

In memoriam JLL

**Apprentissage supervisé et satisfaction de
contraintes**

Michèle Sebag

TAO

CNRS – INRIA – LRI

<http://tao.lri.fr>

Résumé : Apprendre à changer de cadre

- Le message c'est le media

Marshall McLuhan

- Le coût du raisonnement fait partie du raisonnement

Herbert Simon

- Le problème c'est la représentation

JLL

Un conseil pratique

“Saturez votre esprit du problème ; si ça ne suffit pas allez au cinéma”.

Plan

- Apprentissage à partir d'exemples
- Exemple = contrainte
- Changement de représentation
- Vue d'ensemble : la transition de phase
- Apprentissage : programmation par les exemples

Apprentissage supervisé

Input

- Données brutes : signaux, données expérimentales, patients ...
- Labels associés : signaux sismiques/non, pannes, diagnostic,...

	<i>Att</i> ₁	<i>Att</i> ₂	<i>Att</i> ₃	<i>Att</i> ₄	<i>Att</i> ₅	Label
<i>x</i> ₁	Oui	3.2	?	plat	?	+
<i>x</i> ₂	?	-7.3	-4.85	denté	Oui	-
<i>x</i> ₃	Non	2.1	7	?	?	-

Base d'apprentissage

$$\mathcal{E} = \{(x_i, y_i), x_i \in X, y_i \in Y\}$$

Apprentissage supervisé, 2

$$\mathcal{E} = \{(x_i, y_i), x_i \in X, y_i \in Y\}$$

Output : Hypothèses

$$h : X \mapsto Y$$

Critères

- Qualité prédictive

ℓ = fonction de perte

$$h^* = \mathit{Argmin}\{E[\ell(h(x), y)], h \in H\}$$

- Compréhensibilité

h est intelligible, l'expert peut certifier et interagir

Apprentissage et satisfaction de contraintes

Espace de recherche

- Langage des exemples
- Ensemble d'exemples
- Langage d'hypothèses, de requêtes

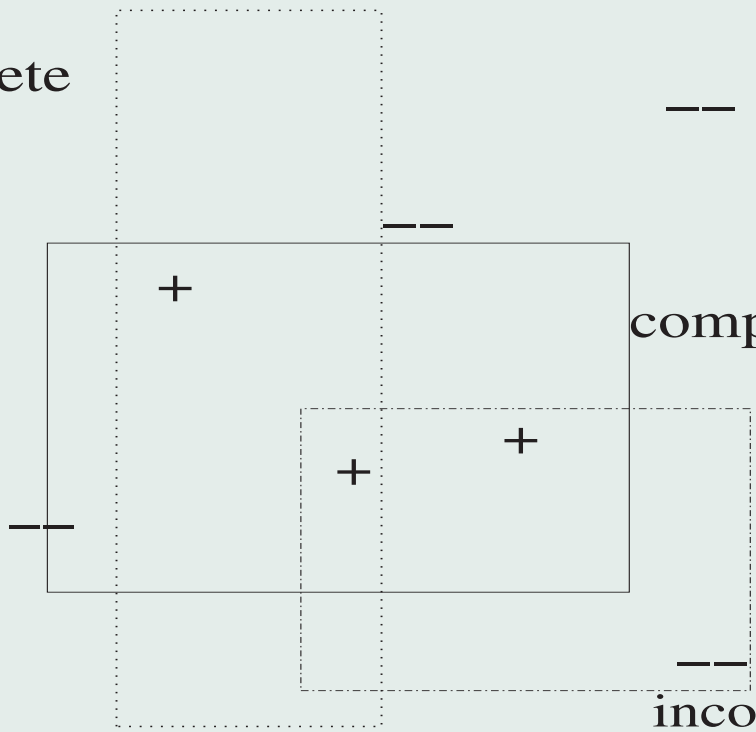
Relation d'ordre partiel \prec

- inclusion des ensembles
- θ -subsumption / query containment
- implication logique

incomplete

complete

incorrecte



L'espace des Versions

Mitchell, 79, 82

Etant donné

- X espace d'exemples
- Exemples positifs, $\mathcal{E}^+ = \{x_1^+, \dots, x_p^+\} \subseteq X$
- Exemples négatifs $\mathcal{E}^- = \{x_1^-, \dots, x_n^-\} \subseteq X$
- H espace d'hypothèses
- \prec relation de généralité sur H
- “Single representation trick” : $X \subseteq H$

Trouver

VS : ensemble des hypothèses h de H qui sont

Complètes
Correctes

couvrent les exemples positifs
ne couvrent pas les exemples négatifs

$$VS(\mathcal{E}^+, \mathcal{E}^-) = \{h \in H / h \text{ complète et correcte}\}$$

Propriétés

h complète, $h \prec h' \implies h'$ complète

h correcte, $h' \prec h \implies h'$ correcte

Définition

S : ensemble des hypothèses complètes les plus spécifiques

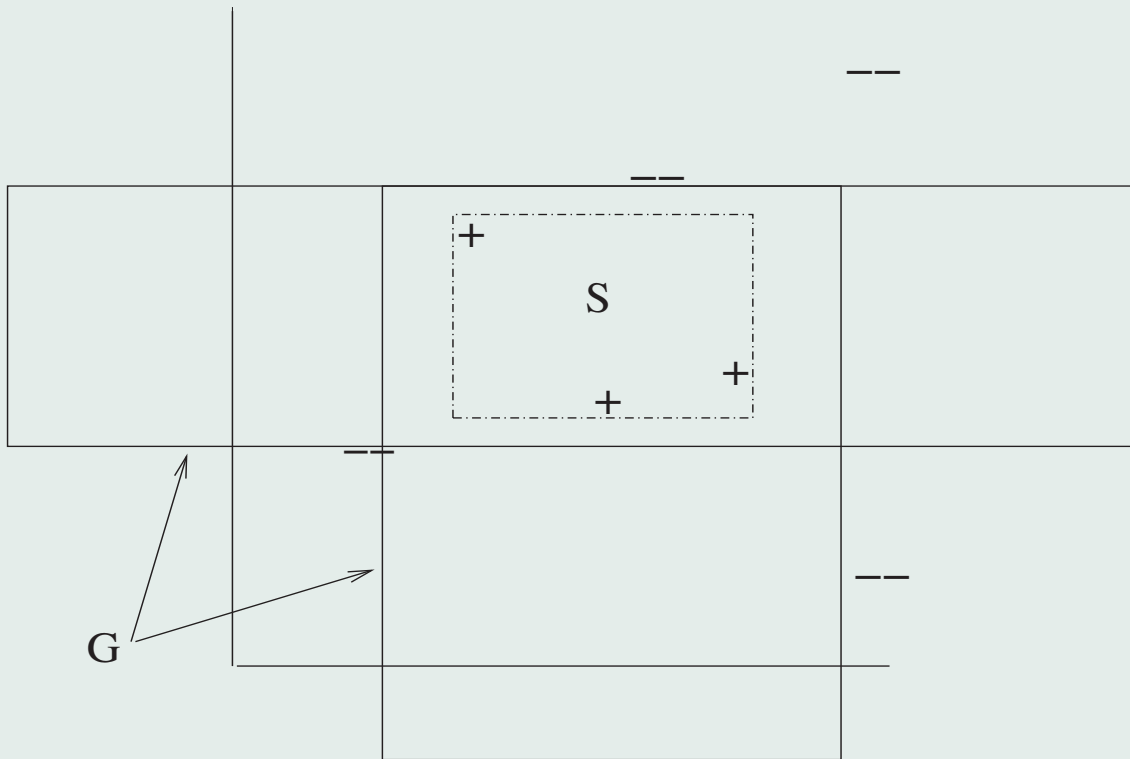
G : ensemble des hypothèses correctes les plus générales

Caractérisation

$$VS = [S, G]$$

Diagnostic ($e \in X$):

$$VS(e) = \begin{cases} + & \text{si il existe } s \in S \text{ t.q. } e \prec s \\ - & \text{si quelque soit } g \in G, e \not\prec g \\ ? & \text{sinon} \end{cases}$$



Quelques retombées

Elagage des exemples

Winston, 1975 : ... *near misses are the really important learning samples. (. ..) Small differences permit the machine to localize some part of its current opinion about a concept for improvement. If one wants the machine to learn that the uprights of an arch cannot marry, one should show it a scene that fails to be an arch only in this respect.*

...Complexité exponentielle...

Données

$$X = \{0, 1\}^{2m}$$

	att_1	att_2	att_3	att_4	att_{2m-1}	att_{2m}
x^+	1	1	1	1	1	1
x_1^-	0	0	1	1	1	1
x_2^-	1	1	0	0	1	1
...								
x_m^-	1	1	1	1	0	0

Espace des Versions

$$S = \{x^+\}$$

$$G = \{att_{\sigma(1)} \wedge \dots \wedge att_{\sigma(m)}, \sigma(i) = 2i \text{ ou } 2i - 1\}$$

$$|G| = 2^m$$

Haussler 89

fragmentation exponentielle de G

...Esprit de JLL...

Intuition

Disjonction de Conjonctions : Exponentiel

Conjonction de Disjonctions : Linéaire

Espace des Versions Disjonctif

Brique de base

$$D(x, x') = \{h \in H / x \prec h, x' \not\prec h\}$$

	<i>Att</i> ₁	<i>Att</i> ₂	<i>Att</i> ₃	<i>Att</i> ₄	<i>Att</i> ₅	Label
<i>x</i>	Oui	3.2	?	plat	?	+
<i>x'</i>	?	-7.3	-4.85	denté	Oui	-

$$D(x, x') = [Att_2 > -7.3] \vee [Att_5 = plat]$$

Proposition

$$CE(x) = \{x' / label(x') \neq label(x)\}$$

$$VS(x, CE(x)) = \bigwedge_{x' \in CE(x)} D(x, x')$$

$$\text{Complexité } |\mathcal{E}^+| \times |\mathcal{E}^-| \times \dim(X)$$

Diagnostic de l'Espace des Versions Disjonctif

Diagnostic ($x \in X$)

- Vote à la majorité des $EV(x')$ tq $x < VS(x', CE(x'))$
bagging, robustesse face au bruit et aux concepts disjonctifs
- $x < VS(x', CE(x'))$ si $x < D(x', x'')$
pour presque tout $x'' \in CE(x')$
robustesse face au bruit des exemples
- $x < D(x', x'')$ si x vérifie M littéraux de $D(x', x'')$
robustesse face à la dispersion des exemples

Plan

- Apprentissage à partir d'exemples
- Exemple = contrainte
- **Changement de représentation**
- Vue d'ensemble : la transition de phase
- Apprentissage : programmation par les exemples

Changement de représentation

Principe

- Soit h_1, \dots, h_T un ensemble d'hypothèses.

Alors

$$\begin{aligned}\Phi : X &\mapsto \{0, 1\}^T \\ x &\mapsto (x \prec h_1, \dots, x \prec h_T)\end{aligned}$$

	h_1	h_2	h_3	h_4	h_5	Label
$\mathcal{E}' :$						
x	?	0	0	0	1	+
x'	1	?	0	1	1	-

Application : META-REGLES

- Apprendre à partir de \mathcal{E}' :

Si h_1 et non h_2 , Alors ..

- Même algorithme, règles plus complexes
- Apprentissage incrémental.

Du recul et des stats

CSP : Constatations

Les problèmes sont exponentiels (pire cas)

Mais en pratique...

Where are the really hard problems ?

Cheeseman 91

Parametres d'ordre

densité des contraintes

dureté des contraintes

Echantillonnage stochastique de CSPs

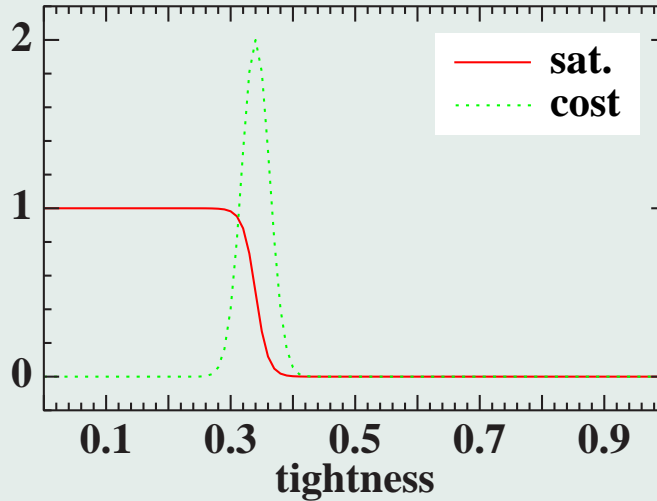
Région OUI
Région NON
Phase transition

Varier dureté à densité fixée

CSPs sous-contraints

CSPs sur-contraints

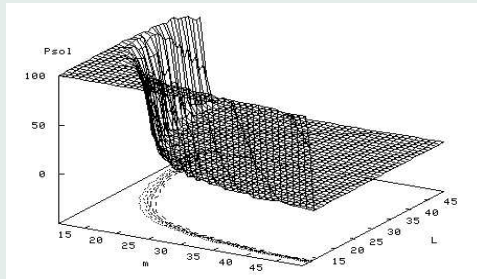
where the real hard pbs are



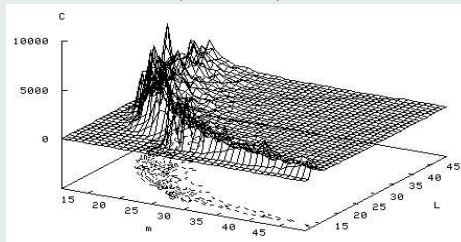
18/12/04, 06:27, michèle, fch: curve2 curve1

Transitions de phase

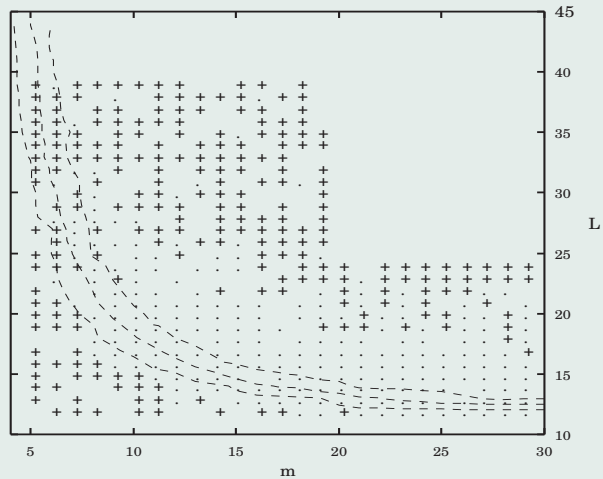
Probabilité de couverture (m, L)



Complexité effective (m, L)



Carte de Compétence : Apprentissage (FOIL)



+ Succès (> 80% sur le test)

· Echec

Programmation par Apprentissage

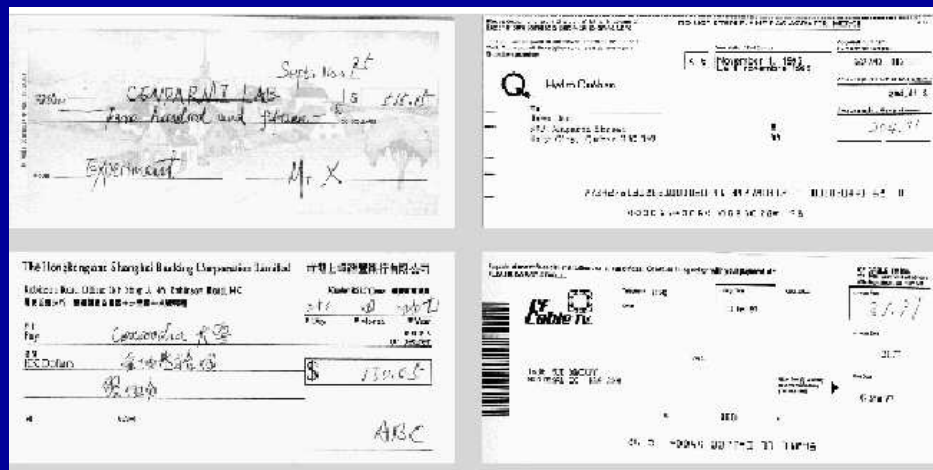
Trois exemples

Dietterich, Oregon State University

- Hand-written recognition
- Wafer test
- Camera allocation in autonomous robotics

Scenario 1: Reading Checks

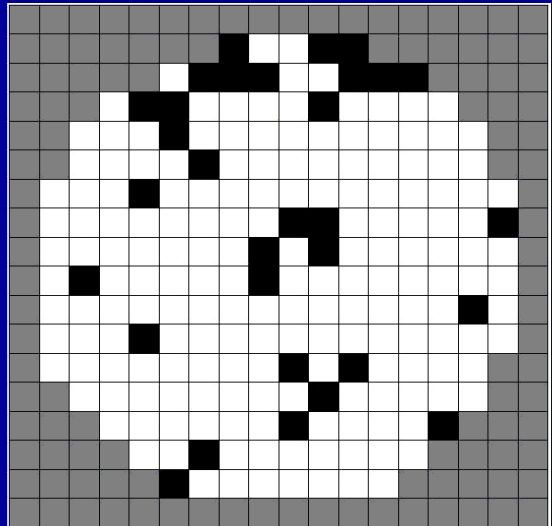
Find and read “courtesy amount” on checks:



Scenario 2

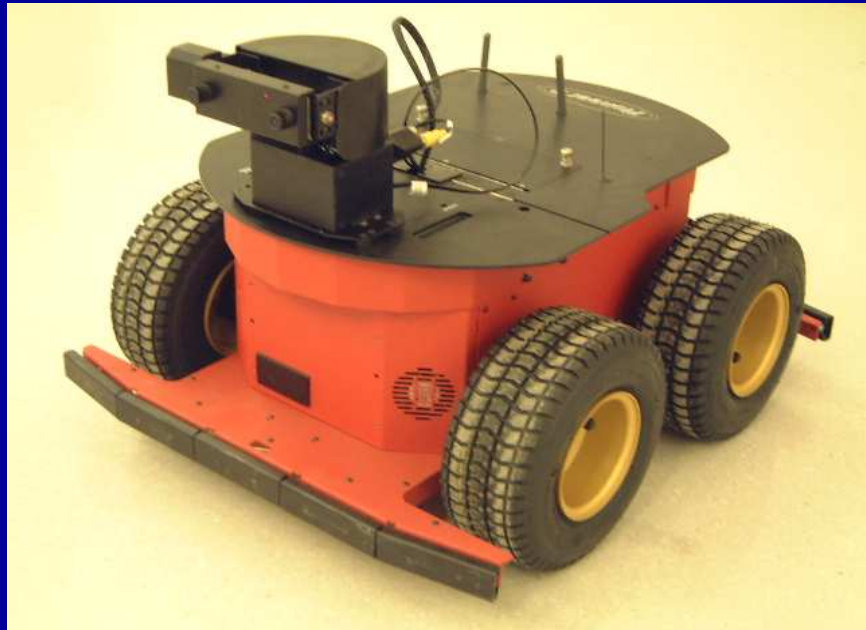
VLSI Wafer Testing

- Wafer test: Functional test of each die (chip) while on the wafer



Scenario 3: Allocating mobile robot camera

Binocular
No GPS



Programmation classique

Analyse : Interviewer experts et utilisateurs pour identifier
quoi et comment faire

Conception : Programmer

Implementer

Tester

Limites : Spécifications ardues; compromis mal définis

Spécifications ardues

Absence d'expertise

fraude téléphonique

Expertise implicite

reconnaissance de caractères

Spécifications évolutives

détection d'intrusion

Personnalisation

filtrage de pourriels

Compromis mal définis

Cas du wafer testing

dépend de la probabilité de panne, du coût du test (Wafer vs SIP)

Cas de la caméra robotique

dépend de la probabilité des obstacles, et de la qualité des amers.

Programmation par Apprentissage

Remplacer les choix du programmeur

reconnaissance de caractères, point d'équilibre pour le test, ...

...par l'apprentissage / la fouille de données

utilisant des modèles appris à partir des données

Défi: utiliser la connaissance du domaine

Où est la difficulté ?

Résoudre le problème

Comprendre le problème

Comprendre le but de l'expert

From Xtrem Programming to Xtrem Solving