Decision Methods and Models

Master's Degree in Computer Science

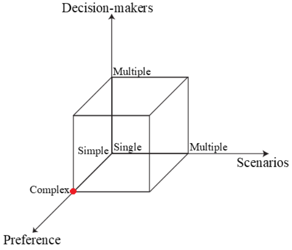
Roberto Cordone
DI - Università degli Studi di Milano

Schedule:	Thursday 16.30 - 18.30 in Aula Magna (CS department)
	Friday 12.30 - 14.30 in classroom 301
Office hours:	on appointment
E-mail:	roberto.cordone@unimi.it
Web page:	https://homes.di.unimi.it/cordone/courses/2024-mmd/2024-mmd.html
Ariel site:	https://myariel.unimi.it/course/view.php?id=4467

Nontransitive preference relations

We assume

- a preference relation Π that is a nontransitive generalisation of Paretian preference
- a certain environment: $|\Omega| = 1 \Rightarrow f(x, \bar{\omega})$ reduces to $f(x)$
- a single decision-maker: $|D| = 1 \Rightarrow \Pi_d$ reduces to Π



The idea here is to

- apply the Paretian preference definition (indicators as costs/benefits)
- extend it with additional impact pairs generated by simple criteria (not individually)

We will lose transitivity!

Why an extension of Pareto?

Paretian preference is

- very natural: many indicators can be turned into costs or benefits
- irrealistically poor: it does not take indifference curves to choose between "incomparable" trips such as

$$\begin{bmatrix} 60 \text{ min} \\ 1\,000 \text{ Euros} \end{bmatrix} \text{ and } \begin{bmatrix} 61 \text{ min} \\ 1 \text{ Euros} \end{bmatrix}$$

61 *minutes is "equal" to* 60

We want

- a richer preference: more comparability
- no number crunching: no combination of estimated values

Relax Paretian preference, allowing preferences with small losses

Roy (1968) proposed a family of methods named:

- ÉLimination: elimination
- Et: and
- Choix: choice
- Traduisant: translating
- la RÉalité: reality

with an extension of Paretian preference called outranking (surclassage)

They focus on finite problems

- impacts and alternatives correspond one to one
- preference (on impacts) and dominance (on solutions) are interchangeable

They are among the legally adopted methods to choose
the most advantageous offer in competitive bids for public works

(*see Italian GU 25/05/2018 pp. 59–66*)

The outranking relation

Human beings have a limited capacity to discriminate similar impacts

The outranking relation starts with a basic definition

$$f \preceq_S f' \quad \Leftrightarrow \quad f_l \geq f'_l - \epsilon_l \quad \text{for all } l \in P$$

where $\epsilon_l \geq 0$ are the discrimination thresholds of the decision-maker: values of f_l differing by less than ϵ_l are considered indifferent

The values ϵ_l are set by specific field experts

$f \preceq_S f'$ does no longer mean "exchanging f with f' is acceptable", but "exchanging f and f' is not a clear loss" (even if all f_l can worsen)

Notice that

- $\epsilon_l = 0$ yields the Paretian preference
- $\epsilon_l = +\infty$ yields complete indifference

Higher values of ϵ_l make the preference richer (set S gets larger)

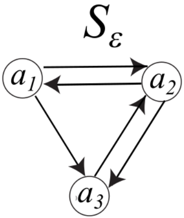
Example: comparability thresholds

Consider the following evaluation matrix and set $\epsilon_l = 0.5$ for all $l \in P$

	f_1	f_2	f_3	f_4
a_1	1	0.7	0.6	0.8
a_2	0.8	0.5	1	1
a_3	0.4	1	0.6	0.6

The outranking relation includes all pairs except for (a_3, a_1)

$$S_\epsilon = \{(a_1, a_1), (a_1, a_2), (a_1, a_3), (a_2, a_1), (a_2, a_2), (a_2, a_3), (a_3, a_2), (a_3, a_3)\}$$



Properties

The basic outranking relation is

- reflexive: $f_l \geq f'_l - \epsilon_l$ for all $l \in P \Leftrightarrow f \preceq_S f'$
- noncomplete: $\begin{bmatrix} 0 \\ 1 \end{bmatrix} \bowtie \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ for any threshold $\epsilon_1 < 1$ or $\epsilon_2 < 1$
- nontransitive: given thresholds $\epsilon_1 = \epsilon_2 = 1$,

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \preceq_S \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 2 \end{bmatrix} \preceq_S \begin{bmatrix} 2 \\ 3 \end{bmatrix} \text{ (only slightly worse)}$$

but $\begin{bmatrix} 0 \\ 1 \end{bmatrix} \not\preceq_S \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ (the worsening is too large)

See the thought experiment about coffee cups in lesson 3

- nonantisymmetric: given thresholds $\epsilon_1 = \epsilon_2 = 1$,

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \preceq_S \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 0 \end{bmatrix} \preceq_S \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

but $\begin{bmatrix} 0 \\ 1 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Refinements

The basic definition is often too loose, but it can be progressively refined

- always keeping the Paretian impact pairs
- introducing additional information to filter out some impact pairs

so that $\Pi_{\mathrm{Pareto}} \subseteq S \subseteq S_\epsilon$

The additional information will use indicator values and indicator weights

- avoiding any numerical combination of the two informations
- always manipulating them separately

Actually, some ELECTRE methods violate this principle

Concordance condition

A second way to define the outranking relation is to

1. assign weights to the indicators:

$$w_l \geq 0 \text{ for all } l \in P \text{ with } \sum_{l \in P} w_l = 1$$

2. treat the indicators as electors
3. hold a weighted election between impact pairs

We compute

- the total weight of the indicators for which f is better than f':

$$w^{+}_{ff'} = \sum_{l \in P : f_l > f'_l} w_l$$

- the total weight of the indicators for which f is indifferent to f':

$$w^{=}_{ff'} = \sum_{l \in P : f_l = f'_l} w_l$$

These two groups of indicators "vote" in favour of $f \preceq f'$

Assume weight vector

w	0.3	0.4	0.2	0.1

for the already considered evaluation matrix

	f_1	f_2	f_3	f_4
a_1	1	0.7	0.6	0.8
a_2	0.8	0.5	1	1
a_3	0.4	1	0.6	0.6

The concordance matrix C collects the "election" results $c_{ff'} = w^+_{ff'} + w^=_{ff'}$

$$C = \begin{bmatrix} 1 & 0.7 & 0.6 \\ 0.3 & 1 & 0.6 \\ 0.6 & 0.4 & 1 \end{bmatrix}$$

The concordance matrix always enjoys the following properties:

- its main diagonal elements are $c_{ff} = 1$ for all f in F
- the sum of transposed element pairs is $c_{ff'} + c_{f'f} = 1 + w^=_{ff'} \geq 1$

Concordance condition

The concordance matrix provides a second outranking relation

$$f \preceq_{S_c} f' \Leftrightarrow c_{ff'} \geq \alpha_c$$

through the user-defined concordance threshold $\alpha_c \in [0, 1]$, that indicates how much the indicators should support the outranking

Notice that

- $\alpha_c = 1$ yields the Paretian preference
- $\alpha_c = 0$ yields complete indifference

This is similar to $\epsilon_l = 0$ or $+\infty$, but uncorrelated

Example

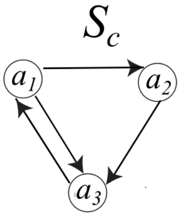
Assume $\alpha_c = 0.5$ for problem

	f_1	f_2	f_3	f_4
w	0.3	0.4	0.2	0.1

a_1	1	0.7	0.6	0.8
a_2	0.8	0.5	1	1
a_3	0.4	1	0.6	0.6

The outranking relation is

$$S_c = \{(a_1, a_1), (a_1, a_2), (a_1, a_3), (a_2, a_2), (a_2, a_3), (a_3, a_1), (a_3, a_3)\}$$



Disordance condition

Yet another way to define the outranking relation is to

1. focus on the indicators that oppose a preference: $\{l \in P : f_l \not\gtrsim f'_l\}$
2. measure the strength of the opposition: $f'_l - f_l$

   (*considering the two values, not the weight!*)
3. compare it with the overall maximum difference: $\max_{l \in P} |f_l - f'_l|$

On this basis, we define the discordance matrix D as

$$d_{ff'} = \frac{\max_{l \in P} [\max(f'_l - f_l, 0)]}{\max_{l \in P} |f'_l - f_l|}$$

setting $d_{ff'} = 0$ when $f = f'$ to avoid the indeterminate form $0/0$

Therefore, the main diagonal elements are zero by definition

Example

The usual evaluation matrix

	f_1	f_2	f_3	f_4
a_1	1	0.7	0.6	0.8
a_2	0.8	0.5	1	1
a_3	0.4	1	0.6	0.6

implies the following discordance matrix D

$$D = \begin{bmatrix} 0 & \frac{(1-0.6)}{(1-0.6)} & \frac{(1-0.7)}{(1-0.4)} \\ \frac{(1-0.8)}{(1-0.6)} \text{ or } \frac{(0.7-0.5)}{1-0.6} & 0 & \frac{(1-0.5)}{(1-0.5)} \\ \frac{(1-0.4)}{(1-0.4)} & \frac{(1-0.6)}{(1-0.5)} \text{ or } \frac{(0.8-0.4)}{(1-0.5)} & 0 \end{bmatrix}$$

that is

$$D = \begin{bmatrix} 0 & 1 & 0.5 \\ 0.5 & 0 & 1 \\ 1 & 0.8 & 0 \end{bmatrix}$$

Discordance condition

The discordance matrix provides a third outranking relation

$$f \preceq_{S_d} f' \Leftrightarrow d_{ff'} \leq 1 - \alpha_d$$

through the user-defined discordance threshold $\alpha_d \in [0, 1]$, that indicates how much the indicators can oppose the outranking at most

Notice that

- $\alpha_d = 1$ yields the Paretian preference
- $\alpha_d = 0$ yields complete indifference

Note: the condition is usually defined as $f \preceq_{S_d} f' \Leftrightarrow d_{ff'} \leq \alpha_d$, but the expression used in this course keeps the nice property above

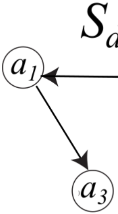
Example

Assuming $\alpha_d = 0.5$ on

$$D = \begin{bmatrix} 0 & 1 & 0.5 \\ 0.5 & 0 & 1 \\ 1 & 0.8 & 0 \end{bmatrix}$$

the outranking relation is

$$S_d = \{(a_1, a_1), (a_1, a_3), (a_2, a_1), (a_2, a_2), (a_3, a_3)\}$$



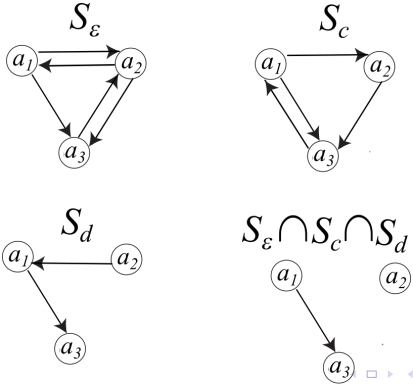
The overall outranking relation

Since the different definitions capture different aspects of the problem, we intersect the three definitions to refine the overall outranking relation

$$S = S_\epsilon \cap S_c \cap S_d$$

In the example, we obtain

$$S = \{(a_1, a_1), (a_1, a_3), (a_2, a_2), (a_3, a_3)\}$$



Threshold tuning

The threshold parameters α_c and α_d are arbitrary

A reasonable tuning to obtain an intermediate number of impact pairs is the average of the coefficients with which they are compared

- $\alpha_c = \frac{\sum_{f \in F} \sum_{f' \in F \setminus \{f\}} c_{ff'}}{|F|(|F|-1)}$
- $\alpha_d = 1 - \frac{\sum_{f \in F} \sum_{f' \in F \setminus \{f\}} d_{ff'}}{|F|(|F|-1)}$

but in this way S becomes dependent on the set of alternatives

Rank reversal becomes possible!

The filtering process: finding the kernel

Once obtained, the outranking relation is used to filter solutions out, keeping into account that it is weaker than the Paretian preference

- not outranked solutions are certainly interesting
- outranked solutions are not necessarily bad
- solutions strictly outranked by good solutions are not very useful
  (*even if good, they can be replaced by the ones that outrank them*)

This directly suggests an algorithm to reduce X to a kernel $K \subseteq X$ that provides a representative subset of solutions (not too bad, not too similar)

Algorithm FindKernel(X, S)

$K := \emptyset;$

While $(K \subset X)$ *do*

{ select all solutions not strictly outranked }

$K := K \cup \{x \in X : \nexists x' \in X : x' \prec_S x\};$

{ remove all solutions strictly outranked by the kernel }

$X := X \setminus \{x \in X : \exists x' \in K : x' \prec_S x\};$

EndWhile;

Return K;

Example

The evaluation matrix

	f_1	f_2	f_3	f_4
A	0.50	0.50	0.50	0.60
B	0.45	0.45	0.45	0.70
C	0.40	0.40	0.40	0.80
D	0.35	0.35	0.35	0.90

with $\epsilon_l = 0.05$ for all $l \in P$ implies the outranking relation

$$S = \{(A, A), (B, A), (B, B), (C, B), (C, C), (D, C), (D, D)\}$$

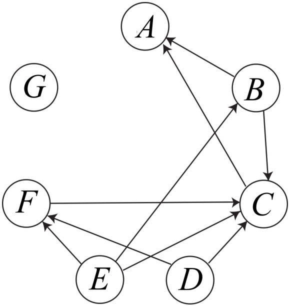
The algorithm performs the following steps

1. $K : \emptyset$
2. no solution strictly outranks D: add it to the kernel ($K = \{D\}$
3. now the kernel outranks C: remove it ($X = \{A, B, D\}$)
4. no solution strictly outranks B: add it to the kernel ($K = \{D, B\}$
5. now the kernel outranks A: remove it ($X = \{B, D\}$)
6. all remaining solutions are in the kernel: stop

The idea is to keep a subset of sufficiently different representatives

Example

Consider the following outranking relation (self-loops removed)



The algorithm performs the following steps

1. $K := \emptyset$
2. add D, E and G to the kernel: $K = \{D, E, G\}$
3. remove B, C and F from the solution set: $X = \{A, D, E, G\}$
4. add A to the kernel: $K = \{A, D, E, G\}$
5. stop

Problem

The definition of kernel has a strong intrinsic limitation:
a kernel is not guaranteed to exist

The "algorithm" could not terminate on cyclic graph

It terminates when the cycles are broken by strict outranking

ADD SOME NICE EXAMPLE

Some solutions have been proposed

- tweak the thresholds ϵ_l, α_c, α_d (*this might be ineffective*)
- find a circuit not strictly outranked and merge it into a macronode (*this always works, but might yield the whole feasible region*)

Building a weak order

If a selection of solutions is not enough, a weak order can be derived

Several methods to this purpose have been proposed, such as

- the forward, backward and combined topological ordering approaches are based on relation S
- the concordance index approach is based on matrix C
- the discordance index approach is based on matrix D

All of these methods usually provide different results
as they have a large amount of arbitrariness

The topological ordering approaches

These methods require an acyclic outranking graph (X, S)

The forward approach builds one of its possible topological orderings

1. define an empty list L_f
2. find an arbitrary solution $x \in X$ not strictly outranked in S
3. append x to L_f
4. remove x from X and all arcs (x, y) from S
5. if the graph is nonempty go to step 2
6. return list L_f

The backward approach proceeds similarly, with two differences

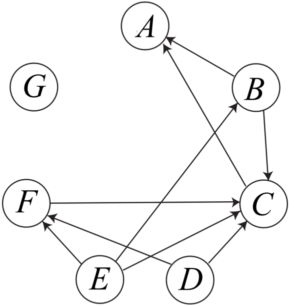
- finds an arbitrary solution that does not outrank any other one
- reverses the list L_b thus obtained before returning it

The combined approach

1. applies the forward approach, obtaining L_f
2. computes the Borda count on L_f, assigning to each solution the number of not preceding ones
3. applies the backward approach, obtaining L_b
4. computes the Borda count on L_b, assigning to each solution the number of not preceding ones
5. sums the two Borda counts to obtain a weak order

Example: the forward approach

Consider the following outranking relation (self-loops removed)



We can summarise its possible forward topological orderings as

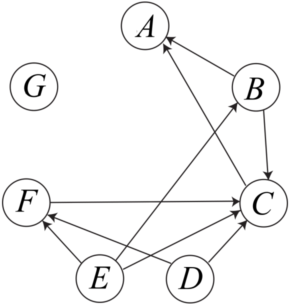
$$L_f = (\{D, E, G\}, \{B, F\}, \{C\}, \{A\})$$

Instead of choosing one, we combine them into a single Borda count, assigning to each solution the arithmetic mean of its possible values

x	A	B	C	D	E	F	G
$B_f(x)$	1	3.5	2	6	6	3.5	6

Example: the backward approach

Consider the following outranking relation (self-loops removed)



The backward topological orderings are

$$L_f = (\{D, E\}, \{B, F\}, \{C\}, \{A, G\})$$

and the resulting Borda count is

x	A	B	C	D	E	F	G
$B_b(x)$	1.5	4.5	3	6.5	6.5	4.5	1.5

Example: the combined approach

Finally, the combined approach yields

x	A	B	C	D	E	F	G
$B_f(x)$	1	3.5	2	6	6	3.5	6
$B_b(x)$	1.5	4.5	3	6.5	6.5	4.5	1.5
$B(x)$	2.5	8	5	12.5	12.5	10	7.5

that implies

$$L_f = (\{D, E\}, \{B, F\}, \{G\}, \{C\}, \{A\})$$

The concordance index approach

The concordance matrix measures the “strength” of the outranking between impact pairs (based on indicator weights)

The idea is to

1. aggregate all values associated to impact f into an index
   - summing the indices on the row
   - subtracting the indices on the column

$$C_f = \sum_{g \in F} c_{fg} - \sum_{g \in F} c_{gf}$$

   By definition $C_f \in [1-n, n-1]$

2. use C_f as a value function to build a weak order

(*the larger the better*)

Example

Given the concordance matrix

$$C = \begin{bmatrix} 1 & 0.7 & 0.6 \\ 0.3 & 1 & 0.6 \\ 0.6 & 0.4 & 1 \end{bmatrix}$$

we obtain

$$\begin{aligned} C_1 &= (1 + 0.7 + 0.6) - (1 + 0.3 + 0.6) &= 0.4 \\ C_2 &= (0.3 + 1 + 0.6) - (0.7 + 1 + 0.4) &= -0.2 \\ C_3 &= (0.6 + 0.4 + 1) - (0.6 + 0.6 + 1) &= -0.2 \end{aligned}$$

that implies

$$a_1 \prec a_2 \sim a_3$$

The discordance index approach

The discordance matrix measures the "opposition" to the outranking between impact pairs (based on indicator values)

The idea is to

1. aggregate all values associated to impact f into an index
   - summing the indices on the row
   - subtracting the indices on the column

$$D_f = \sum_{g \in F} d_{fg} - \sum_{g \in F} d_{gf}$$

   By definition $D_f \in [1-n, n-1]$

2. use D_f as a cost function to build a weak order

   (*the smaller the better*)

Of course, you can combine C_f and D_f if you are keen on fiddling

Example

Given the discordance matrix

$$D = \begin{bmatrix} 0 & 1 & 0.5 \\ 0.5 & 0 & 1 \\ 1 & 0.8 & 0 \end{bmatrix}$$

we obtain

$$\begin{aligned} D_1 &= (0 + 1 + 0.5) - (0 + 0.5 + 1) &&= 0 \\ D_2 &= (0.5 + 0 + 1) - (1 + 0 + 0.8) &&= -0.3 \\ D_3 &= (1 + 0.8 + 0) - (0.5 + 1 + 0) &&= 0.3 \end{aligned}$$

that implies

$$a_2 \prec a_1 \prec a_3$$