Handling missing values in discrete MCDA problems

Bertrand MARESCHAL & Mohamed EL HADOUCHI
05/08/2015

Introduction

This note describes and compares several ways to handle missing values in the evaluation table of discrete multicriteria problems.
We consider multicriteria decision problems including a set of *n* actions:

$$A = \left\{ a_1, a_2, \ldots, a_i, \ldots, a_n \right\} \quad (1)$$

and *k* criteria that have to be maximized:

$$f_1, f_2, \ldots, f_j, \ldots, f_k \quad (2)$$

In such a context it happens regularly that some evaluations $f_j(a_i)$ are not available, for different reasons:

- Data can be unavailable or too difficult to collect for practical reasons (insufficient time or resources, technical difficulties, …).
- A criterion can be irrelevant for some of the actions.

Most current multicriteria decision aid (MCDA) methods assume that the whole multicriteria evaluation table is available for analysis. This is thus a practical problem and many different approaches can be used to solve it.
In a first step we distinguish:

- External approaches which are independent from the MCDA method that is used.
- Internal approaches which are method-dependent and rely on the mathematics of the method used. For this category of approaches we focus on the PROMETHEE outranking methods.

External approaches

These approaches can be applied independently from the MCDA method being used.
Two main categories can be distinguished: elimination and replacement.

Elimination

Elimination is rather straightforward: criteria and/or actions for which evaluations are missing are simply removed from the analysis.
This approach can be meaningful in two cases:

1. If there are many missing evaluations for a single criterion, this criterion, even if it is important for the decision maker, cannot be used to discriminate well the whole set of actions.
2. If there are many missing evaluations for a single action, it is very difficult to compare this action to the other ones.

In these cases it can seem legitimate to remove such criteria and/or actions from the analysis.
Otherwise the elimination approach is rather brutal and can lead to an important loss of information.

Advantages

- Easy to use.
- No additional information required.

Drawbacks

- Modifies the original problem.
- Potential large loss of information.

Substitution

The idea behind substitution is to simply replace missing evaluations by either arbitrary or computed values.

Substitution by an arbitrary value

Many times missing evaluations are replaced by an arbitrary value such as the default values in an MCDA software spreadsheet (usually 0). This is of course not recommended as this arbitrary value is usually a very bad or very good one and is by no way representative of the actual missing value. This approach will thus introduce unwanted biases in the analysis.

Substitution by central value

A better approach is to replace the missing evaluation by a central value such as the arithmetic average of the available evaluations for the criterion. Other central values can be considered as well: median, mode, other types of averages, … In any case an assumption is made that the missing evaluation is a central one rather than an extreme (very good or very bad) one. It thus also introduces an unwanted bias.

Substitution by a computed value

More sophisticated replacement strategies can be considered. For instance the replacement value could be computed as a function of the available evaluations on other criteria. Multiple linear regression can for instance be used to build such a function taking into account the observed dependencies between the criteria.

Advantages

- Easy to use.

Drawbacks

- Arbitrary assumption about the actual value of the missing evaluation.
- Potential large loss of information.

Internal approaches

Using the characteristics of the specific MCDA method that is used it becomes possible to develop better missing values handling strategies by adapting the method rather than the data. This is particularly true for the outranking MCDA methods that rely on the pairwise comparison of the actions. Indeed no assumptions have to be made directly on the performance of the individual actions but rather on how they compare to each other. In this paper we consider the case of the PROMETHEE methods and propose two internal approaches. These could be easily adapted to other outranking methods such as the ELECTRE methods.

Advantages

- Less assumptions are made.
- Less induced biases.

Drawbacks

- Method specific.
- Can reduce the potential of the method.

The (pairwise) valued outranking relation at the heart of the PROMETHEE method is defined by the multicriteria preference index:

$$\pi(a,b)=\sum_{j=1}^{k} w_j P_j(a,b) \qquad (3)$$

that measures with which degree action a is preferred to action b taking into account all the criteria f_j (j=1,…,k), their associated preference functions P_j and their weights w_j.
The multicriteria net flow is then computed in order to obtain the PROMETHEE II complete ranking:

$$\phi(a)=\frac{1}{n-1}\sum_{b\neq a}\left[\pi(a,b)-\pi(b,a)\right] \qquad (4)$$

and can be disaggregated into a weighted sum of unicriterion net flows:

$$\phi(a)=\sum_{j=1}^{k} w_j \phi_j(a) \qquad (5)$$

where:

$$\phi_j(a)=\frac{1}{n-1}\sum_{b\neq a}\left[P_j(a,b)-P_j(b,a)\right] \qquad (6)$$

The relationship (5) is central to the GAIA descriptive method and to weight sensitivity analysis tools such as the Walking Weights and the Weight Stability Intervals.
The preference function also has the following properties:

1. The preference degree is always a number between 0 and 1 :

$$0 \leq P_j(a,b) \leq 1 \qquad (7)$$

2. Both $P_j(a,b)$ and $P_j(b,a)$ cannot be simultaneously positive:

$$P_j(a,b)\times P_j(b,a)=0 \qquad (8)$$

As the preference function $P_j(a,b)$ is a function of the difference between $f_j(a)$ and $f_j(b)$ it can only be computed when both evaluations are available. In case of missing values (either $f_j(a)$ or $f_j(b)$ or both) we consider the two following possibilities.

Pairwise elimination

A first possibility is to cancel criteria for which values are missing in the computation of the multicriteria preference index (3). Formula (3) then becomes:

$$\pi(a,b)=\sum_{j\in F(a,b)} w_j P_j(a,b) \Big/ \sum_{j\in F(a,b)} w_j \qquad (9)$$

where $F(a,b)$ is the subset of criteria for which evaluations are available for both a and b. Note that (9) holds only when $F(a,b)$ is not empty. If $F(a,b)$ is empty the preference degree should be set equal to 0.
A drawback to this approach is that the weighing of the criteria is different from one pairwise comparison to another when missing values are present and that the linear relationship (5) between the unicriterion net flows and the multicriteria net flow doesn't hold anymore. This makes it impossible to perform a GAIA analysis and severely reduces the possibilities for developing advanced weight sensitivity analysis tools such as the Weight Stability Intervals.

Pairwise substitution

A less restrictive approach is to set both preference degrees equal to 0 when missing values make it impossible to compute the actual value:

$$P_j(a,b) = P_j(b,a) = 0 \qquad (10)$$

In this way no assumption is made that either a or b performs better on criterion f_j when either $f_j(a)$ or $f_j(b)$ are missing.
This approach is not equivalent to the external replacement approach as a missing value is considered here as equal to each available value for each pairwise comparison. It is thus also less subjective than the replacement approach.
Another advantage, compared to the pairwise elimination, is that relation (5) still holds and that the GAIA analysis as well as the weight sensitivity analyses can still be performed as usual.
This is this approach that has been first implemented in the Decision Lab 2000 software and that is available today in the Visual PROMETHEE software.

Caveat

As for any missing values treatment it is best suitable when the number of missing values is small with respect to the size of the multicriteria table. Indeed the more missing values the less information available in the table.
Let us take three extreme cases to illustrate this.

All data are missing

If all evaluations are missing it is impossible to discriminate the actions. All the approaches described in this paper lead to the same result: all actions are indifferent.

All data are missing for a single criterion

In that case and for any approach all the actions are considered as indifferent for that particular criterion. The criterion thus becomes irrelevant and the final ranking is based only on the other criteria.

All data are missing for a single action

For substitution approaches the action will have an arbitrary or central place in the ranking. For internal approaches, the action will also be located in the middle of the ranking.