

AUTOMATIC ENHANCEMENT OF THE REFERENCE SET FOR MULTI-CRITERIA SORTING IN THE FRAME OF THESEUS METHOD

Eduardo FERNANDEZ* Jorge NAVARRO** Eduardo SALOMON**

Abstract. Some recent works have established the importance of handling abundant reference information in multi-criteria sorting problems. More valid information allows a better characterization of the agent's assignment policy, which can lead to an improved decision support. However, sometimes information for enhancing the reference set may be not available, or may be too expensive. This paper explores an automatic mode of enhancing the reference set in the framework of the THESEUS multi-criteria sorting method. Some performance measures are defined in order to test results of the enhancement. Several theoretical arguments and practical experiments are provided here, supporting a basic advantage of the automatic enhancement: a reduction of the vagueness measure that improves the THESEUS accuracy, without additional efforts from the decision agent. The experiments suggest that the errors coming from inadequate automatic assignments can be kept at a manageable level.

Keywords: Multi-criteria decision; Sorting; Outranking methods; Knowledge management

1. Introduction

This paper addresses the multi-criteria sorting problem, which is considered a particular case of classification problems involving preferences (cf. [10], [11]). Unlike nominal classification problems in which the classes (groups, categories) have been defined in a nominal way, sorting refers to problems in which the categories have been defined in an ordinal way ([3]). Among different decision aiding situations concerning sorting problems, our interest in this paper is focused on a particular case in which i) a set of ordered categories (from the worst to the best in the specific decision aiding context) has been

*Corresponding author, Autonomous University of Sinaloa

** Autonomous University of Sinaloa

defined (or agreed) by a decision agent; ii) there is a universe of objects that are characterized by multiple criteria; iii) the decision agent has assigned (or agreed with the assignment of) a set of objects (reference set, training set) which are considered representative of his/her assignment policy; iv) the decision agent assesses (or agrees with) a predicate $M(a,b)$ which, based on the multi-criteria description of the objects a,b , establishes the truth value of a certain preference statement. $M(a,b)$ may come from an outranking model, a value function, a system of rules, or other multi-criteria decision model. $M(a,b)$ may be assessed directly by the decision agent, or inferred from the reference set through a preference-disaggregation analysis (e.g. [4], [6], [8], [11], [13]); v) the decision agent wants to use the reference set as a basic knowledge in order to obtain appropriate prescriptions with a simultaneous minimization of his/her participation in new assignments.

The preference information contained in $M(x,y)$ and in the reference set can be considered *preference knowledge*. Multi-criteria sorting methods use such information in order to suggest new assignments in a compatible way with the assignment policy that may be implicit in this knowledge. More valid knowledge should be preferred. Reference sets of bigger size should provide more information about the agent's assignment policy, thus allowing more appropriate assignments.

Here, we are interested in outranking-based sorting methods. Unlike other multi-criteria approaches, outranking methods can handle ordinal and imprecise information, incomparability and veto effects (cf. [9]). Some recent papers have addressed the size of the reference set and its importance in achieving a good characterization of the categories ([2], [7], [9]). If the object to be sorted were incomparable with most actions in the reference set, an outranking-based sorting method would suggest ill-defined (vague) assignments. To illustrate this point, let us consider the below trivial example taken from [7]. Table 1 shows a reference set which is built from a very simple assignment policy. This can be formulated as: 'If its average criterion value is close to 3 or less than 3, the object should be assigned to C_1 . If it is close to 4, assign the object to C_2 . If it is close to 5 or greater than 5, assign the object to C_3 '.

In the following b_{ki} denote the i -th object assigned to category C_k .

Table 1. A reference set

Reference element	Multi-criteria description (increasing preferences)	Category
b_{11}	(3, 3, 3, 3)	C_1
b_{21}	(4, 4, 4, 3)	C_2
b_{31}	(4, 6, 5, 4)	C_3

Let $\sigma(x,y)$ denote the credibility of 'x outranks y'. It will be calculated as in ELECTRE III and ELECTRE TRI (e.g. [15]). Let us accept that the ELECTRE III model parameters (weights, indifference thresholds, preference thresholds, veto thresholds) are set to: $w_j=0.25$; $q_j=0$; $p_j=1$; $v_j=2$; ($j=1,\dots,4$). Consider a new object $x=(2, 2, 2, 6)$ to be sorted. Note that $\sigma(b_{k1}, x)=0$, $k=1, 2, 3$, $\sigma(x, b_{11})=0.25$, $\sigma(x, b_{k1})=0$, $k=2, 3$. So, x is incomparable with any reference action. This leads to an ill-determined assignment of (2, 2, 2, 6). Now, suppose that the reference set is expanded by introducing $b_{12}=(2, 2, 3, 5)$, assigned to C_1 . Note that $\sigma(x, b_{12})=0.75 = \sigma(b_{12}, x)$. The valued indifference relation proposed by Fernandez et al. in [5],[6] $i(x, b_{12})= \min(\sigma(x, b_{12}), \sigma(b_{12}, x))$ suggests assigning x to C_1 ,

prescription which corresponds to the assignment policy. So, the new reference element introduced new information useful for a better characterization of the assignment policy.

The ability to capture the agent's assignment policy and make well-defined assignments depends on the reference set. Expanding the reference set may be a way to improve results. In this paper, the process of including new reference elements aimed at making more appropriate assignments is called 'enhancement of the reference set'.

The process of enhancement may be performed in three different ways:

1. A decision agent provides new assignment examples;
2. In automatic mode, without agent's participation. Thus, the method itself decides which new objects should be incorporated into the reference set;
3. A semi-automatic way in which the proper sorting method suggests new assignments to agent approval.

In the first way, the new reference examples may be provided by:

- a) former assignments accepted as valid by the decision agent;
- b) assignment of a limited set of objects for which the decision agent is comfortable expressing his/her decision policy.

To the best of our knowledge, the paper by Greco et al. ([12]) is the deepest analysis about enhancement processes in the scientific literature. They proposed an interesting approach in which inclusions of new elements in the reference set are performed in an interactive way. Their assignments should be compatible with the previous assignments in the frame of an additive value function model. The consistency preference model-reference set can be increased by removing objects from the set or modifying their assignments.

However, in some problems the decision maker (*DM*) may be inaccessible or may be a mythical person. Sometimes there is no person to ask about new assignments; the reference set may be obtained from past data as in bankruptcy prediction, corporate risk assessment, or data-based scholar success assessment. If there is no identifiable decision maker, if (s)he is not accessible, or if (s)he does not want to provide more assignment examples, the cardinality of the reference set could be increased by the self method (automatic mode). An examination of whether the automatic enhancement mode provides good results should be welcomed.

Here we are interested in studying the automatic enhancement mode in the framework of the THESEUS method proposed in [7]. The main goal is to find out if the automatic enhancement increases the THESEUS capacity to suggest better assignments. This paper is structured as follows: Some notation and premises are described in Section 2. A brief outline of the THESEUS method is given in 2.2. Some theoretical arguments and the proposal of enhancement process are presented in Section 3, followed by detailed numerical examples (Section 4). Finally, we draw some concluding remarks.

2. Some background

2.1. Basic premises and notation

Some premises follow:

i) There is a finite set of ordered categories $Ct = \{C_1, \dots, C_M\}$, ($M \geq 2$); C_M is assumed to be the preferred category. The term ‘preferred’ is related to each particular sorting problem (for instance, ‘more quality’, ‘more consensual’, ‘less risky’).

ii) Let U be the universe of objects x described by a coherent set of N criteria, denoted $G = \{g_1, g_2, \dots, g_j, \dots, g_N\}$, with $N \geq 3$. This set has been defined in order to evaluate any object considered for assignment to a certain category ([1],[16]).

iii) There is a decision agent who has (or agrees with) a certain decision policy defined on a subset U' of the universe. According to this policy, there should be a certain function $F: U' \rightarrow Ct$ such that for each $x \in U'$ the decision agent accepts that $F(x)$ is the most appropriate assignment of x . F expresses the assignment policy from the agent (or accepted by him/her).

iv) There is a set of reference objects or training examples T , which is composed of elements $b_{kh} \in U$ assigned to category C_k , ($k = 1, \dots, M$).

v) The decision agent agrees with a fuzzy outranking relation $\sigma(x, y)$ defined on $U \times U$. Its value models the degree of credibility of the statement ‘ x is at least as good as y ’ from the agent’s perspective in each particular sorting problem (for instance, ‘ x is at least as consensual as y ’, or ‘ x is at most as risky as y ’).

Remarks:

- The decision agent may be a real decision maker (who has his/her preferences and decision policy), or be an actor in charge of solving the problem, accepting the assignment examples;
- The reference set may come from the decision agent or be accepted by him/her when the information comes from past data;
- The reference set contains implicit multi-criteria preferences among objects. A high level of compatibility between σ and T is desirable. In order to hold a full compatibility, σ should satisfy:
For each $(b_{kh}, b_{k'l})$, if b_{kh} is at least as good as $b_{k'l}$ with a sufficient high degree of credibility, THEN $k \geq k'$.
Looking for compatibility with the reference information, the σ parameters may be inferred by using preference-disaggregation methods as in [6],[8]; these approaches minimize inconsistencies with the above implication.

Let us consider a real value $\lambda > 0.5$.

Definition 1: Given $\sigma(x, y)$, the following crisp binary relations are defined on the universe:

$(x, y) \in S(\lambda)$ iff $\sigma(x, y) \geq \lambda$ (λ -outranking)

$(x, y) \in P(\lambda)$ iff $\sigma(x, y) \geq \lambda \wedge \sigma(y, x) < 0.5$ (λ -strict preference)

$(x, y) \in Q(\lambda)$ iff $\sigma(x, y) \geq \lambda \wedge 0.5 \leq \sigma(y, x) < \lambda$ (λ -weak preference)

$(x, y) \in I(\lambda)$ iff $\sigma(x, y) \geq \lambda \wedge \sigma(y, x) \geq \lambda$ (λ -indifference).

$(x, y) \in R(\lambda)$ iff $\sigma(x, y) < \lambda \wedge \sigma(y, x) < \lambda$ (λ -incomparability).

Definition 2: We say that a reference set T covers Ct if for each $C_k \in Ct$ there is at least one $b_{kh} \in T$.

2.2. The THESEUS method

The THESEUS method is based on comparing a new object to be assigned with reference objects through models of preference and indifference relations (cf. [7]). The assignment is not a consequence of the object intrinsic properties; it is rather the result of comparisons with other objects whose assignments are known. In the following $C(x)$ denotes a potential assignment of object x . According to THESEUS, $C(x)$ should satisfy:

$$\forall x \in U, \forall b_{kh} \in T$$

$$\begin{aligned} xP(\lambda)b_{kh} &\Rightarrow C(x) \succeq C_k \\ b_{kh}P(\lambda)x &\Rightarrow C_k \succeq C(x) \end{aligned} \quad (1.a)$$

$$\begin{aligned} xQ(\lambda)b_{kh} &\Rightarrow C(x) \succeq C_k \\ b_{kh}Q(\lambda)x &\Rightarrow C_k \succeq C(x) \end{aligned} \quad (1.b)$$

$$xI(\lambda)b_{kh} \Rightarrow (C(x) \succeq C_k) \wedge (C_k \succeq C(x)) \Leftrightarrow C(x) = C_k \quad (1.c)$$

(\succeq denotes the statement “is not worse than” on the set of categories, which is related to the decision-aiding context).

Note that $C(x)$ is a variable whose domain is the set of ordered categories. Equations (1.a-c) express the necessary consistency amongst the preference model, the reference set and the appropriate assignments of x . The assignment $C(x)$ should be as compatible as possible with the current knowledge about the assignment policy.

THESEUS uses the inconsistencies with Equations (1.a-c) to compare the possible assignments of x . More specifically:

- The set of $P(\lambda)$ -inconsistencies for x and $C(x)$ is defined as $D_{P=} = \{(x, b_{kh}), (b_{kh}, x), b_{kh} \in T \text{ such that (1.a) is FALSE}\}$;
- The set of $Q(\lambda)$ -inconsistencies for x and $C(x)$ is defined as $D_{Q=} = \{(x, b_{kh}), (b_{kh}, x), b_{kh} \in T \text{ such that (1.b) is FALSE}\}$;
- The set of $I(\lambda)$ -inconsistencies for x and $C(x)$ is defined as $D_{I=} = \{(x, b_{kh}), (b_{kh}, x), b_{kh} \in T \text{ such that (1.c) is FALSE}\}$.

Suppose that $C(x) = C_k$ and consider $b_{jh} \in T$. Some cases in which $xI(\lambda)b_{jh} \wedge |k-j|=1$ might be explained by “discontinuity” of the description; x may be close to the upper (lower) boundary of C_k and b_{jh} may be close to the lower (upper) boundary of C_j . They will be called second-order $I(\lambda)$ -inconsistencies and grouped in the set D_{2I} . The set $D_{1I} = D_{I=} - D_{2I}$ contains the so-called first-order $I(\lambda)$ -inconsistencies, which are not consequences of the described discontinuity effect. Let n_P, n_Q, n_{1I}, n_{2I} denote the cardinality of the above-defined inconsistency sets, and $N_1 = n_P + n_Q + n_{1I}, N_2 = n_{2I}$.

THESEUS suggests an assignment that minimizes the above inconsistencies with lexicographic priority favouring N_1 , which is the most important criterion ([7]). The basic assignment rule is:

For each $x \in U$ and given a minimum credibility level $\lambda > 0.5$

- a) Starting with $k=1$ ($k=1, \dots, M$) and considering each $b_{kh} \in T$, calculate $N_1(C_k)$;
- b) Identify the set $\{C_j\}$ whose elements hold $C_j = \text{argmin} N_1(C_k)$;
- c) Select $C_{k^*} = \text{argmin} N_2(C_i)$;
 $\{C_j\}$
- d) If C_{k^*} is a single solution, assign x to C_{k^*} ; other situations are approached below.

The suggestion may be a single category or a sequence of categories. The first case is called a well-defined assignment; otherwise, the obtained solution highlights the highest category (C_H) and the lowest category (C_L) which is appropriate for assigning the object, but fails in determining the most appropriate. Such solution will be called "a vague assignment". In this sense, the following uniqueness theorem is important:

Theorem: A THESEUS solution C_{k^*} is unique only if one of the following conditions is held:

- i) there is (l, m) such that $xS(\lambda)b_{k^*}$ and also $b_{k^*m}S(\lambda)x$;
- ii) there is a b_{ML} such that $xS(\lambda)b_{ML}$;
- iii) there is a b_{1n} such that $b_{1n}S(\lambda)x$.

In case ii) $C_{k^*} = C_M$; in iii) $C_{k^*} = C_1$.

A proof of a similar result was given in [7].

In simple words, in order to be unique the THESEUS assignment should be 'bounded' by two objects in the same category, or be 'bounded' by objects belonging to extreme categories.

3. The proposal

3.1. Theoretical bases

Definition 3: If $x \in U$ is assigned by THESEUS to an element (well-specified) of Ct , we say that the assignment of x is THESEUS well-defined.

Definition 4: Consider the subset U'' of the universe that includes all the objects whose assignments are THESEUS well-defined. The THESEUS assignment function F_{TH} is defined as $F_{TH}: U'' \rightarrow Ct$ in which $F_{TH}(x)$ is the assignment of x suggested by THESEUS.

Definition 5: Under the above notation, for a finite universe U the THESEUS vagueness measure is defined as

$$V_g = 1 - \left[\sum_{j=1}^{card(U)} z_j(x_j) / card(U) \right] \quad (2)$$

where $x_j \in U$ and $z_j = 1$ if x_j is THESEUS-well defined; otherwise, $z_j = 0$.

Definition 6: The λ -assignment range for C_k in the set T is defined as

For $1 < k < M$, $A_{\lambda,k}(T) = \{x \in U \text{ such that there is } (b_{kb}, b_{kn}) \in T \times T \text{ which satisfies } xS(\lambda)b_{kl} \wedge b_{kn}S(\lambda)x\}$;

For $k = 1$, $A_{\lambda,1}(T) = \{x \in U \text{ such that there is } b_{1n} \in T \text{ which satisfies } b_{1n}S(\lambda)x\}$;

For $k = M$, $A_{\lambda,M}(T) = \{x \in U \text{ such that there is } b_{Ml} \in T \text{ which satisfies } xS(\lambda)b_{Ml}\}$.

The above concept can be interpreted as the set of objects of the universe for which there are positive arguments for assigning them to C_k . If $x \in A_{\lambda,k}(T)$, assigning this object to a category different from C_k would be inconsistent with Eq. 1. $x \in A_{\lambda,k}(T)$ is a necessary condition for assigning x to C_k (uniqueness theorem). Hence, if there is no C_k such that $x \in A_{\lambda,k}(T)$, then the assignment of x is not well-defined.

The following proposition was proved in [7]:

Proposition 1: Let T' be equal to $T \cup \{b_{kl}\}$ and $T \cap \{b_{kl}\} = \emptyset$. Then $A_{\lambda,k}(T) \subseteq A_{\lambda,k}(T')$.

That is, $A_{\lambda,k}(T)$ may be different from $A_{\lambda,k}(T')$; the addition of b_{kl} may enhance $A_{\lambda,k}$ and increase the cardinal of the THESEUS assignment function. So, if T is enhanced with new reference elements, from Proposition 1 and the uniqueness theorem the THESEUS vagueness measure should decrease.

Definition 7: Under the above notation, for a finite universe U the THESEUS accuracy measure is defined as

$$A_c = 1 / \text{card}(U) \sum_{j=1}^{\text{card}(U)} y_j(x_j) \quad (3)$$

where $x_j \in U$ and $y_j = 1$ if $F_{TH}(x_j) = F(x_j)$; otherwise, (including when x_j does not belong to $U' \cap U'$), $y_j = 0$.

In case of non-finite universe, the vagueness and accuracy measures can be extended as:

$$V_g = 1 - \left[\lim_{N\text{sample} \rightarrow \infty} \sum_{j=1}^{N\text{sample}} z_j(x_j) / N\text{sample} \right] \quad (4)$$

$$A_c = \lim_{N\text{sample} \rightarrow \infty} 1 / N\text{sample} \sum_{j=1}^{N\text{sample}} y_j(x_j) \quad (5)$$

where x_j comes from a sample of the universe whose cardinality is $N\text{sample}$. V_g and A_c can be estimated from a random sample of large size.

Remark: As was discussed above, the domain of the THESEUS assignment function is influenced by $\text{card}(T)$. As much greater $\text{card}(T)$ is through an enhancement process, more $y_j = 1$ to be added in Eqs.3 and 5, thus improving the THESEUS accuracy measure. Let us formalize this issue by introducing some probability-based arguments.

Let us denote by

P_{ac} : the probability of obtaining an accurate assignment;

$P_{well-def}$: the probability of obtaining a well-defined assignment;

$P(acc/well-defined)$: the conditional probability of obtaining an accurate assignment given it is well-defined.

The probability $P(acc/well-defined)$ depends on two factors: a) how exactly the relations $P(\lambda)$, $Q(\lambda)$, $I(\lambda)$ reflect real preference and indifference relations of the decision-maker; b) the quality of information in the reference set. Moreover, the discontinuity effect, present in the reference set, should make $P(acc/well-defined)$ be lower than 1. Suppose that a well-defined assignment of x is given by $xI(\lambda)b_{kl}$. If b_{kl} is close to the “border” between C_k and C_{k+1} , may be $F(x) = C_{k+1}$, although THESEUS suggests assigning x to C_k .

Since A_c and V_g can be interpreted as probabilities and $P_{ac} = P(acc/well-defined)$. $P_{well-def}$, we can write

$$A_c \approx P(acc/well-defined) \cdot (1 - V_g) \quad (6)$$

If we assume that $P(acc/well-defined)$ is not a decreasing function on card (T), then A_c should be increased when the vagueness measure is reduced as a consequence from an enhancement process of the reference set.

3.2. Enhancing the reference set

In this section, our proposal of automatic enhancement is justified by analogy with an enhancement of the reference set conducted by a real decision maker, whose preferences are compatible with the information contained in that set.

First of all, let us revisit Premise iii) of Section 2. The assignment policy may not be clearly defined in the DM 's mind. Indeed, this policy should be based on certain relational system of preferences, which is reflected in different assignments. Given $x \in U$, $F(x)$ can be estimated for other objects of the universe in a way compatible with the DM 's preferences. F should contain all the pairs $(x, F(x))$ for which a) the DM feels comfortable with the assignment, and b) the assignment is compatible with the DM 's system of preferences taking into account the information from other previously known/accepted assignments. Thus, the assignment function should be compatible with the DM 's system of preferences, as well as with the set of categories and some assignments previously made by the DM or accepted by him/her. To some extent, this claim for consistency among new assignments, previous assignments and DM preferences was suggested by Greco et al. in [12], but in the framework of additive value models. Here, this is extended to relational models.

More formally:

Assumption1: The DM agrees with certain indifference, strict, weak preference and outranking relations on a subset of U denoted by I , P , Q , and S respectively. They were defined by Roy in [16] as follows:

- Indifference corresponds to the existence of clear and positive reasons that justify equivalence between the two actions (objects).
- Strict preference corresponds to the existence of clear and positive reasons that justify significant preference in favor of one (identified) of the two actions. The statement ‘ x is strictly preferred to y ’ is denoted by xPy .

- Weak preference corresponds to the existence of clear and positive reasons in favor of x over y , but these are not sufficient to justify strict preference. Indifference and strict preference cannot be distinguished appropriately. This is denoted by xQy .
- Outranking: It corresponds to the existence of clear and positive reasons that justify the statement “ x is at least as good as y ”, but with no significant division being established among the situations of strict preference, weak preference and indifference. Notation: xSy .

Assumption 2: The assignment of x is compatible with the DM 's preference relations and the previously accepted assignments if

$\forall b_{kh}$ whose assignment has been performed/approved by the DM :

$$x \text{ outranks } b_{kh} \Rightarrow \text{the category of } x \text{ is at least as good as } C_k \quad (7)$$

$$b_{kh} \text{ outranks } x \Rightarrow C_k \text{ is at least as good as the category of } x$$

This assumption claims for a basic consistency between preferences and assignments that are expression of preferences.

Note that (7) matches with Eq. 1 when S is replaced by $S(\lambda)$. So, the THESEUS assigning manner given by Eq. 1 is a model of the compatible assignment policy defined by Assumption 2. Hence, in order to be unique, an assignment derived from (7) should fulfil the Uniqueness Theorem of THESEUS solutions. Thus, x is assigned to C_{k^*} only if i) there is a pair (l,m) such that xSb_{k^*l} and also $b_{k^*m}Sx$; ii) there is a b_{ML} such that xSb_{ML} ; or iii) there is a b_{1n} such that $b_{1n}Sx$. In case ii) $C_{k^*} = C_M$; in iii) $C_{k^*} = C_1$. Informally, x is assigned to C_{k^*} when the DM finds good arguments to make this assignment based on his/her particular outranking relation and the previous assignments (s)he performed or approved, in a way compatible with (7). So, an enhancement process on a finite universe (conducted by a real DM that agrees with Assumption 2) should obey the following logic:

S-based enhancement

$\forall x_j \in U$ and $x_j \notin T^{(i)}$

x_j is incorporated to $T^{(i)}$ if one of the below propositions is true:

- a) There is $b_{1h} \in T^{(i)}$ such that $b_{1h}Sx_j$;
- b) There is $b_{Ml} \in T^{(i)}$ such that x_jSb_{Ml} ;
- c) There is $(b_{kh}, b_{kl}) \in T^{(i)} \times T^{(i)}$ such that x_jSb_{kh} and $b_{kl}Sx_j$.

In case a), assign x_j to C_1 ; in b), assign x_j to C_M ; in case c), assign x_j to C_k .

The superscript i denotes the i -th iteration of the enhancement process. This process finishes when there is no x_j satisfying the above conditions.

In case of non-finite universe, the above algorithm can be applied by restricting x_j to a large-cardinality finite subset of the universe.

In order to perform an automatic process that simulates the S -based enhancement, the DM should be replaced by a model of his/her outranking relation. Let us introduce the following assumption:

Assumption 3: The cutting level λ and the parameters of the model $\sigma(x,y)$ can be identified in order to achieve $(P(\lambda) \cup Q(\lambda)) \subset S$.
From this assumption

$$(b_{kh}P(\lambda)x_j \text{ or } b_{kh}Q(\lambda)x_j) \Rightarrow b_{kh}Sx_j; \quad (8.a)$$

$$(x_jP(\lambda)b_{kh} \text{ or } x_jQ(\lambda)b_{kh}) \Rightarrow x_jSb_{kh}; \quad (8.b)$$

$$(x_jP(\lambda)b_{kh} \text{ or } x_jQ(\lambda)b_{kh}) \text{ and } (b_{kl}P(\lambda)x_j \text{ or } b_{kl}Q(\lambda)x_j) \Rightarrow x_jSb_{kh} \text{ and } b_{kl}Sx_j. \quad (8.c)$$

Using (8.a, 8.b, 8.c) in the above S -based enhancement, the following procedure for the enhancement in automatic mode is straightforward:

Automatic enhancement procedure (T, T^{enh})

Initialize i

$$T^{enh(i)} = T$$

$$T^{enh(i+1)} = \Phi$$

DO WHILE $T^{enh(i+1)} \neq T^{enh(i)}$

$$\forall x_j \in U \text{ and } x_j \in T^{enh(i)}$$

x_j is incorporated to $T^{enh(i+1)}$

End of Loop

$$\forall x_j \in U \text{ and } x_j \notin T^{enh(i)}$$

x_j is incorporated into $T^{enh(i+1)}$ if one of the below propositions is true:

- d) There is $b_{1h} \in T^{enh(i)}$ such that $b_{1h}P(\lambda)x_j$ or $b_{1h}Q(\lambda)x_j$;
- e) There is $b_{Ml} \in T^{enh(i)}$ such that $x_jP(\lambda)b_{Ml}$ or $x_jQ(\lambda)b_{Ml}$;
- f) There is $(b_{kh}, b_{kl}) \in T^{enh(i)} \times T^{enh(i)}$ such that $(x_jP(\lambda)b_{kh} \text{ or } x_jQ(\lambda)b_{kh})$ and $(b_{kl}P(\lambda)x_j \text{ or } b_{kl}Q(\lambda)x_j)$.

In case d), assign x_j to C_1 ; in e), assign x_j to C_M ; in case f), assign x_j to C_k .

End of Loop

End of DO

$$T^{enh} = T^{enh(i+1)}$$

End of procedure

In plain language: an element x is accepted as a new reference example if there are convincing arguments in favor of Conditions a), b) or c) of the S -based enhancement process. The automatic enhancement finishes when no new element holds Conditions d), e) or f), which are sufficient for a), b), c). The procedure only works on finite universes. Infinite ones should be approximated by large-size finite subsets. The simplest way may be the following: i) For each criterion, define a range of variation that covers the universe; ii) Use the indifference threshold in order to make a partition of each criterion range; if the indifference threshold is zero, then use a fraction of the strict preference threshold; iii) If P_i denotes the partition in the i -th criterion, take the sample of the universe as the Cartesian product $P_1 \times P_2 \times \dots \times P_N$, where N is the number of criteria.

Under Assumption 3, all x_j that satisfy $d)$, $e)$ or $f)$, also satisfy $a)$, $b)$, $c)$. The reciprocal is not true. So, the automatically enhanced reference set should be a subset of the S -based enhanced set. Thus, Assumption 3 guarantees certain correspondence between the automatic enhancement and the enhancement conducted by the DM . However, in practice there could be pairs (x_j, b_{kh}) and (b_{kh}, x_j) belonging to $(P(\lambda) \cup Q(\lambda))$ but not to S . For these pairs Conditions $d)$, $e)$ or $f)$ may be true although $a)$, $b)$, $c)$ are false. In such cases, the assignment of x_j performed by the automatic enhancement procedure may be different from $F(x_j)$. If $F(x_j)$ does not coincide with the prescription of the enhancement procedure, the error can be propagated in its next iterations. So, the fulfillment of Assumption 3 may be a central issue, which can be addressed by two different ways: i) increasing λ , thus reducing the cardinality of $(P(\lambda) \cup Q(\lambda))$; or ii) applying a preference disaggregation analysis procedure (e.g. [4, 6, 8]) in order to achieve a more appropriate settlement of model's parameters. The first way should be carefully used, since reducing cardinality of $(P(\lambda) \cup Q(\lambda))$ could decrease the cardinality of T^{enh} .

Replacing S by $S(\lambda)$, the S -based enhancement procedure may be extended to perform a semi-automatic enhancement: the method suggests new incorporations to the reference set, and the DM approves/rejects its suggestions. Thus, the new assignment examples should be coincident with F . σ parameters should be re-inferred from time to time and consistency analyses should be introduced.

4. Two illustrative numerical tests

First example:

The universe is generated by a multi-criteria functional model which simulates the assignment policy from a decision agent. Let us consider the model $F = 0.2 (F_1 + F_2 + F_3 + F_4 + (F_1 \cdot F_2 \cdot F_3 \cdot F_4)^{0.25})$. For both the criteria F_i and the decision categories, the following scale is considered: *Very Poor* = 1, *Poor* = 2, *Below Average* = 3; *Average* = 4; *Above Average* = 5; *Good* = 6; *Very Good* = 7. F_i takes every integer value in the interval [1, 7]. So, $card(U) = 7^4 = 2401$.

Function F takes on values from [1, 7]. If a fractional value is obtained for F , it is rounded either up or down, and an integer F_n is obtained. If the decimal part were exactly 0.5, F_n would be randomly chosen amongst $F-0.5$ and $F+0.5$. The categories of the different objects are determined by their F_n values ($F_n = 1, 2, \dots, 7$ correspond to *Very Poor*, *Poor*, ... *Very Good*).

The following experiment was replicated 20 times, each with different value of N' :

- i. A covering reference set T of size N' is randomly generated;
- ii. λ and the parameters of the fuzzy outranking relation are settled;
- iii. THESEUS V_g and A_c are calculated according to Eq. 2 and 3;
- iv. $P(acc/well-defined)$ is calculated from Equation 6;
- v. T^{enh} is obtained by applying the automatic enhancement procedure;
- vi. V_g and A_c are calculated by using T^{enh} ;
- vii. $P(acc/well-defined)$ is calculated from Equation 6.

$\sigma(x,y)$ is calculated as in ELECTRE-III, with the simplification proposed by Mousseau and Dias ([14]) in calculating the discordance index. The parameter settlement follows:

Table 2. Average results (non-enhanced vs. enhanced reference set)

Card (T)	$V_g(T)$	$A_c(T)$	P_b	$C_{H^+}C_L \geq 2$ (%)	Card(T^{enh})	$V_g(T^{enh})$	$A_c(T^{enh})$	P_a	$C_{H^+}C_L \geq 2$ (%)
7	0.907	0.071	0.763	78.63	10.95	0.906	0.072	0.762	78.63
10	0.851	0.118	0.792	67.94	15.45	0.840	0.128	0.796	66.87
25	0.611	0.33	0.848	24.96	61.60	0.484	0.444	0.862	18.64
50	0.359	0.562	0.877	6.28	169.50	0.184	0.735	0.900	2.71
100	0.164	0.762	0.911	0.52	419.90	0.050	0.887	0.934	0.16

Weights $w_i = 0.25, i=1, \dots, 4$;
 Indifference thresholds $q_i = 0.2, i=1, \dots, 4$;
 Preference thresholds $p_i = 0.7, i=1, \dots, 4$;
 Veto thresholds $v_i = 3, i=1, \dots, 4$;
 Pre-veto thresholds $u_i = 2, i=1, \dots, 4$.
 The cutting threshold λ was set as 0.7.
 The average results are shown in Table 2 and Table 3

Note: C_H, C_L refer to the highest and lowest category suggested by THESEUS assignment rule. P_b denotes $P(\text{acc/well-defined})$ before the enhancement; P_a denotes the same probability after the enhancement process.

Table 3. Average number of inappropriate assignments by the enhancement process

Card (T)	Card (T ^{enh})	% of inappropriate assignments in T ^{enh}
7	10.95	0
10	15.45	0
25	61.60	0
50	169.50	0
100	419.90	0

Remarks:

- According to Table 3, no errors are introduced by the enhancement process in the enhanced reference set;
- The vagueness measure decreases with $\text{card}(T)$ (Table 2, second column);
- In case of vague assignments, the range of suggested categories is narrower when $\text{card}(T)$ increases (fifth column);
- The THESEUS accuracy increases with $\text{card}(T)$ (third column);
- Improvements of $P(\text{acc/well-defined})$ are pointed out in column fourth when $\text{card}(T)$ is increased. With more reference information, THESEUS assignments may be more justified, and a well-defined assignment may be based on more arguments, being thus more reliable;
- The vagueness measure decreases when the enhancement process takes place (compare columns 2 and 7);
- In case of vague assignments, the range of suggested categories is reduced by the enhancement process (compare columns 5 and 10);
- The THESEUS accuracy is improved by the enhancement process (compare columns 3 and 8); in some cases, increases above 30% are achieved when the before-enhancement accuracy is compared to the after-enhancement accuracy;
- $P(\text{acc/well-defined})$ is slightly improved by the enhancement process (compare columns 4 and 9 in Table 2). A greater cardinality of the enhanced reference set produces a slight increment of that probability.

The distribution of the objects which are incorporated into the reference set is far from being uniform. Compared with Condition f) of the automatic enhancement procedure, the fulfillment of Conditions d) and e) is easier. Hence, in proportion, more objects are incorporated into categories C_1 and C_7 . Table 4 shows results of the enhancement process per category. For comparison, the distribution of the universe per category is given by the first row. Each reference set was randomly generated, taking care to cover the set of

Table 4. Distribution per category

	$Card(C_1)$	$Card(C_2)$	$Card(C_3)$	$Card(C_4)$	$Card(C_5)$	$Card(C_6)$	$Card(C_7)$
UNIVERSE	15	195	645	885	529	121	11
$Card(T) = 7$	1/4	1/1	1/1	1/1	1/1	1/1	1/4
$Card(T) = 10$	1/2	1/1	2/2	2/2	2/2	1/1	1/2
$Card(T) = 25$	1/3	3/5	8/12	9/58	2/4	1/1	1/4
$Card(T) = 50$	1/4	4/4	14/50	14/32	12/68	4/5	1/4
$Card(T) = 100$	2/3	8/13	22/92	35/161	20/45	12/37	1/4

categories. Their cardinality per category (before enhancement/after enhancement) is given by the different columns.

Note that proportionally to the universe C_1 and C_7 are more populated in the enhanced set. With $Card(T) = 7$ and $Card(T) = 10$, the enhancement is weak and limited to those categories. This effect tends to be reduced for bigger T since its intermediate categories are initially more populated, thus increasing the probability of fulfilling the enhancement condition f).

Second example:

Here each object in the universe is characterized by four criteria (F_1, F_2, F_3, F_4) which take integer values on the interval $[1, 7]$. So, $card(U) = 7^4 = 2401$.

A fuzzy outranking relation (with its parameters settled as in the first example) is defined on the universe. Like before, λ was set as 0.7.

Using σ -values, a merit quantity of each object with respect to the entire universe is calculated as $F(x_j) = card(U_{st}) - card(U_w)$, where

$$U_{st} = \{y_j \in U \text{ such that } \sigma(x_j, y_j) \geq \lambda\} \text{ and } U_w = \{y_j \in U \text{ such that } \sigma(y_j, x_j) \geq \lambda\}$$

The DM's assignment policy is simulated by using the merit F as follows:

F_{min} and F_{max} denote the minimum and maximum values of F on the universe. The range $F_{max} - F_{min}$ is divided in seven intervals of equal length $[F_{min}, a_1), [a_1, a_2), \dots, [a_6, F_{max}]$. Then:

If $F(x_j) \in [F_{min}, a_1)$, x_j is assigned to *Very Poor*;

If $F(x_j) \in [a_1, a_2)$, x_j is assigned to *Poor*;

.

.

.

If $F(x_j) \in [a_6, F_{max}]$, x_j is assigned to *Very Good*.

Like in the first example, the following experiment was replicated 20 times:

- i. A covering reference set T of size N' is randomly generated according to the above simulated assignment policy;
- ii. THESEUS V_g and A_c are calculated according to Eq. 2 and 3;
- iii. $P(acc/well-defined)$ is calculated from Equation 6;
- iv. T^{enh} is obtained by applying the automatic enhancement procedure
- v. THESEUS V_g and A_c are calculated by using T^{enh} ;
- vi. $P(acc/well-defined)$ is calculated from Equation 6.

The average results are provided by Table 5 and Table 6.

Table 5. Average results (non-enhanced vs. enhanced reference set)

Card (T)	$V_g(T)$	$A_c(T)$	P_b	$C_{H-C_L} \geq 2$ (%)	$\text{Card}(T^{\text{enh}})$	$V_g(T^{\text{enh}})$	$A_c(T^{\text{enh}})$	P_a	$C_{H-C_L} \geq 2$ (%)
7	0.852	0.111	0.75	76.45	53.35	0.830	0.122	0.721	75.45
10	0.803	0.147	0.746	66.73	60.15	0.775	0.164	0.728	65.24
25	0.605	0.305	0.772	35.34	84.00	0.561	0.334	0.762	32.82
50	0.386	0.497	0.809	10.52	142.55	0.297	0.568	0.807	7.36
100	0.182	0.691	0.845	1.87	300.70	0.103	0.758	0.845	0.80

Note: C_H , C_L refer to the highest and lowest category suggested by THESEUS assignment rule. P_b denotes $P(acc/well-defined)$ before the enhancement; P_a denotes the same probability after the enhancement process.

Table 6. Average number of inappropriate assignments by the enhancement process

Card (T)	Card (T^{enh})	Inappropriate assignments in T^{enh}	% of erroneous information with respect to $card(T^{enh})$
7	53.55	1.00	1.87
10	60.15	0.85	1.41
25	84.00	2.05	2.44
50	142.55	4.15	2.91
100	300.70	12.20	4.05

As in the first example, the following remarks come from Table 5:

- The vagueness measure decreases with $card(T)$;
- In case of vague assignments, the range of suggested categories is narrower when $card(T)$ increases;
- The THESEUS accuracy increases with $card(T)$;
- From $card(T)=10$, $P(acc/well-defined)$ increases with the reference set size;
- The vagueness measure decreases when the enhancement process takes place;
- In case of vague assignments, the range of suggested categories is reduced by the enhancement process;
- The THESEUS accuracy is improved by the enhancement process.

According to Table 6 the automatic enhancement introduces some erroneous information in the reference set. This means that the models $S(\lambda)$ does not represent as well as in the first example the real outranking relation S from the DM . Therefore, $P(acc/well-defined)$ is lower, as shown by the fourth columns in Table 2 and Table 5. The inaccurate information introduced in the enhanced reference set may reduce $P(acc/well-defined)$ as shown in Table 5, columns 4 and 9. Anyway, the decrement of $P(acc/well-defined)$ due the enhancement process is always lower than 4%. When $card(T) = 50, 100$, although the proportion of erroneous assignments is higher, $P_a \cong P_b$ is a consequence of the bigger size of T^{enh} , since more reference information should provide greater reliability to THESEUS well-defined assignments.

The distribution per category is provided by Table 7.

Table 7. Distribution per category (second example)

	$Card(C_1)$	$Card(C_2)$	$Card(C_3)$	$Card(C_4)$	$Card(C_5)$	$Card(C_6)$	$Card(C_7)$
UNIVERSE	116	249	496	679	496	249	116
$Card(T) = 7$	1/36	1/1	1/1	1/1	1/1	1/1	1/4
$Card(T) = 10$	1/26	1/1	2/2	1/1	2/2	2/6	1/36
$Card(T) = 25$	1/60	3/3	4/4	8/8	4/4	3/3	2/16
$Card(T) = 50$	3/60	5/5	10/12	17/33	9/9	3/3	3/24
$Card(T) = 100$	3/33	12/51	19/47	24/52	29/53	9/12	4/34

C_1 and C_7 are strongly privileged by the enhancement process. With $Card(T)=7$ and $Card(T)=10$, the enhancement is almost limited to those categories, but strong in comparison to the universe. There is a tendency to reduce this privilege, but it is slighter than in the first example.

5. Some conclusions

Enhancing the reference set is a way of enlarging the preference knowledge in multi-criteria sorting problems. This improves the reference set capacity to suggest more appropriate assignments. Some assumptions guarantee certain correspondence between the automatic enhancement proposed here and the enhancement conducted by a real *DM*, whose assignment function is compatible with his/her particular outranking relation and the previous assignments (s)he performed or approved.

The automatic enhancement mode can be applied by other multi-criteria sorting methods that use several assignment examples by each category. It is only necessary to have a computable model of a preference predicate ' x seems to be preferred to y '. Here, the study was based on the THESEUS method, but some results may be generalized.

THESEUS performs better when the cardinality of the reference set is increased. Several theoretical arguments and practical experiments were given in this paper supporting the following concluding remarks:

- the THESEUS vagueness measure decreases when $card(T)$ increases;
- the THESEUS accuracy increases with $card(T)$;
- $P(acc/well-defined)$ (the conditional probability of obtaining an accurate assignment given it is well-defined) tends to increase when more valid information is included in the reference set;
- the THESEUS automatic enhancement mode can be based on λ -strict preference and λ -weak preference relations, which model decision maker's asymmetric preference relations;
- the automatic enhancement mode has several basic advantages: i) it allows a reduction of the vagueness measure what improves the THESEUS accuracy; ii) even if vagueness is not eliminated, the range of suggested categories is reduced by the enhancement process; iii) these improvements can be achieved without the decision maker and without new reference examples;
- its main drawback concerns the potential introduction of incorrectly assigned new objects into the enhanced reference set; nevertheless, these errors seem to be kept at a manageable level;
- the automatic enhancement may influence $P(acc/well-defined)$. Incorrectly assigned reference elements affect this probability. On the other hand, more valid information in the reference set helps a better characterization of the *DM*'s assignment policy, thus positively influencing on $P(acc/well-defined)$. In the analyzed examples $P(acc/well-defined)$ shows a slight variation in a range of -4% to 2.5%;
- the accuracy increases with $P(acc/well-defined)$ and when the vagueness measure is reduced. Since $P(acc/well-defined)$ is approximately constant, the automatic enhancement process increases the THESEUS accuracy. In the analyzed examples,

- when $P(\text{acc/well-defined})$ is reduced due to the introduction of some erroneous information during the enhancement process, the vagueness measure decreases more rapidly, thus allowing an improvement of the accuracy;
- if $P(\lambda)$ and $Q(\lambda)$ were close to the real decision maker's asymmetric preferences, $P(\text{acc/well-defined})$ would tend to be improved by the enhancement process, which would be even more effective; this could be handled by increasing the cutting threshold λ . However, this should be carefully performed, since reducing cardinality of $(P(\lambda) \cup Q(\lambda))$ could decrease the cardinality of T^{enh} and hence increase vagueness;
- in the automatic enhancement mode, there is a clear tendency to favor the least preferred and the most preferred categories. Thus, the enhancement should be more effective when most of the universe is concentrated in these extreme classes.

Although in the examples the DM has been simulated by a decision model, the practical application of the automatic enhancement has similar features. The proposal is useful under the following conditions:

- A) There are no sufficient previous assignments and the DM cannot or does not want to provide or approve more reference assignments;
- B) The DM agrees with Assumption 2, which claims for consistency among new assignments, previous assignments and his/her system of preferences;
- C) The outranking model parameters can be elicited in such a way that the model of the DM 's asymmetric preference relation is basically a subset of his/her outranking relation.

Several aspects of this proposal can be extended to the semi-automatic enhancement mode. Comparisons between both modes, and discussions about the reasonable limit of the enhancement process, and how the cognitive effort from the DM can be reduced and consistency analyses should be performed in the semi-automatic mode, are topics of a new paper.

Acknowledgement

We acknowledge support from CONACyT project no. 57255.

References

- [1] Almeida-Dias J., Figueira J., Roy B., Electre Tri-C: A multiple criteria sorting method based on characteristic reference actions, *European Journal of Operational Research*, **204**, 3, 2010, 565-580.
- [2] Almeida-Dias J., Figueira J., Roy B., A multiple criteria sorting method where each category is characterized by several reference actions: The ELECTRE TRI-NC method, *European Journal of Operational Research*, **217**,3, 2012, 567-579.
- [3] Doumpos M., Zopounidis C., *Multicriteria Decision Aid Classification Methods*, Kluwer Academic Publishers, Dordrech-Boston-London, 2002.
- [4] Doumpos M., Marinakis Y., Marinaki M., Zopounidis C., An evolutionary approach to construction of outranking models for multicriteria classification: The case of the ELECRE-TRI method, *European Journal of Operational Research*, **199**,2, 2009, 496-505.

-
- [5] Fernandez E., Navarro J., Duarte A., Multicriteria sorting using a valued preference closeness relation, *European Journal of Operational Research*, **185**, 2, 2008, 673-686.
- [6] Fernandez E., Navarro J., Bernal S., Multicriteria sorting using a valued indifference relation under a preference-disaggregation paradigm, *European Journal of Operational Research*, **198**, 2, 2009, 602-609.
- [7] Fernandez E., Navarro J., A new approach to multicriteria sorting problems based on fuzzy outranking relations: The THESEUS method, *European Journal of Operational Research*, **213**, 2, 2011, 405-413.
- [8] Fernandez E., Navarro J., Mazcorro G., Evolutionary multi-objective optimization for inferring outranking model's parameters under scarce reference information and effects of reinforced preference, *Foundations of Computing and Decision Sciences*, **37**, 3, 2012, 163-197.
- [9] Figueira J., Greco S., Roy B., Słowiński R., ELECTRE Methods: Main features and Recent Developments, In: Zopounidis C., Pardalos P.(eds.), *Handbook of Multicriteria Analysis. Applied Optimization*, Springer, Heidelberg-Dordrecht-London-New York, 2010, 51-89.
- [10] Greco S., Matarazzo B., Słowiński R., The use of rough sets and fuzzy sets in Multiple-Criteria Decision Making, In: Gal T., Stewart T., Hanne T.(eds.), *Advances in Multiple Criteria Decision Making*, Kluwer Academic Publishers, Boston, 1999, 14.1-14.59.
- [11] Greco S., Matarazzo B., Słowiński R., Rough sets theory for multicriteria decision analysis, *European Journal of Operational Research*, **129**, 1, 2001, 1-47.
- [12] Greco S., Mousseau V., Słowiński R., Multiple criteria sorting with a set of additive value functions, *European Journal of Operational Research*, **207**, 3, 2010, 1455-1470.
- [13] Jacquet-Lagrange E. Siskos J., Assessing a set of additive utility functions for multicriteria decision making: The UTA method, *European Journal of Operational Research*, **10**, 2, 1982, 151-164.
- [14] Mousseau V., Dias LC., Valued outranking relations in ELECTRE providing manageable disaggregation procedures, *European Journal of Operational Research*, **156**, 2, 2004, 467-482.
- [15] Roy B., The outranking approach and the foundations of ELECTRE methods, *Theory and Decision*, **31**, 1, 1991, 49-73.
- [16] Roy B., *Multicriteria methodology for Decision Aiding*, Kluwer Academic Publisher, Dordrecht-Boston-London, 1996.

Received September, 2013