

Handling incomplete information in the framework of THESEUS multicriteria sorting method

Jorge Navarro¹, Eduardo Fernandez², Rafael Olmedo³

¹Universidad Autónoma de Sinaloa, Culiacan, Mexico

jnavarro@uas.uasnet.mx, eddyf@uas.uasnet.mx, rafael@uas.edu.mx

Abstract

In this paper we face the multicriteria sorting problem considering that we have incomplete information. Multicriteria sorting is a particular case of classification problems. It consists in the assignment of some actions to some pre-defined classes. Classification refers to problems where the classes (groups, categories) have been defined in a nominal way. We use the term “incomplete information” to indicate the absence of a value in some criterion of the object to be assigned to a class (category).

Keywords: multicriteria sorting problem, incomplete information, decision support.

1. Introduction

In this paper we face the multicriteria sorting problem. Multicriteria sorting is a particular case of classification problems. It consists in the assignment of some actions to some pre-defined classes. Classification refers to problems where the classes (groups, categories) have been defined in a nominal way. On the other hand, sorting refers to problems in which the categories have been defined in an ordinal way (cf. [9]). Moreover, multicriteria sorting refers to problems in which all objects to be classified or sorted are described by a set of criterion functions. We use the term “incomplete information” to indicate the absence of a value in some criterion of the object to be assigned to a class (category).

The DM's rationality is limited by the lack of availability of information, his/her cognitive limitation and the amount of time he/she has to make a decision. Simon [5, 6] points out that the majority of people is not completely rational and, in many situations, acts on their emotional, not entirely rational, impulses. Simon's principle of “bounded rationality” suggests that a DM rather uses heuristic methods to make decisions than rigid rules of optimization, since due to the complexity of situation or his/her own inability to process all the alternatives he/she always lacks necessary information to make an optimal decision.

For inconsistent data or vague knowledge, Artificial Intelligence has a tool to conceptualize, analyze and organize various types of data: rough set theory [17]. Several approaches have been developed to handle incomplete information systems, such as those with unknown values of some attributes. This theory proves to be adequate for reasoning in incomplete information systems.

In [12], reduction of knowledge that eliminates information not essential from the DM's point of view was proposed. Different ways of handling incomplete information have been suggested, such as removal of objects with unknown values from the original system. Other methods transform an incomplete system into complete where objects from the original incomplete information system are represented by other objects in the target system [13]. Taking only as a hypothesis a fact that the unknown value of an attribute is from a known domain, Kryszkiewicz [12] shows how to find, from an incomplete decision table, decision-making rules with a minimal number of conditions. This method requires no change in the original system and is able to reduce knowledge by efficiently eliminating only the information that is not essential from the DM's point of view. By handling the concept of dominance, this approach takes into account decision maker's preferences, but the increase in the amount of information decreases its ability of making appropriate assignments.

Another approach is based on a polyvalent logic that admits six, partially ordered, truth values [11] —actually, generalization of a trivalent logic system created by Kleene [14, 10] —and the existence of the highest and lowest level for each pair, forming a lattice [1]. It provides a process for drawing conclusions while working with incomplete information. Encheva [15] proposes a method based on formal concept analysis and the six-valued logic [7] that includes: true, false, contradiction, unknown and an intermediate level of truth [8] between unknown and true (or false) denominated unknown true (or unknown false). To obtain knowledge in cases with incomplete information it applies Armstrong's axioms [16].

Among the methods that support multicriteria decision making, those based on fuzzy outranking relation stand out due to their ability to handle ordinal, qualitative and uncertain information. The uncertainty is handled by using thresholds that identify its limits and allow handling its intensities. Despite all this, these methods have not been used in problems of multicriteria sorting with incomplete information. The present paper attempts to fill that void.

2. Brief description of THESEUS method

Let us consider the following premises:

- There is a set of ordered categories $C_t = \{C_1, \dots, C_M\}$ ($M \geq 2$); where we suppose that C_M is the best category.
- There is a universe U of objects described by a set of independent criteria $G = \{g_1, \dots, g_N\}$. This set has been defined to evaluate any object; $a = (g_1(a), \dots, g_N(a)) = (a_1, \dots, a_N)$.
- There is a decision maker (DM) with certain assignment policy, which is implicit in a set of T reference objects composed of elements $b_{k,h} \in U$ assigned to the category C_k , ($k = 1, \dots, M$).

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

Definition 1: The following crisp binary relations are given on the universe:

$$\begin{aligned}
 (x, y) \in S(\lambda) &\text{ iff } \sigma(x, y) \geq \lambda \quad (\lambda\text{-outranking}) \\
 (x, y) \in P(\lambda) &\text{ iff } \sigma(x, y) \geq \lambda \wedge \sigma(y, x) < 0.5 \quad (\lambda\text{-strict preference}) \\
 (x, y) \in Q(\lambda) &\text{ iff } \sigma(x, y) \geq \lambda \wedge 0.5 \leq \sigma(y, x) \quad (\lambda\text{-weak preference}) \\
 (x, y) \in I(\lambda) &\text{ iff } \sigma(x, y) \geq \lambda \wedge \sigma(y, x) \geq \lambda \quad (\lambda\text{-indifference}) \\
 (x, y) \in R(\lambda) &\text{ iff } \sigma(x, y) < \lambda \wedge \sigma(y, x) < \lambda \quad (\lambda\text{-incomparability})
 \end{aligned}$$

In [4], we proposed the THESEUS method, which is based on comparing a new object to be assigned with reference objects through models of preference and indifference relations. 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 and $C(b)$ is the actual assignment of b . According to THESEUS, $C(x)$ should satisfy:

$$\begin{aligned}
 \forall x \in U, \forall b \in T \\
 xP(\lambda)b \Rightarrow C(x) \succeq C(b) \\
 bP(\lambda)x \Rightarrow C(b) \succeq C(x) \quad (1.a) \\
 xQ(\lambda)b \Rightarrow C(x) \succeq C(b) \\
 bQ(\lambda)x \Rightarrow C(b) \succeq C(x) \quad (1.b) \\
 xI(\lambda)b \Rightarrow C(x) \succeq C(b) \wedge C(b) \succeq C(x) \Leftrightarrow C(x) = C(b) \quad (1.c)
 \end{aligned}$$

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

Some $I(\lambda)$ -inconsistencies might be explained by ‘discontinuity’ of the description based on the set of categories. The cases in which $xI(\lambda)b \wedge |k - j| = 1$ 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 (cf. [4]). The basic assignment rule is

- Assign the minimum credibility level $\lambda > 0.5$;
- Starting with $k = 1$ ($k = 1, \dots, M$) and considering each $b_{k,h} \in T$, calculate $N_1(C_k)$;
- Identify the set $\{C_j\}$ whose elements hold $C_j = \text{argmin } N_1(C_k)$;
- Select $C_{k^*} = \text{argmin } N_2(C_i)$;
 $\{C_j\}$
- If C_{k^*} is a single solution, assign x_j 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.

3. Our Proposal

Let us suppose that the object x should be assigned; it possesses a subset $S_x = \{s_1, \dots, s_r\}$ of criteria (i.e. $S_x \subset G$) such that the information on object x in S_x criteria is unknown.

In the case that information on criteria for assigning the object x is known, we use the THESEUS method [4]. The same method is employed, although in a modified version, if the criteria values for its assigning are unknown. We propose three alternatives to modify the THESEUS method.

3.1. Modification of Sigma

The value $\sigma(a, b)$ indicates the credibility level of the statement “the alternative a is at least as good as the alternative b ” and is calculated in the same way as in ELECTRE III (cf. [2]).

$$\sigma(a, b) = C(a, b) \wedge \sim D(a, b)$$

where $C(a, b)$ is the level of concordance of the statement “ a is at least as good as b ” based on the values of a and b criteria; $D(a, b)$ represents the level of discordance with that statement (against evidences).

To calculate $C(a, b)$ agreement indices $c_j(a, b)$ are defined for each criterion j from the index $c_j(a_j, b_j)$, defined as:

$$\begin{cases} c_j(a_j, b_j) = 0 \Leftrightarrow p_j < b_j - a_j \\ 0 < c_j(a_j, b_j) < 1 \Leftrightarrow q_j < b_j - a_j \leq p_j \\ c_j(a_j, b_j) = 1 \Leftrightarrow b_j - a_j \leq q_j \end{cases}$$

where q_j and p_j are indifference and preference thresholds, respectively.

To calculate $D(a, b)$ discordance indices $d_j(a, b)$ are defined for each criterion j from the index $d_j(a_j, b_j)$, defined as:

$$\begin{cases} d_j(a_j, b_j) = 0 \Leftrightarrow b_j - a_j \leq u_j \\ 0 < d_j(a_j, b_j) < 1 \Leftrightarrow u_j < b_j - a_j \leq v_j \\ d_j(a_j, b_j) = 1 \Leftrightarrow v_j < b_j - a_j \end{cases}$$

where u_j and v_j are pre-veto and veto thresholds, respectively.

Modification:

If $y \in U$ is a complete information object, we define the concordance index $c_j(x, y)$ as follows:

If $j \notin S_x$, $c_j(x, y)$ is calculated in the traditional fashion.

If $j \in S_x$,

$$c_j(x, y) = \frac{\sum_{\forall z \in G_j} c_j(y_j, z)}{\text{Card}(G_j)}$$

The discordance indices $d_j(x, y)$ and $d_j(y, x)$ are defined as follows:

If $j \notin S_x$, $d_j(x, y)$ and $d_j(y, x)$ is calculated in the traditional fashion.

If $j \in S_x$,

$$d_j(x, y) = \frac{\sum_{\forall z \in G_j} d_j(z, y_j)}{\text{Card}(G_j)}$$

$$d_j(y, x) = \frac{\sum_{\forall z \in G_j} d_j(y_j, z)}{\text{Card}(G_j)}$$

Once the calculations of concordance and discordance indices for all the criteria are obtained, the integration of these values to calculate $\sigma(a, b)$ is done as follows:

$$C_j(a, b) = \sum c_j(a, b)$$

$$D_j(a, b) = \max \{d_j(a, b)\}$$

$$\sigma(a, b) = C(a, b)(1 - D(a, b))$$

Lastly, the assigning of x is performed using THESEUS method and the particular model of σ explained above.

3.2. Majority

In this alternative of the modified THESEUS method, the object x is assigned one value for each unknown attribute value and assigned to a category according to the THESEUS method. The above procedure is repeated for each variant in the domains of unknown attribute values, where the categories associated with assignments are stored in a common file. Once all the variants are used, the category(ies) that has(ve) the majority (fashion) in the set of ranking categories that are being stored, is selected as the assignment of x .

3.3. Mean Value

Here we give up using exhaustive testing in which all values of the unknown criteria are tested. We try to make the burden of assigning mainly rest upon information on known criteria and the influence of unknown values minimal. That is achieved by assigning the intermediate value to the criteria with incomplete information and afterwards, by applying THESEUS method to find the object assignment.

4. Experimentation

We have considered objects with 5 criteria. The domain of each criterion coincides with the categories of object assignment {worst, very bad, bad, regular, good, very good, excellent}.

In total, 4 experiments were made: 2 experiments used T references size 50 (NR_1_50, NR_2_50); 2 experiments used references size 100 (NR_1_100, NR_2_100). Reference sets were generated using automatic generation method developed by [3]. In this method, outranking model parameters and object attributes are randomly generated; object assignments are compatible with THESEUS method.

In order to contrast different ways of assigning objects with incomplete information, each $x \in T$ was assigned considering T' as reference and THESEUS as sorting method. We assume that the correct assignment of the object is the one calculated by THESEUS method.

Incomplete Information

For every object $x \in T$, 5 assignments were made for each of the variants, one per each suppressed attribute. That is to say, based on an object x objects x^1, x^2, x^3, x^4 and x^5 are generated, where x^i is identical to x with the only difference that x^i lacks information in the i -th attribute. Every object x^i is assigned in every variant and is contrasted with the assignment generated by THESEUS with complete information.

Let us define the following terms:

- *Vagueness of the assignment* - Vagueness measures the rank amplitude of the set of categories suggested by a method.

- *Match* - term used to refer to absolute coincidence between the category suggested by the method and the real object classification.
- *Weakness* - this term refers to the partial coincidence between the method's and real object classification.
- *Failure* - is the absence of coincidence between the method's and real object classification.
- *Level of Failure* - is the distance between the assignments of two methods, i.e., it is the number of categories that separates the assignment suggested by the method from the real object classification. This level is only greater than zero when there is a failure.

Vagueness results obtained by each method for each reference are shown in tables 1, 2, 3 and 4.

Table 1: Vagueness with NR_1_50

Vagueness	0	1	2	3	4	5	6
Method							
Mean Value	245	5	0	0	0	0	0
Majority	0	239	11	0	0	0	0
Modified Sigma	248	2	0	0	0	0	0

Table 2: Vagueness with NR_2_50

Vagueness	0	1	2	3	4	5	6
Method							
Mean Value	200	18	27	5	0	0	0
Majority	0	249	1	0	0	0	0
Modified Sigma	247	3	0	0	0	0	0

Table 3: Vagueness with NR_1_100

Vagueness	0	1	2	3	4	5	6
Method							
Mean Value	484	16	0	0	0	0	0
Majority	59	436	5	0	0	0	0
Modified Sigma	499	1	0	0	0	0	0

Table 4: Vagueness with NR_2_100

Vagueness	0	1	2	3	4	5	6
Method							
Mean Value	498	2	0	0	0	0	0
Majority	0	499	1	0	0	0	0
Modified Sigma	499	1	0	0	0	0	0

The results of contrasting in relation to the real object classification are shown in tables 5, 6, 7 and 8.

Table 5: Contrasting with NR_1_50

Result	Marches	Weakness	Failure	Level of failure
Method				
Mean Value	189	5	56	1:50 2:6
Majority	0	197	53	1:36 2:11 3:6
Modified Sigma	197	2	51	1:44 2:6 3:1

Table 6: Contrasting with NR_2_50

Result	Marches	Weakness	Failure	Level of failure
Method				
Mean Value	168	17	65	1:61 2:4 3:
Majority	0	191	59	1:32 2:5 3:10 4:12
Modified Sigma	174	2	74	1:47 2:23 3:4

Table 7: Contrasting with NR_1_100

Result	Marches	Weakness	Failure	Level of failure
Method				
Mean Value	401	5	94	1:72 2:22 3:0
Majority	55	372	73	1:43 2:30 3:0
Modified Sigma	404	1	95	1:67 2:28 3:0

Table 8: Contrasting with NR_2_100

Result	Marches	Weakness	Failure	Level of failure
Method				
Mean Value	417	2	81	1:72 2:9 3:0
Majority	0	430	70	1:49 2:17 3:4
Modified Sigma	425	3	72	1:65 2:7 3:0

Meaning of $N_1:N_2$ coding given in the Level of Failure column:

- N_1 represents the number of categories that separate method's from the real classification.
- N_2 represents the total of objects whose assignment presents the level of failure N_1 .

5. Conclusions

- In the four experiments carried out, the method based on the modification of sigma calculation had the best performance: less vagueness and a higher number of successes. In other words, the recommended assignments are more precise and have a greater amount of coincidences with the real classification when they are obtained using complete information.
- The performance of the mean value method was very similar to that of the modified sigma, the only difference being a bit lower performance.
- The majority method had the poorest performance.
- More experimentation is needed to obtain more solid conclusions.

6. References

- [1] B. A. Davey and H.A. Priestley. *Introduction to Lattices and Order*. Cambridge mathematical text books. Cambridge University Press, 2002.
- [2] B. Roy. The outranking approach and the foundations of electre methods. *Theory and Decision*, 31(1):49-73, 1991.
- [3] E. Covantes. “Robustness analysis of an evolutionary based indirect parameter elicitation method” (in Spanish), Master Thesis in preparation, Autonomous University of Sinaloa, 2013.
- [4] E. Fernandez and J. Navarro. A new approach to multi-criteria sorting based on fuzzy outranking relations: The THESEUS method. *European Journal of Operational Research*, 213(2):405-413, 2011.
- [5] H. A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1):99-118, 1955.
- [6] H. A. Simon. A mechanism for social selection and successful altruism. *Science*, 250(4988):1665-1668, 1990.
- [7] J. García-Duque, M. López-Nores, J. J. Pazos-Arias, A. Fernández-Vilas, R. P. Díaz-Redondo, A. Gil-Solla, Y. Blanco-Fernández, and M. Ramos-Cabrer. A Six-valued Logic to Reason about Uncertainty and Inconsistency in Requirements Specifications. *J. Log. and Comput.*, 6(2):227-255, April 2006.
- [8] K. M. Sim. Bilattices and Reasoning in Artificial Intelligence: Concepts and Foundations. *Artificial Intelligence Review*, 15(3):219-240, 2001.
- [9] M. Doumpos and C. Zopounidis. *Multicriteria Decision Aid Classification Methods*. Applied Optimization. Springer, 2002.
- [10] M. Fitting. Kleene’s Logic, Generalized. *Journal of Logic and Computation*, 1(6):797-810, 1991.
- [11] M. Moussavi, O. Garcia. A six valued logic and its application to artificial intelligence. In *Proceedings of the Fifth Southeastern Logic Symposium*. 1989.
- [12] M. Kryszkiewicz. Rough set approach to incomplete information systems. *Information Sciences*, 112(1-4):39-49, 1998.
- [13] M. R. Chmielewski, J. W. Grzymala-busse, N. W. Peterson, and S. Than. The rule induction system LERS-A version for personal computers, *Foundations of Computing and Decision Sciences* 18, 1993.
- [14] S. C. Kleene. *Introduction to metamathematics*. Bibliotheca mathematica. North-Holland Pub. Co., 1964.
- [15] S. Encheva. Decision making with incomplete information. In *Proceedings of the 10th WSEAS international conference on communications, electrical #38; computer engineering, and 9th WSEAS international conference on Applied electromagnetics, wireless and optical communications*, ACELAE’11, page 109-113, Stevens Point, Wisconsin, USA, 2011. World Scientific and Engineering Academy and Society (WSEAS).
- [16] W. W. Armstrong. Dependency Structures of Data Base Relationships. In *IFIP Congress*, page 580-583, 1974.
- [17] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Norwell, MA, USA, 1992.