KKU ENGINEERING JOURNAL 2016;43(S3):376-379 Research Article

KKU Engineering Journal

https://www.tci-thaijo.org/index.php/kkuenj/index

Comparison of rank-based weighting methods for multi-criteria decision making

Panitas Sureeyatanapas*

Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand.

Received April 2016
Accepted June 2016

Abstract

In engineering and management science, decision makings are always influenced by several criteria. In general, degrees of contribution of each criterion to the decision, or the weights of criteria, are not equal. Among various weighting methods suggested in the decision literature, rank-based methods which convert a criteria ranking order into numerical weights have been claimed as a good compromise choice between ease of implementation and quality of the decision result. While previous studies compared quality of several rank-based methods through computer simulations, this study conducts an empirical experiment using actual decision data in order to validate theoretical conclusions. Four weighting methods are compared based upon the hit percentage and the rank-order conformity. The results support previous studies that rank order centroid (ROC) method tends to perform the best due to its steepness and non-linear function of the weights which are likely to be mostly consistent with decision makers' behaviour.

Keywords: Multiple criteria decision making, Criteria weights, Multi-attribute decision, Rank-order weights, Rank-based weighting, Rank order centroid

1. Introduction

In many cases of engineering and management decision problems, the best alternative or option cannot be straightforwardly determined due to conflict among several criteria. Multiple criteria decision analysis (MCDA) methods have been widely employed to facilitate such situations. Most MCDA methods are employed under the condition that each criterion plays a role in determining the decision result according to its weight, which is relative to those of other criteria. Accurate weights should be able to generate the most intuitive and satisfied conclusion. The weights, however, are highly influenced by different elicitation and/or computation methods. As such, a reasonable way to assign the weights is important for satisfied decision making [1].

Various methods to determine the weights have been suggested in the MCDA literature. Among those, rank-based weighting methods which convert a criteria ranking order into numerical weights have been claimed that they outweigh, in many aspects, the methods that require a decision maker (DM) to subjectively assign scores to reflect degrees of criteria importance, such as AHP, SWING, Point allocation, or Direct rating [2-7]. A number of compelling reasons have been claimed. Firstly, DMs normally feel uncomfortable in assigning precise weights, or they lack adequate understanding to do that, such that their judgements tend to be vague or perfunctory. On the other hand, thinking merely about the priority of the criteria is much easier since ranking criteria is an initial step for many weighting methods. When DMs are more confident to only prioritise the criteria, the weights derived from this information should be more reliable. Moreover, in group decision making, consensus among DMs regarding precise weights of various criteria is an unrealistic requirement while reaching agreement on a ranking order is more likely to happen. Last but not least, rank-based weighting is suitable for situations with time limitation as well as when DMs have limited knowledge or information to conduct a complex elicitation method. From the literature, many formulas used to convert a ranking order into weights have been proposed, such as the rank sum (RS), rank exponent (RE), rank reciprocal (RR), and rank order centroid (ROC).

Stillwell et al. [5] examine, based on three decision cases, quality of four rank-based methods (RS, RR, RE, and Decision rule rank) and the equal weighting (EW). The quality is here defined as correlations between decision results produced by each method and the true weights (the ratio weights) under an assumption that DMs can surely assign strengths of their preferences to the criteria. At the end, they conclude that rank-based methods produce higher correlations with the true weights than equal weights do. Barron and Barrett [2], again, compare quality of four weighting methods (ROC, RS, RR, and EW) using computer simulation which generates the data on a random basis. The quality is here reflected by three measures, 'hit rate', 'average value loss', and 'average proportion of maximum value range achieved'. The results show that ROC outperforms the others in most scenarios and in every measure. Jia et al. [7] claim that previous studies usually depend on a strong assumption that the true weights are

*Corresponding author. Tel.: +6692 280 5193
Email address: panisu@kku.ac.th
doi: 10.14456/kkuenj.2016.134

certainly known. They, therefore, compare quality of several weighting methods (EW, RS, and ROC) to the direct rating, which requires DMs to directly quantify their preferences, under a simulation condition that DM's judgements of criteria weights are uncertain or subject to random error. The results show that the direct rating tends to give better quality of decision results when the uncertainty is set as small, whilst ROC provides comparable results to the ratio weights when a large degree of error is placed. The superiority of ROC over other rank-based methods is also subsequently confirmed by Ahn and Park [6] under different simulation conditions.

Roszkowska [4] demonstrates the weight functions of RS, RR, and ROC by varying the number of criteria from two to seven. It shows that ROC gives the largest gap between the weights of the most important criterion and the least. RS provides the flattest weight function in the linear form. For RR, the weight of the most important one descends most aggressively to that of the second, but the function continues to move flatter. Recently, Wang and Zionts [8] propose another rank-based method called generalised sum of ranks (GRS) and compare it to RS, RR, and ROC using simulation. Unfortunately, GRS only provides similar performance to ROC.

From a review of literature, most studies analysed the performance of several weighting methods based on simulation experiments under an unrealistic assumption that the true weight is known. It is considered here that an empirical study which employs real-life cases and actual decision data is required, and the question still remains open as, among several rank-based methods, which one is able to determine the most accurate weights that eventually produce the most satisfactory decision in practices.

2. Methodology

The selection of car was used here to represent a simple decision problem. The samples (DMs) are 16 postgraduate students of the department of Industrial Engineering, Khon Kaen University. The lists of criteria were elicited from the pilot study by asking them to list all criteria that influence the selection. All given criteria were then grouped and renamed in order to eliminate the redundancy of the words. The final list of the criteria includes (i) Aesthetic, (ii) Price, (iii) Engine performance, (iv) Brand image, (v) After-sale service, (vi) Customer excitement features, and (vii) Safety and Security Features.

The data collection started by briefing the DMs about the meaning of each criterion. They were then asked to rank the importance of the criteria, or the degree of contribution of each criterion to their decision. Next, the list of cars (with details on every criterion) was provided to the DMs, and they were asked to choose only four cars which appealed to their interest. They were then asked to assess each car through the concept of utility function. The utility scores were elicited by asking the DMs to quantify their own preference on each alternative, regarding each particular criterion, into the score which is defined as a function between zero to a hundred. Greater degree of the utility score means the greater degree of the DM's preference [9]. Finally they were asked to arbitrarily rank the alternatives based on their own feelings. The given ranking orders were eventually compared to those obtained from mathematical calculation based on the simple additive aggregation method, as shown in Eq. (1).

$$u(a_l) = \sum_{n=1}^{N} W_n \cdot u\left(X_n^{a_l}\right) \quad (1)$$

where $u\left(X_n^{a_l}\right)$ represents the utility of alternative a_l for criterion n. $u\left(X_n^{a_1}\right) > u\left(X_n^{a_2}\right)$ when the performance of alternative a_1 for criterion n is preferred to that of alternative a_2. W_n is the relative weight of criterion n. This algorithm is fundamental to many MCDA aggregation methods.

For this study, the weight, W_n, was determined according to EW and well-known rank-based methods, including RS, RR, and ROC, as shown in Eq. (2) – (5), respectively [4-5].

$$W_n(EW) = \frac{1}{N} \quad (2)$$

$$W_n(RS) = \frac{N - r_n + 1}{\sum_{k=1}^{N} N - r_k + 1} \quad (3)$$

$$W_n(RR) = \frac{1/r_n}{\sum_{k=1}^{N}(1/r_k)} \quad (4)$$

$$W_n(ROC) = \frac{1}{N} \cdot \sum_{k=n}^{N} \frac{1}{r_k} \quad (5)$$

where $n = 1, 2, \ldots, N$, r_n is the rank of n^{th} criterion, and $r_n = n$. Note that the most important criterion is ranked first (n = 1), while the least important one has $n = N$. From Eq. (2) – (5), the weights of the seven criteria (N = 7) are shown in Table 1, and Figure 1 then illustrates their distributions.

Table 1 The weights given by different weighting methods in case of seven criteria

Criteria (n)	EW	RS	RR	ROC
n = 1	0.143	0.250	0.386	0.370
n = 2	0.143	0.214	0.193	0.228
n = 3	0.143	0.179	0.129	0.156
n = 4	0.143	0.143	0.096	0.109
n = 5	0.143	0.107	0.077	0.073
n = 6	0.143	0.071	0.064	0.044
n = 7	0.143	0.036	0.055	0.020

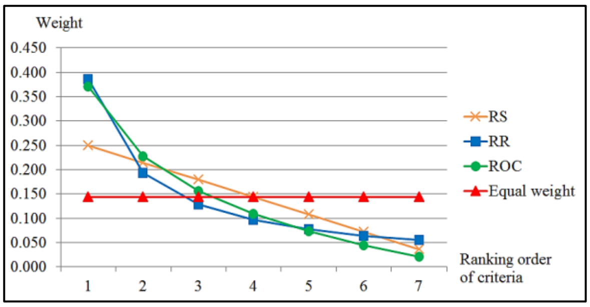

Figure 1 The weight distributions given by different weighting methods in case of seven criteria

Two indicators, (i) Hit percentage and (ii) Rank-order conformity, were employed for comparing the superiority of the weighting methods. The first reflects the ability of each method to give the most satisfactory choice. It is assessed through the percentage of all cases that a weighting method selects the same best alternative as a DM's intuitive feeling does [6, 8]. The conformity then reflects the similarity of the ranking order derived from a particular method and that arbitrarily given by a DM. This is calculated by using a conformity measure (C_m), following Eq. (6).

$$C_m = \sum_{a_l}\left(R_{m,a_l} - R_{i,a_l}\right)^2 \quad (6)$$

Applied from Rezaei [10], C_m denotes the sum of the Euclidean distance, for a particular DM, between the ranks intuitively assigned to alternative a_l (R_{i,a_l}) and its ranks derived from method m (R_{m,a_l}). C_ms from different weighting methods were compared using descriptive statistics. Mood's median test was also used to infer a significant difference among them.

3. Results and discussion

Table 2 shows the hit percentages and the conformity measure of the four weighting methods. It shows that ROC gives the highest percentage, at 81.2%. This indicates that more than four out of five of the DMs chose the same choice, based on their intuitive feeling, to the decision using ROC weights. RR and RS hold the second and the third ranks, while EW shows the poorest performance by giving less than half of the results fitting the DMs' judgements.

Table 2 Performances of the four weighting methods

Indicators	EW	RS	RR	ROC
Hit percentage (%)	43.7	62.5	75.0	81.2
Median C_m	3.75	2.0	2.0	2.0
Mean C_m	4.4	4.0	3.0	2.8
Standard deviation C_m	4.2	4.3	4.4	3.6

In terms of the average conformity measure (C_m), ROC seems to be the best method by giving the least average score, while RR, RS, and EW hold the second-, the third-, and the fourth-ranked scores, respectively. On the other hand, RR, RS, and ROC perform the same in terms of the median values, but EW still shows up as the poorest performer. Nevertheless, when considering the Mood's median test result from Minitab software, the p-value is 0.187 indicating that the conformity performances of the four methods are not significantly different, at 0.05 significant level. The reason of the insignificance might relate to the number of criteria and alternatives being considered. As remarked by Jia et al. [7], when comparing the accuracy from different weighting methods, the difference always increases when the number of criteria and alternatives is larger. Small sample size (only 16 DMs) might be also a possible reason.

Overall, the results appear to be consistent with those of previous simulation studies which conclude that ROC tends to give the most accurate decision and all the rank-based methods are superior to EW [2, 6-7]. The poor performance of EW is also supported by Stillwell et al. [5] who stress that EW gives less information to differentiate alternatives while adequate discrimination power can be captured by the rank orders. An only advantage of EW is that effort to elicit the weights is not required. The superiority of ROC and RR, especially in terms of the hit percentage, might relate to the fact that these methods give a large gap between the weights of the most important criterion and the second, as seen in Figure 1. This characteristic tends to be consistent with when people make an intuitive decision. Studies relating to decision behaviour suggest that functions of criteria weights influencing people's choices are generally steep and non-linear [7, 11-12]. It appears that, among several criteria, DMs tend to intuitively prefer the alternative that is substantially superior in terms of the most important criterion. If that score of an alternative is comparable to others, DMs might shift their attention to the second and the third most important criteria. However, this study does not conclude that ROC weights are always more accurate than RR. The accuracy depends on a DM's belief about the distribution of true weights. Weight functions of both are non-linear but ROC weight function is steeper while the gap between the most and the second most important criteria from RR is larger.

4. Conclusions

A conclusion here is that ROC shows up a theoretical justification that it performs better than other rank-based weighting methods in terms of a choice accuracy which can be reflected by the highest hit percentage. In other words, the result from ROC is considered to be highly accurate since it can satisfy most of the DMs for the selected choices. Clearly, making decision using equal weights is not suggested although it dominates others in terms of time and effort required. Rank-based weighting is considered as a better choice in MCDA particularly when DMs do not feel comfortable to assign precise scores or to estimate the degree of importance of each criterion. Furthermore, cognitive effort in weighting criteria is not intense. From this experiment, delays in making judgements are not obvious, compared to the author's experiences in asking people to weight decision criteria using the direct rating method [1, 13]. The results from this study do not only facilitate decision making for a personal problem, such as buying a car, but are also applicable to other engineering and management decision problems such as choosing a plant location, designing a business strategy, or choosing materials for new product development.

However, there are a few suggestions for future research. Previous studies generally assume that a DM selects the best alternative based upon an additive aggregation [2, 5-8] whilst they ignore the existence of the multiplicative aggregation approach in real-life decision practices. This point should be considered in subsequent studies. Furthermore, as claimed by Wang and Zionts [8], a limitation of rank-based weighting methods is that the strength of DM's preferences are not utilised. Subsequent studies may empirically compare performances of ROC, which has been found as the best rank-based one, to other weighting methods that need DMs to arbitrarily assign scores such as direct rating, trade-off, AHP, SMART, SWING, or to methods determining the weights based on assessment data of alternatives such as Entropy. This would confirm whether the rank-based method is really a good compromise choice between ease of elicitation and quality of the weight obtained [5, 7]. An obvious limitation of this study, last but not least, is from the point that the number of criteria and alternatives are fixed, and this might not guarantee that ROC is always the most accurate method for other cases. For subsequent studies, varying the number of criteria and alternatives is suggested in order to investigate the sensitivity of the ROC performances.

5. References

[1] Sureeyatanapas P, Yang JB, Bamford D. Analysis of criteria weights for the assessment of corporate sustainability: a case study in sugar manufacturing. KKU Engineering Journal 2016;43(3):135-145.

[2] Barron FH, Barrett BE. Decision quality using ranked attribute weights. Management Science 1996;42(11):1515-1523.

[3] Kirkwood CW, Sarin RK. Ranking with partial information: a method and an application. Operations Research 1985;33(1):38-48.

[4] Roszkowska E. Rank ordering criteria weighting methods - a comparative overview. Optimum Studia Ekonomiczne 2013;5(65):14-33.
[5] Stillwell WG, Seaver DA, Edwards W. A comparison of weight approximation techniques in multiattribute utility decision making. Organizational Behavior and Human Performance 1981;28(1):62-77.
[6] Ahn BS, Park KS. Comparing methods for multiattribute decision making with ordinal weights. Computers & Operations Research 2008;35(5):1660-1670.
[7] Jia J, Fischer GW, Dyer JS. Attribute weighting methods and decision quality in the presence of response error: a simulation study. Journal of Behavioral Decision Making 1998;11(2):85-105.
[8] Wang J, Zionts S. Using ordinal data to estimate cardinal values. Journal of Multi-Criteria Decision Analysis 2015;22(3-4):185-196.
[9] Winston WL. Operations research applications and algorithms. 4th ed. Toronto: Thomson Learning; 2004.
[10] Rezaei J. Best-worst multi-criteria decision-making method. Omega 2015;53:49-57.
[11] Fischer GW, Hawkins SA. Strategy compatibility, scale compatibility, and the prominence effect. Journal of Experimental Psychology: Human Perception and Performance 1993;19(3):580-597.
[12] Tversky A, Sattath S, Slovic P. Contingent weighting in judgment and choice. Psychological Review 1988;95(3):371-384.
[13] Sureeyatanapas P, Yang JB, Bamford D. Evaluation of corporate sustainability. Frontiers of Engineering Management 2014;1(2):176-194.