Capacitive MR-Sort model

Preference modeling and learning

Olivier Sobrie[1,2] - Vincent Mousseau[1] - Marc Pirlot[2]

[1]École Centrale de Paris - Laboratoire de Génie Industriel
[2]University of Mons - Faculty of engineering

November 20, 2014

1 **Introductory example**

2 MR-Sort

3 Capacitive MR-Sort

4 Learning a Capacitive MR-Sort model

5 Experimentations

6 Comments and Conclusion

Introductory example

- Admission/Refusal of student
- Students are evaluated in 4 courses
- Admission condition : score above 10/20 in all the courses of one the minimal winning coalitions.

Minimal winning coalitions

- {math, physics}
- {math, chemistry}
- {chemistry, history}

Maximal loosing coalitions

- {math, history}
- {physics, chemistry}
- {physics, history}

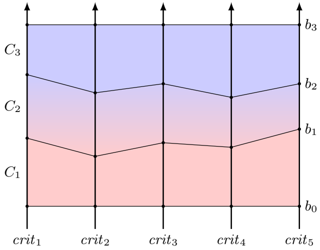
	Math	Physics	Chemistry	History	A/R
James	11	11	9	9	A
Marc	11	9	11	9	A
Robert	9	9	11	11	A
John	11	9	9	11	R
Paul	9	11	9	11	R
Pierre	9	11	11	9	R

MR-Sort I

Characteristics

- Allows to sort alternatives in ordered classes on basis of their performances on monotone criteria
- MCDA method based on outranking relations
- Simplified version of ELECTRE TRI

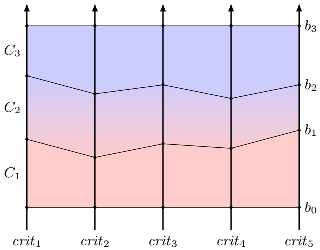
Parameters



- Profiles performances ($b_{h,j}$ for $h = 1, ..., p-1; j = 1, ..., n$)
- Criteria weights ($w_j \geq 0$ for $n = 1, ..., n$)
- Majority threshold (λ)

MR-Sort II

Parameters



- Profiles performances ($b_{h,j}$ for $h = 1, ..., p-1; j = 1, ..., n$)
- Criteria weights ($w_j \geq 0$ for $n = 1, ..., n$)
- Majority threshold (λ)

Assignment rule

$$a \in C_h \iff \sum_{j:a_j \geq b_{h-1,j}} w_j \geq \lambda \text{ and } \sum_{j:a_j \geq b_{h,j}} w_j < \lambda$$

MR-Sort applied to the examples

- Profile fixed at 10/20 on each criterion
- Admission condition : score above 10/20 in all the courses of one the minimal winning coalitions :
  - {math, physics}
  - {math, chemistry}
  - {chemistry, history}

$$\Rightarrow \begin{cases} w_{\text{math}} + w_{\text{physics}} \geq \lambda \\ w_{\text{math}} + w_{\text{chemistry}} \geq \lambda \\ w_{\text{chemistry}} + w_{\text{history}} \geq \lambda \end{cases}$$

- Maximal loosing coalitions :
  - {math, history}
  - {physics, chemistry}
  - {physics, history}

$$\Rightarrow \begin{cases} w_{\text{math}} + w_{\text{history}} < \lambda \\ w_{\text{physics}} + w_{\text{chemistry}} < \lambda \\ w_{\text{physics}} + w_{\text{history}} < \lambda \end{cases}$$

- $w_{\text{math}} + w_{\text{physics}} + w_{\text{chemistry}} + w_{\text{history}} = 1$

MR-Sort applied to the examples

- Profile fixed at 10/20 on each criterion
- Admission condition : score above 10/20 in all the courses of one the minimal winning coalitions :
  - {math, physics}
  - {math, chemistry}
  - {chemistry, history}

$$\Rightarrow \begin{cases} w_{\text{math}} + w_{\text{physics}} \geq \lambda \\ w_{\text{math}} + w_{\text{chemistry}} \geq \lambda \\ w_{\text{chemistry}} + w_{\text{history}} \geq \lambda \end{cases}$$

- Maximal loosing coalitions :
  - {math, history}
  - {physics, chemistry}
  - {physics, history}

$$\Rightarrow \begin{cases} w_{\text{math}} + w_{\text{history}} < \lambda \\ w_{\text{physics}} + w_{\text{chemistry}} < \lambda \\ w_{\text{physics}} + w_{\text{history}} < \lambda \end{cases}$$

- $w_{\text{math}} + w_{\text{physics}} + w_{\text{chemistry}} + w_{\text{history}} = 1$
- $w_{\text{math}} + w_{\text{physics}} \geq \lambda$ and $w_{\text{chemistry}} + w_{\text{history}} \geq \lambda \Rightarrow \lambda \leq \frac{1}{2}$

MR-Sort applied to the examples

- Profile fixed at 10/20 on each criterion
- Admission condition : score above 10/20 in all the courses of one the minimal winning coalitions :
  - {math, physics}
  - {math, chemistry}
  - {chemistry, history}

$$\Rightarrow \begin{cases} w_{\text{math}} + w_{\text{physics}} \geq \lambda \\ w_{\text{math}} + w_{\text{chemistry}} \geq \lambda \\ w_{\text{chemistry}} + w_{\text{history}} \geq \lambda \end{cases}$$

- Maximal loosing coalitions :
  - {math, history}
  - {physics, chemistry}
  - {physics, history}

$$\Rightarrow \begin{cases} w_{\text{math}} + w_{\text{history}} < \lambda \\ w_{\text{physics}} + w_{\text{chemistry}} < \lambda \\ w_{\text{physics}} + w_{\text{history}} < \lambda \end{cases}$$

- $w_{\text{math}} + w_{\text{physics}} + w_{\text{chemistry}} + w_{\text{history}} = 1$
- $w_{\text{math}} + w_{\text{physics}} \geq \lambda$ and $w_{\text{chemistry}} + w_{\text{history}} \geq \lambda \Rightarrow \lambda \leq \frac{1}{2}$
- $w_{\text{math}} + w_{\text{history}} < \lambda$ and $w_{\text{physics}} + w_{\text{chemistry}} < \lambda \Rightarrow \lambda > \frac{1}{2}$

MR-Sort applied to the examples

- Profile fixed at 10/20 on each criterion
- Admission condition : score above 10/20 in all the courses of one the minimal winning coalitions :
  - {math, physics}
  - {math, chemistry}
  - {chemistry, history}

$$\Rightarrow \begin{cases} w_{\text{math}} + w_{\text{physics}} \geq \lambda \\ w_{\text{math}} + w_{\text{chemistry}} \geq \lambda \\ w_{\text{chemistry}} + w_{\text{history}} \geq \lambda \end{cases}$$

- Maximal loosing coalitions :
  - {math, history}
  - {physics, chemistry}
  - {physics, history}

$$\Rightarrow \begin{cases} w_{\text{math}} + w_{\text{history}} < \lambda \\ w_{\text{physics}} + w_{\text{chemistry}} < \lambda \\ w_{\text{physics}} + w_{\text{history}} < \lambda \end{cases}$$

- $w_{\text{math}} + w_{\text{physics}} + w_{\text{chemistry}} + w_{\text{history}} = 1$
- $w_{\text{math}} + w_{\text{physics}} \geq \lambda$ and $w_{\text{chemistry}} + w_{\text{history}} \geq \lambda \Rightarrow \lambda \leq \frac{1}{2}$
- $w_{\text{math}} + w_{\text{history}} < \lambda$ and $w_{\text{physics}} + w_{\text{chemistry}} < \lambda \Rightarrow \lambda > \frac{1}{2}$
- Impossible to represent all the coalitions with a MR-Sort model

Capacitive MR-Sort

Characteristic

- Take criteria interactions into account
- Improvement of the expressivity of the model
- Non Compensatory Sorting Model [Bouyssou and Marchant, 2007]

Capacity

- $F = \{1, ..., n\}$: set of criteria
- A capacity is a function $\mu : 2^F \to [0, 1]$ such that :
  - $\mu(B) \geq \mu(A)$, for all $A \subseteq B \subseteq F$ (monotonicity) ;
  - $\mu(\emptyset) = 0$ and $\mu(F) = 1$ (normalization).

New assignment rule

$$a \in C_h \iff \mu(\{j \in F : a_j \geq b_{h-1,j}\}) \geq \lambda \quad \text{and} \quad \mu(\{j \in F : a_j \geq b_{h,j}\}) < \lambda$$

Learning a Capacitive MR-Sort model - MIP I

Mixed Integer Programming

- Objective : Finding a model compatible with as much example as possible
- MIP to learn an MR-Sort model in [Leroy et al., 2011]
- Limitation to 2-additive capacities
- For Capacitive MR-Sort, more constraints and binary variable are required

Table: Max number of constraints

	MIP MR-Sort	MIP Capacitive MR-Sort
# binary variables	$n(2m+1)$	$n(2m+1+2m(m+1))$
# constraints	$2n(5m+1)+n(p-3)+1$	$2n(5m+1)+n(p-3)+1+2m(n^2+1)+n^2$

- Too much variables and constraints to be used with large datasets

Learning a Capacitive MR-Sort model - MIP II

Application to the introductory example

- Admission condition : score above 10/20 in all the courses of one these coalitions :
  - {math, physics}
  - {math, chemistry}
  - {chemistry, history}
- MIP is able to find a model matching all the rules

J	$m(J)$
{math}	0
{physics}	0
{chemistry}	0
{history}	0

$\lambda = 0.3$

J	$m(J)$
{math, physics}	0.3
{math, chemistry}	0.3
{math, history}	0
{physic, chemistry}	0
{physic, history}	0
{chemistry, history}	0.4

Learning a Capacitive MR-Sort model - Meta I

Metaheuristic to learn a Capacitive MR-Sort model

- Objective : Finding a model compatible with as much example as possible
- Being able to handle large datasets

Recall : Metaheuristic to learn parameters of a MR-Sort model

- Sobrie, O., Mousseau, V., and Pirlot, M. (2012). Learning the parameters of a multiple criteria sorting method from large sets of assignment examples.
In *DA2PL 2012 Workshop From Multiple Criteria Decision Aid to Preference Learning*, pages 21–31.
Mons, Belgique
- Sobrie, O., Mousseau, V., and Pirlot, M. (2013). Learning a majority rule model from large sets of assignment examples.
In Perny, P., Pirlot, M., and Tsoukiás, A., editors, *Algorithmic Decision Theory*, pages 336–350. Springer

Learning a Capacitive MR-Sort model - Meta II

Recall : Metaheuristic to learn a MR-Sort model

- Principle (genetic algorithm) :
  - Initialize a population of MR-Sort models
  - Evolve the population by iteratively
    - Optimizing weights (profiles fixed) with a LP
    - Improving profiles (weights fixed) with a heuristic
    - Selecting the best models and reinitializing the others
  - ... to get a "good" MR-Sort model in the population
- Stopping criteria :
  - If one of the models restores all examples
  - Or after N iterations

Learning a Capacitive MR-Sort model - Meta II

Recall : Metaheuristic to learn a MR-Sort model

- Principle (genetic algorithm) :
  - Initialize a population of MR-Sort models
  - Evolve the population by iteratively
    - Optimizing weights (profiles fixed) with a LP
    - Improving profiles (weights fixed) with a heuristic
    - Selecting the best models and reinitializing the others
  - ... to get a "good" MR-Sort model in the population
- Stopping criteria :
  - If one of the models restores all examples
  - Or after N iterations

Metaheuristic to learn a Capacitive MR-Sort model

- Adaptation of the LP to learn capacities and adaptation of the heuristic

Learning a Capacitive MR-Sort model - Meta III

Linear Program to learn the capacities and the majority threshold

- Fixed profiles
- Expression of the capacities with the Möbius transform
  $\mu(A) = \sum_{B \subseteq A} m(B)$, for all $A \subseteq F$, with $m(B)$ defined as :
  $m(B) = \sum_{C \subseteq B} (-1)^{|B|-|C|} \mu(C)$
- Limitation to 2-additive capacities in view of limitting the number of variables and constraints

$$\mu(A) = \sum_{i \in A} m(\{i\}) + \sum_{\{i,j\} \in A} m(\{i,j\})$$

- Minimization of a slack that tends to be equal to 0 when all examples are correctly assigned

Learning a Capacitive MR-Sort model - Meta III

Linear Program to learn the capacities and the majority threshold

$$
\begin{cases}
\min & \sum_{a \in A} (x'_a + y'_a) & \\
\sum_{j:a_j \geq b_{h-\mathbf{1},j}}^{n} \left(m_j + \sum_{k:a_k \geq b_{h-\mathbf{1},k}}^{j} m_{j,k} \right) - x_a + x'_a = \lambda & & \forall a \in A_h, \forall h \in P \backslash \{1\} \\
\sum_{j:a_j \geq b_{h,j}}^{n} \left(m_j + \sum_{k:a_k \geq b_{h,k}}^{j} m_{j,k} \right) + y_a - y'_a = \lambda - \varepsilon & & \forall a \in A_h, \forall h \in P \backslash \{p-1\} \\
\sum_{j=\mathbf{1}}^{n} m_j + \sum_{j=\mathbf{1}}^{n} \sum_{k=\mathbf{1}}^{j} m_{j,k} = 1 & & \\
m_j + \sum_{k \in J} m_{j,k} \geq 0 & & \forall j \in F, \forall J \subseteq F \backslash \{j\} \\
\lambda \in [0.5; 1] & & \\
m_j \in [0, 1] & & \forall j \in F \\
m_{j,k} \in [-1, 1] & & \forall j \in F, \forall k \in F, k < j \\
x_a, y_a, x'_a, y'_a \in \mathbb{R}_{\mathbf{0}}^{+} & & a \in A.
\end{cases}
$$

Learning a Capacitive MR-Sort model - Meta IV

Heuristic to adjust the profiles

- Fixed Möbius indices and majority threshold
- Principle of the heuristic : moving the profile in view of increasing the number of alternatives correctly assigned
- Multiple iterations over each profile and each criterion
- Same principles as in [Sobrie et al., 2013], adapted for capacities instead of weights

Experimentations I

Dataset	#instances	#attributes	#categories
DBS	120	8	2
CPU	209	6	4
BCC	286	7	2
MPG	392	7	36
ESL	488	4	9
MMG	961	5	2
ERA	1000	4	4
LEV	1000	4	5
CEV	1728	6	4

- Instances split in two parts : learning and generalization sets
- Binarization of the categories

Source : [Tehrani et al., 2012]

Experimentations II

Average Classification Accuracy

Dataset	META MR-Sort	META Capa-MR-Sort
DBS	0.8400 ± 0.0456	0.8306 ± 0.0466
CPU	0.9270 ± 0.0294	0.9203 ± 0.0315
BCC	0.7271 ± 0.0379	0.7262 ± 0.0377
MPG	0.8174 ± 0.0290	0.8167 ± 0.0468
ESL	0.8992 ± 0.0195	0.9018 ± 0.0172
MMG	0.8303 ± 0.0154	0.8318 ± 0.0121
ERA	0.6905 ± 0.0192	0.6927 ± 0.0165
LEV	0.8454 ± 0.0221	0.8445 ± 0.0223
CEV	0.9217 ± 0.0067	0.9187 ± 0.0153

- 50% of the dataset used as learning set
- Results are not convincing, overfitting ?

Experimentations II

Average Classification Accuracy

Dataset	META MR-Sort	META Capa-MR-Sort
DBS	0.9318 ± 0.0036	0.9247 ± 0.0099
CPU	0.9761 ± 0.0000	0.9694 ± 0.0072
BCC	0.7737 ± 0.0013	0.7700 ± 0.0077
MPG	0.8418 ± 0.0000	0.8418 ± 0.0000
ESL	0.9180 ± 0.0000	0.9180 ± 0.0000
MMG	0.8491 ± 0.0011	0.8508 ± 0.0005
ERA	0.7142 ± 0.0028	0.7158 ± 0.0004
LEV	0.8650 ± 0.0000	0.8650 ± 0.0000
CEV	0.9225 ± 0.0000	0.9225 ± 0.0000

- Full dataset used as learning set
- Results are not convincing

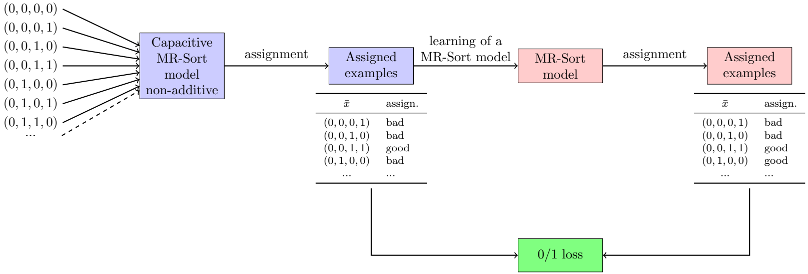
Comments I

What to conclude after the experiments ?

- Expressivity of the model is not so much improved ?
- Algorithm not well adapted ?
- To what extent MR-Sort approximates non-additive learning sets ?

Comments II

To what extent MR-Sort approximates non-additive learning sets ?



- Generation of 2^n binary vectors of performances
- Generation of Capacitive MR-Sort model non-additive and assignment
- Learning of a MR-Sort model from assignment
- Test with all the non-additive models

Comments III

To what extent MR-Sort approximates non-additive learning sets ?

n	$D(n)$	% non-additive	0/1 loss		
			min.	max.	avg.
4	168	11 %	1/16	1/16	1/16
5	7 581	57 %	1/32	3/32	1.26/32
6	7 828 354	97 %	1/64	8/64	2.73/64

- Few alternatives are incorrectly assigned

Conclusion

- For problems involving small number of criteria (< 7), we don't win so much in expressivity with Capacitive MR-Sort
- Metaheuristic can be improved to better deal with interactions
- Tests with datasets in which there exist interactions between criteria

Thank you for your attention !

References I

Bouyssou, D. and Marchant, T. (2007).
An axiomatic approach to noncompensatory sorting methods in MCDM, I : The case of two categories.
European Journal of Operational Research, 178(1) :217–245.

Leroy, A., Mousseau, V., and Pirlot, M. (2011).
Learning the parameters of a multiple criteria sorting method.
In Brafman, R., Roberts, F., and Tsoukiàs, A., editors, *Algorithmic Decision Theory*, volume 6992 of *Lecture Notes in Computer Science*, pages 219–233. Springer Berlin / Heidelberg.

References II

Sobrie, O., Mousseau, V., and Pirlot, M. (2012).
Learning the parameters of a multiple criteria sorting method from large sets of assignment examples.
In *DA2PL 2012 Workshop From Multiple Criteria Decision Aid to Preference Learning*, pages 21–31.
Mons, Belgique.

Sobrie, O., Mousseau, V., and Pirlot, M. (2013).
Learning a majority rule model from large sets of assignment examples.
In Perny, P., Pirlot, M., and Tsoukiás, A., editors, *Algorithmic Decision Theory*, pages 336–350. Springer.

Tehrani, A. F., Cheng, W., Dembczynski, K., and Hüllermeier, E. (2012).
Learning monotone nonlinear models using the Choquet integral.
Machine Learning, 89(1-2) :183–211.