On the Strategies of Criteria Discretization and Subintervals Design in UTA-like Methods

Mohammad Ghaderi[1], Francisco Ruiz[2], Nuria Agell[1]

1 ESADE Business school, University of Ramon Llull

2 Technical University of Catalonia (UPC)

ESADE
Ramon Llull University
GREC - RESEARCH GROUP ON KNOWLEDGE ENGENEERING

EWG-MCDA
EURO Working Group on Multicriteria Decision Aiding

MCDA81
26-28 mars 2015
Annecy

Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial

Index

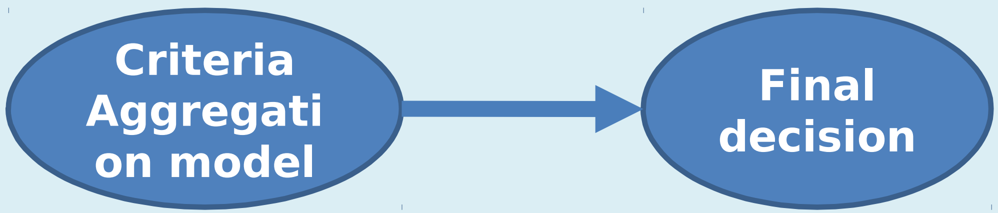
- Introduction
- Learning Non-Monotonic Preferences
- Discretization Techniques for UTA-like Methods
- Comparing Discretization Techniques
- Conclusion

Introduction

E

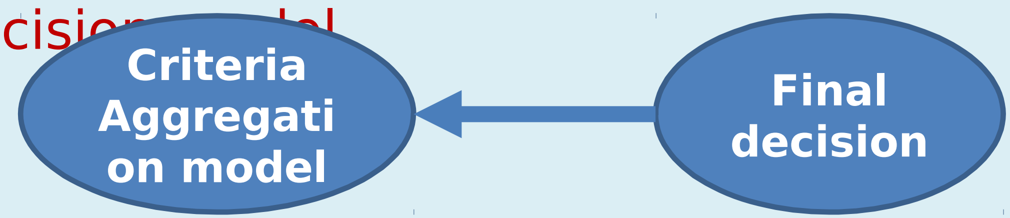
Aggregation approach

It aims at building the final decision using a priori known criteria aggregation model.



Disaggregation approach

The final decision is known a priori and it is decomposed to reveal the underlying decision model



Introduction:
Motivation and Objective

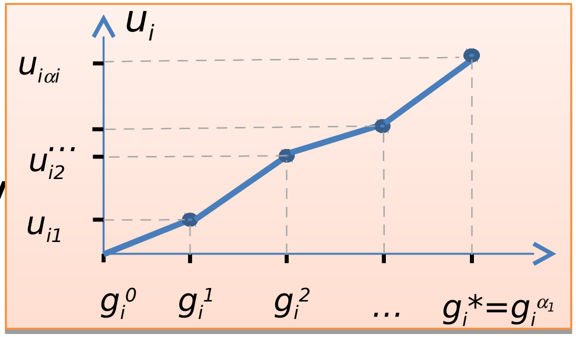
- How to learn non-monotonic preferences?
- Does discretization matter? How does it impact the performance of UTA-like preference learning method?
- How a “good” discretization technique for ranking-based UTA-like method should be?

- ➢ Define a new supervised discretization method for the criteria variables:
  - to be applied in disaggregation method
  - to enhance preference learning
- ➢ Compare the results of the defined discretization method with some existing ones by means of a new UTA-like method (UTA-RR)

UTA (*UTilitiés Additives*)

[Jacquet-Lagrèze and Siskos 1982]

- Additive marginal utility function
- Reference set A_R
- Ranking
- Monotone piecewise linear marginal utility
- Linear programming technique
- Post-optimality process



$$Min \sum_{a \in A_R} \left(\sigma^+(a) + \sigma^-(a) \right)$$

$$\left. \begin{array}{c} U(a) + \sigma^+(a) - \sigma^-(a) \geq U(b) + \sigma^+(b) - \sigma^-(b) + \varepsilon \leftrightarrow a \succ b \\ U(a) + \sigma^+(a) - \sigma^-(a) = U(b) + \sigma^+(b) - \sigma^-(b) \leftrightarrow a \sim b \end{array} \right\} \forall a \in A_R$$

$$u_i(g_i^{j+1}) - u_i(g_i^j) \geq 0, \forall i, \forall j = 0, \ldots, \gamma_i - 1$$

$$\sum_i u_i(g_i^{\gamma_i}) = 1$$

$$U(a) = \sum_i u_i(x_i^a)$$

Non-monotonicity

Existing Methodologies

UTA-NM (Kleiger, T., 2009)

- Removes the monotonocity costraints
- Many binary variables
- The penalization to prevent overfitting entails an excessive computational cost

Doumpos (2012)

- To model quasiconcave, quasiconvex, and S-type marginal utility functions
- Non-linear Mixed Integer Programming model

Proposed Methodology: UTA-RR

- Remove monotonicity constraint

But…

- How to ensure normalization?
- How to estimate maximal shares of the criteria into the comprehensive value?

$$Min \sum_{a \in A_R} \left(\sigma^+(a) + \sigma^-(a) \right)$$

$$\left. \begin{array}{l} U(a) + \sigma^+(a) - \sigma^-(a) \geq U(b) + \sigma^+(b) - \sigma^-(b) + \varepsilon \leftrightarrow a \succ b \\ U(a) + \sigma^+(a) - \sigma^-(a) = U(b) + \sigma^+(b) - \sigma^-(b) \leftrightarrow a \sim b \end{array} \right) \forall a \in A_R$$

$$\text{Re}\,move: \; u_i(g_i^{j+1}) - u_i(g_i^{j}) \geq 0, \forall i, \forall j = 1, \ldots, \gamma_i - 1$$

$$\text{Re}\,move: \; \sum_i u_i(g_i^{\gamma_i}) = 1$$

Proposed Methodology : UTA-RR

Initial Solution

$$Min \sum_{a \in A_R} \left(\sigma^+(a) + \sigma^-(a) \right)$$

$$\left. \begin{array}{l} U(a) + \sigma^+(a) - \sigma^-(a) \geq U(b) + \sigma^+(b) - \sigma^-(b) + \varepsilon \leftrightarrow a \succ b \\ U(a) + \sigma^+(a) - \sigma^-(a) = U(b) + \sigma^+(b) - \sigma^-(b) \leftrightarrow a \sim b \end{array} \right\} \forall a \in A_R$$

$$A_i W_i \geq 0, \forall \text{i}$$

$$A_i W_i \leq 1, \forall \text{i}$$

$$W_i = \left(u_i(g_i^0), u_i(g_i^1) - u_i(g_i^0), \ldots, u_i(g_i^{\alpha_i}) - u_i(g_i^{\alpha_i - 1}) \right)^{\text{T}}$$

A_i an $\alpha_\text{i} * \alpha_\text{i}$ lower triangular matrix of elements 1

$$g_i^{j_*} = \arg\min_j u_i(g_i^j)$$

$$g_i^{j^*} = \arg\max_j u_i(g_i^j)$$

f^* : optimal value of the objective function

Proposed Methodology : UTA-RR
Iterative part

Two types of constraints are imposed in each iteration to ensure normalization

- Restrictive Constraint

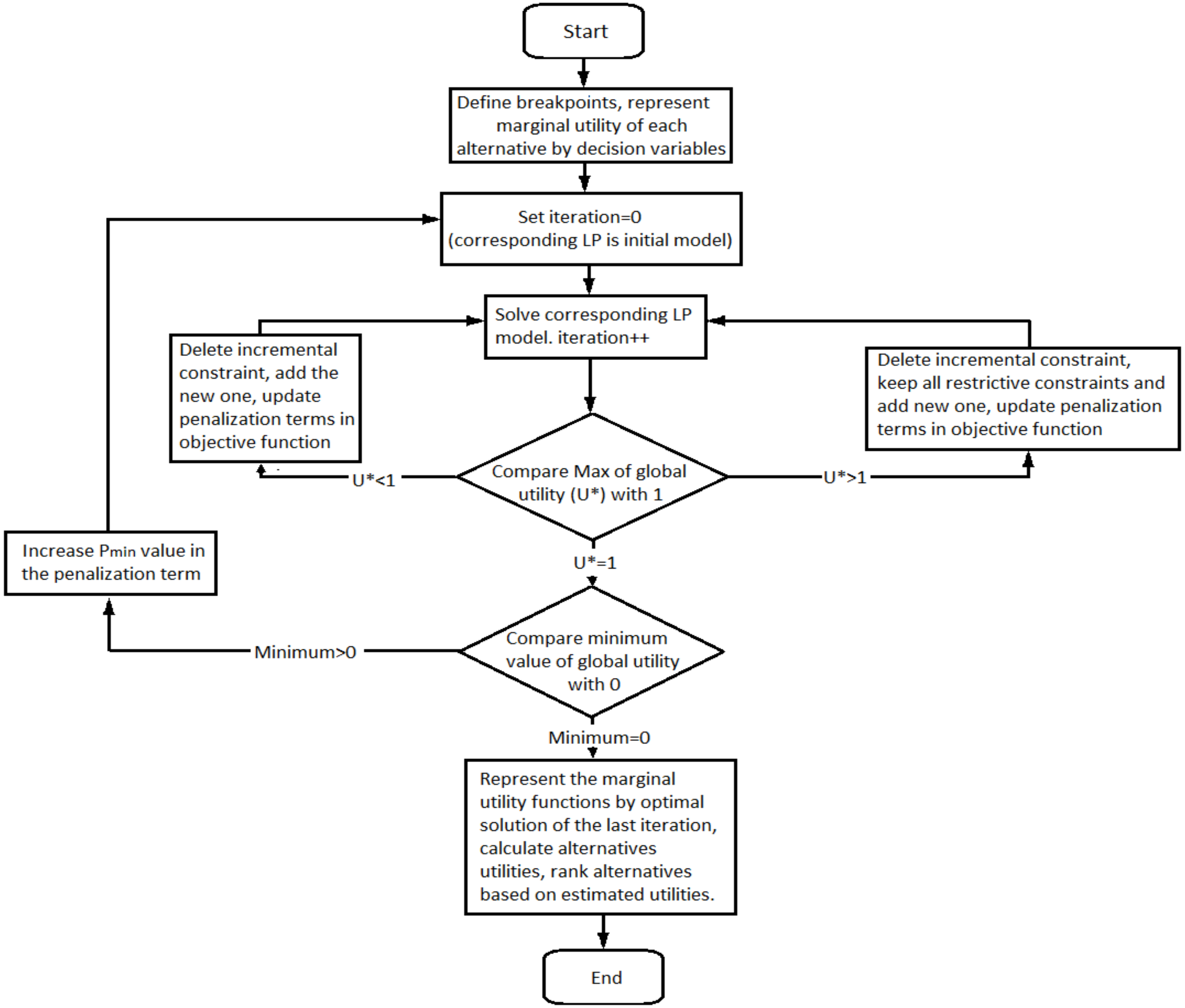
  - In the case that optimal comprehensive value of the last iteration exceed one
- Incremental Constraint

  - In the case that optimal comprehensive value of the last iteration be less than one
- Two penalization terms are added to the model, to prevent random changes over iterations

Concretely, in iteration $t+1$:

$$Min \sum_{a \in A_R} \left(\sigma^+(a) + \sigma^-(a) \right) + p_1 f^t (1 - \sum_i u_i^t (g_i^{t*})) + p_2 f^t \sum_i u_i^t (g_{i*}^t)$$

Proposed Methodology



Discretization Techniques for UTA-like Methods

- Equal-Width
- Equal-Frequency
- K-Means Clustering
- Kernel-based Discretization
- Proposed Supervised Discretization

Discretization Techniques for UTA-like Methods

Proposed Discretization Technique

- *Characterization:* Supervised, Global, Top-Down, Direct

- *Initial intuition:* The observations into the same subinterval are better to have similar rankings

Why this intuition is misleading?

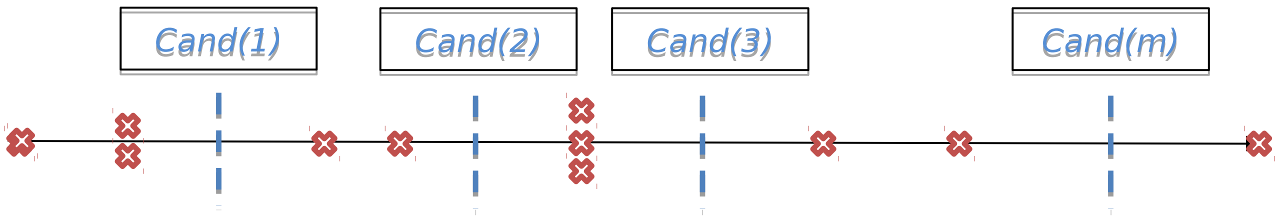
- Ranking-based UTA methods build (in)equality constraints by comparing consecutive elements in the ranking
- Consecutive rankings into the same subinterval
  1. Prevents accommodating the comparison information in m.v.f
  2. Increases sparsity of the coefficient matrix in the corresponding LP

Discretization Techniques for UTA-like Methods

Proposed Discretization Technique

- *Intuition:* Construct subintervals to maximize variation of rankings in each subinterval

Example: Partitioning an attribute into 3 subintervals



$X = \{x_1, \ldots, x_n\}$ Set of Obeservations

$r(X) = \{r(x_1), \ldots, r(x_n)\}$ Corresponding rankings

$$\underset{i,j>i}{\text{Max}}\ F(i,j) = \underset{cand(i)<X}{Range}(r(X)) + \underset{cand(i)<X<cand(j)}{Range}(r(X)) + \underset{cand(j)>X}{Range}(r(X))$$

Discretization Techniques for UTA-like Methods

Proposed Discretization Technique

- F is discrete, not convex, and the number of the points that we need to compute the function F for *k* subintervals grows proportional to

$$\binom{|M|/2}{k}$$

$$M = unique(X)$$

Discretization Techniques for UTA-like Methods

Proposed Discretization Technique

- Heuristic:
  1. Create set of candidate landmarks
  2. For each candidate *i*, compute *Range(r(X))* for $X<cand(i)$, denote by $F_{left}(i)$, and for $X>cand(i)$, denote by $F_{right}(i)$, considering only X values within the same interval of cand(i)
  3. Compute $F(i)=\min(F_{left}(i),F_{right}(i))$, and select the candidate with the highest F value as the new landmark, and partition the attribute based on the obtained landmark.
  4. Repeat the process for the new subintervals until obtaining *k-1* landmarks

Comparing Discretization Techniques

Dataset: A ranking of 28 cars, each described by 3 attributes

- Select a discretization technique

1. For k_1=3:7, k_2=3:7, k_3=3:7, discretize the three attributes
2. Use UTA-RR to learn the value functions using this discretization. **Leave-one-out cross-validation** is used to measure generalizability of the extracted v.fs, and **Kendall-Tau** is used to measure accuracy of the v.fs. Therefore 125 Tau values and 125 CV values for each discretization technique is obtained.

- . Repeat the above process for all the 5 discretization techniques

Comparing Discretization Techniques

Fuzzy preference relations are used to compare each pairs of discretization techniques.

• Fuzzy preference relations are used to compare each pairs of discretization techniques.

Given two discretization techniques *a* and *b:*

Given two discretization techniques *a* and *b:*

- *Degree of credibility*

- *Degree of credibility*

$$d_{Tau}(a,b) = \frac{|\{K \,|\, Tau(a) - Tau(b) \geq 0\}|}{|K|} \quad ; \quad K = \{k_1, k_2, k_3\}, |K| = 125$$

- *Strict preference*

- *Strict preference*

$$\mu^s(a,b) = \begin{cases} d(a,b) - d(b,a) & d(a,b) > d(b,a); \\ 0 & otherwise. \end{cases}$$

- *Non-Domination degree*

- *Non-Domination degree*

$$\min_b \{1 - \mu^s(b,a)\}$$

Comparing Discretization Techniques

Results

Comparison based on Tau (accuracy)

Disc. Technique	Non-Dom. deg.
Eq-w	1
Eq-f	0.92
Kmeans	0.568
Kernel	0
Proposed	0.72

Comparison based on CV (generaliz.)

Disc. Technique	Non-Dom. deg.
Eq-w	0.8
Eq-f	0.784
Kmeans	0.552
Kernel	0.08
Proposed	1

- Dataset is **uniformly distributed** over the attributes.
- The same process has been repeated for a **skewed dataset**.

Comparing Discretization Techniques

Results (for the skewed dataset)

Comparison based on Tau (accuracy)

Disc. Technique	Non-Dom. deg.
Eq-w	1
Eq-f	0.173333
Kmeans	0.533333
Kernel	0
Proposed	0.68

Comparison based on CV (generaliz.)

Disc. Technique	Non-Dom. deg.
Eq-w	0.626667
Eq-f	0.666667
Kmeans	0.853333
Kernel	0.053333
Proposed	1

Conclusion & Future Studies

- Learning non-monotonic preferences
- Importance of discretization in UTA-like methods

- New measures of variation: Range, variance, IQR, etc
- Improvement on the proposed heuristic
- What about maximizing number of subintervals between the performance level of two consecutive alternatives in the ranking list?
- Improving UTA-RR by restricting slope-change degree to ensure interpretability

Thank you for your attention!

ESADE
Ramon Llull University
GREC - RESEARCH GROUP ON KNOWLEDGE ENGENEERING

UPC
Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial