Why pairwise comparison methods may fail in MCDM rankings

PL Kunsch

Abstract This article is based on previous author's articles on a correct way for ranking alternatives in Multi-criteria Decision-Making (MCDM). A straightforward but robust Weighed-Sum Procedure (WSP) is presented. It consists in combining additively ranks or scores evaluated against a dimensionless scale, common to all criteria. MCDM methodologies based on pair-wise comparison methods are on the contrary not providing the correct rankings in many practical situations. The roots of the observed failures are made clearer for the well-known Promethee method by means of a simple example.

1 Introduction

Practical Multi-criteria Decision-Making (MCDM) problems are mostly concerned with ranking n alternatives $(a_i), i = 1, 2, ..., n$. Assume: (1) the (nxK) pay-off table (P_{ik}) containing the dimensioned evaluations of n alternatives for K criteria, $(C_k), k = 1, 2, ..., K$, assumed with no loss of generality not to depend on the decision-maker(s) (DMs) preferences. (2) The preference relationship between the K criteria provided by DMs, for example by ranking the criteria with the possibility of indifference. It has been shown by the author [2] [3] that the following Weighed-Sum Procedure (WSP) provides all possible global rankings, and only them, compatible with these preferences:

- To start with, the pay-off table (P_{ik}) with different scales and units is replaced by a dimensionless (nxK) scoring table (S_{ik}) in which the alternatives are evaluated on a common dimensionless rank, or integer-score scale. Commensurability is gained, as required for solving the MCDM problem: the columns of (S_{ik}) may be combined the adequate way;
- In the second step the global ranking of the alternatives is obtained by combining additively the ranks or scores for all criteria, in accordance with the preference

P.L. Kunsch
Vrije Universiteit Brussel, Pleinlaan 2 BE-1050 Brussels, e-mail: pkunsch@vub.ac.be

relationship. Rounding off rules are used in order to identify indifference ties between two or more alternatives.

These two steps are elaborated in the sections 2 and 3 respectively. In section 4 some failure mechanisms of pair-wise comparison methods are demonstrated by using the popular *Promethee* method [1].

2 The choice of a common scale for the criteria

A straightforward scoring scale is the ordinal scale in which the alternatives are ranked from the best ones, which receive the rank=score 1, to the least good ones which receive ranks=scores $1 \leq s_{ik} = l_{ik} \leq n$, where l_{ik} is the rank position of a_i in C_k. Some technical rounding rules are used for identifying indifference ties from the pay-off table evaluations, e.g., evaluations differing by less than 1 % in absolute are considered as giving indifference between two alternatives. Commensurability is gained, as required for solving the MCDM problem: the rank vectors S_k, i.e., the columns of (S_{ik}) may be combined. Note that the commensurability properties of the ordinal rank scale are three:

1. The scale is dimensionless;
2. The scale is unique and common to all criteria;
3. The values on the scale are enumerable and finite; they are integer numbers; these numbers are $\leq n$ in the ranking case.

The last property means that the raw data from the pay-off table have been put into n classes: the classes of the first ranked, second ranked, etc. Several classes may be empty in case of ties. Any different scale choice for making the criteria evaluations commensurable should at least share these three properties. While the ordinal scale is unique up to the choice of indifference thresholds, the choice of a cardinal scale is no longer unique. A quite sensible choice for fulfilling the above properties is the (inverse) Likert scale [2] [3] with 5 scores, in which 1 is the best score (rather than 5 in the direct Likert scale):

$$L_5 = [1 = Excellent 2 = Good 3 = Satisfactory 4 = Sufficient 5 = JustAcceptable] \quad (1)$$

L_5 indeed provides five distinct classes in which to put the (P_{ik}) evaluations. Some classes may be empty, also the first one (score 1), which would not be possible with ordinal ranking. To achieve the scoring classification DMs have to provide lower and upper limits to each of the five classes. The choice of only five classes is a sensible one because the human mind has difficulties in making judgements on too many objects at the time. Also because evaluations have many times no perfect accuracy, assuming non-integer scores for gaining a thinner granularity is often unnecessary, but may be considered. Because some alternatives might become arbitrarily bad a new class *Unacceptable* is added to L_5. This addition defines the basic 6-point German Grading Scale G_6:

$$G_6 = L_5 \cup [6 = Unacceptable] \tag{2}$$

With this scale a preliminary screening is possible in order to eliminate unacceptable alternatives from the global ranking.

3 The global WSP rankings

The second step of the WSP requires normalised weights within $(Kx1)$ column vectors Ws compatible with the preferences of the DMs. For instance assume that within a group of DMs with similar preferences regarding five criteria C_k , $k = 1, 2, \cdots, 5$ a complete ranking is provided as follows:

$$C_1 \succ C_4 \succ C_2 = C_5 \succ C_3 \tag{3}$$

where $\succ$ indicate strict preference and $=$ indicates indifference. Sets of normalised random numbers ordered like in (3) are generated. The global $(nx1)$ score vector $G(W)$ is calculated for each set of normalised weights contained in the vector W, and the score marix S as follows:

$$G(W) = SW \tag{4}$$

The global ranking with respect to W is herewith provided with the possibility of indifference by using adequate rounding rules. When exact values for the weights are known (a rare case, except for complete indifference between the criteria) a unique computation is sufficient. In the case the weights are random numbers, e.g., corresponding to the ranking (3), statistics on the ranks occupied by the alternatives are obtained by means of Monte-Carlo calculations. Let us consider for illustration the following case with equal weights. In a company six typists $T_i, i = 1, 2, \cdots, 6$ must be ranked in order to determine promotions and salary increases with respect to Typing Speed C_1 and Accuracy C_2 – both of equal importance – as measured by objective indicators. Using the supposedly known limits of the six classes on two $G_6 = L_5$ scales (none of the candidates gets excluded from the ranking) the following $(2x6)$ transposed score table S' is obtained:

$$S' = \begin{vmatrix} 1 & 2 & 2 & 2 & 3 & 3 \\ 4 & 3 & 1 & 3 & 1 & 2 \end{vmatrix} \qquad W = (0.5, 0.5) \tag{5}$$

From (4) the global scores and ranking are obtained:

$$(2.5, 2.5, 1.5, 2.5, 2, 2.5) \rightarrow T_3 \succ T_5 \succ T_1 = T_2 = T_4 = T_6 \tag{6}$$

This ranking is clearly correct according to the very ancient additive formula for evaluating pupils at school. Note that each typist gets a combined score with does not depend on the scores obtained by the other candidates. This thus excludes the possibility of illogical rank reversals (RRs). RRs appear when the ranking of two alternatives are reversed or changed from preference to indifference, when third irrelevant alternatives are introduced, removed, or replaced by less desirable alternatives. RRs affect all pair-wise comparison methods [5]. In the next section the roots of this anomaly are analysed by means of an example, by using the Promethee method [1], which is representative of outranking methods based on pair-wise comparisons.

4 A comparison with Promethee

The Promethee methodology is now briefly introduced [1]. The relative or absolute difference in performance of each couple of alternatives $d_k(a,b) \geq 0$ regarding each criterion C_k is evaluated by means of a preference function $0 \leq P_k(a,b) \leq 1$, always vanishing when $d_k(a,b) < 0$. The most common type of preference function used in Promethee is the linear type represented in Fig. 1.

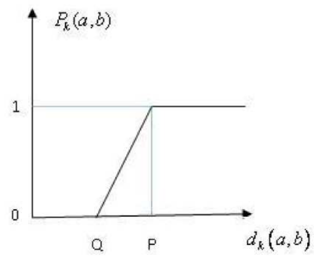

Fig. 1 The linear preference function of Promethee between a pair of alternatives (a,b) in function of $d_k(a,b) \geq 0$. Q is the indifference threshold and P the strict preference threshold.

The global preference index of a over b is given by a weighed sum using the normalised weight vector $(w_k), k = 1,2,\cdots,K$:

$$\pi(a,b) = \sum_{k=1}^{K} w_k P_k(a,b) \leq 1 \tag{7}$$

From (7) the net flow of alternative a is calculated as follows:

$$0 \leq \phi(a) = \frac{1}{(n-1)} \sum_{b \neq a} [\pi(a,b) - \pi(b,a)] \leq 1 \tag{8}$$

The net flow of alternative a represents the degree of preference of this alternative over all other alternatives in the considered set. A global ranking is established by ordering the net flows in decreasing order. The ranking will in general depend on the choice of the technical thresholds Q, P. To illustrate the problems that may arise, let us consider the typist example introduced in section 3. Because the scores are integer numbers we start by considering that $Q = P = 0$, or any value $0 \leq P \leq 1$ for the preference function $P_k(a,b), k = 1,2,...K$ in Fig. 1. This is equivalent to working with ordinary criteria with no thresholds. This results in the following Promethee net flow vector for the six typists, up to a sign inversion from maximum to minimum and an affine transformation not affecting the ranking. The latter is established accordingly:

$$(3.50, 3.75, 2.25, 3.75, 3.50, 4.25) \rightarrow T_3 \succ T_1 = T_5 \succ T_2 = T_4 \succ T_6 \tag{9}$$

Promethee's global ranking is wrong, compare with (6)[1]. Though integer scoring data in the pay-off table have been used right from the beginning, similar conclusions would be obtained with dimensioned absolute data: working with ordinary criteria illicitly changes the pay-off table. In the present case with integer scores and ordinary criteria:

1. The pay-off table is replaced by the table of ranks;
2. In case of a tie of between p alternatives the average rank at position l is used, being equal to:

$$r_l^{avg} = l + \frac{(p-1)}{2} \tag{10}$$

Promethee thus replaces the scoring table (5) by the following one:

$$S_1^{\prime(P)} = \begin{vmatrix} 1 & 3 & 3 & 3 & 5.5 & 5.5 \\ 6 & 4.5 & 1.5 & 4.5 & 1.5 & 3 \end{vmatrix} \tag{11}$$

giving the ranking (9) from (4). This easily induces RRs because the Promethee-adjusted score of any alternative in some criterion depends on the tie multiplicity p. In this example, the global Promethee ranking gives $T_1 = T_5$ in (9). Moving now the third alternative from the first to the second position in criterion C_2 (thus given a worse score to this alternative), the new scoring table becomes:

$$S_2^{\prime(P)} = \begin{vmatrix} 1 & 3 & 3 & 3 & 5.5 & 5.5 \\ 6 & 4.5 & 2 & 4.5 & 1 & 3 \end{vmatrix} \rightarrow T_3 \succ T_5 \succ T_1 \succ T_2 = T_4 \succ T_6 \tag{12}$$

and now $T_5 \succ T_1$ due to this change of a third alternative. Thus the global ranking is not only wrong, but it is also sensitive to changes in the multiplicities of the existing ties, creating an important risk of causing undesirable RRs. This result confirms the conclusion of a recent article [5] exploring RRs in outranking methods, particularly *Electre* [4]. The basic root of this anomaly is recognised there as being the artificial interdependency created between alternatives because of the pair-wise

[1] For example why is $T_2 \succ T_6$ in (9) while the scores in the two equally important criteria are $(2,3)$ and $(3,2)$ respectively, see (5)?

comparisons. Though Promethee is very different from Electre, this general observation is also verified in our example. Alternatives which are very close together for some criterion C_k – or even are building a tie like in the typist example – mutually influence themselves. The final scores change and with them possibly the global ranking, compare (6) with (9) or (12). The only way for eliminating this artificial interference between alternatives is avoiding pair-wise comparisons altogether. Instead direct comparisons of alternatives must be made with independent anchoring values. This is verified for the typist example by using the preference function in Fig. 1 with $Q = 0$, and a linear preference relationship covering the whole range of evaluations; the maximum range of scores$= 3$, so that $P \geq 3$ must be chosen. Doing so, it is observed that the net flows (8) are now equivalent to the global scores in (6): the pair-wise comparisons are made unnecessary, because they reduce to a direct comparison of the alternatives with the best one with score 1, for each criterion.

5 Conclusion

When solving MCDM problems one should be aware of difficulties arising from pair-wise comparisons. The presented typist example, though simple and tested here only against Promethee has a general validity: it is not just pinpointing an isolated atypical anomaly of some most popular outranking methods. A fundamental flaw is inherent to all pair-wise comparison methods, of which Rank Reversals are only one manifestation [5], [2], [3]. The correct way is making direct comparisons of individual alternatives with independent anchoring points representing the limits of performance classes. The German Grading Scale, which has been successfully used for centuries in schools, is very useful for serving this purpose.

References

1. Brans, J.P., Mareschal, B.: PROMETHEE methods. In Figueira J., Greco S., Ehrgott M. (eds.) Multiple Criteria Decision Analysis: state of the art surveys, pp. 163-195. Springer, New York (2005)
2. Kunsch, P.L.: A Statistical Approach to Complex Multi-Criteria Decisions. In: Ruan D. (ed.) Computational Intelligence in Complex Decision Systems, pp. 147-182. Atlantis Press, Paris (2010)
3. Kunsch, P.L.: A Statistical multi-criteria procedure with stochastic preferences. International Journal of Multicriteria Decision Making **1**, 49-73 (2010)
4. Roy B.: The Outranking Approach and the Foundations of the ELECTRE Methods. Theory and Decision **31**, 49-73 (1991)
5. Wang, X., Triantaphyllou, E.: Ranking irregularities when evaluating alternatives by using some ELECTRE methods, Omega **36**, 45-63 (2008)