Fuzzy Weighted Attribute Combinations Based Similarity Measures

Giulianella Coletti[a] Davide Petturiti[b] Barbara Vantaggi[c]

[a]Dip. Matematica e Informatica, Università di Perugia
[b]Dip. Economia, Università di Perugia
[c]Dip. S.B.A.I., "La Sapienza" Università di Roma

ISIPTA '17 - ECSQARU 2017

Lugano, Switzerland – 10-14 July 2017

Motivation: similarity and attribute interactions (1)

Similarity measures

- Quantitative expression of

  "what is in common between two objects"

- Usually take into account (at most) the significance value of different features individually

Capacities and the Choquet integral

- Model and evaluate the measure of a specific *"concept"* (such as utility, power, coalition effort)
- Take into account the significance value of different features and their mutual (positive or negative) interactions

Most known similarity measures for binary data

⇒ Most known similarity measures are based on cardinalities $|A \cap B|$, $|A \setminus B|$, $|B \setminus A|$ and $|A^c \cap B^c|$

	Similarity	Expression
Type 1	Jaccard	$\frac{\lvert A\cap B\rvert}{\lvert A\setminus B\rvert+\lvert B\setminus A\rvert+\lvert A\cap B\rvert}$
	Dice	$\frac{2\lvert A\cap B\rvert}{\lvert A\setminus B\rvert+\lvert B\setminus A\rvert+2\lvert A\cap B\rvert}$
	Tversky	$\frac{\lvert A\cap B\rvert}{\alpha\lvert A\setminus B\rvert+\beta\lvert B\setminus A\rvert+\lvert A\cap B\rvert},\ \alpha,\beta>0$
	Ochiai	$\frac{\lvert A\cap B\rvert}{\sqrt{\lvert A\setminus B\rvert+\lvert A\cap B\rvert}\sqrt{\lvert B\setminus A\rvert+\lvert A\cap B\rvert}}$
	Kulczynski 2	$\frac{1}{2}\left(\frac{\lvert A\cap B\rvert}{\lvert A\setminus B\rvert+\lvert A\cap B\rvert}+\frac{\lvert A\cap B\rvert}{\lvert B\setminus A\rvert+\lvert A\cap B\rvert}\right)$
Type 2	Sokal and Michener (Euclidean)	$\frac{\lvert A\cap B\rvert+\lvert A^c\cap B^c\rvert}{\lvert A\setminus B\rvert+\lvert B\setminus A\rvert+\lvert A\cap B\rvert+\lvert A^c\cap B^c\rvert}$
	Russel and Rao	$\frac{\lvert A\cap B\rvert}{\lvert A\setminus B\rvert+\lvert B\setminus A\rvert+\lvert A\cap B\rvert+\lvert A^c\cap B^c\rvert}$
	De Baets	$\frac{\alpha(\lvert A\setminus B\rvert+\lvert B\setminus A\rvert)+\beta\lvert A\cap B\rvert+\gamma\lvert A^c\cap B^c\rvert}{\alpha'(\lvert A\setminus B\rvert+\lvert B\setminus A\rvert)+\beta\lvert A\cap B\rvert+\gamma\lvert A^c\cap B^c\rvert},\ \alpha,\alpha',\beta,\gamma>0$

⇒ Only single features are taken into account and all are given the same "importance"

Motivation: similarity and attribute interactions (2)

GOAL

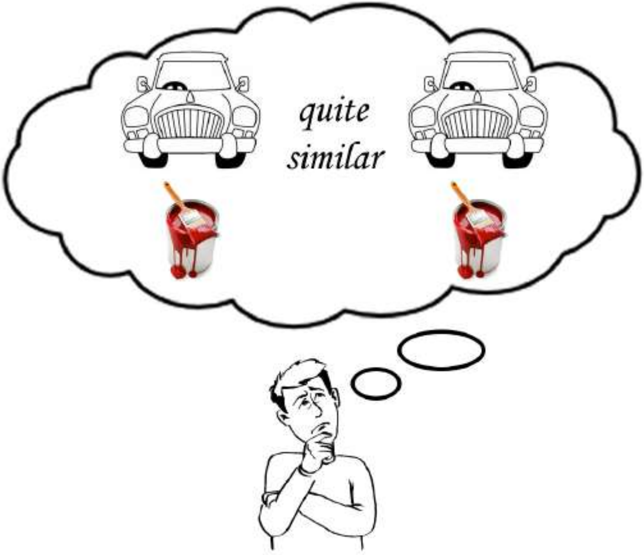
Propose similarity measures able to consider weights which can be interpreted as the “significance” (positive or negative) of groups of attributes

We define similarity measures based on a weight capacity and the Choquet integral, generalising the Jaccard similarity measure

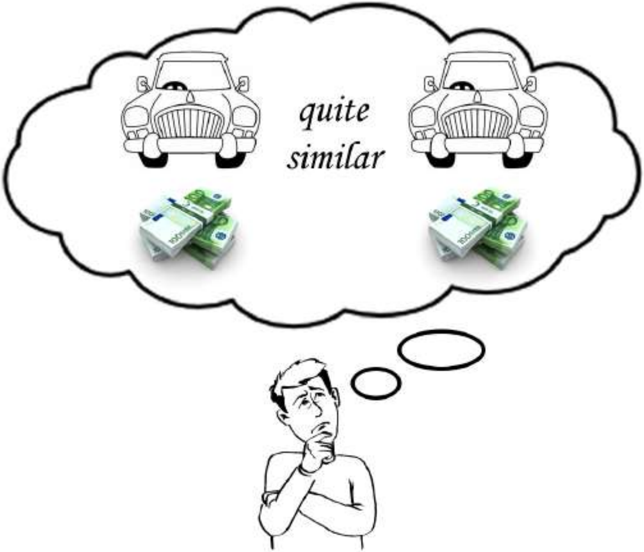
- **Crisp data:** attributes can be only present or absent
- **Fuzzy data:** attributes are present with a degree $\alpha \in [0, 1]$

Interactions among attributes: Cars comparison

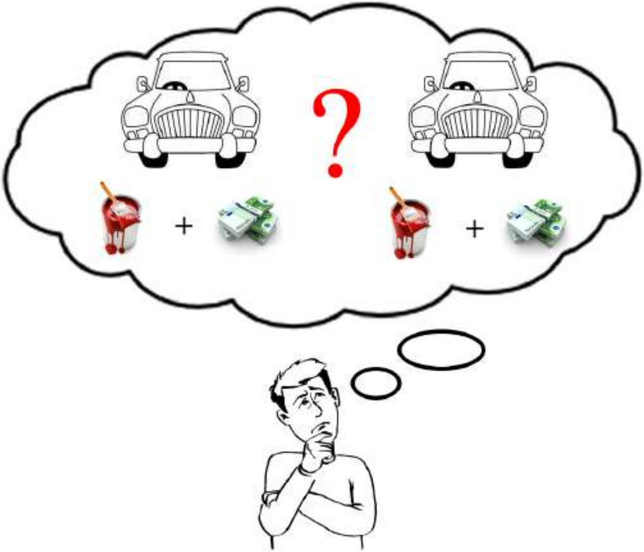
Interactions among attributes: Cars comparison



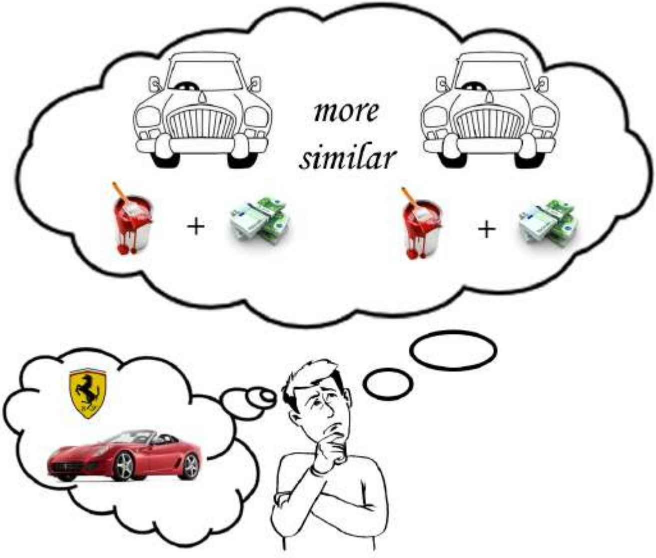
Interactions among attributes: Cars comparison



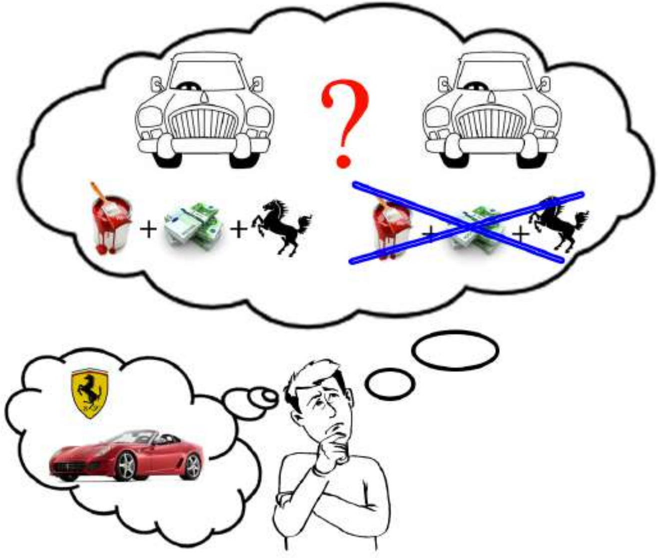
Interactions among attributes: Cars comparison



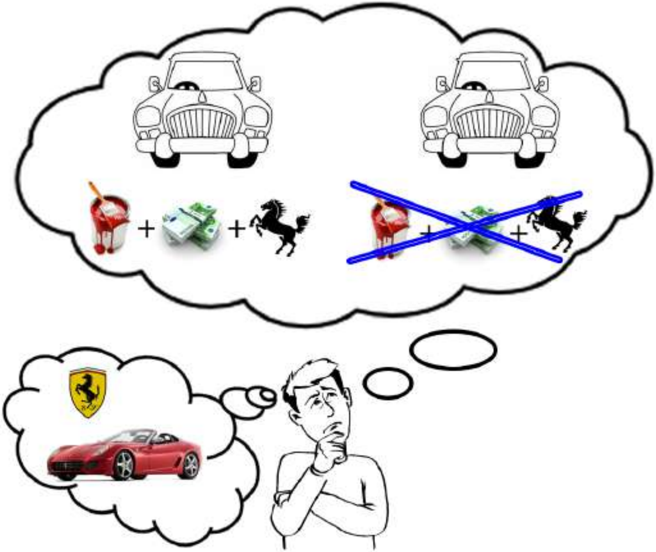
Interactions among attributes: Cars comparison



Interactions among attributes: Cars comparison



Interactions among attributes: Cars comparison



Significance assessment

Consider:

- $N = \{1, \dots, n\}$, a finite index set
- $\wp(N)$, the powerset of N

Significance assessment

A function $\sigma : \wp(N) \to \mathbb{R}$ satisfying the following conditions:
(S1) $\sigma(\emptyset) = 0$;
(S2) $\sum_{\{i\} \subseteq B \subseteq A} \sigma(B) \geq 0$, for every $A \in \wp(N)$ and every $i \in A$.

- $\sigma(A)$ is a weight of "significance" of the set $A \in \wp(N)$
- $\sigma(\{i\}) \geq 0$ for every $i \in N$
- $\sigma(A)$ can be positive or negative when $|A| > 1$

Weight capacity

Weight capacity

Define $\mu : \wp(N) \to [0, +\infty)$, for every $A \in \wp(N)$, as:

$$\mu(A) = \sum_{B \subseteq A} \sigma(B).$$

The function μ satisfies the following properties:
(C1) $\mu(\emptyset) = 0$;
(C2) $A \subseteq B \Longrightarrow \mu(A) \leq \mu(B)$, for every $A, B \in \wp(N)$.

- σ is the **Möbius inversion** of μ
- If $\sum_{A \in \wp(N)} \sigma(A) = 1$, then μ is **normalized**, that is $\mu(N) = 1$
- If $\sigma \geq 0$, the corresponding μ is **totally monotone**
- If $\sigma(A) = 0$ for every $A \in \wp(N)$ such that $|A| > 1$, then the corresponding μ is **additive**

Choquet integral

Consider:

- μ, a weight capacity on $\wp(N)$
- $X \in [0,1]^N$

Choquet integral of X with respect to μ

$$\mathbf{C}_\mu(X) = \sum_{i=1}^{n} \left[X(\pi(i)) - X(\pi(i-1))\right] \mu(\{\pi(i), \ldots, \pi(n)\}),$$

where π is a permutation of N such that $X(\pi(1)) \leq \ldots \leq X(\pi(n))$ and $X(\pi(0)) := 0$.

- If $X \in \{0,1\}^N$ then X can be identified with a subset of N (still denoted with X) and so $\mathbf{C}_\mu(X) = \mu(X)$.

Crisp data

Consider:

- $N = \{1, \ldots, n\}$, a finite set of crisp attribute indices
- Every attribute can be present or absent
- $\mathcal{C} = \{0, 1\}^N$, set of all (crisp) object descriptions

Any object description is regarded as a (crisp) subset of N, which is identified with its **indicator function**, so, we simply denote it as a function $X : N \to \{0, 1\}$

Jaccard similarity measure

$$S_J(X, Y) = \frac{|X \cap Y|}{|X \setminus Y| + |Y \setminus X| + |X \cap Y|} = \frac{|X \Delta Y|}{|X \Delta Y| + |X \cap Y|} = \frac{|X \cap Y|}{|X \cup Y|}.$$

Towards a generalization

1) cardinality $\rightsquigarrow$ weighted mean ($\equiv$ additive capacity μ)

The **weighted mean**

- differentiates the importance of single attributes
- does not care of interactions among attributes

2) weighted mean $\rightsquigarrow$ Choquet integral ($\equiv$ capacity μ)

The **Choquet integral**

- differentiates the importance of single attributes
- cares of interactions among attributes
- σ distinguishes between positive and negative interactions

- Three possible generalized Jaccard similarity measures
- In any case the **maximality condition** holds

An example (1)

Apartments in New York described by the following crisp attributes indexed by $N = \{1, 2, 3, 4\}$:

1: the apartment is located in a skyscraper;

2: the apartment has a terrace;

3: the apartment has a panoramic view;

4: the apartment is equipped with a lift;

$\wp(N)$	1	2	3	4	12	13	14	23	24	34	123	124	134	234	1234
σ	0.3	0.2	0.3	0.2	0.4	−0.1	−0.1	−0.1	0	0	−0.1	0	0	0	0
μ	0.3	0.2	0.3	0.2	0.9	0.5	0.4	0.4	0.4	0.5	0.9	1	0.6	0.6	1

Negative significance assessment

Some combinations of attributes are penalized with a negative σ since their common presence is not discriminative for the similarity of two objects

An example (2)

Subset	1	2	3	4
X	1	0	1	0
Y	0	0	1	1
X'	1	0	1	1
Y'	0	1	1	1

$$\frac{\mu(X \cap Y)}{\mu(X \setminus Y) + \mu(Y \setminus X) + \mu(X \cap Y)} = \frac{3}{8} \quad < \quad \frac{1}{2} = \frac{\mu(X' \cap Y')}{\mu(X' \setminus Y') + \mu(Y' \setminus X') + \mu(X' \cap Y')}$$

$$\frac{\mu(X \cap Y)}{\mu(X \Delta Y) + \mu(X \cap Y)} = \frac{3}{7} \quad > \quad \frac{5}{14} = \frac{\mu(X' \cap Y')}{\mu(X' \Delta Y') + \mu(X' \cap Y')}$$

$$\frac{\mu(X \cap Y)}{\mu(X \cup Y)} = \frac{1}{2} \quad = \quad \frac{1}{2} = \frac{\mu(X' \cap Y')}{\mu(X' \cup Y')}$$

Depending on the particular functional form chosen for generalizing the Jaccard similarity measure, we reach completely different similarity orderings between the pairs (X, Y) and (X', Y')

Fuzzy data

Consider:

- $N = \{1, \ldots, n\}$, a finite set of fuzzy attribute indices
- Every attribute can be present with a degree $\alpha \in [0, 1]$
- $\mathcal{F} = [0, 1]^N$, set of all fuzzy object descriptions
- $\mathcal{C} = \{0, 1\}^N$, set of all crisp object descriptions

Any object description is regarded as a fuzzy subset of N, which is identified with its **membership function**, so, we simply denote it as a function $X : N \to [0, 1]$

Fuzzy set-theoretic operations

- $(\cdot)^c = 1 - (\cdot)$, fuzzy complement
- T, S, pair of dual t-norm and t-conorm such as

$$\begin{aligned} T_M(x,y) &= \min\{x,y\}, & S_M(x,y) &= \max\{x,y\}, \\ T_P(x,y) &= x \cdot y, & S_P(x,y) &= x + y - x \cdot y, \\ T_L(x,y) &= \max\{x+y-1,0\}, & S_L(x,y) &= \min\{x+y,1\}. \end{aligned}$$

Fuzzy set-theoretic operations

For every $X, Y \in \mathcal{F}$, define pointwise on N:

- $X \cap Y = T(X, Y)$
- $X \setminus Y = T(X, Y^c)$
- $Y \setminus X = T(Y, X^c)$
- $X \Delta Y = S(X \setminus Y, Y \setminus X)$
- $X \cup Y = S(X, Y)$

Similarity measures for fuzzy data

Consider:

- μ, a weight capacity
- σ, corresponding significance assessment

Fuzzy Weighted Attribute Combinations Based Similarities

For every $X, Y \in \mathcal{F}$:

$$
\begin{aligned}
S_1^{\mu}(X, Y) &= \frac{\mathbf{C}_{\mu}(X \cap Y)}{\mathbf{C}_{\mu}(X \setminus Y) + \mathbf{C}_{\mu}(Y \setminus X) + \mathbf{C}_{\mu}(X \cap Y)} \\
S_2^{\mu}(X, Y) &= \frac{\mathbf{C}_{\mu}(X \cap Y)}{\mathbf{C}_{\mu}(X \Delta Y) + \mathbf{C}_{\mu}(X \cap Y)} \\
S_3^{\mu}(X, Y) &= \frac{\mathbf{C}_{\mu}(X \cap Y)}{\mathbf{C}_{\mu}(X \cup Y)}
\end{aligned}
$$

If the denominator of S_i^{μ} vanishes, we set $S_i^{\mu}(X, Y) := 0$

Some immediate properties

- Under an additive μ [Scozzafava, Vantaggi 2008], the similarity measures S_1^μ, S_2^μ and S_3^μ coincide on $\mathcal{C}^2$, but are generally different on $\mathcal{F}^2 \setminus \mathcal{C}^2$
- Even if μ is additive, the maximality condition $S_i^\mu(X,X) \geq S_i^\mu(X,Y)$, for $i = 1, 2$, may fail
- In general, there is no dominance relation between S_1^μ, S_2^μ and S_3^μ if we consider the whole $\mathcal{F}^2$ and an arbitrary weight capacity μ
- In general, T'-transitivity (possibly $T' \neq T$), i.e., for every $X, Y, Z \in \mathcal{F}$, $S_i^\mu(X,Z) \geq T'(S_i^\mu(X,Y), S_i^\mu(Y,Z))$, for $i = 1, 2$, may fail

Proposition

If the weight capacity $\mu : \wp(N) \to [0, +\infty)$ is additive, then the similarity measure S_3^μ is T_L-transitive.

General failure of dominance

Consider:

- $N = \{1, 2, 3\}$
- T, S, any pair of dual t-norm and t-conorm
- μ_1, μ_2, the weight capacities on $\wp(N)$ below

$\wp(N)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	N
μ_1	0	0.1	0.1	0.1	0.2	0.2	0.2	1
μ_2	0	0.25	0.25	0.5	0.5	0.5	0.5	1

Fuzzy subset	1	2	3
X	0	1	1
Y	1	1	0

- $S_3^{\mu_1}(X, Y) = 0.1 < 0.3333 = S_1^{\mu_1}(X, Y) = S_2^{\mu_1}(X, Y)$
- $S_1^{\mu_2}(X, Y) = S_3^{\mu_2}(X, Y) = 0.25 < 0.3333 = S_2^{\mu_2}(X, Y)$

Superadditive and subadditive capacities

Restricting S_1^μ, S_2^μ and S_3^μ on $\mathcal{C}^2$:

- If μ is **superadditive**, i.e., for every $A, B \in \wp(N)$ with $A \cap B = \emptyset$ it holds

$$\mu(A \cup B) \geq \mu(A) + \mu(B),$$

  then $S_1^\mu(X, Y) \geq S_2^\mu(X, Y) \geq S_3^\mu(X, Y)$ for every $X, Y \in \mathcal{C}$

- If μ is **subadditive**, i.e., for every $A, B \in \wp(N)$ with $A \cap B = \emptyset$ it holds

$$\mu(A \cup B) \leq \mu(A) + \mu(B),$$

  then $S_1^\mu(X, Y) \leq S_2^\mu(X, Y) \leq S_3^\mu(X, Y)$ for every $X, Y \in \mathcal{C}$

Failure of dominance under super/subadditivity

Consider:

- $N = \{1, 2, 3\}$
- μ_1, μ_2, the super/subadditive capacities on $\wp(N)$ below

$\wp(N)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	N
μ_1	0	0	0	0	0	0	0	1
μ_2	0	1	1	1	1	1	1	1

Fuzzy subset	1	2	3
X	0.2	0.4	0.3
Y	0.9	0.7	0.8

For $T = T_M$, $S = S_M$

- $S_1^{\mu_1}(X, Y) = 0.2222 < S_3^{\mu_1}(X, Y) = 0.2857 < S_2^{\mu_1}(X, Y) = 0.6666$
- $S_1^{\mu_2}(X, Y) = 0.2666 < S_3^{\mu_2}(X, Y) = 0.4444 < S_2^{\mu_2}(X, Y) = 0.5714$

For $T = T_L$, $S = S_L$

- $S_3^{\mu_1}(X, Y) = 0.1 < S_1^{\mu_1}(X, Y) = 0.25 < S_2^{\mu_1}(X, Y) = 1$
- $S_3^{\mu_2}(X, Y) = 0.1 < S_1^{\mu_2}(X, Y) = 0.125 < S_2^{\mu_2}(X, Y) = 1$

Failure of T'-transitivity

Consider:

- $N = \{1, 2, 3\}$
- $T = T_M$ and $S = S_M$
- μ_1, the superadditive capacity on $\wp(N)$ below

$\wp(N)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	N
μ_1	0	0	0	0	0	0	0	1

Fuzzy subset	1	2	3
X	0.47	0.87	0.95
Y	0.46	0.99	0.56
Z	0.98	0.23	0.21

$S_1^{\mu_1}(X,Z) = 0.75$, $S_1^{\mu_1}(X,Y) = 0.884615$, $S_1^{\mu_1}(Y,Z) = 0.875$,

$S_2^{\mu_1}(X,Z) = 0.913043$, $S_2^{\mu_1}(X,Y) = 0.978723$, $S_2^{\mu_1}(Y,Z) = 0.954545$,

$S_3^{\mu_1}(X,Z) = 0.241379$, $S_3^{\mu_1}(X,Y) = 0.978723$, $S_3^{\mu_1}(Y,Z) = 0.375$,

$$S_i^{\mu_1}(X,Z) < T_L(S_i^{\mu_1}(X,Y), S_i^{\mu_1}(Y,Z)) \leq T_M(S_i^{\mu_1}(X,Y), S_i^{\mu_1}(Y,Z))$$

A paradigmatic example (1)

We consider 3 students x, y, z evaluated with respect to 3 subjects [Grabisch 1995]: mathematics (1), physics (2) and literature (3), whose final marks are given on a scale from 0 to 20:

Student	1	2	3
x	18	16	10
y	10	12	18
z	14	15	15

$\rightsquigarrow$

Fuzzy subset	1	2	3
X	0.9	0.8	0.5
Y	0.5	0.6	0.9
Z	0.7	0.75	0.75

A paradigmatic example (2)

It is common knowledge that “usually” students good at mathematics are also good at physics, and vice versa.

Take the capacity $\mu : \wp(N) \to [0,1]$ given below:

$\wp(N)$	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	N
σ	0	0.45	0.45	0.3	-0.4	0.15	0.15	-0.1
μ	0	0.45	0.45	0.3	0.5	0.9	0.9	1

- μ is neither **superadditive** nor **subadditive**

A paradigmatic example (3)

Taking $T = T_M$ and $S = S_M$:

S_1^μ :

	X	Y	Z
X	0.5538	0.4866	0.5298
Y	0.4866	0.5354	0.5281
Z	0.5298	0.5281	0.5775

S_2^μ :

	X	Y	Z
X	0.7680	0.5266	0.6368
Y	0.5266	0.6974	0.6318
Z	0.6368	0.6318	0.7322

S_3^μ :

	X	Y	Z
X	1	0.6124	0.7591
Y	0.6124	1	0.8034
Z	0.7591	0.8034	1

Denote with $\precsim_i$ the weak order induced by the similarity measure S_i^μ on $\{X, Y, Z\}^2$, for $i = 1, 2, 3$:

$$\begin{matrix}(X,Y)\\(Y,X)\end{matrix} \prec_1 \begin{matrix}(Y,Z)\\(Z,Y)\end{matrix} \prec_1 \begin{matrix}(X,Z)\\(Z,X)\end{matrix} \prec_1 (Y,Y) \prec_1 (X,X) \prec_1 (Z,Z)$$

$$\begin{matrix}(X,Y)\\(Y,X)\end{matrix} \prec_2 \begin{matrix}(Y,Z)\\(Z,Y)\end{matrix} \prec_2 \begin{matrix}(X,Z)\\(Z,X)\end{matrix} \prec_2 (Y,Y) \prec_2 (Z,Z) \prec_2 (X,X)$$

$$\begin{matrix}(X,Y)\\(Y,X)\end{matrix} \prec_3 \begin{matrix}(X,Z)\\(Z,X)\end{matrix} \prec_3 \begin{matrix}(Y,Z)\\(Z,Y)\end{matrix} \prec_3 \begin{matrix}(X,X)\\(Y,Y)\\(Z,Z)\end{matrix}$$

Conclusions and future perspectives

- The use of the Choquet integral and a weight capacity μ ($\equiv$ a significant assessment σ) increases the expressive power of the studied similarity measures: we can incorporate positive or negative interactions among the attributes
- We have an exponential (with respect to $|N|$) number of parameters to specify
- The most "natural" procedure to obtain μ (or σ) is through the elicitation by a field expert
- A learning procedure can be envisaged [Baioletti, Coletti, Petturiti 2012] analogous to metric function learning for learning a Mahalanobis distance

THANK YOU FOR YOUR ATTENTION