

**FERA, 21 octobre 2005**

# **Relationel + Probabilités, Pedro Domingos**

**Michèle Sebag**

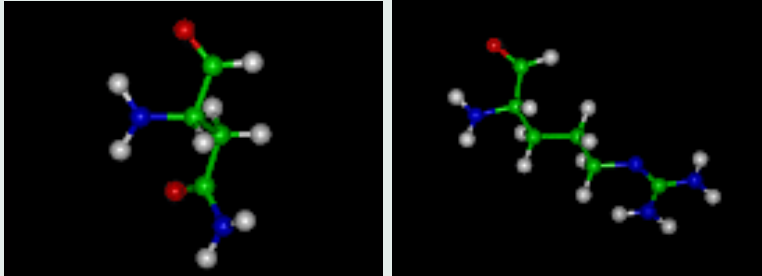
Groupe TAO

<http://tao.lri.fr>

CNRS - INRIA - Université Paris-Sud

# Données relationnelles

## Exemple



$mutagene(m) \leftarrow atm(m, m_1, carbone), atm(m, m_2, carbone), lien(m_1, m_2), ..$

## Caractéristique des problèmes relationnels

Alignement de deux exemples non unique  
mais exponentiel

*comme dans  $\mathbb{R}^n$*

## Remarque

Une fois résolu, tout problème relationnel peut se mettre sous forme propositionnelle.

# Historique

## Programmation Logique Inductive

Muggleton & De Raedt 94

### Eléments du paysage

- Concept cible
- Exemples
- Background knowledge
- Relation de généralité

### Problème

Fume(Paul). Fume(Jean).

Ami(Anne, Hélène). Ami(Anne, Jean). Ami(Jean,Paul). Soeur(Anne, Paul). etc

# Programmation Logique Inductive, 2

- $tc : Fume$

## Learning from interpretations

Exemple

BK

*Close World Assumption*

Fume(Jean)

Ami(Jean,Paul),...

## Learning from entailments

Exemple

Fume(Jean)  $\leftarrow$  Ami(Jean,Paul), Fume(Paul), Ami(Jean,Anne), Chien(Anne,Medor)

BK

Ami(X,Y)  $\leftarrow$  Ami(X,Z), Chien(Y,Z)

But

Trouver  $H$  tq  $H, BK \models E$

# Programmation Logique Inductive, 3

## Relation de généralité

- Implication logique : ingérable
- Theta-subsumption ( $\prec$ )

Plotkin 70

## Définition

Clause  $C_1$   $\theta$ -subsume  $C_2$   
substitutions

Soit  $\Sigma : Var(C_1) \mapsto Var(C_2) \cup Cste(C_2)$

$$C_1 \prec C_2 \quad : \quad \exists \sigma \in \Sigma \quad tq \quad C_1 \sigma \subset C_2$$

## Exemple

$C$   $mut(X) : -atm(X, X_1, carbon), atm(X, X_2, hydr), bond(X_1, X_2)$

$E$   $mut(m) : -atm(m, m_1, hydr), atm(m, m_2, carb), atm(m, m_3, hydr), bond(m_2, m_3)$

$\sigma = \{X/m, X_1/m_2, X_2/m_3\}, \quad C\sigma \subset E$

# Discussion

## Limites PLI

- CWA
- Traitement du bruit
- Traitement des données numériques
- Transition de phase

Ce qui n'est pas connu est faux.

Pas d'err. ds la chaine logique

voir CLP

voir CSP

## Rq sur les SVM relationnels

ILP: Comparer Exemple / Hypothèse

RSVM: Comparer Exemple / Exemple

Propositionnalisation

# Relationnel + Probas : Beaucoup d'essais

Voir le tutoriel Pfeffer - Getoor, IJCAI 2005.

## Programmes Logiques Stochastiques

- Muggleton 96.

## Réseaux Bayesiens Relationnels

- Koller ; Getoor ; Pfeffer.

## Bayesian Logic Programs

- DeRaedt and Kersting, 00.

# Points clé

## Déduction focalisée ou tous azimuts

supervised ML vs réseau bayésien.

## Gestion de la disjonction

tc si cause 1, proba  $p_1$

tc si cause 2, proba  $p_2$

causes 1 et 2 vérifiées, proba(tc) ?

## Efficacité algorithmique

savoir compter le nombre d'objets en cause ; mais pas deux fois  
réutilisation, modularité, hiérarchie

Pas de cycles!

# Markov Networks / Random Fields

Della Pietra, 97; McCallum, 03

## Paysage

- Graphe (noeuds = variables)
- Cliques → Fonctions de potentiel

$$P(X = x) = \frac{1}{Z} \prod_{c \text{ clique}} \phi_c(\text{clique } c | x)$$

Modèles log linéaires :

$$P(X = x) = \frac{1}{Z} \exp \left\{ \sum_j w_j f_j(x) \right\}$$

## Points clé

- Facteur de normalisation  $Z$
- Apprentissage de la structure (les  $f_j$ )
- Apprentissage des poids (les  $w_j$ )

# Markov Logic Networks

Richardson Domingos, 03

MLN: (formules logiques  $F_i$ , poids  $w_i, i = 1..|F|$ ).

## Paysage

Noeuds  $\rightarrow$  atomes logiques.

Cliques  $\rightarrow$  Formules  $F_i$

$$P(X = x) = \frac{1}{Z} \exp \left\{ \sum_{i=1}^{|F|} w_i \times n_i(x) \right\}$$

où  $n_i(x)$ : nb d'instanciations de  $F_i$  pour  $x$ .

# Exemple

$$\neg Fr(x, y) \vee \neg Fr(y, z) \vee Fr(x, z)$$

amis d'amis = amis  
0.7

$$(\exists y Fr(x, y)) \vee Sm(x)$$

sans amis, fumeur  
2.3

$$\neg Fr(x, y) \vee Sm(x) \vee \neg Sm(y)$$

deux amis fument  
1.1

$$\neg Fr(x, y) \vee \neg Sm(x) \vee Sm(y)$$

ou ne fument pas  
1.1

# Fonctionnement / Hypothèses de travail

## Principe

MLN = un template(x,y)

→ MN = MLN  $\sigma$

$\sigma = \{x/Anne, y/Paul\}$

## Restrictions

i) monde fini ; nb de constantes fini

permet aplatissage quantificateurs

ii) fonctions connues  $f(cst) = cst$

remplacement : réécriture

## Propriétés

MLN généralise tout.

# Markov Logic Networks, Apprentissage

## Apprentissage de structure

Les  $F_i$

Comme en ILP:

partir du minimum (les prédicats),  
et spécialiser tant que ça améliore.

## Apprentissage des poids

Optimiser la log vraisemblance.

concave

i) Calculer le nombre d'instanciations (possibles, vraies)

NP

ii) Approximation MCMC

trop lent

iii) Optimiser pseudo-vraisemblance

# Pseudo-vraisemblance

Richardson Domingos 04

Monde fini:  $x = (x_1, \dots, x_n)$  atomes

Couverture de Markov, Markov blanket

$$x_l \rightarrow B_l = \{x_k, \exists F, \exists \sigma / x_l x_k \subset F \sigma\}$$

Pseudo-vraisemblance

$$P_w(X_l = x_l | B_l = b_l) =$$

$$\frac{\exp\{w_i f_i(X_l = x_l, B_l = b_l)\}}{\exp\{w_i f_i(X_l = 0, B_l = b_l)\} + \exp\{w_i f_i(X_l = 1, B_l = b_l)\}}$$

# Pseudo-vraisemblance, 2

## Gradient

$$\begin{aligned} \frac{\partial}{\partial w_i} \log P_w(X = x) &= \sum_l n_i(x) \\ &\quad - P_w(X_l = 0 | B_l(X_l)) n_i(x_{|X_l=0}) \\ &\quad - P_w(X_l = 1 | B_l(X_l)) n_i(x_{|X_l=1}) \end{aligned}$$

En pratique, gradient. BFGS

# Pseudo-vraisemblance, 2

## Limites

Kok Domingos 05

Trop de poids aux prédicats de grosse arité

## Pénalisations

diverses...

## Heuristiques

- se restreindre aux prédicats de  $F_i$
- compter  $n_i(x)$ ,  $n_i(x|X_l=1)$ ,  $n_i(x|X_l=0)$  une seule fois
- ignorer  $x$  / flipper un  $x_l$  ne change pas  $F_i(x)$
- échantillonner
- lazy évaluation

# Expérimentations

## U-Washington database

22 prédicats (professeur, étudiant, cours, directeur-thèse...)

1158 constantes

Pedro, Matt,...

> 4 millions atomes possibles

(types)

3 212 atomes

KB manuelle possible (incohérente)

## Résultats

- Terrible : ILP : Claudien, Foil,...

- Mauvais : propositionnel, Naive Bayes

- Pas Bon, MLN + ILP (poids sur structure ILP), Bayesian Network

- Bon : MLN + structure automatique; éventuellement ajout de clause manuelle.

# A suivre

## Remarques

Conditional Log Likelihood  $\leftrightarrow$  AUC

AUC très mauvais...

## Dernière

approche discriminante  $>$  approche générative

Singla Domingos, 05