

Machine Learning et Intelligence Artificielle

Christel VRAIN

LIFO
Université d'Orléans

Journée Cascimodot 2018

Outline

1 Machine Learning et Intelligence Artificielle

2 Machine Learning et autres domaines de l'Intelligence Artificielle

- Programmation par Contraintes
- Quelques mots sur le clustering
- Unifier clustering conceptuel et à base de distance
- Clustering descriptif

Outline

1 Machine Learning et Intelligence Artificielle

2 Machine Learning et autres domaines de l'Intelligence Artificielle

- Programmation par Contraintes
- Quelques mots sur le clustering
- Unifier clustering conceptuel et à base de distance
- Clustering descriptif

Terminologie

- Apprentissage supervisé

Given

- ▶ k classes
- ▶ un ensemble d'apprentissage $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$

⇒ Trouver une dépendance fonctionnelle $\mathbf{x} \xrightarrow{h} y$

- Apprentissage non supervisé

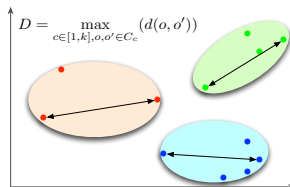
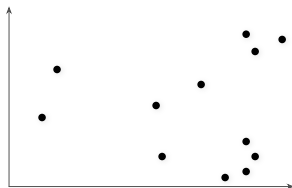
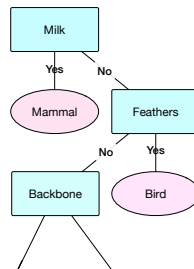
Etant données des observations $o_1, \dots, o_n$

Clustering - Classification non supervisée: Trouver un ensemble de clusters compacts et bien séparés

Pattern Mining: Recherche de motifs intéressants

Classification supervisée / non supervisée

<i>name</i>	<i>hair</i>	<i>milk</i>	<i>domestic</i>	<i>class</i>
<i>bear</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>mammal</i>
<i>cat</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>mammal</i>
<i>dolphin</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>mammal</i>
<i>honeybee</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>insect</i>
<i>moth</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>insect</i>



Intelligence Artificielle

- GDR Aspects Algorithmiques et Formels de l'Intelligence Artificielle
www.gdria.fr
 - ▶ Représentation des connaissances et modélisation des raisonnements
 - ▶ Incertitude, modèles graphiques, réseaux Bayésiens
 - ▶ Contraintes et SAT
 - ▶ Apprentissage automatique
 - ▶ Planification et recherche heuristique
 - ▶ Systèmes multi-agents et décision collective
- Conférences nationales et internationales
 - ▶ Conférences généralisées : IJCAI, ECAI
 - ▶ Conférences spécialisées: ICML, NIPS, ECML, CP
- Des thématiques où l'Intelligence Artificielle joue un rôle important: TALN

Place de l'Apprentissage Automatique en Intelligence Artificielle

- IJCAI1985 : 6 sessions sur l'Apprentissage
- IJCAI 2018 : plus de 25 sessions sur l'Apprentissage

- Neural Networks (2)
- Deep Learning
- Kernel Methods
- Online Learning
- Multi-Instance, Multi-View, Multi-Label Learning (2)
- Transfer - Adaptation - Multitask Learning (2)
- Learning Theory
- Learning Preferences or Ranking
- Learning Generative Models
- Semi-supervised learning
- Probabilistic Machine Learning

- Feature Selection, Learning Sparse Models
- Dimensionality Reduction and Manyfold Learning
- Knowledge Representation and Learning
- Constraints, Satisfiability and Learning

- Heuristic Search and Learning
- Planning and Learning
- Reinforcement Learning
- Reinforcement Learning and Online Learning
- Reinforcement Learning and Applications
- Agents and Learning

- Deep Learning for Computer vision
- Time series and data streams (2)
- Unsupervised Learning and Applications
- Machine Learning and Recommender Systems
- Interpretability

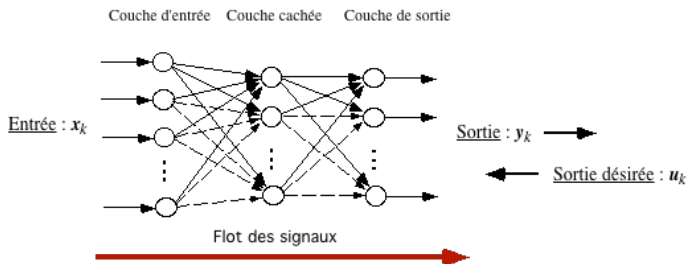
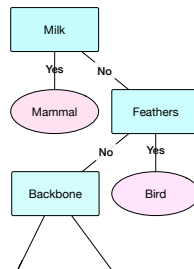
Formal Concept Analysis ? Inductive Logic Programming ? Analogy ? Case based reasoning ?

L'Apprentissage Automatique dans les années 80 en quelques mots

- **Deux courants**: apprentissage symbolique / connexionnisme
- **Mécanismes principaux**: induction comme généralisation, abduction, analogie / rétro-propagation du gradient
- Principalement **deux cadres** pour l'Apprentissage Symbolique
 - ▶ Classification supervisée : apprentissage de concepts
 - ▶ Clustering conceptuel: un cluster doit être un concept
- Apprentissage = recherche d'une bonne hypothèse dans un espace de recherche structuré par une relation de généralité
- **Critère d'optimisation**
 - ▶ Minimisation du nombre d'erreurs
 - ▶ Category utility
- **Systèmes**:
AQ (R.S. Michalski), Version space (T. Mitchell), ID3 (R. Quinlan), Cobweb (D. Fisher)

Apprentissage supervisé

<i>name</i>	<i>hair</i>	<i>milk</i>	<i>domestic</i>	<i>class</i>
<i>bear</i>	<i>T</i>	<i>T</i>	<i>F</i>	<i>mammal</i>
<i>cat</i>	<i>T</i>	<i>T</i>	<i>T</i>	<i>mammal</i>
<i>dolphin</i>	<i>F</i>	<i>T</i>	<i>F</i>	<i>mammal</i>
<i>honeybee</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>insect</i>
<i>moth</i>	<i>T</i>	<i>F</i>	<i>F</i>	<i>insect</i>



Explicabilité ?

Evolution de l'Apprentissage Automatique

- Data Mining

- ▶ Découverte de motifs fréquents → Exploration d'espaces de recherche discrets
- ▶ Différents types de données : séquences d'ADN, réseaux sociaux, réseaux d'interaction, ...

- Théorie de l'Apprentissage : théorie de l'apprentissage statistique

- De nouvelles méthodes

- ▶ Boosting - Bagging : combinaison de classifieurs faibles
- ▶ Machines à vecteur support
 - ★ Maximisation de la marge
 - ★ Kernel trick
- ▶ Deep Learning

- Nouveaux problèmes d'apprentissage

- ▶ Apprentissage de métriques
- ▶ Apprentissage de préférences
- ▶ Apprentissage de classement

- Nouveaux paradigmes

- ▶ Apprentissage semi-supervisé
- ▶ Apprentissage multi-vues
- ▶ Apprentissage par transfert

L'optimisation numérique semble partout.

- Exemples

- ▶ Classification supervisée

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^m \ell(h(\mathbf{x}_i), y_i) + \Omega_{\text{hyperparameters}}(\mathcal{H})$$

- ▶ Apprentissage génératif

$$\ln(\mathbf{P}(\mathbf{x}|\text{parameters}))$$

- ▶ Clustering

$$\sum_{j=1}^k \sum_{\mathbf{x}_i \in \mathcal{C}_j} \|\mathbf{x}_i - \mu_j\|_2^2$$

- ▶ Apprentissage de métrique, Changement de représentation, ...

- Peu d'optimisation discrète

- Algèbre linéaire

Outline

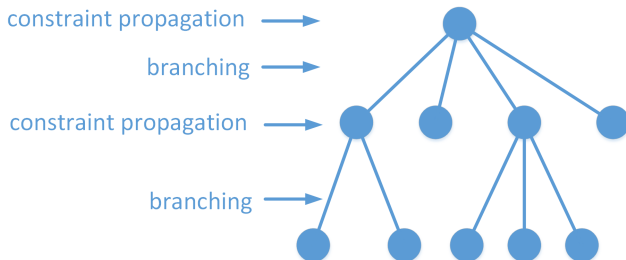
1 Machine Learning et Intelligence Artificielle

2 Machine Learning et autres domaines de l'Intelligence Artificielle

- Programmation par Contraintes
- Quelques mots sur le clustering
- Unifier clustering conceptuel et à base de distance
- Clustering descriptif

Programmation par Contraintes (CP)

- Modélisation déclarative d'un problème en spécifiant des **variables** et des **contraintes**
- Recherche de solutions en alternant **propagation de contraintes** et **branchement**



Un exemple en CP

$$SEND + MOST = MONEY$$

- Trouver toutes les affectations de chiffres aux lettres telles que

$$\begin{array}{rcccc} & S & E & N & D \\ + & M & O & S & T \\ \hline M & O & N & E & Y \end{array}$$

→ CSP : problème de satisfaction de contraintes

- Trouver une affectation de chiffres aux lettres telles que
 - ▶ $SEND + MOST = MONEY$
 - ▶ $MONEY$ est maximal

→ COP: problème d'optimisation de contraintes

Modélisation en CP

$$SEND + MOST = MONEY$$

- Variables $S, E, N, D, M, O, T, Y \in \{0, \dots, 9\}$
- Variable V capturant la valeur de la fonction à optimiser
- Contraintes:
 - ▶ $S \neq 0, M \neq 0$
 - ▶ $\text{alldifferent}(S, E, N, D, M, O, T, Y)$
 - ▶ Une contrainte linéaire

$$\begin{aligned} & (1000 \times S + 100 \times E + 10 \times N + D) \\ + & (1000 \times M + 100 \times O + 10 \times S + T) \\ = & 10000 \times M + 1000 \times O + 100 \times N + 10 \times E + Y \end{aligned}$$

- ▶ $V = 10000 \times M + 1000 \times O + 100 \times N + 10 \times E + Y$
- Fonction objectif : *maximiser* V

Solvers

- CP solvers: itération de
 - ▶ Propagation de contraintes : enlever des valeurs inconsistantes des domaines des variables

$$\begin{aligned}D_S &= \{9\} \\D_E &= \{2, 3, 4, 5, 6, 7\} \\D_M &= \{1\} \\D_O &= \{0\} \\D_N &= \{3, 4, 5, 6, 7, 8\} \\D_D &= D_T = D_Y = \{2, 3, 4, 5, 6, 7, 8\}\end{aligned}$$

- ▶ Branchement: création de branches dans l'arbre de recherche

$$\begin{aligned}\text{Arbre binaire: } E = 2 \text{ and } E \neq 2 \\ \rightarrow \\ D_E = \{2\} \quad \text{and} \quad D_E = \{3, 4, 5, 6, 7\}\end{aligned}$$

- ▶ Pour un problème d'optimisation, ajouter une contrainte chaque fois qu'une solution est trouvée.

Application au clustering: deux cadres

- Clustering à base de distances :
 - ▶ fondé sur une distance ou une dissimilarité entre objets
 - ▶ approprié pour les données numériques (distance euclidienne)
 - ▶ différents types d'attributs doivent être encapsulés dans une distance
- Clustering conceptuel:
 - ▶ introduit dans les années 80's [*Michalski & Stepp, 1983, Fisher, 1985*]
 - ▶ actuellement fondée sur l'analyse formel de concepts
 - ▶ fondé sur des attributs qualitatifs (pas d'attributs numériques)
 - ▶ non prise en compte de distance entre objets (pas de notion de compacité)

Clustering Conceptuel

Un cluster pour être intéressant doit être un concept.

- Deux composantes dans un cluster: un ensemble d'objets et une description.
 - ▶ une conjonction de propriétés booléennes: $2pattes \wedge plumes$
 - ▶ des descriptions plus complexes: $(couleur = bleu \vee rouge) \wedge (longueur > 2)$
- Le clustering conceptuel est lié à la notion de généralisation, où la généralisation d'un ensemble d'observations est l'ensemble des propriétés qu'elles satisfont.
- Le clustering conceptuel permet de choisir les attributs pertinents, ceux qui définissent le concept.
- En clustering à base de distance tous les attributs doivent être encapsulés dans une distance. La sélection d'attributs doit être faite avant.

Clustering à base de distance

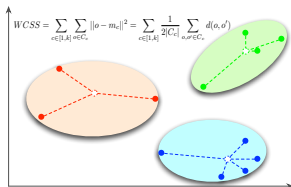
Un exemple classique: k-moyenne

Minimiser le critère *within-cluster sum of squares* (WCSS) défini par (m_c représente le centre du cluster C_c),

$$WCSS(\Delta) = \sum_{c \in [1, k]} \sum_{o_i \in C_c} d(o_i, m_c)^2$$

équivalent dans un espace euclidien à minimiser

$$WCSS(\Delta) = 1/2 \sum_{c \in [1, k]} \frac{1}{|C_c|} \sum_{o_i, o_j \in C_c} d(o_i, o_j)^2.$$



Clustering sous contraintes

- Clustering est en général NP-difficile.
 - Les méthodes classiques sont heuristiques et cherchent un optimum local.
 - Le résultat obtenu ne correspond pas toujours à celui attendu
- Intégrer des connaissances utilisateur sous forme de contraintes
- Clustering sous contraintes
- ▶ Contraintes sur les clusters: taille du cluster, diamètre, ...
 - ▶ Contraintes sur des paires of points: must-link, cannot-link
- Les méthodes sont en général heuristiques et dédiées à certains types de contraintes.

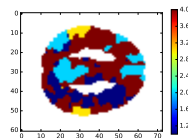
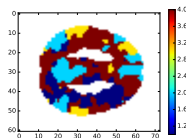
fMRI brain imaging data

- Chaque image correspond à des prises 2D variant avec le temps.
- 1730 voxels dont les niveaux d'oxydation du sang sont mesurés à intervalles de 3ms sur 200+ pas de temps
- Construction d'un graphe complètement connecté de 1730 nœuds - poids des arêtes : valeur absolue de la corrélation de Pearson entre les voxels en fonctions du temps.
- Création d'un clustering initial (spectral clustering normalisé et sélection du meilleur résultat de 10 runs de k-means sur le plongement spectral)

Constraints:

$x\text{-diameter} \leq 15$ pour cluster 3 (jaune)

$y\text{-diameter} \leq 30$ pour cluster 2 (bleu)



Clustering sous contraintes en Programmation par Contraintes

Dao, Duong, Vrain, AIJ 2017

- Entrées: une base de données ou une mesure de dissimilarité entre paires de points
- Clusters définis par des affectations de points à leur cluster:
 $G[o] = c, c \in [1, k]$
- Critère d'Optimisation, e.g. minimiser le diamètre maximum
- Contraintes sont exprimées
 - ▶ pour représenter une partition
 - ▶ pour casser les symétries
 - ▶ pour représenter les contraintes utilisateur: taille, diamètre, séparation, ...

$$G_1 = 1 \text{ and } G_i \leq \max_{j \in [1, i-1]} (G_j) + 1, \text{ for } i \in [2, n]$$
$$\#\{i \mid G_i = k_{min}\} \geq 1$$

Comment rendre un clustering interprétable?

- Après le clustering
 - expliquer les clusters
 - ▶ Caractérisation
 - ▶ Généralisation
 - ▶ Statistiques
- Avant le processus de clustering : intégrer la connaissance de l'expert avant le clustering
 - actionable clustering
- Pendant le processus de clustering. Deux hypothèses
 - ▶ Clustering et explications sont dans le même espace de représentation are in the same representation space → clustering conceptuel
 - ▶ Clustering et explications sont dans des espaces de représentation différents.

Clustering et explications

- Dans beaucoup de domaines, les données ont des attributs pour former des clusters compacts mais
 - ▶ ils ne peuvent pas *expliquer* le clustering
 - ▶ les données sont aussi décrites par un ensemble (éventuellement *peu dense et bruité*) de descripteurs/étiquettes/tags utiles pour des explications that are useful for explanation

Cadre	Variables/attributs	Descripteurs/tags
Twitter network	mention/retweet graph	hashtag usage
Images	SIFT features	tags

But

- Trouver un équilibre entre obtenir des clusters **compacts** (w.r.t. à une **distance** entre objets) et
 - ▶ consistants avec les attentes des experts
 - ▶ **interprétables/explicables** aux experts
- buts:
 - 1 intégrer des connaissances des experts pour obtenir des clusters plus proches de leurs attentes
 - 2 découvrir **simultanément** des explications pendant le processus de clustering
- Comment ?
 - ▶ combiner clustering classique et clustering conceptuel
 - ▶ formuler les connaissances des experts sous forme de contraintes
 - ▶ utiliser la programmation par contraintes pour résoudre le problème

Un exemple sur des voitures

- 193 objets
 - propriétés techniques (22 attributs) :
 - ▶ motorisation (diesel ou non)
 - ▶ roues motrices (4, 2 avant, 2 arrière)
 - ▶ puissance (entre 48 et 288)
 - ▶ ...
- Après discrétisation, 64 attributs qualitatifs
- un attribut numérique : prix
- ⇒ Trouver 3 clusters compacts en termes de prix mais explicables.

Exemple de voitures

- Cadre conceptuel

- (e) concepts + maximizing min. taille des clusters

- (f) concepts + maximizing min. taille des concepts

Distribution du prix pas convaincante

- Cadre à base de distance

- (g) minimiser diamètre maximum

Pas de concepts convaincants

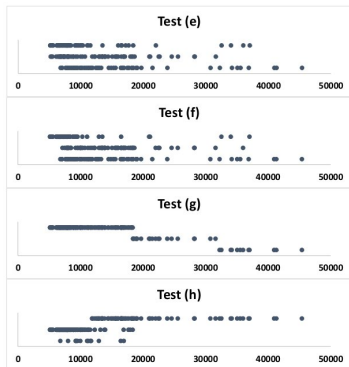
- Cadre unifié

- (h) concepts + minimiser diamètre maximum

Une meilleure modélisation des 3 classes de voiture avec des concepts fondés sur la taille, la puissance de moteur, la consommation de fuel, ...

⇒ des critères d'optimisation en conflit

⇒ une notion de concept trop forte pour trouver des descriptions intéressantes

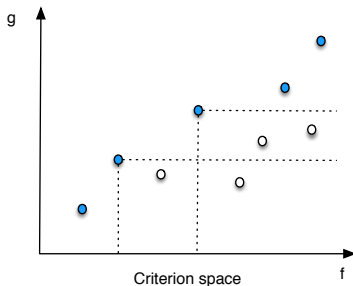


Descriptive clustering formulation

Dao, Kuo, Ravi, Vrain, Davidson, IJCAI 2018

- Data: n observations décrites par *des attributs numériques* X et *des descripteurs/tags booléens interprétables* D
 - Trouver des clusters
 - ▶ compacts dans une modalité (e.g. descripteurs SIFT pour images)
 - ▶ descriptifs dans une autre modalité (e.g. tags)
 - Ces objectifs **ne sont pas compatibles**
- calcul d'un front de Pareto permettant un compromis entre les 2 objectifs
- f : **fondé sur les attributs** (compacité)
 - g : **fondé sur les descripteurs** (interprétabilité)

Solutions Pareto optimales et front de Pareto



- Partition P' domine P iff elle est meilleure dans un critère et non pire ailleurs
- P est une solution Pareto optimale ssi elle n'est pas dominée.
- front de Pareto = $\{(f(P), g(P)) \mid P \text{ est une solution Pareto optimale}\}$

Contraintes sur descriptions

- Chaque cluster est décrit par un ensemble non vide de tags

$$\forall c = 1, \dots, k, \sum_{i=1}^n S_{cp} \geq 1$$

- Un objet dans un cluster doit satisfaire *une grande partie* de ses descripteurs (jusqu'à α exceptions):

$$\forall c = 1, \dots, k, \forall i = 1, \dots, n, \\ Z_{ic} = 1 \implies \sum_{p=1}^r S_{cp}(1 - D_{ip}) \leq \alpha$$

- Une étiquette est incluse dans la description d'un cluster ssi *la plupart* de ses instances la vérifient (jusqu'à β exceptions) :

$$\forall c = 1, \dots, k, \forall p = 1, \dots, r, \\ S_{cp} = 1 \iff \sum_{i=1}^n Z_{ic}(1 - D_{ip}) \leq \beta$$

- Avec des bases de données denses, on peut poser $\alpha = \beta = 0$

Des critères d'optimisation fondés sur les descriptions

Tags peuvent être denses ou non, bruités $\Rightarrow$ différentes fonctions à optimiser

1 Max-min complete tag agreement (MMCTA):

- ▶ les tags d'un cluster sont partagés par *toutes* ses instances ($\alpha = \beta = 0$)
- ▶ maximiser l'ensemble des tags de chaque cluster (taille du plus petit)
- ▶ Usage: beaucoup de tags et peu de bruit

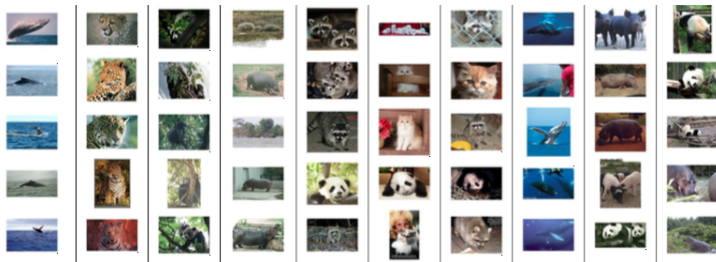
2 Minimize tag disagreement (MTD)

- ▶ minimiser $\alpha + \beta$
- ▶ Usage: quand les tags sont bruitées

3 Max-min neighborhood agreement (MMNA):

- ▶ chaque paire d'instances dans un même cluster doit partager *au moins* q tags
- ▶ maximiser q
- ▶ Usage: tags très peu denses

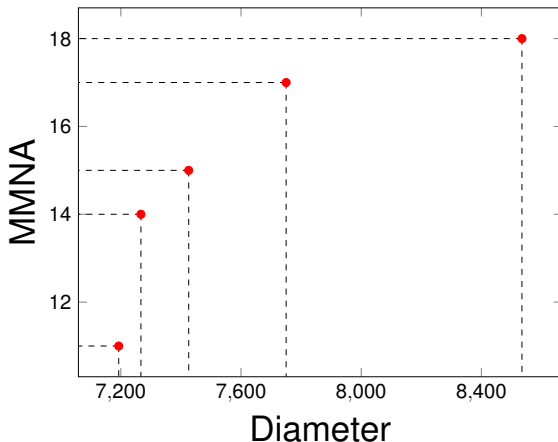
Clustering d'images étiquetées



[Lampert C.H, Nickisch H., Harmeling S., CVPR 2009]

- 30000 images de 50 classes d'animaux
- Chaque image est décrite par 2000 descripteurs SIFT et un nombre variable de tags (black, fast, timid, etc.)
- Les classes des animaux ne sont pas données en entrée à l'algorithme
- Echantillonnage aléatoire de 100 images des 10 premières classes d'animaux

Compromis entre compacité et capacité de description



Premier point de Pareto: Diamètre minimisé. MMCTA=4. MMNA=11

Cl#	Composition par animaux	Description par étiquettes
C1	1 grizzly bear, 2 dalmatian, 1 horse, 2 blue whale	big, fast, strong, muscle, new-world, smart
C2	5 antelope, 2 grizzly bear, 2 beaver, 5 dalmatian, 5 persian cat, 5 horse, 6 german shepherd, 3 siamese cat	furry, chewteeth, fast, quadrapedal, newworld, ground
C3	69 beaver, 64 dalmatian, 42 persian cat, 29 blue whale, 42 siamese cat	tail, fast, newworld, timid, smart, solitary
C4	100 killer whale, 69 blue whale, 1 siamese cat	tail, fast, fish, smart
C5	95 antelope, 97 grizzly bear, 29 beaver, 29 dalmatian, 53 persian cat, 94 horse, 94 german shepherd, 54 siamese cat	furry, chewteeth, fast, quadrapedal, newworld, ground

MMCTA: min complete tag agreement, MMNA: min neighborhood agreement

Troisième point de Pareto. MMCTA=9, MMNA=15

Cl#	Composition par animaux	Description par étiquettes
C1	2 antelope, 4 dalmatian, 2 horse, 3 german shepherd, 4 siamese cat	furry, lean, longleg, tail, chewteeth, walks, fast, muscle, quadrapedal, active, agility, newworld, oldworld, ground
C2	2 beaver, 1 persian cat, 1 horse, 1 german shepherd	furry, tail, chewteeth, fast, quadrapedal, agility, newworld, ground, smart
C3	100 grizzly bear, 98 beaver, 99 persian cat, 1 siamese cat	furry, paws, chewteeth, claws, fast, quadrapedal, fish, newworld, ground, smart, solitary
C4	100 killer whale, 100 blue whale	spots, hairless, toughskin, big, bulbous, flippers, tail, strainteeth, swims, fast, strong, fish, plankton, arctic, ocean, water, smart, group
C5	98 antelope, 96 dalmatian, 97 horse, 96 german shepherd, 95 siamese cat	furry, lean, longleg, tail, chewteeth, walks, fast, muscle, quadrapedal, active, agility, newworld, oldworld, ground

MMCTA: min complete tag agreement, MMNA: min neighborhood agreement

Cinquième point de Pareto: MMNA maximisé. MMCTA=15, MMNA=18

Cl#	Composition par animaux	Description par étiquettes
C1	100 antelope, 100 dalmatian	furry, big, lean, longleg, tail, chewteeth, walks, fast, strong, muscle, quadrapedal, active, agility, newworld, oldworld, ground, timid, group
C2	100 horse, 99 german shepherd, 98 siamese cat	black, brown, gray, patches, furry, lean, longleg, tail, chewteeth, walks, fast, muscle, quadrapedal, active, agility, newworld, oldworld, ground, smart, domestic
C3	100 grizzly bear, 100 beaver, 1 siamese cat	brown, furry, paws, chewteeth, claws, fast, muscle, quadrapedal, active, nocturnal, fish, newworld, ground, smart, solitary
C4	100 killer whale, 100 blue whale	spots, hairless, toughskin, big, bulbous, flippers, tail, straintooth, swims, fast, strong, fish, plankton, arctic, ocean, water, smart, group
C5	100 persian cat, 1 german shepherd, 1 siamese cat	gray, furry, pads, paws, tail, chewteeth, meatteeth, claws, walks, fast, quadrapedal, agility, meat, newworld, oldworld, ground, smart, solitary, domestic

MMCTA: min complete tag agreement, MMNA: min neighborhood agreement

Conclusion

Importance des paradigmes déclaratifs pour encapsuler des connaissances utilisateur

- Cadre développé applicable à d'autres types de contraintes et de critères d'optimisation
- Problèmes en général NP-complet → Passage à l'échelle
 - ▶ Smart data
 - ▶ Echantillonnage
 - ▶ Relâcher la recherche d'une solution optimale
 - ▶ Intégration dans des sous-problèmes où la dimensionalité est plus petite.
- Application à des données temporelles, à la chémo-informatique
- Apprentissage de contraintes et de préférences
- D'autres fertilisations croisées entre Machine Learning et les autres domaines de l'IA
 - ▶ Inductive Logic Programming (Relational learning) et Deep Learning
 - ▶ Groupe de travail Apprentissage et Raisonnement

Thanks

Remerciements à :

- T. B. H. Dao (University of Orléans)
- K.C. Chuong (PhD student, University of Orléans)
- I. Davidson (University of Davis California)
- T. Guns (University of Bruxel)
- P. Gañçarski (University of Strasbourg)
- C. T. Kuo (PhD Student, University of Davis California)
- S. Ravi (University of Albany and now Virginia Tech)
- W. Lesaint (University of Orléans)

Merci pour votre attention !