



Clustering Multicritère

Alexandru Olteanu^{1,2}

Raymond Bisdorff¹

Patrick Meyer²

¹ University of Luxembourg

² TELECOM Bretagne

6 October 2011 @ HEIG-VD

Introduction

Problématiques d'aide multicritère à la décision:

- Meilleur choix:
 - déterminer la meilleure alternative;
- Classement:
 - classer les alternatives;
- Tri:
 - affecter des alternatives à des catégories ordonnées prédéfinies;
- Clustering:
 - affecter des alternatives à des catégories non définies a priori;
 - [Nemery et al:05], [De Smet et al:09].

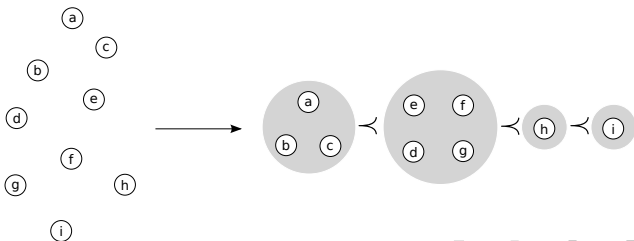
Formulation du problème

Formulation du problème

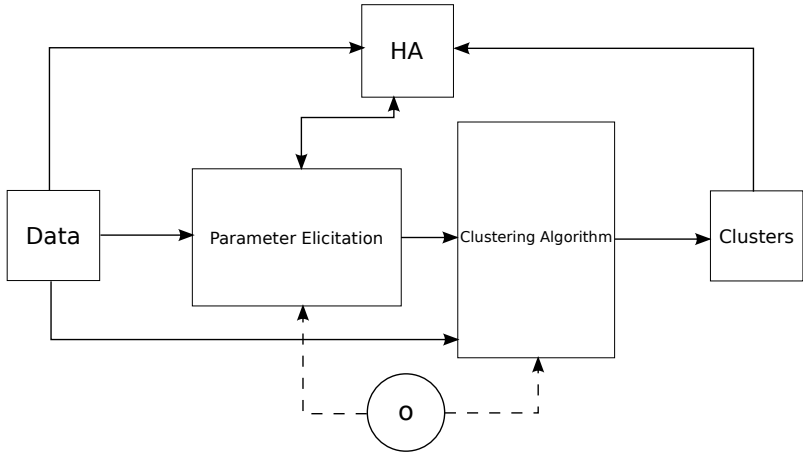
Entrées:

- Alternatives: $X = \{x, y, \dots\}$, $|X| = n$;
- Critères: $I = \{i, j, \dots\}$, $|I| = m$:
 - échelle nominale, ordinale, d'intervalle, de ratio;
- Données: $x = (x_i, x_j, \dots)$, $\forall x \in X$, (données manquantes);
- Relation de surclassement: σ ;

Sortie:



Aperçu de la méthode



La relation de surclassement

- Une alternative a **surclasse** une alternative b lorsqu'une majorité pondérée de critères valide le fait que a est au moins aussi bonne que b et qu'il n'y a pas de critère pour lequel a est sérieusement moins bonne que b [Roy].
 - Relation de surclassement: $o : X \times X \rightarrow [-1, 1]$
 - Affirmations de surclassement:
 - a **surclasse** b: $o(a, b) > 0$;
 $a \rightarrow b$
 - a **ne surclasse pas** b: $o(a, b) < 0$;
 $a \leftarrow b$
 - a et b sont dans un état de surclassement **indéterminé** : $o(a, b) = 0$;
 $a \leftrightarrow b$
- Graphe de surclassement orienté: $o^* : X \times X \rightarrow \{-1, 0, 1\}$

La relation de surclassement

■ Affirmations d'indifférence:

- a est **indifférent** de b: $o(a, b)$ et $o(b, a)$ positifs;
a — b
- a **n'est pas indifférent** de b: $o(a, b)$ ou $o(b, a)$ négatif;
a b
- a et b sont dans un état d'indifférence **indéterminé** : sinon.
a --- b

→ Graphe d'indifférence:

$$q^*(x, y) := \begin{cases} +1 & , \text{ si } o(x, y) > 0 \text{ et } o(y, x) > 0; \\ -1 & , \text{ si } o(x, y) < 0 \text{ ou } o(y, x) < 0; \\ 0 & , \text{ sinon.} \end{cases}$$

Algorithme de clustering

Algorithme de clustering

Définition du clustering :

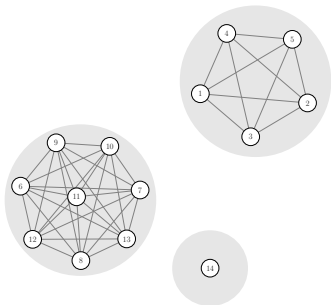
- en fouille de données:
 - regrouper des objets qui sont **similaires** et séparer ceux qui ne sont **pas similaires**;
- en AMCD:
 - regrouper des alternatives qui sont **indifférentes** en tenant compte des préférences du décideur, et séparer celles qui ne sont **pas indifférentes**;

Clustering Algorithm

Idéalement, les alternatives d'un cluster sont:

- **indifférentes** à **toutes** les alternatives **du** cluster;
- **non indifférentes** à **toutes** les alternatives **à l'extérieur** du cluster;

Relation d'indifference (q') induite par le résultat du clustering:



- Solution du problème: la partition qui $\min \sum |q^* - q'|$.

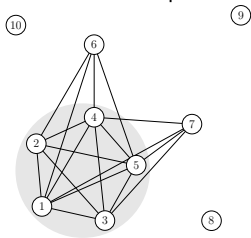
L'algorithme de clustering

Approche algorithmique:

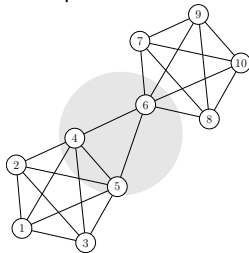
- Approche exacte;
 - Enumérer toutes les partitions [Orlov:02] et retenir celle qui $\min \sum |q^* - q'|$ (**très difficile**);
- Approche approximative;
 - Noyau d'un cluster:
 - alternatives indifférentes;
 - représentant du cluster;
 - en grande partie **indifférent** aux alternatives du cluster;
 - en grande partie **non indifférent** aux alternatives en dehors du cluster;

Noyau de cluster

Caractéristique:



Propriété de localité:



■ Approche approximative;

1. Rechercher les noyaux:

- Énumération des cliques maximales du graphe d'indifférence;
- Sélection des maxima locaux d'une fonction de mérite;

2. Expansion des noyaux:

- ajouter les objets au cluster correspondant au noyau auquel ils sont le plus indifférent;

Exemple

Thierry's choice problem

Introduction

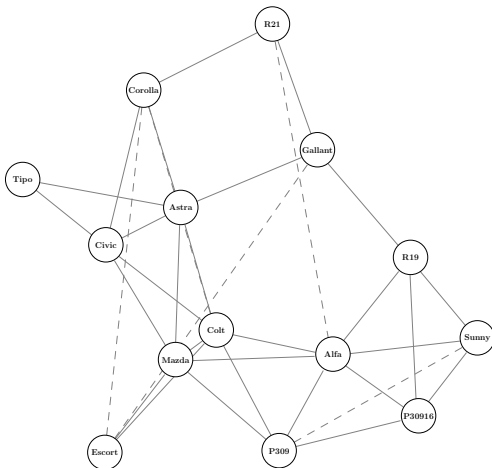
Formulation du problème

Solution du problème

Exemple

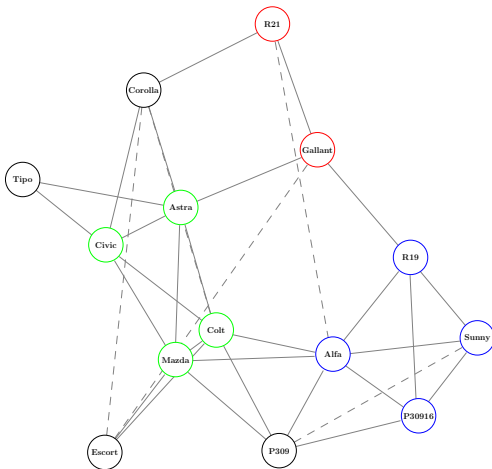
Critères	Coût ↓	Accélération ↓	Reprise ↓	Freinage ↑	Tenue de route ↑
Weights	1.0	2.0	1.0	0.5	0.5
l	350	0.2	0.4	0.12	0.12
π	700	0.4	0.7	0.35	0.35
Fiat Tipo	18342	30.7	37.2	2.33	3.00
Alfa 33	15335	30.2	41.6	2.00	2.50
Nissan Sunny	16973	29.0	34.9	2.66	2.50
Mazda 323	15460	30.4	35.8	1.66	1.50
Mitsubishi Colt	15131	29.7	35.6	1.66	1.75
Toyota Corolla	13841	30.8	36.5	1.33	2.00
Honda Civic	18971	28.0	35.6	2.33	2.00
Open Astra	18319	28.9	35.3	1.66	2.00
Ford Escort	19800	29.4	34.7	2.00	1.75
Renault 19	16966	30.0	37.7	2.33	3.25
Peugeot 309 16V	17537	28.3	34.8	2.33	2.75
Peugeot 309	15980	29.6	35.3	2.33	2.75
Mitsubishi Gallant	17219	30.2	36.9	1.66	1.25
Renault 21	21334	28.9	36.7	2.00	2.25

Thierry's choice problem



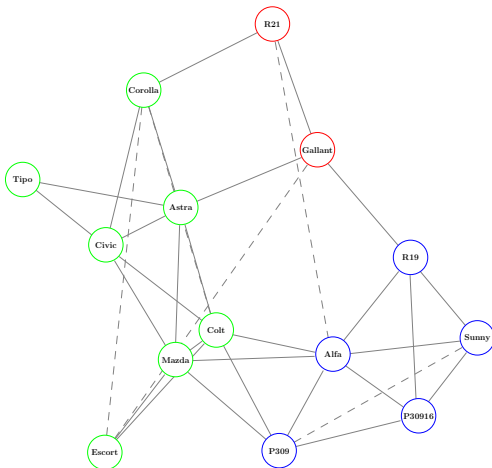
Graphe d'indifférence original;

Thierry's choice problem



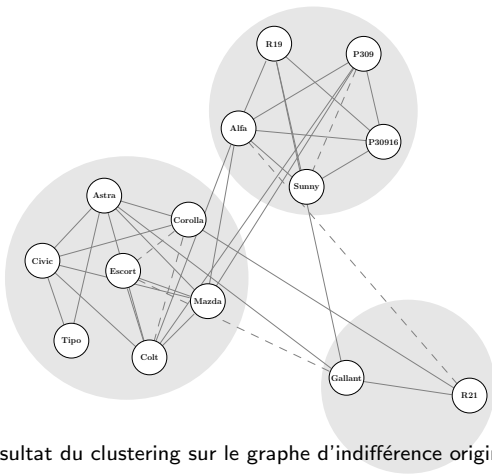
Noyaux de cluster sur le graphe d'indifférence original;

Thierry's choice problem



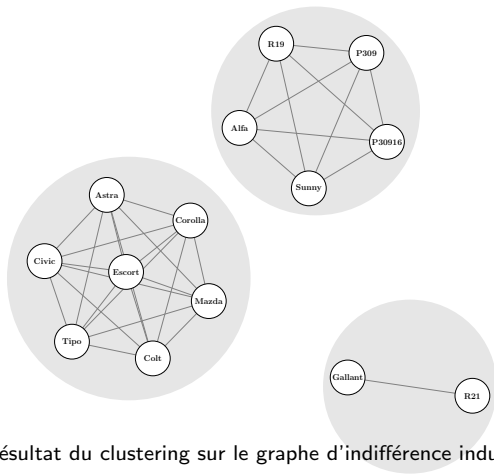
Résultat du clustering sur le graphe d'indifférence original;

Thierry's choice problem

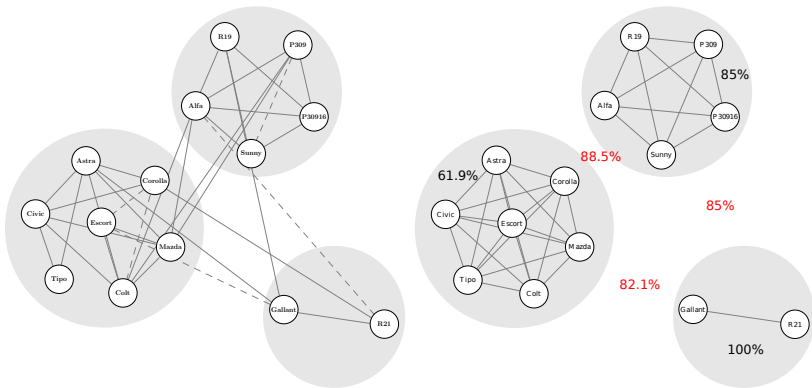


Résultat du clustering sur le graphe d'indifférence original;

Thierry's choice problem

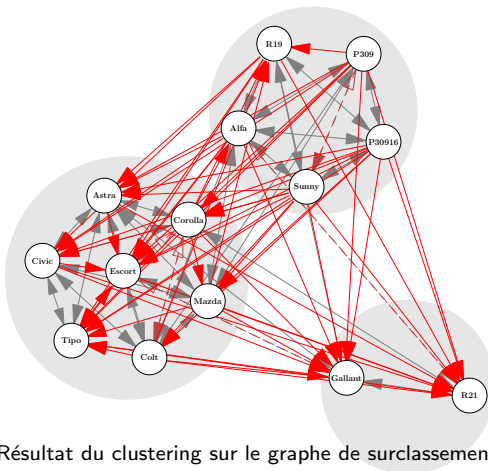


Thierry's choice problem



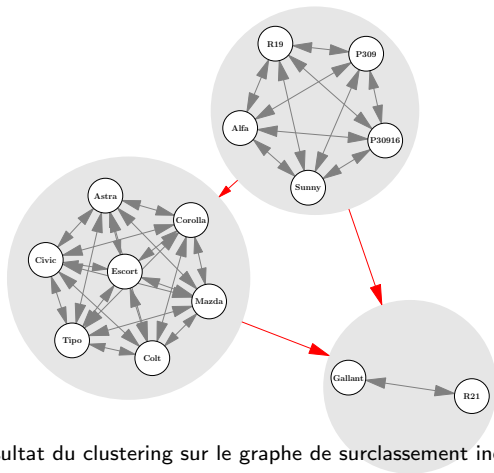
Graphes d'indifférence original et induit; Proximité:80.7%;

Thierry's choice problem

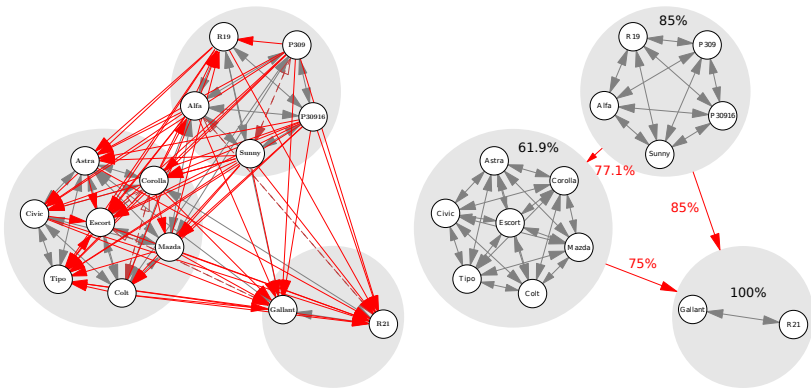


Résultat du clustering sur le graphe de surclassement;

Thierry's choice problem



Thierry's choice problem



Graphes de surclassement original et induit; Proximité:75.2%;

Résultats comparatifs

# Clusters	K-means	PAM	PAM on q^*	GAMMA-P
1	33.5	33.5	33.5	-
2	35.0	46.7	40.0	-
3	45.4	48.1	45.8	75.2
4	52.1	51.3	50.9	-
5	57.0	55.5	54.9	-
6	60.5	58.5	58.0	-
7	61.9	61.0	60.0	-
8	63.6	62.3	61.6	-
9	64.8	65.7	63.6	-
10	65.6	66.9	64.0	-
11	65.6	65.8	64.8	-
12	65.8	66.0	65.4	-
13	65.4	67.6	65.4	-
14	66.4	66.4	66.4	-

Résultats moyens de 100 essais: $|o^* - o'|$ (%);

Thank you