

# Leçons de l'Action Spécifique

## Extraction de Connaissance et Fouille de Données

GafoDonnées – 2000-2002

Michèle Sebag, LRI, Paris-Sud

Réunion GdR I3, 12 juin 2003

# Grandes lignes

1. L'animation de la recherche
2. Quelques directions scientifiques
3. Retour à l'animation de la recherche

## Définition d'une AS

1. Objectifs ponctuels
2. Groupe restreint
3. Liens Réseaux Thématiques Pluri-Disciplinaires

# Politiques scientifiques

1. Le son de l'Europe
2. Le coût de l'évaluation
3. Pôles d'excellence vs Animation de la communauté
4. Quelles niches écologiques ?

# Le contenu de GafoDonnées

1. Motifs fréquents → Jean-François Boulicaut et Bruno Crémilleux
2. Mesures de qualité → Henri Briand et Fabrice Guillet
3. Bio-Informatique → Nouvelle AS, F. Denis et J.-D. Zucker
4. Autres :
  - Fouille du cube
  - Multi-média
  - Langage d'hypothèses et de requêtes
  - Fouille de données visuelle
  - ...

# Bases de données et cartes de Kohonen

## Objectif

Classification / Clustering / Apprentissage non supervisé

## Analyse de données - apprentissage statistique

k-Means

Nuées dynamiques

Information bottleneck

Tishby, Pereira, Bialek 1999

## Perspective “Masse de données” : pourquoi Kohonen

Un seul passage

Système dynamique : un attracteur, une classe.

# Bases de données et distributions

## Objectif

Apprendre un modèle

La bd est-elle l'univers ou un échantillon ?

## Etat de l'art

Motifs fréquents, règles d'approximation

Réseaux Bayesiens, HMM, ...

Machines à Vecteurs Supports

## Perspective “Masse de données”

Probabilistic Models for Query Approximation with Large Sparse Binary Data Sets,

D. Pavlov, H. Mannila, P. Smyth, UAI 2000

# Motifs fréquents et changements de représentation

## Objectif

Se projeter dans un espace avec de “bonnes caractéristiques”

## Etat de l’art

Analyse en composantes principales

Analyse sémantique latente

Analyse en composantes indépendantes

A. Hyvarinen, J. Karhunen and E. Oja, 2001

## Le cahier des charges : FISICA

Jouteau et al. EGC 2003

Des motifs rares

Orthogonaux

Suffisants

# Motifs fréquents et espaces semi-structurés

## Objectif

Etendre Apriori pour XML

## Pré-requis

Quelle définition de l'inclusion

Niveaux de complexité: de polynomial à NP.

## Compromis

Relaxer la définition ?

Relaxer la complétude ?

# Langages d'hypothèses et langages de requêtes

## Objectif

Eviter de faire le travail en double

## Exemples

Query packs

Blockeel, ILP 2001)

Transition de phase

Giordana Saitta, MLJ 2000

Propositionnalisation

Kramer ILP 1999; Roth ILP 2001

Formalisation logique probabiliste

Koller UAI 1999; Kersting & De Raedt, MLJ 2003

# Du QUOI au COMMENT

## ...revenons à la politique scientifique

### Conditions nécessaires

Pluri-disciplinarité

Applications communes

Intégration

### Les instruments

RTP

AS

Equipes pluri-laboratoires

Instituts sans murs