getting Order Independence in Incremental Learning.

European Conference on Machine Learning (ECML-93), (Ed. P. Brazdil), Vienna, Austria (1993).

Springer-Verlag, LNAI-667, pp. 196-212.

Propriétés des trajectoires d'apprentissage

Systèmes insensibles à l'ordre



Espace mémoire :

- Calcul du **Max** : $\max(t)$
- Calcul de la **Moyenne** : $\text{moy}(t)$ et t
- **ID5R** :
- **Espace des versions** :

[Cor93b]

A. Cornuéjols.

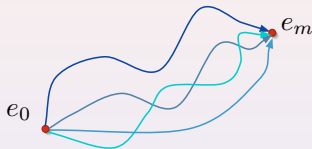
Getting Order Independence in Incremental Learning.

European Conference on Machine Learning (ECML-93), (Ed. P. Brazdil), Vienna, Austria (1993).

Springer-Verlag, LNAI-667, pp. 196-212.

Propriétés des trajectoires d'apprentissage

Systèmes insensibles à l'ordre



Espace mémoire :

- Calcul du **Max** : $\max(t)$
- Calcul de la **Moyenne** : $\text{moy}(t)$ et t
- **ID5R** : $\mathcal{O}(t)$
- **Espace des versions** :

[Cor93b]

A. Cornuéjols.

Getting Order Independence in Incremental Learning.

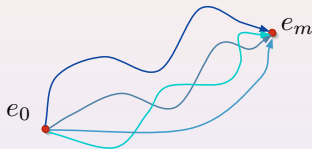
European Conference on Machine Learning (ECML-93), (Ed. P. Brazdil), Vienna, Austria (1993).

Springer-Verlag, LNAI-667, pp. 196-212.

Propriétés des trajectoires d'apprentissage

Systèmes insensibles à l'ordre

Espace mémoire :



- Calcul du **Max** : $\max(t)$
- Calcul de la **Moyenne** : $\text{moy}(t)$ et t
- **ID5R** : $\mathcal{O}(t)$
- **Espace des versions** : $\leq \mathcal{O}(t)$

[Cor93b]

A. Cornuéjols.

Getting Order Independence in Incremental Learning.

European Conference on Machine Learning (ECML-93), (Ed. P. Brazdil), Vienna, Austria (1993).

Springer-Verlag, LNAI-667, pp. 196-212.

Propriétés des trajectoires d'apprentissage

Systèmes insensibles à l'ordre

Espace mémoire :



- Calcul du **Max** : $\max(t)$
- Calcul de la **Moyenne** : $\text{moy}(t)$ et t
- **ID5R** : $\mathcal{O}(t)$
- **Espace des versions** : $\leq \mathcal{O}(t)$

Oubli \equiv exemples supplémentaires !

[Cor93b]

A. Cornuéjols.

Getting Order Independence in Incremental Learning.

European Conference on Machine Learning (ECML-93), (Ed. P. Brazdil), Vienna, Austria (1993).

Springer-Verlag, LNAI-667, pp. 196-212.

Propriétés des trajectoires d'apprentissage

Trajectoires et systèmes dynamiques

Trajectoire \iff **Lagrangien** (moindre action)

Application

Symétrie / permutation des entrées \iff ***courant associé***

Théorème de Noether

[Cor93a] **A. Cornuéjols**
Training Issues in Incremental Learning.
AAAI Press, 1993.

[Cor06] **A. Cornuéjols**
Machine Learning : The Necessity of Order.
In Order to Learn : How ordering processes and sequencing effects in machines illuminate human learning and vice-versa, E. Lehtinen and F. Richter (Eds.), Cambridge University press, 2006.

Plan

- 1 Un point de vue sur l'apprentissage
 - Etat de l'art et limites
 - Des applications
 - Gain d'information et transition de phase
- 2 Pour une science de la dynamique de l'apprentissage
 - Le cadre i.i.d.
 - Les effets de séquences
 - Contributions
- 3 Bilan et perspectives
 - Bilan
 - Perspectives

Dans le cadre actuel : approche statistique

Encore d'immenses **problèmes pratiques** :

- Apprentissage semi-supervisé
- Types de bruit
- Grandes dimensions / nb exemples
- Déséquilibre des classes
→ **apprentissage actif**
- Très grandes bases de données
- Données semi-structurées

Conceptuellement, loin d'être fini :

*Conditions de gain
d'information ?*

Possibilité de **modification
dynamique de** $\mathcal{L}_{\mathcal{H}}$ et de $\mathcal{L}_{\mathcal{X}}$

Données non i.i.d. : vers une dynamique de l'apprentissage

Inévitabilité des effets de séquence

*Étude de l'apprentissage
comme un système
dynamique*

Développer des outils pour des **espaces de programmes**

- Produit scalaire
- Changement de référentiel
- Lagrangien

Propriétés déterminantes :

- **capacités de calcul et de mémoire limitées**
- ... la structure de la connaissance

Données non i.i.d. : vers une dynamique de l'apprentissage

Mieux comprendre la nature de l'information
et de sa circulation



*En bas, quiétude
immobile. La clé est
peut-être sous le
réverbère.*

En haut, le ciel.

*Espace, transparence
et mystère. Nuages et
turbulences.*

L'empire des lumières.
Magritte, 1954

Plan

4

Annexe

- Systèmes d'inférence grammaticale
- Changement de référentiel et analogie
- Changement de référentiel et effet tunnel

Table des Appendices

4

Annexe

- Systèmes d'inférence grammaticale
- Changement de référentiel et analogie
- Changement de référentiel et effet tunnel

Gain d'information en inférence grammaticale

Algorithmes avec stratégies de recherche

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

RPNI [OG92][Lan92]

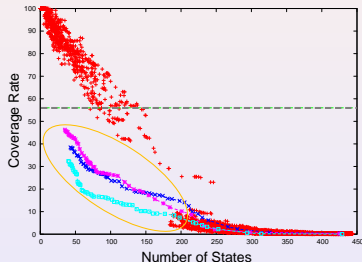
- Choix des états *en largeur d'abord*

EDSM [Lan et al.98][Lan98]

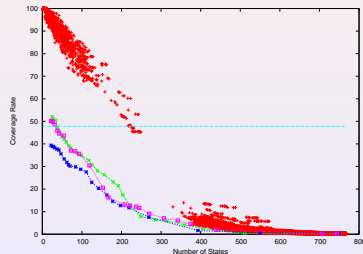
- Choix des états *dont la fusion conduit au maximum de fusions pour détermination*

Gain d'information en inférence grammaticale

Algorithmes avec stratégies de recherche : **résultats**



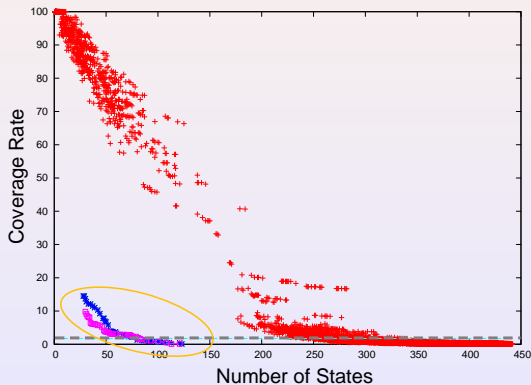
RPNI



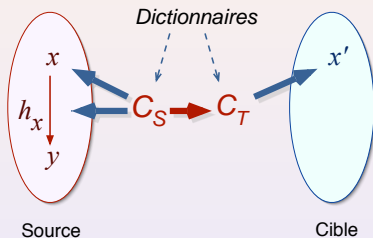
EDSM

Gain d'information en inférence grammaticale

Algorithmes avec stratégies de recherche : **résultats**

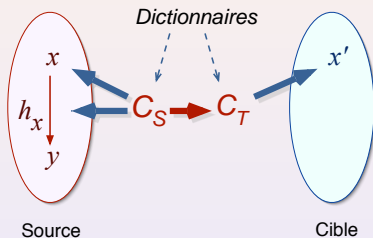


Changement de référentiel et analogie



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

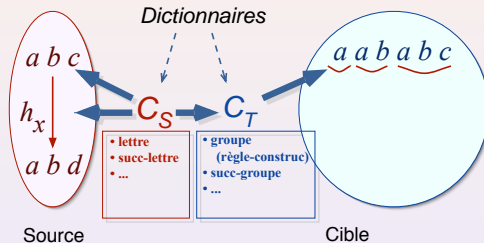
Changement de référentiel et analogie



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

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), Sètes, France, 1996, pp.233-247.

A. Cornuéjols and J. Ales-Bianchetti.

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

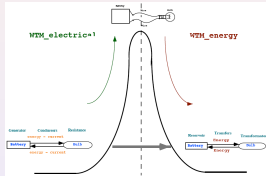
Workshop Advances in Analogy research : Integration of theory and data from cognitive, computational and neural sciences, Sofia, Bulgaria, 1998. New Bulgarian University Series (Eds. K. Holyoak, D. Gentner and B. Kokinov), pp. 365-372.

[ECML-94 (wkp)], [COLT-94 (wkp)], [Dagstuhl,94], [Book-chapter,96]

Changement de référentiel

Transferts entre domaines conceptuels

L'effet tunnel cognitif



Apprentissage du concept d'énergie chez des lycéens

[CTC00] A. Cornuéjols, A. Tiberghien 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]