Monotone but not boring: some encounters with non-monotonicity

Prof. dr. Bernard De Baets

Ghent University, Belgium

invited talk presented at the Eleventh Internat. Conference on Intelligent Systems Design and Applications

Córdoba (SP), Nov 24, 2011

Starting observations

1. In many modelling problems, there exists a **monotone relationship** between some of the **input variables** and the **output variable**
2. Monotonicity is a common property of **evaluation** and **selection procedures**
3. This monotone relationship may **not** be **fully present** in the **observed input-output data** due to data imperfections
4. Monotonicity is a **global property** in contrast to a local property such as continuity
5. In case the monotonicity property applies, any **violation** of it is simply **unacceptable**

1. Fuzzy rule-based modelling

Soil erosion

Phenomenon: loss of soil by erosion increases with increasing slope angle and decreasing soil coverage with vegetation

(Geoderma, Mitra et al., 1998)

slope angle class	forest	pasture
very large	moderately high	high
large	moderately low	moderately high
medium	low	moderately high
small	low	moderate
very small	low	low

land use class

Increasing, non-smooth rule base

Citrus sudden death

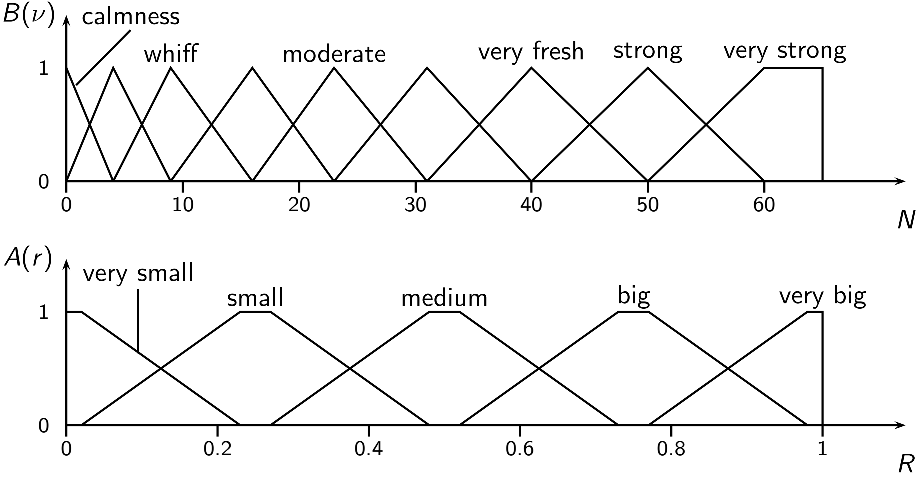
Phenomenon: distance covered by the vector transmitting citrus sudden death to sweet orange trees increases with increasing wind intensity

(Ecological Modelling, M. da Silva Peixoto et al., 2008)

IF	N IS calmness	THEN	R IS very small
IF	N IS breeze	THEN	R IS very small
IF	N IS whiff	THEN	R IS small
IF	N IS weak	THEN	R IS medium
IF	N IS moderate	THEN	R IS medium
IF	N IS fresh	THEN	R IS big
IF	N IS very fresh	THEN	R IS big
IF	N IS strong	THEN	R IS very big
IF	N IS very strong	THEN	R IS very big

Increasing, smooth rule base

Citrus sudden death



Fuzzy rule-based model

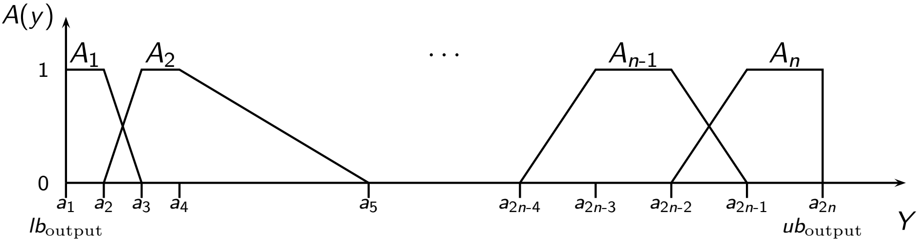
Model characteristics:

- m input variables X_ℓ and a single output variable Y
- rules of the form

$$R_s\text{: IF } X_1 \text{ IS } B^1_{j_{1,s}} \text{ AND } \ldots \text{ AND } X_m \text{ IS } B^m_{j_{m,s}}$$
$$\text{THEN } Y \text{ IS } A_{i_s}$$

- linguistic values $B^\ell_{j_{\ell,s}}$ of X_ℓ: **trapezoidal**; **fuzzy partition**
- linguistic values A_{i_s}: **trapezoidal**; **fuzzy partition** (bounded domain)
- natural ordering on the linguistic values of each variable

Fuzzy partition



Mamdani–Assilian fuzzy models

Observation

Mamdani–Assilian fuzzy models with a monotone rule base do not necessarily result in a monotone input-output mapping

Monotone input-output behaviour

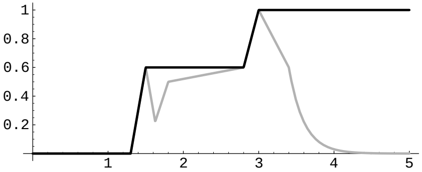
If the original rule base is **complete** and **increasing**, then the input-output mapping can only be **increasing** in the following cases:

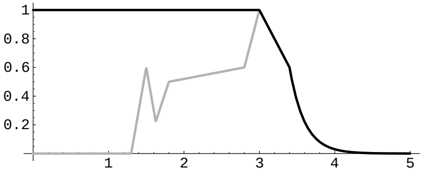
1. **Center-of-Gravity defuzzification**:
    - one input variable: basic t-norms $T_\mathbf{M}$, $T_\mathbf{P}$ and $T_\mathbf{L}$
    - two or three input variables: $T_\mathbf{P}$ and a **smooth** rule base
2. **Mean-of-Maxima defuzzification**:
    - one input variable: basic t-norms $T_\mathbf{M}$, $T_\mathbf{P}$ and $T_\mathbf{L}$
    - two or more input variables: $T_\mathbf{M}$ or $T_\mathbf{P}$, and a **smooth** rule base

Implication-based fuzzy models

Cumulative modifiers:

- at-least modifier: $\text{ATL}(C)(x) = \sup\{C(t) \mid t \leq x\}$
- at-most modifier: $\text{ATM}(C)(x) = \sup\{C(t) \mid t \geq x\}$





Implication-based fuzzy models

Connectives: left-continuous t-norm and its residual implicator

Modified rule bases:

- **ATL/ATM rule base**: applying ATL/ATM to all antecedents and consequents
- **ATLM rule base**: union of the above

Monotone input-output mapping

If the original rule base is **increasing**, then the input-output mapping is **increasing** in the following cases:

1. ATL rule base and **First-of-Maxima defuzzification**
2. ATM rule base and **Last-of-Maxima defuzzification**
3. ATLM rule base and **Mean-of-Maxima defuzzification**

References

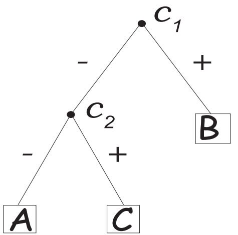
1. E. Van Broekhoven and B. De Baets, *Fast and accurate center of gravity defuzzification of fuzzy system outputs defined on trapezoidal fuzzy partitions*, Fuzzy Sets and Systems **157** (2006), 904–918.
2. E. Van Broekhoven and B. De Baets, *Monotone Mamdani–Assilian models under Mean of Maxima defuzzification*, Fuzzy Sets and Systems **159** (2008), 2819–2844.
3. E. Van Broekhoven and B. De Baets, *Only smooth rule bases can generate monotone Mamdani–Assilian models under COG defuzzification*, IEEE Trans. Fuzzy Systems **17** (2009), 1157–1174.
4. M. Štěpnička and B. De Baets, *Implication-based models of monotone fuzzy rule bases*, Fuzzy Sets and Systems, submitted.

2. Multi-class classification

Toy example

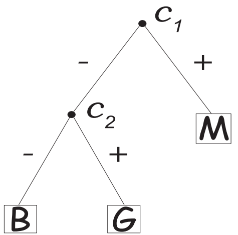
Classification problem:

	c_1	c_2	c_3	class label
a_1	−	−	+	A
a_2	+	−	−	B
a_3	−	+	+	C
a_4	+	+	−	B



Monotone classification problem:

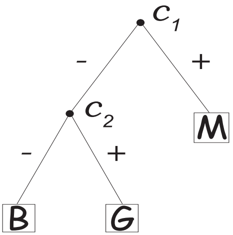
	c_1	c_2	c_3	evaluation
a_1	−	−	+	Bad
a_2	+	−	−	Moderate
a_3	−	+	+	Good
a_4	+	+	−	Moderate



Toy example

Monotone classification problem:

	c_1	c_2	c_3	evaluation
a_1	−	−	+	Bad
a_2	+	−	−	Moderate
a_3	−	+	+	Good
a_4	+	+	−	Moderate
a_5	−	+	−	**Good**
a_6	+	+	+	**Moderate**



Research question

How to produce guaranteed **monotone** classification results, even when the set of learning examples is **not monotone**?

Multi-class classification

- Problem: to attach labels from a finite set $\mathcal{L}$ to the elements of some set of **objects** Ω
- Each object $a \in \Omega$ is represented by a **feature vector**

$$\mathbf{a} = (c_1(a), c_2(a), \ldots, c_n(a))$$

  in the **feature space** $\mathcal{X}$
- **Collection of learning examples**: **multiset**

$$(\mathcal{S}, d) \equiv \{\langle \mathbf{a}, d(a)\rangle \mid a \in \mathcal{S}\}$$

  where:
  - $\mathcal{S} \subseteq \Omega$ is a given set of objects
  - $d : \mathcal{S} \to \mathcal{L}$ is the associated decision function
  - notation: $\mathcal{S}_{\mathcal{X}} = \{\mathbf{a} \mid a \in \mathcal{S}\}$

Multi-class classification

- **Goal** of supervised classification algorithms:
  - extend the function d to Ω in the most reasonable way
  - concentrate on finding a function $\lambda\colon \mathcal{X} \to \mathcal{L}$ that minimizes the expected loss on an independent set of test examples
- Different approaches:
  - **instance-based**, such as nearest neighbour methods
  - **model-based**, such as classification trees
- **Distribution classifiers**: output is a PMF over $\mathcal{L}$
  - mathematically: $\tilde{\lambda}\colon \mathcal{X} \to \mathcal{F}(\mathcal{L})$
  - selecting a single label: Bayesian decision (label with the highest probability is returned)

Multi-criteria evaluation

- In many cases, $\mathcal{L}$ exhibits a natural ordering and could be treated as an ordinal scale (chain): **ordinal classification/regression**
- Often, objects are described by (true) **criteria** $(c_i, \leq_{c_i})$ (chains)
- The **product ordering** turns $\mathcal{X}$ into a **partially ordered set** $(\mathcal{X}, \leq_{\mathcal{X}})$ (poset)
- Multi-criteria evaluation: quality assessment, environmental data, social surveys, etc.

Natural monotonicity constraint

An object *a* that scores at least as good on all criteria as an object *b* must be classified (ranked) at least as good as object *b*

Monotone classification

Monotone classifier

Classifier + **basic monotonicity constraint**:

$$\mathbf{x} <_{\mathcal{X}} \mathbf{y} \Rightarrow \lambda(\mathbf{x}) \leq_{\mathcal{L}} \lambda(\mathbf{y})$$

(supervised ranking/ordered sorting, monotone ordinal regression)

Monotone distribution classifier

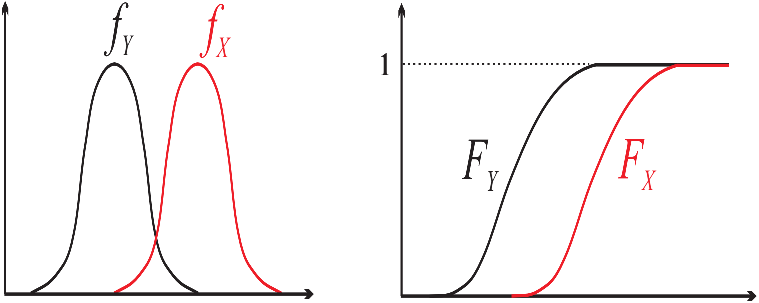
Distribution classifier + **stochastic monotonicity constraint**:

$$\mathbf{x} <_{\mathcal{X}} \mathbf{y} \Rightarrow \tilde{\lambda}(\mathbf{x}) \preceq_{\mathrm{SD}} \tilde{\lambda}(\mathbf{y})$$

(First order) Stochastic Dominance (SD):

$$f_X \preceq_{\mathrm{SD}} f_Y \;\Leftrightarrow\; F_X \geq F_Y$$

Stochastic dominance



Selecting a single label

- Bayesian decision potentially breaks the desired monotonicity and is **no longer acceptable** in this case
- The well-known relationship

$$f_X \preceq_{\mathrm{SD}} f_Y \Rightarrow \mathrm{E}[f_X] \leq \mathrm{E}[f_Y]$$

  cannot be used as it requires the transformation of the ordinal scale into a numeric scale
- **Set of medians** (interval) of f_X:

$$\mathrm{med}(f_X) = \{\ell \in \mathcal{L} \mid \mathcal{P}\{X \leq \ell\} \geq 1/2 \wedge \mathcal{P}\{X \geq \ell\} \geq 1/2\}$$

  - reduces in the continuous case to the median m
  - only endpoints of the interval have non-zero probability

Selecting a single label from the set of medians

- The set of medians reduces the PMF to an interval. Does there exist an ordering on intervals that is compatible with FSD?

$$[k_1, \ell_1] \leq_{\mathcal{L}}^{[2]} [k_2, \ell_2] \Leftrightarrow \left(k_1 \leq_{\mathcal{L}} k_2 \wedge \ell_1 \leq_{\mathcal{L}} \ell_2\right)$$

- New relationship:

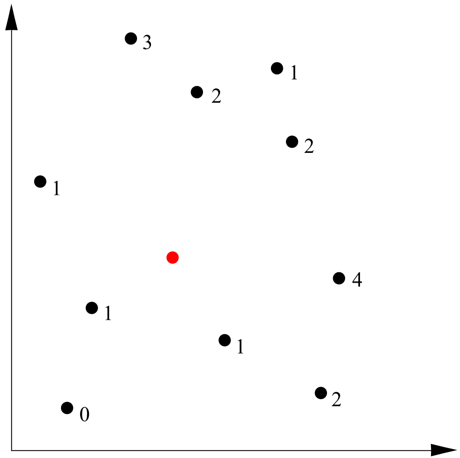
$$f_X \preceq_{\mathrm{SD}} f_Y \Rightarrow \mathrm{med}(f_X) \leq_{\mathcal{L}}^{[2]} \mathrm{med}(f_Y)$$

Selecting a single label

1. Pessimistic median (lower)
2. Optimistic median (upper)
3. Midpoint (or smaller/greater of the two midpoints) [not meaningful]

turn a monotone distibution classifier into a monotone classifier

How to label a new point?



Minimal and maximal extensions

1. **Minimal Extension**: $\lambda_{\text{min}} : \mathcal{X} \to \mathcal{L}$
   - assigns best label of **"objects below"**:

     $$\lambda_{\text{min}}(\mathbf{x}) = \max\{d(s) \mid \mathbf{s} \in \mathcal{S}_{\mathcal{X}} \wedge \mathbf{s} \leq_{\mathcal{X}} \mathbf{x}\}$$

   - if no such object: $\lambda_{\text{min}}(\mathbf{x}) = \min(\mathcal{L})$
2. **Maximal Extension**: $\lambda_{\text{max}} : \mathcal{X} \to \mathcal{L}$
   - assigns worst label of **"objects above"**:

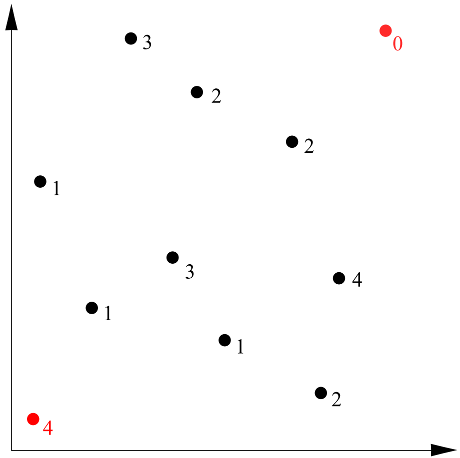
     $$\lambda_{\text{max}}(\mathbf{x}) = \min\{d(s) \mid \mathbf{s} \in \mathcal{S}_{\mathcal{X}} \wedge \mathbf{x} \leq_{\mathcal{X}} \mathbf{s}\}$$

   - if no such object: $\lambda_{\text{max}}(\mathbf{x}) = \max(\mathcal{L})$

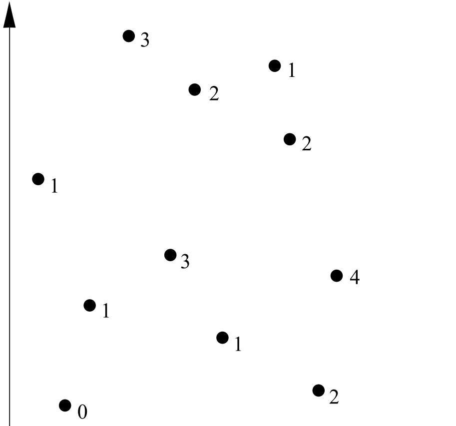
Monotone classifiers

1. λ_{min} and λ_{max} are monotone classifiers
2. **Interpolation**: midpoint leads to a monotone classifier

Things can go dead wrong



A non-monotone data set



Noise in multi-criteria evaluation

- $(\mathcal{S}, d)$ is called **monotone** if for all x and y in $\mathcal{S}$

$$\mathbf{x} = \mathbf{y} \Rightarrow d(x) = d(y)$$

(absence of doubt/ambiguity)

and

$$\mathbf{x} <_{\mathcal{X}} \mathbf{y} \Rightarrow d(x) \leq_{\mathcal{L}} d(y)$$

(absence of reversed preference)

- Non-monotonicity defines a **symmetric** and **transitive** relation on $\mathcal{S}$

Monotone extensions

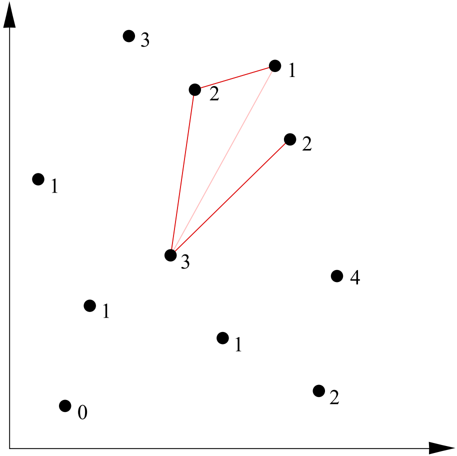
If the data set is monotone, then

1. λ_{min} and λ_{max} are monotone extensions of d to $\mathcal{X}$
2. any monotone extension λ of d to $\mathcal{X}$: $\lambda_{\text{min}} \leq_{\mathcal{L}} \lambda \leq_{\mathcal{L}} \lambda_{\text{max}}$

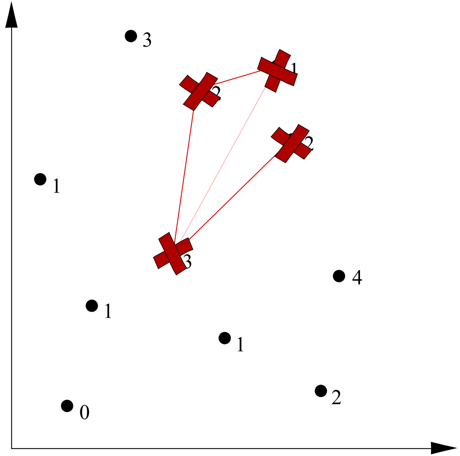
How to handle noise?

1. **Non-invasive approach**: keep the data set as is
   - excludes the use of some monotone classification algorithms (such as TOMASO)
   - restricts the accuracy of any monotone classifier (independence number)
2. **Data set reduction**: **identify** the noisy objects and **delete** them
3. **Data set relabelling**: **identify** the noisy objects and **relabel** them

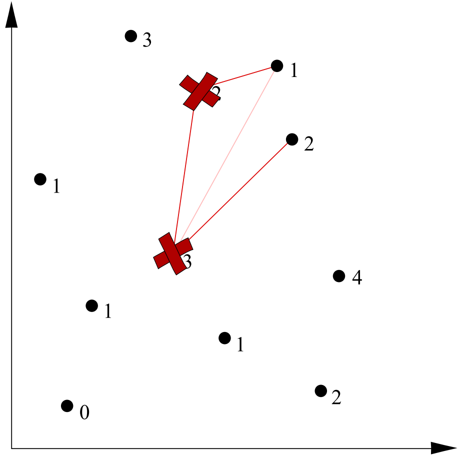
A non-monotone data set



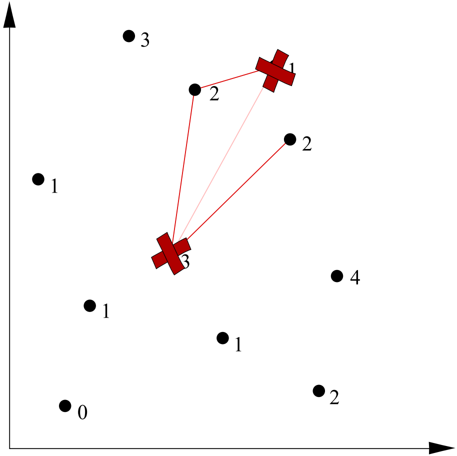
A non-monotone data set



A non-monotone data set



A non-monotone data set



The maximum independent set problem

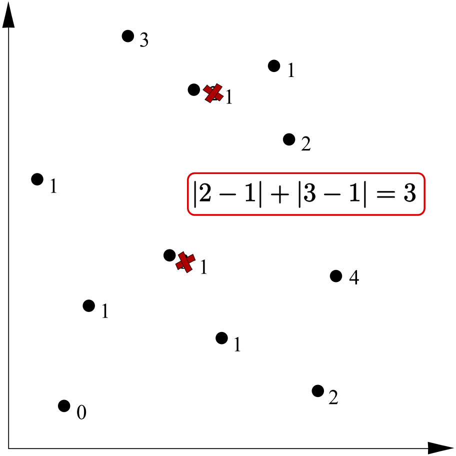
The non-monotonicity relation corresponds to a **comparability graph**:

- A monotone subset corresponds to an **independent set** of this graph
- **Maximal independent set** = independent set that is not a subset of any other independent set
- **Maximum independent set** = independent set of biggest cardinality (= **independence number** α)
- A maximum independent set in a comparability graph can be determined using **network flow theory** (cubic time complexity)

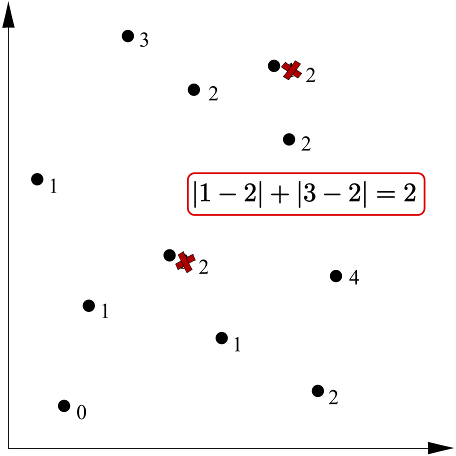
Let $(\mathcal{S}', d)$ be a maximal independent set. For $x \notin \mathcal{S}'$, it holds that

$$d(x) <_{\mathcal{L}} \lambda_{\min}(\mathbf{x}) \leq_{\mathcal{L}} \lambda_{\max}(\mathbf{x}) \quad \text{or} \quad \lambda_{\min}(\mathbf{x}) \leq_{\mathcal{L}} \lambda_{\max}(\mathbf{x}) < d(x)$$

Which maximum independent set to select?



Which maximum independent set to select?



$$|1-2|+|3-2|=2$$

Relabelling options

Universal tool: weighted maximum independent set problems and network flow theory

1. **Optimal ordinal relabelling**: relabelling a minimum number of objects, of which all corona objects are relabelled to a minimum extent
2. **Optimal cardinal relabelling** (identifying $\mathcal{L}$ with the first n integers): minimal relabelling loss
   - zero-one loss: maximum independent set
   - broad class of loss functions, including L1 loss and squared loss
3. **Optimal hierarchical cardinal relabelling** (single pass):
   - minimizing loss while relabelling a minimal number of objects
   - relabelling a minimal number of objects while minimizing loss

Distribution representation of a data set

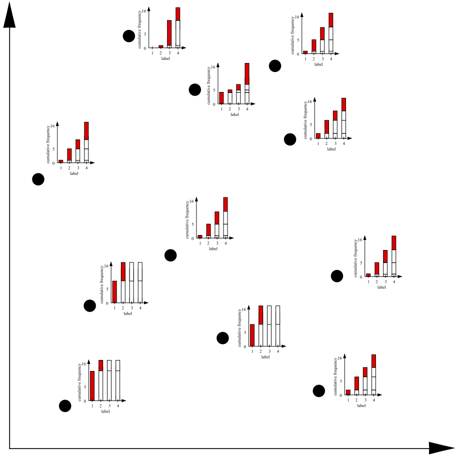
- Collection of learning examples $(\mathcal{S}, d)$
- For each $\mathbf{x} \in \mathcal{S}_{\mathcal{X}}$, a CDF $\hat{F}(\mathbf{x}, \cdot) \colon \mathcal{L} \to [0, 1]$ is built from the collection of learning examples

$$\hat{F}(\mathbf{x}, \ell) = \frac{|\{a \in \mathcal{S} \mid \mathbf{a} = \mathbf{x} \wedge d(a) \leq_{\mathcal{L}} \ell\}|}{|\{a \in \mathcal{S} \mid \mathbf{a} = \mathbf{x}\}|}$$

(cumulative relative frequency distribution)

- The distribution data set $(\mathcal{S}_{\mathcal{X}}, \hat{F})$

A distribution data set



Stochastic minimal and maximal extensions

1. **Minimal Extension**: $F_{\min} : \mathcal{X} \times \mathcal{L} \to [0,1]$

$$F_{\min}(\mathbf{x},\ell) = \min\{\hat{F}(\mathbf{s},\ell) \mid \mathbf{s} \in \mathcal{S}_{\mathcal{X}} \wedge \mathbf{s} \leq_{\mathcal{X}} \mathbf{x}\}$$

   - if no such object: $f_{\min}(\mathbf{x}, \min(\mathcal{L})) = 1$

2. **Maximal Extension**: $F_{\max} : \mathcal{X} \times \mathcal{L} \to [0,1]$

$$F_{\max}(\mathbf{x},\ell) = \max\{\hat{F}(\mathbf{s},\ell) \mid \mathbf{s} \in \mathcal{S}_{\mathcal{X}} \wedge \mathbf{x} \leq_{\mathcal{X}} \mathbf{s}\}$$

   - if no such object: $f_{\max}(\mathbf{x}, \max(\mathcal{L})) = 1$

Monotone distribution classifiers

1. $F_{\min}$ and $F_{\max}$ are monotone distribution classifiers
2. **Interpolation**: for any $S \in [0,1]$, the mapping

$$\tilde{F} = S\,F_{\min} + (1-S)\,F_{\max}$$

is also a monotone distribution classifier

Monotone distribution data sets

- $(\mathcal{S}_{\mathcal{X}}, \hat{F})$ is called **monotone** if for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{S}_{\mathcal{X}}$

$$\mathbf{x} <_{\mathcal{X}} \mathbf{y} \Rightarrow \hat{F}(\mathbf{x}, \cdot) \preceq_{\mathrm{SD}} \hat{F}(\mathbf{y}, \cdot)$$

- **Reversed preference**:

$$\mathbf{x} <_{\mathcal{X}} \mathbf{y} \text{ while not } \hat{F}(\mathbf{x}, \cdot) \preceq_{\mathrm{SD}} \hat{F}(\mathbf{y}, \cdot)$$

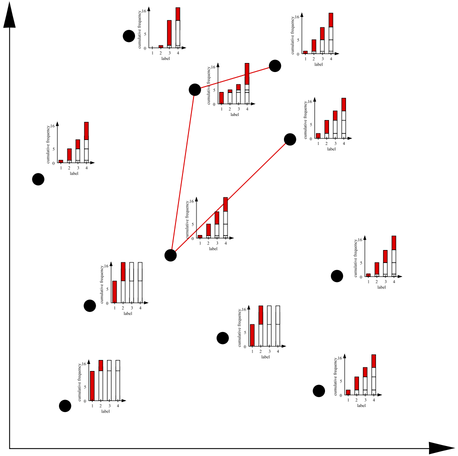
Monotone extensions

If the distribution data set is monotone, then

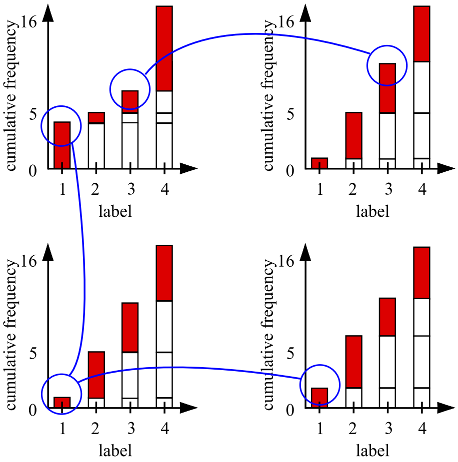
1. $F_{\min}$ and $F_{\max}$ are monotone extensions of $\hat{F}$ to $\mathcal{X}$
2. any monotone extension F of d to $\mathcal{X}$:

$$F_{\min}(\mathbf{y}, \cdot) \preceq_{\mathrm{SD}} F(\mathbf{y}, \cdot) \preceq_{\mathrm{SD}} F_{\max}(\mathbf{y}, \cdot)$$

A non-monotone distribution data set



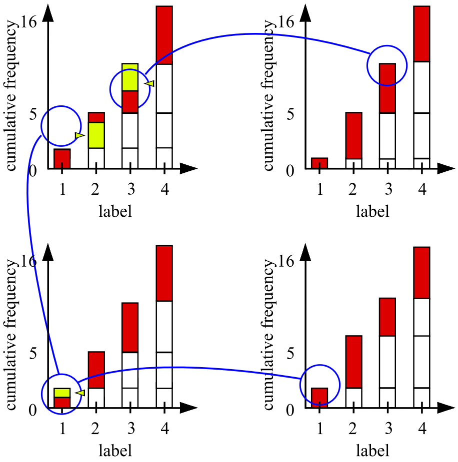
A non-monotone distribution data set



How to handle noise?

1. **Non-invasive approach**: keep the data set as is
2. **Data set reduction**: identify the noisy distributions and delete them
   - the non-monotonicity relation is **not transitive** (maximum independent set problem is NP-complete)
   - deleting entire distributions is quite invasive
   - deleting a single object affects the entire distribution and is hard to realize
3. **Data set relabelling**:
   - transitivity of non-monotonicity still holds at the label level
   - **L1-optimal relabelling** is possible using network flow algorithms
   - does not affect the frequency of feature vectors

After relabelling: a monotone distribution data set



A non-invasive approach

- Aim: to build a monotone distribution classifier from a possibly non-monotone distribution data set
- Weighted sums of F_{min} and F_{max} are solutions to this problem
- Aim: to identify more general interpolation schemes, depending on both the element $\mathbf{x}$ and the label ℓ
- For given $\mathbf{x}$ and ℓ:
  - **monotone situation**: $F_{\text{min}}(\mathbf{x}, \ell) \geq F_{\text{max}}(\mathbf{x}, \ell)$
  - **reversed preference situation**: $F_{\text{min}}(\mathbf{x}, \ell) < F_{\text{max}}(\mathbf{x}, \ell)$

The main theorem

OSDL generic theorem

Given two $\mathcal{X} \times \mathcal{L} \to [0,1]$ mappings s and t, the mapping $\tilde{F} : \mathcal{X} \times \mathcal{L} \to [0,1]$

$$\tilde{F}(\mathbf{x},\ell) = \begin{cases} s(\mathbf{x},\ell)F_{\min}(\mathbf{x},\ell) + \big(1 - s(\mathbf{x},\ell)\big)F_{\max}(\mathbf{x},\ell) \\ \qquad\qquad \text{if } F_{\min}(\mathbf{x},\ell) \geq F_{\max}(\mathbf{x},\ell) \\ t(\mathbf{x},\ell)F_{\min}(\mathbf{x},\ell) + \big(1 - t(\mathbf{x},\ell)\big)F_{\max}(\mathbf{x},\ell) \\ \qquad\qquad \text{if } F_{\min}(\mathbf{x},\ell) < F_{\max}(\mathbf{x},\ell) \end{cases}$$

is a **monotone distribution classifier** if and only if

1. s is decreasing in 1st and increasing in 2nd argument
2. t is increasing in 1st and decreasing in 2nd argument

Realization 1: OSDL

If one does not want to distinguish between the monotone and the reversed preference situation (s and t are identical), then the simple interpolation scheme is the only one

OSDL

If $s(\mathbf{x}, \ell) = t(\mathbf{x}, \ell)$ for all $\mathbf{x}$ and ℓ, then

$$s(\mathbf{x}, \ell) = t(\mathbf{x}, \ell) = S$$

for some constant $S \in [0, 1]$

Measuring support

1. The mapping $N_{\min} : \mathcal{X} \times \mathcal{L} \to \mathbb{N}$:

$$N_{\min}(\mathbf{x}, \ell) = |\{\langle \mathbf{y}, d(y)\rangle \in (\mathcal{S}, d) \,|\, \mathbf{y} \leq_{\mathcal{X}} \mathbf{x} \wedge d(y) >_{\mathcal{L}} \ell\}|$$

   counts the number of instances that indicate that **x** should receive a label strictly greater than ℓ

2. The mapping $N_{\max} : \mathcal{X} \times \mathcal{L} \to \mathbb{N}$:

$$N_{\max}(\mathbf{x}, \ell) = |\{\langle \mathbf{y}, d(y)\rangle \in (\mathcal{S}, d) \,|\, \mathbf{x} \leq_{\mathcal{X}} \mathbf{y} \wedge d(y) \leq_{\mathcal{L}} \ell\}|$$

   counts the number of instances that indicate that **x** should receive a label at most ℓ

3. Both are strictly positive in the reversed preference situation

Realizations 2 and 3: one parameter $S \in [0,1]$

Balanced OSDL

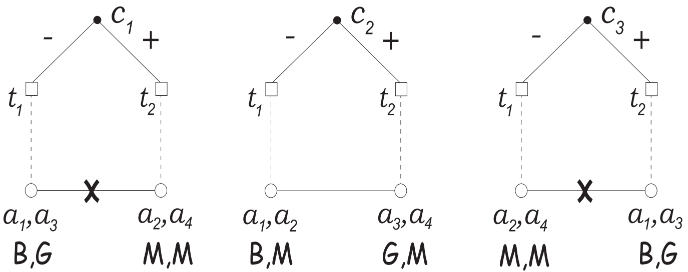
1. $s(\mathbf{x},\ell) = S$
2. $$t(\mathbf{x},\ell) = \frac{N_{\min}(\mathbf{x},\ell)}{N_{\min}(\mathbf{x},\ell) + N_{\max}(\mathbf{x},\ell)}$$

Double-balanced OSDL

1. $$s(\mathbf{x},\ell) = \begin{cases} \dfrac{N_{\max}(\mathbf{x},\ell)}{N_{\min}(\mathbf{x},\ell) + N_{\max}(\mathbf{x},\ell)} & \text{if } N_{\min}(\mathbf{x},\ell) + N_{\max}(\mathbf{x},\ell) \neq 0 \\ S & \text{otherwise} \end{cases}$$
2. $$t(\mathbf{x},\ell) = \frac{N_{\min}(\mathbf{x},\ell)}{N_{\min}(\mathbf{x},\ell) + N_{\max}(\mathbf{x},\ell)}$$

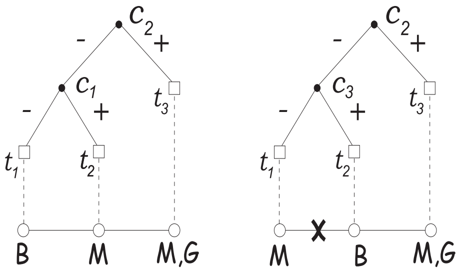
Towards ranking trees

	c_1	c_2	c_3	class label
a_1	$-$	$-$	$+$	Bad
a_2	$+$	$-$	$-$	Moderate
a_3	$-$	$+$	$+$	Good
a_4	$+$	$+$	$-$	Moderate



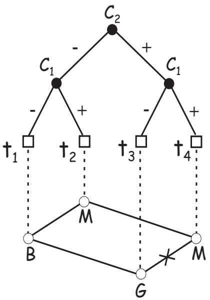
Towards ranking trees

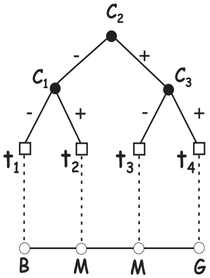
	c_1	c_2	c_3	class label
a_1	$-$	$-$	$+$	Bad
a_2	$+$	$-$	$-$	Moderate
a_3	$-$	$+$	$+$	Good
a_4	$+$	$+$	$-$	Moderate



Towards ranking trees

	c_1	c_2	c_3	class label
a_1	$-$	$-$	$+$	Bad
a_2	$+$	$-$	$-$	Moderate
a_3	$-$	$+$	$+$	Good
a_4	$+$	$+$	$-$	Moderate





Basic principles

- **Growing principle**: strive for purity by **minimizing solvable doubt** and **reversed preference**
- **Interdependence of the leaves** due to monotonicity: split of one leaf may have an effect on all the other leaves
- **Pruning**: theory of minimal cost-complexity pruning collapses
- **Labelling rule**: poset structure on the leaves allows for the use of OSDL as labelling rule

Topics not discussed

- Other approaches to monotone classification
- Experimental comparison
- Performance measures
- Random generation of monotone data sets

References

Monotone classification:

1. K. Cao-Van and B. De Baets, *Growing decision trees in an ordinal setting*, Internat. J. Intelligent Systems **18** (2003), 733–750.
2. S. Lievens, B. De Baets and K. Cao-Van, *A probabilistic framework for the design of instance-based supervised ranking algorithms in an ordinal setting*, Annals of Operations Research **163** (2008), 115–142.
3. S. Lievens and B. De Baets, *Supervised ranking in the WEKA environment*, Information Sciences **180** (2010), 4763–4771.

References

Relabelling:

1. M. Rademaker, B. De Baets and H. De Meyer, *Optimal monotone relabelling of partially non-monotone ordinal data*, Optimization Methods and Software, in press.
2. M. Rademaker, B. De Baets and H. De Meyer, *Loss optimal monotone relabelling of noisy multi-criteria data sets*, Information Sciences **179** (2009), 4089–4096.
3. M. Rademaker and B. De Baets, *Optimal restoration of stochastic monotonicity w.r.t. cumulative label frequency loss functions*, Information Sciences **181** (2011), 747–757.

Monotone data set generation:

1. K. De Loof, B. De Baets and H. De Meyer, *On the random generation and counting of weak order extensions of a poset with given class cardinalities*, Information Sciences **177** (2007), 220–230.
2. K. De Loof, B. De Baets and H. De Meyer, *On the random generation of monotone data sets*, Information Processing Letters **107** (2008), 216–220.

3. Decision making

The Lar region in Iran

Lar region in Iran

- 75 km north-east of Tehran
- Ecological, economical and socio-cultural value
  - Flora and fauna
  - Water supply
  - Extensive stock farming
  - Tourism
  - Nomads
- Region under heavy ecological pressure

Management of the region

- Four proposed management plans involving 12 criteria
- For each of the 12 criteria, each of the 31 stakeholders
  - defines a **linear order** on the 4 plans
  - expresses how strongly (s)he prefers one plan over another: **intensities** ranging from **very weak** to **very strong**
- Overall problem: establish a linear order on the 4 plans
- Subproblems: for each of the 12 criteria establish
  - **social order**: linear order on the 4 plans (social choice problem)
  - **social intensities of preferences**: express how strongly one plan is preferred over another
- Solution procedure: translate into a monotonicity problem

Data representation

- A set $\mathcal{S}$ of N stakeholders
- A set $\mathcal{A}$ of k alternatives
- An ordinal scale $\mathcal{L}$ of m intensities of preference: $\ell_1 < \cdots < \ell_m$
- Each stakeholder S_i delivers:
  - a linear order $\succ_i$ on $\mathcal{A}$
  - a mapping $P_i : \succ_i \rightarrow \mathcal{L}$ assigning intensities to couples of alternatives
  - **consistency** in the form of **monotonicity conditions**:
    - P_i is increasing (w.r.t. $\succ_i$) in its first argument
    - P_i is decreasing (w.r.t. $\succ_i$) in its second argument

Example: data of stakeholder S_i

- Linear order of S_i: $a \succ_i b \succ_i c \succ_i d$
- Intensities of preferences P_i:

	a	b	c	d
a		strong	strong	strong
b			weak	strong
c				moderate
d				

Extended representation

- We extend the ordinal scale $\mathcal{L}$ to the ordinal scale
$$\mathcal{L}^* = \{-\ell_m, \ldots, -\ell_1, \ell_0, \ell_1, \ldots, \ell_m\}$$
containing:
    - signed intensities of preferences
    - zero intensity of preference ℓ_0
- We extend P_i to $P_i' : \mathcal{A} \times \mathcal{A} \to \mathcal{L}^*$ as follows:
    - if $a \succ_i b$, then $P_i'(a,b) = P_i(a,b)$ and
    $$P_i'(b,a) = -P_i(a,b)$$
    - $P_i'(a,a) = \ell_0$
- Notation: $\ell_{-i} := -\ell_i$

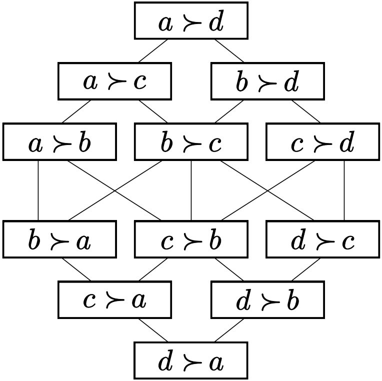
Example revisited

- Linear order of S_i: $a \succ_i b \succ_i c \succ_i d$
- Intensities of preferences P_i':

	a	b	c	d
a	**none**	strong	strong	strong
b	**−strong**	**none**	weak	strong
c	**−strong**	**−weak**	**none**	moderate
d	**−strong**	**−strong**	**−moderate**	**none**

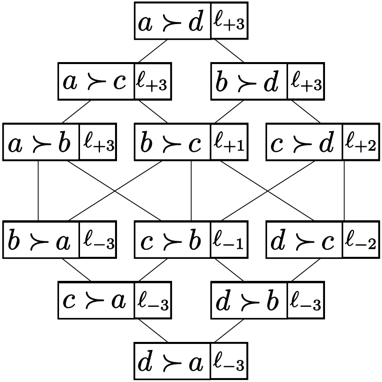
Poset representation

Linear order of S_i: $a \succ_i b \succ_i c \succ_i d$



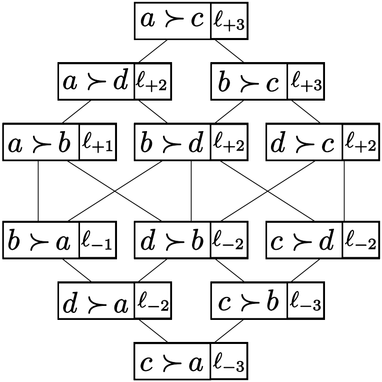
Poset representation with intensities

Linear order of S_i: $a \succ_i b \succ_i c \succ_i d$



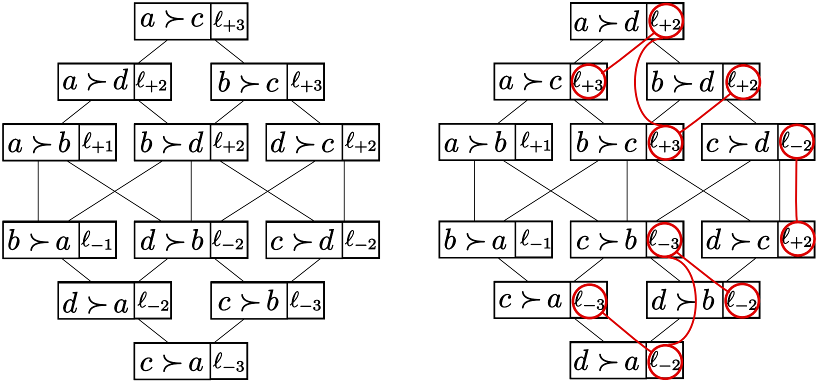
Poset representation with intensities

Linear order of S_j: $a \succ_j b \succ_j d \succ_j c$



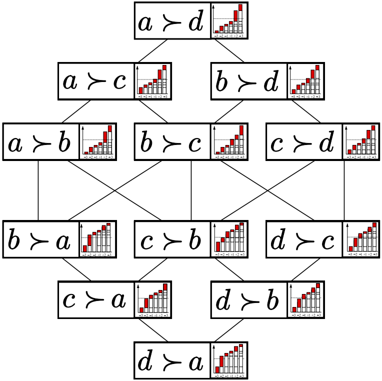
Poset representation with intensities

Intensities of preferences of S_j on the poset of S_i:



Poset representation with distribution of intensities

$F_{(a,b)}(\ell_i)$ = number of times the intensity for $a \succ b$ is at most ℓ_i



Social order and social intensities of preferences

Suppose a social order has been selected

Optimistic or pessimistic median intensities (upper half)

- **not** necessarily increasing
- **not** necessarily positive

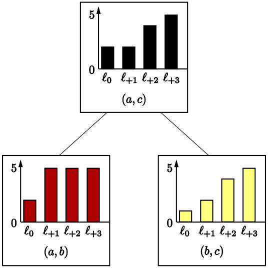
Case of stochastic monotonicity (upper half)

Medians are

- increasing
- **not** necessarily positive

Classical majorities

- Interpretation:
  - $F_{(a,b)}(\ell_0)$ is the number of stakeholders **against** $a \succ b$
  - $F_{(b,a)}(\ell_0)$ is the number of stakeholders **in favour** of $a \succ b$
  - $F_{(a,b)}(\ell_0) + F_{(b,a)}(\ell_0) = N$
- Classical majority $a \succeq_M b$: $F_{(a,b)}(\ell_0) \leq F_{(b,a)}(\ell_0)$



Condorcet order

- **Condorcet majority cycle**:

  if $a \succ_M b$ and $b \succ_M c$, but $c \succ_M a$

- If $a \succ_M b$, $b \succ_M c$ and $a \succ_M c$:
  - **Condorcet order** $a \succ_M b \succ_M c$
  - it may hold that $a \succ_M c$ received the weakest support

In case of a Condorcet order

Medians are

- positive
- **not** necessarily increasing (and hence no stochastic monotonicity)

Optimization problem

- For any linear order $\succ$ on $\mathcal{A}$, we determine G such that:
  1. it generates **cumulative frequency distributions**: for any (a,b), $G_{(a,b)}$ is increasing and $G_{(a,b)}(\ell_m) = N$
  2. it preserves **symmetry**: $G_{(b,a)}(\ell_{-i}) = N - G_{(a,b)}(\ell_{i-1})$
  3. it renders $\succ$ a **Condorcet order**: for any $a \succ b$ it holds that $G_{(a,b)}(\ell_0) \leq G_{(b,a)}(\ell_0)$
  4. it generates **stochastically monotone** distributions
  5. it is as close as possible to F, i.e. it has **minimal error**

     $$d(F,G) = \sum_{a,b\in\mathcal{A}} \sum_{\ell\in\mathcal{L}} \left|F_{(a,b)}(\ell) - G_{(a,b)}(\ell)\right|$$

     (implicitly assumes L_1-distance on $\mathcal{L}$)
- **Social order**: linear order $\succ$ for which $d(F,G)$ is minimal

Optimization problem

- The optimization problem can be translated into a **weighted maximum independent set problem**
- Using an intelligent scheme, it can be solved efficiently using nearest and farthest maximum cuts in flow networks

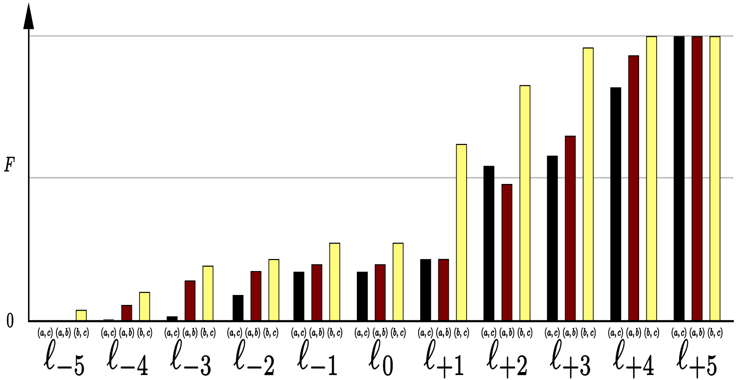
In case of a Condorcet order

If the Condorcet order is not stochastically monotone, then

- there does not exist a stochastically monotone linear order
- it is not necessarily optimal

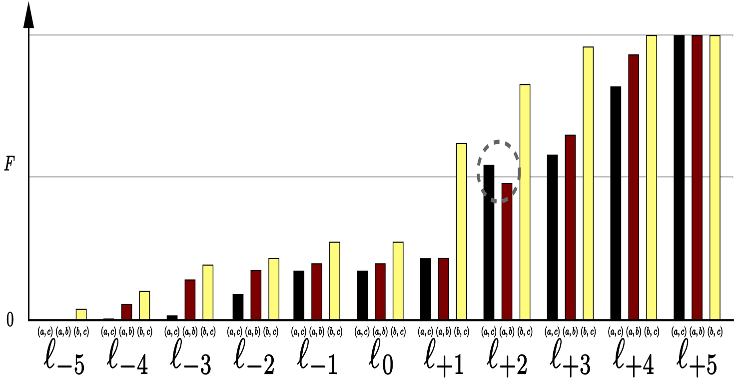
Illustration

Condorcet order $a \succ_M b \succ_M c$



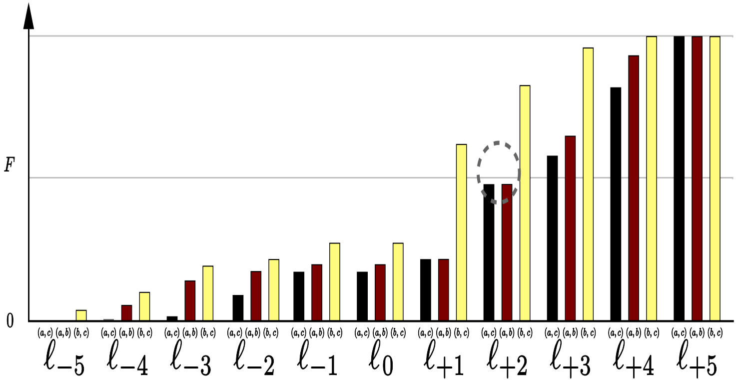
Illustration

Condorcet order $a \succ_M b \succ_M c$



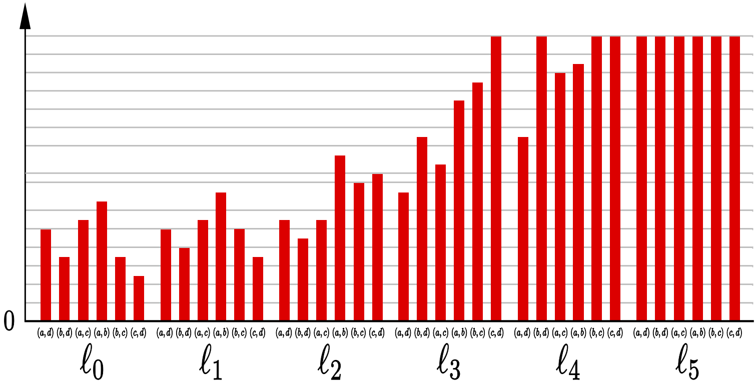
Illustration

Condorcet order $a \succ_M b \succ_M c$



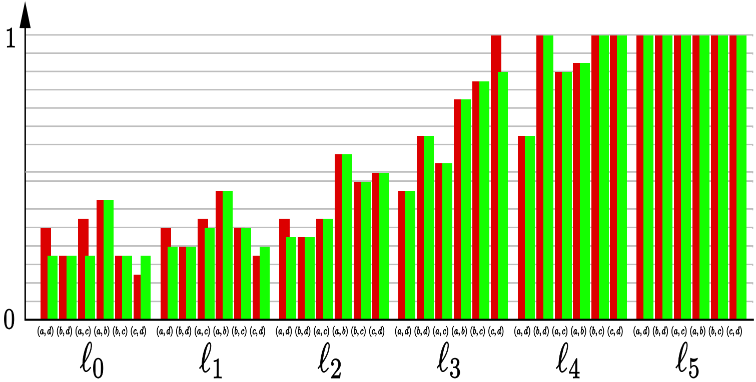
Real-life example: the Wildlife criterion in Lar

Condorcet order $a \succ_M b \succ_M c \succ_M d$



Real-life example: the Wildlife criterion in Lar

Condorcet order $a \succ_M b \succ_M c \succ_M d$



Discussion

Transparent methodology:

- it builds distributions **before** taking the median (credo: "first process the data, then defuzzify")
- it does so in an **optimal** way
- it allows to simulate the effect of the inclusion/exclusion of **minorities**

A new social choice function

- Reduced setting:
  - each stakeholder S_i delivers only a linear order $\succ_i$, no intensities of preferences
  - intensity ℓ_1: vote in favour
  - intensity ℓ_{-1}: vote against
- **Social order**: linear order for which a minimum number of preferences needs to be reversed

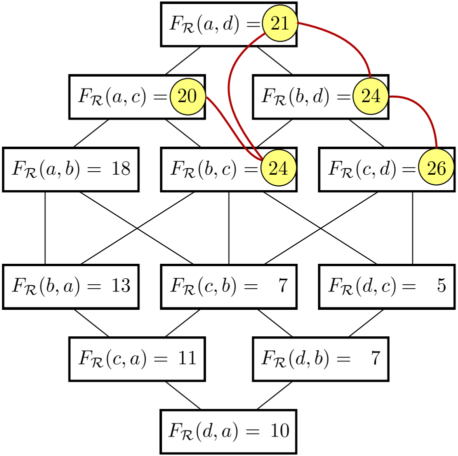
In case of a Condorcet order

If the Condorcet order is not monotone, then

- there does not exist a monotone linear order
- it is not necessarily optimal

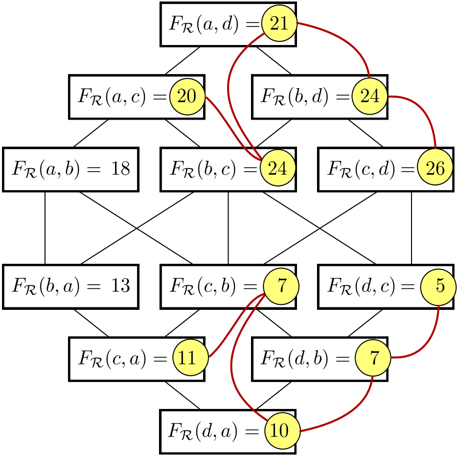
A final example: a non-optimal Condorcet order

Condorcet order: $a \succ b \succ c \succ d$



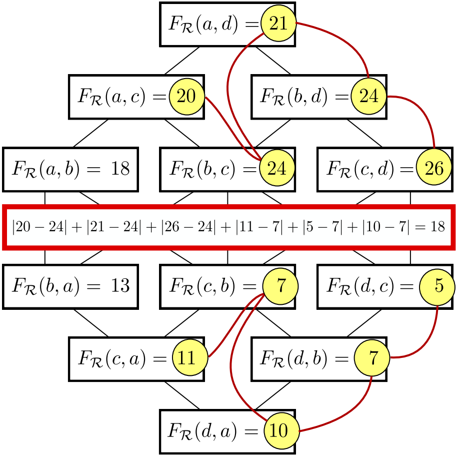
A final example: a non-optimal Condorcet order

Condorcet order: $a \succ b \succ c \succ d$



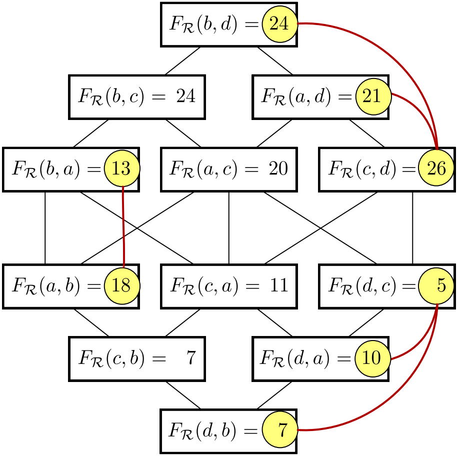
A final example: a non-optimal Condorcet order

Condorcet order: $a \succ b \succ c \succ d$



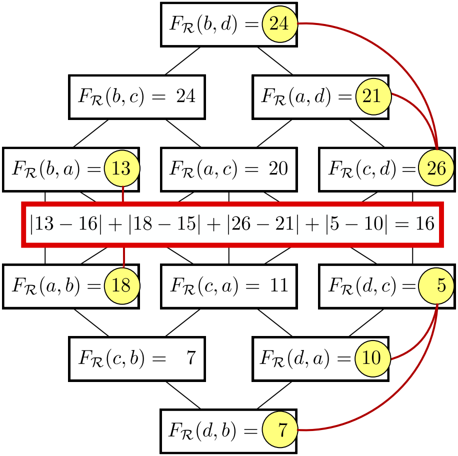
A final example: a non-optimal Condorcet order

Condorcet order: $b \succ a \succ c \succ d$



A final example: a non-optimal Condorcet order

Condorcet order: $b \succ a \succ c \succ d$



References

1. K. Zendehdel, M. Rademaker, B. De Baets and G. Van Huylenbroeck, *Qualitative valuation of environmental criteria through a group consensus based on stochastic dominance*, Ecological Economics **67** (2008), 253–264.
2. K. Zendehdel, M. Rademaker, B. De Baets and G. Van Huylenbroeck, *Improving tractability of group decision making on environmental problems through the use of social intensities of preferences*, Environmental Modelling and Software **24** (2009), 1457–1466.
3. K. Zendehdel, M. Rademaker, B. De Baets and G. Van Huylenbroeck, *Environmental decision making with conflicting social groups: A case study of the Lar rangeland in Iran*, J. Arid Environments **74** (2010), 394–402.
4. M. Rademaker and B. De Baets, *Aggregation of monotone reciprocal relations with application to group decision making*, Fuzzy Sets and Systems **184** (2011), 29–51.
5. M. Rademaker and B. De Baets, *A ranking procedure based on a natural monotonicity constraint*, Information Fusion, submitted.

Closing observations

1. In many modelling problems, there exists a **monotone relationship** between some of the **input variables** and the **output variable**
2. **Resolution of non-monotonicity** can be translated into an optimization problem
3. The key lies in a **cumulative approach** and **network flow theory**
4. **Group decision making** problems can be cast in this framework
5. A new avenue of research

Thank you for your attention!

bernard.debaets@ugent.be