STEVEN

DEC 2008

8

14/11/08

International Journal of Information Technology & Decision Making
Vol. 7, No. 4 (2008) 1–13

World Scientific
www.worldscientific.com

DETERMINING AGGREGATE CRITERIA WEIGHTS FROM CRITERIA RANKINGS BY A GROUP OF DECISION MAKERS

HESHAM K. ALFARES* and SALIH O. DUFFUAA†

Systems Engineering Department, PO Box 5067
King Fahd University of Petroleum and Minerals
Dhahran 31261, Saudi Arabia
*alfares@kfupm.edu.sa
†salihod@kfupm.edu.sa

In this paper, we present an empirical methodology to determine aggregate numerical criteria weights from group ordinal ranks of multiple decision criteria. Assuming that such ordinal ranks are obtained from several decision makers, aggregation procedures are proposed to combine individual rank inputs into group criteria weights. In this process, we use previous empirical results for an individual decision maker, in which a simple function provides the weight for each criterion as a function of its rank and the total number of criteria. Using a set of experiments, weight aggregation procedures are proposed and empirically compared for two cases: (i) when all the decision makers rank the same set of criteria, and (ii) when they rank different subsets of criteria. The proposed methodology can be used to determine relative weights for any set of criteria, given only criteria ranks provided by several decision makers.

Keywords: Multi-criteria; decision making/process; group decisions.

1. Introduction

This work was originally motivated by a real-life situation at the academic department of the authors. A few years ago, several faculty members applied for sabbatical-year leaves during the subsequent academic year. Naturally, it was not possible due to staffing requirements and also strict academic regulations to grant all applicants permission to leave the department at the same time. Therefore, the chairman requested all faculty members to list (in the order of priority) the factors they thought were most relevant for evaluating and comparing sabbatical leave applications. By the time the lists were received from all faculty members, the issue had been already resolved by a friendly gentleman's agreement. However, it became apparent that a methodology was needed in order to assign weights to each factor in the given lists and also to aggregate the weights into an overall department weight for each factor. Developing such a methodology on the basis of empirical data is specifically the purpose of this paper.

The situation mentioned above is not the only case where weight assignment and aggregation is needed. Determining criteria weights is a central problem in multi-criteria decision making (MCDM). Weights are used to express the relative importance of criteria in MCDM. The determination and aggregation of weights are required when applying MCDM methods such as goal programming, the Analytic Hierarchy Process (AHP), and the weighted score method. In practice, it is difficult even for a single decision maker to supply relative numerical weights of different decision criteria. Naturally, obtaining criteria weights from several decision makers is more difficult. Quite often, decision makers are much more comfortable in simply assigning ordinal ranks to the different criteria under consideration. In such cases, relative criteria weights can be derived from criteria ranks supplied by decision makers. The methodology presented in this paper is useful in assisting decision makers to determine criteria weights from criteria ranking, and it is helpful in alternative selection when these weights are used with one of the techniques of MCDM.

The objective of this paper is to combine individual criteria rankings supplied by different decision makers into aggregate group weights for all criteria. In determining criteria weights for any individual, we assume that a universal functional relationship exists between criteria ranks and average weights. In the following section, we present empirical evidence from the literature that supports this assumption. This empirically developed functional relationship was presented in an earlier work by the authors.[1] Moreover, given criteria ranks by several decision makers, we assume that this functional relationship can be used to combine the various rank inputs into a set of aggregate (group) criteria weights.

Subsequent sections of this paper are organized as follows. The relevant literature is reviewed in Sec. 2. Problem definition and experimental design are introduced in Sec. 3. Weight aggregation methodologies are presented in Sec. 4 when all the decision makers rank the same set of criteria, and in Sec. 5 when they rank different subsets of criteria. Finally, results are discussed and conclusions are given in Sec. 6.

2. Literature Review

Bouyssou provides a recent and comprehensive review of MCDM literature.[2] In this paper, we focus on the following MCDM aspects: (a) deriving criteria weights from ordinal ranks, and (b) aggregating individual weight inputs for group decision making.

2.1. *Deriving criteria weights from ranks*

Marichal and Roubens determine criteria weights from partial ranking of the alternatives, individual criteria, or criteria pairs.[3] Hinloopen *et al.* integrate the assessment of the scores (cardinal input) and rankings (ordinal input) of the decision-makers' preference structure.[4] Relative criteria importance is represented by a set of cardinal weights or ranks. Salo and Punkka describe Rank Inclusion in Criteria Hierarchies,[5] a MCDM method in which ranks are given to a set of attributes, and the best alternative is chosen using certain dominance relations and decision rules. Kangas uses simulation to assess the risks of using stochastic multi-criteria acceptability analysis (SMAA) with incomplete criteria weight information.[6] The results indicate the need to have at least complete rank order of criteria in order to minimize the risk of making the wrong decisions.

Doyle *et al.*[7] and Bottomley *et al.*[8] report empirical results that indicate that the rank-weight relationship is basically linear. Doyle *et al.* also use numerical experiments to show the existence of a theoretical straight-line relationship between rank and average weight. In the empirical experiments of Doyle *et al.*,[7] the slope of the linear function depends on the number of criteria being ranked. Bottomley and Doyle find that Max100 weight elicitation procedure,[9] in which the most important criterion is given a weight of 100, has the highest reliability, rank-weight linearity, and subject preference.

Assuming that rank is inversely related to weight (rank 1 means highest weight), the weights must be a nonincreasing function of the rank. Paelinck's theorem describes the set of weights that satisfy a particular criteria ranking.[10] Specific functions for assigning weights have been suggested by several authors. Stillwell *et al.* propose three functions: rank reciprocal (inverse), rank sum (linear), and rank exponent weights.[11] Solymosi and Dompi[12] and Barron[13] propose rank-order centroid weights. Lootsma[14] and Lootsma and Bots[15] suggest two types of geometric weights. Roberts and Goodwin develop rank-order distribution (ROD) weights,[16] which approximate to the rank sum weights as the number of criteria increases. Recently, Alfares and Duffuaa propose an empirically developed linear rank-weight function whose slope depends on the number of criteria.[1]

The empirical model of Alfares and Duffuaa[1] is compatible with the empirical and theoretical results of Doyle et al.[7] and Bottomley *et al.*[8,9] The model in Ref. 1, which is based on the Max100 procedure, is a linear rank-weight function for any number of decision criteria. In this paper, this linear function is used to combine rank inputs from several decision makers into group criteria weights. Ital

2.2. *Aggregating individual weights for group decisions*

In this paper, we use the term "aggregation" to specifically mean combining weights supplied by different individuals into group weights for all of the criteria. Lansdowne compares several well-known vote aggregation methods.[17] Wei *et al.* describe a minimax procedure that employs linear programming,[18] to determine a compromise weight for multi-criteria group decision making that minimizes conflict among the different individual preferences. Barzilai and Lootsma use an aggregation procedure based on geometric means to calculate the global scores for a group of participants.[19] Lahdelma and Salminen develop the Stochastic Multi-criteria Acceptability Analysis II (SMAA-2) to support discrete group decision making.[20] Weight vectors for any rank are analyzed to determine rank acceptabilities, which are in turn combined using meta-weights. Lahdelma *et al.* use SMAA with ordinal criteria, to convert criterion-wise rankings of alternatives into cardinal information to select a waste treatment facility location.[21]

Using simulation to compare methods for aggregating individual rankings of alternatives, Hurley and Lior confirm the superiority of trimmed median rank in the presence of bias.[22] Mateos *et al.* use simulation and the centroid function to aggregate utility functions and attribute weight intervals from several decision makers.[23] Gonzalez-Pachon and Romero consider aggregating partial rankings of the alternatives,[24] where each individual rank is not a fixed value but a specific range. Interval goal programming is used to minimize the social choice function, i.e. the total aggregated disagreement. Lootsma defines the relative importance of any pair of criteria under two widely used MCDM methods.[25] The first is the geometric-mean aggregation rule in the multiplicative AHP technique, and the second is the arithmetic-mean aggregation rule in the Simple Multi-Attribute Rating Technique (SMART) of Von Winterfeldt and Edwards.[26]

Xu and La analyze MCDM problems where only value ranges of criteria weights are given,[27] but not their exact individual weights. A projection method is proposed to determine criteria weights and to select the most appropriate alternative(s). Given incomplete linguistic preference relations, Xu uses the extended arithmetic averaging operator for group decision making.[28] Ahmad *et al.* combine AHP and Data Envelopment Analysis (DEA) to assess the performance of Small-to-Medium-Sized Manufacturing Enterprises (SMEs).[29] By eliminating the weaknesses and emphasizing the strengths of each of these two methods, the integrated AHP/DEA model provides superior MCDM solutions.

From this literature review, it is evident that this is the first paper to empirically aggregate criteria rank inputs from several individual decision makers, in order to develop numerical criteria weights representing the preferences of the whole group.

3. Problem Definition and Experimental Design

In this paper, we consider a group MCDM problem with l decision alternatives, m decision makers, and n decision criteria. Given the performance score $a_{j,k}$ of alternative k $(k = 1, 2, \ldots, l)$ in terms of criterion j $(j = 1, 2, \ldots, n)$, the overall score of alternative k is given by:

$$P_k = \sum_{j=1}^{n} W_j a_{j,k}, \quad k = 1, 2, \ldots, l. \tag{1}$$

Our objective is to determine criteria weights $(W_1, \ldots, W_n)$ for all MCDA contexts in which Eq. (1) is applicable. Each decision maker (DM) i $(i = 1, 2, \ldots, m)$ may select and rank a subset of n_i criteria $(n_i \leq n)$ that he or she deems to be relevant, giving each criterion j a rank $r_{i,j}$, $(r_{i,j} = 1, \ldots, n_i)$. Given the ranks of criteria (subsets) provided by all DMs, we aim to develop aggregate (group) weights for all n criteria.

A set of experiments were performed to develop and evaluate an empirical methodology to convert ordinal criteria rankings from several DMs into aggregate criteria weights. The experimental design aims to test whether relationships between ranks and aggregate weights change according to the given criteria or the specific group of DMs. Therefore, the experiments involved two groups of DMs (students and faculty) and two sets of criteria applicable in two contexts (student learning and instructor evaluation). The student sample was composed of 111 college students from different years and in different fields of study. Naturally, the faculty sample was much smaller, containing only 23 faculty members. The survey given to this sample of students and faculty was administered in two consecutive parts as follows:

Part I: The participants were asked the two following questions:

Question 1. List in the order of priority (most important to least important) factors that hinder students' learning and retaining course materials.

Question 2. List in the order of priority (most important to least important) factors that affect the evaluation of course instructors.

After listing these factors, the participants were requested to give weights to all factors in each list. Following the Max100 method suggested by Bottomley and Doyle,[9] a weight of 100% must be given to the most important (first) factor.

Part II In the second part of the survey, the participants were provided with two prepared lists of standard criteria shown in Table 1: 12 factors hindering student learning (Question 1), and 16 factors affecting instructor evaluation (Question 2). The participants were asked to rank each set of factors based on their importance.

Part II of the survey was administered only after finishing Part I, in order to avoid suggesting any factors for Part I. After ranking the factors in each list, the participants were required to assign weights to each factor, starting with a weight of 100% for the most important (first ranked) factor.

Aggregation involves combining ordinal rankings of criteria given by several individuals in order to determine the overall group weight for each criterion. Before starting to develop aggregation methodologies, we tested whether the survey responses differed according to the type of decision makers. Therefore, the Wilcoxon signed-rank test for paired observations (applied to each group's mean rankings on each criterion) was used to test whether the differences between the two sets of decision makers (students and faculty) are significant. The effects of the different decision makers on the aggregate weights were found to be insignificant at significance level $\alpha = 0.05$. Therefore, all inputs from the two categories of participants (students and faculty) were combined.

Table 1. Ready made lists of criteria given to participants in Part II of the survey.

No.	Question 1. Learning hindrances	Question 2. Instructor evaluation
1	Large class size	Encourages student participation and questions
2	Lack of student motivation	Available and helpful in office hours
3	Grading system	Prepared for class
4	Current teaching methods	Speaks clearly
5	Faculty unavailability outside class time	Has clear presentation
6	High study and homework demands	Motivates students
7	Course load (many courses per term)	Seems knowledgeable in course subject
8	Emphasis on theory in class	Uses educational aids and presentations
9	Students' poor English proficiency	Fair in grading
10	Lack of practical cases	Concerned about student's understanding
11	Fast pace of material coverage	Explains concepts clearly using examples
12	Students' poor study habits	Prompt in attending and leaving class
13		Gives tests to measure students understanding
14		Assigns homework and gives quizzes regularly
15		States objectives of each class
16		Grades tests and assignments promptly

thick rule

In order to develop aggregate criteria weights, we utilized the empirical rank-weight relationship of Alfares and Duffuaa.[1] This linear relationship specifies the average weight for each rank for an individual DM, assuming a weight of 100% for the first-ranked (most important) factor. For any set of n ranked factors, the percentage weight of a factor ranked as r is given by

$$w_{r,n} = 100 - s_n(r-1), \tag{2}$$

where

$$s_n = 3.195 + \frac{37.758}{n}, \quad 1 \le n \le 21, \quad 1 \le r \le n, \quad r \text{ and } n \text{ are integer.} \tag{3}$$

The upper limit ($n \le 21$) is meant to prevent Eq. (2) from assigning negative weights to criteria ranked greater than 21. Obviously, this range of up to 21 criteria is sufficient for all practical MCDM purposes. Although we may combine the two groups of participants (students and faculty), we obviously cannot combine the criteria of the two questions (students learning and instructor evaluation) because the aggregation data is criterion-specific. Moreover, for each question, we must separately analyze the data obtained from the two parts of the survey. Therefore, we evaluated two different sets of aggregation methods using two different types of data:

(1) Data with the same criteria for all decision makers (Part II of the survey).
(2) Data with different criteria for each decision maker (Part I of the survey).

4. Aggregate Weights for the Same Ranked Criteria from all DMs

It must be noted that in all the aggregation methods presented in Secs. 4 and 5, there is in an implicit last step in which weights are normalized to make their sum equal to 100%. In this section, we consider methods to determine aggregate (group) weights if all the decision makers rank the same set of criteria. Let us assume we have m individuals and n criteria that are common to all individuals ($n_1 = n_2 = \cdots = n_m = n$). We also assume that each individual i assigns a rank of $r_{i,j}$ to criterion j. This kind of data is provided by the two ready-made lists of 12 or 16 criteria provided in Part II of the survey ($n = 12$ for Question 1, $n = 16$ for Question 2). For the data of each question, we compared the three following aggregation methods.

4.1. *Method S1*

In this method, we first convert individual ranks into individual weights for each factor, and then calculate the average weight for each factor among all individuals. The two steps are given as follows:

(1) For each individual i, use Eq. (2) to convert ranks $r_{i,j}$ into individual weights $w_{i,j}$ for all n criteria:

$$w_{i,j} = 100 - s_n(r_{i,j} - 1), \quad i = 1, \ldots, \ m, \ j = 1, \ldots, n. \tag{4}$$

(2) Calculate the aggregate weight of each criterion by averaging its weights obtained from all m individuals:

$$W_j = \frac{1}{m} \sum_{i=1}^{m} w_{i,j}, \quad j = 1, \ldots, n. \tag{5}$$

The two steps of this method can be reversed; we may average the ranks first, and then convert average ranks into average (aggregate) weights. The same values of relative aggregate weights will be obtained.

4.2. *Method S2*

This method is similar to Method $S1$; thus, Eq. (4) is used in the first step. However, in the second step, the geometric mean of individual weights (the mth root of the product of the m individual weights) is used to determine aggregate weights as proposed by Barzilai and Lootsma[19]:

$$W_j = \sqrt[m]{w_{1,j} \times w_{2,j} \times \cdots \times w_{m,j}}, \quad j = 1, \ldots, n. \tag{6}$$

4.3. *Method S3*

This method is similar to method $S1$ performed in reverse order. However, in the first step, the geometric mean of individual ranks is used instead of the simple arithmetic mean to determine the average rank of each criterion

$$\bar{r}_j = \sqrt[m]{r_{1,j} \times r_{2,j} \times \cdots \times r_{m,j}}, \quad j = 1, \ldots, n. \tag{7}$$

In the second step, Eq. (4) is used to convert average rank $\bar{r}_j$ into average (aggregate) weight W_j for each of the n criteria:

$$W_j = 100 - s_n(\bar{r}_j - 1), \quad j = 1, \ldots, n. \tag{8}$$

4.4. *Illustration and comparison of methods for the same criteria*

A small numerical example is used to illustrate the steps of the three methods described above. Table 2 shows a solved example for aggregation of weights when the same set (number) of criteria is ranked by all decision makers. The example involves three decision makers (DM1, DM2, and DM3) and four decision criteria (A, B, C, and D). As explained above, two alternative calculation sequences are possible for applying method $S1$.

Comparison of the three above aggregation methods is based on how close they estimate the relative actual sum of weights given by all the participants. Taking Question 1 of Part II of the survey as an example, $n = 12$. Therefore, we used Eq. (3) to find the slope value $s_{12} = 6.3416$ and normalized weights to make $\sum_{j=1}^{n} w_j = 100\%$. The results of applying the three methods to Question 1 of Part II of the survey, including the mean absolute percentage error (MAPE) values, are shown in Table 3.

Similar results are obtained for Question 2, showing that method $S1$ consistently has the minimum absolute percentage errors. On the basis of these results, we can conclude that method $S1$ is the best for aggregation when all individuals rank the same set of criteria.

Table 2. Example of applying three aggregation methods for the same set of ranked criteria.

	Criterion	A	B	C	D
Given	DM1 rank	1	2	3	4
	DM1 rank	2	1	3	4
	DM3 rank	1	2	4	3
Method $S1$	DM1 weight	100	87.37	74.73	62.10
	DM1 weight	87.37	100	74.73	62.10
	DM3 weight	100	87.37	62.10	74.73
	Arithmetic average weight	95.79	91.58	70.52	66.31
	Percent weight	29.55	28.25	21.75	20.45
Method $S1$ (alternative)	Arithmetic average rank	1.33	1.67	3.33	3.67
	Average weight	95.79	91.58	70.52	66.31
	Percent weight	29.55	28.25	21.75	20.45
Method $S2$	Geometric average weight	95.60	91.39	70.26	66.05
	Percent weight	29.57	28.27	21.73	20.43
Method $S3$	Geometric average rank	1.26	1.59	3.30	3.63
	Average weight	96.72	92.58	70.92	66.72
	Percent weight	29.58	28.32	21.69	20.41

← thick rule

Table 3. Actual and calculated aggregate percent weights of 12 criteria of Question 2 in Part II of the survey (same set of criteria).

Factor No.	Actual	Method $S1$	Method $S2$	Method $S3$
1	7.41	7.76	7.55	7.93
2	8.80	9.02	9.20	8.89
3	10.08	9.96	10.10	9.89
4	9.19	9.40	9.38	9.43
5	7.77	7.64	7.63	7.58
6	8.55	8.59	8.66	8.49
7	8.90	8.99	9.12	8.91
8	7.87	7.79	7.75	7.82
9	8.52	8.66	8.63	8.75
10	7.63	7.34	7.32	7.29
11	7.06	6.88	6.87	6.74
12	8.21	7.99	7.78	8.28
MAPE		2.13	2.43	2.43

5. Aggregate Weights for Different Ranked Criteria from each DM

The data in this section are collected in Part I of the survey. At the beginning, the participants listed a total of 53 factors (criteria) that they considered as significant for both questions. Since this number is too large, we decided to concentrate only on the criteria that we judged to be most important on the basis of their frequency. As a result, we ended up with only 16 criteria for Question 1 and 10 criteria for Question 2.

In order to determine aggregate criteria weights for each question in Part I, we applied four different aggregation methods. These methods are similar to those used for aggregation when the same set of criteria is ranked by all individuals. However, adjustments are made to accommodate two new facts. First, the number of criteria changes from a constant n for all DMs to a variable n_i that depends on the individual DM i. Second, since some criteria are listed by more DMs, criteria frequency must be taken into consideration. Therefore, the aggregate weight of each criterion is determined by both its rank(s) and its frequency.

5.1. *Method D1*

This method involves the two following steps:

(1) For each individual i, convert ranks $r_{i,j}$ into individual weights $w_{i,j}$ for all n_i criteria. The slope of the linear conversion function, $-s_{n_i}$, is determined by the number of criteria listed by the individual, i.e. n_i. Therefore, $w_{i,j} = 100 - s_{n_i}(r_{i,j} - 1)$ if criterion j is listed by individual i, otherwise $w_{i,j} = 0$.
(2) Calculate the aggregate weight of each criterion as the arithmetic mean of weights obtained from all individuals.

5.2. *Method D2*

This method involves three steps:

(1) Calculate the arithmetic mean of ranks for each criterion (only among individuals listing the given criterion).
(2) Convert the average rank into an average weight for each criterion based on estimated slope $-s_n$ (n is the total number of criteria listed by all individuals).
(3) Multiply the average criterion weight by the corresponding frequency (number of individuals listing the given criterion).

5.3. *Method D3*

This method involves three steps:

(1) Convert individual ranks $r_{i,j}$ into individual weights $w_{i,j}$ for all n_i criteria.
(2) Compute the geometric mean of weights for each criterion (only among individuals listing the given criterion).
(3) Multiply the geometric mean of criterion weight by the corresponding frequency.

5.4. *Method D4*

This method involves three steps:

(1) Calculate the geometric mean of ranks for each criterion (only among individuals listing the given criterion).
(2) Convert the geometric mean of ranks into an average weight for each criterion based on estimated slope $-s_n$.
(3) Multiply the average criterion weight by the corresponding frequency.

5.5. *Illustration and comparison of methods for different criteria*

A small numerical example is shown in Table 4 to illustrate the steps of the four aggregation methods described above. Similar to the example of Sec. 4, this example involves three decision makers and four decision criteria.

Taking Question 2 of Part I of the survey as an example, $n = 10$. Thus, we used Eq. (3) to find the absolute fitted slope value $s_{10} = 6.9709$. Applying the four methods to the 10 criteria of Question 2 of Part I of the survey, we obtain the normalized weights shown in Table 5. Based on the MAPE values shown in Table 5, Method $D2$ seems to be the best for determining aggregate weights when the number of ranked criteria varies among different individuals. Similar results are obtained from the data of Question 1 of Part I of the survey.

6. Discussion and Conclusions

Although the proposed methodology is based on strong empirical evidence, it still has inherent limitations. First, by giving an equal weight to each individual's rank,

Table 4. Example of applying four aggregation methods for different sets of ranked criteria.

	Criterion	A	B	C	D
Given	DM1 rank	1	2	3	4
	DM1 rank	2	1		
	DM3 rank	1	2		3
Method $D1$	DM1 weight	100	87.37	74.73	62.10
	DM1 weight	77.93	100		
	DM3 weight	100	84.22		68.44
	Total weight	277.93	271.58	74.73	130.53
	Percent weight	36.82	35.98	9.90	17.29
Method $D2$	Arithmetic average rank	1.33	1.67	3.00	3.50
	Average weight	95.79	91.58	74.73	68.41
	Total weight	287.365	274.731	74.73	136.827
	Percent weight	37.14	35.51	9.66	17.69
Method $D3$	Geometric average weight	92.02	90.28	74.73	65.19
	Total weight	276.068	270.835	74.73	130.38
	Percent weight	36.71	36.01	9.94	17.34
Method $D4$	Geometric average rank	1.26	1.59	3.00	3.46
	Average weight	96.72	92.58	74.73	68.87
	Total weight	290.148	277.735	74.73	137.734
	Percent weight	37.18	35.59	9.58	17.65

Table 5. Actual and calculated aggregate percent weights for 10 criteria of Question 2 in Part I of the survey (different sets of criteria).

Criterion no.	Actual	Method $D1$	Method $D2$	Method $D3$	Method $D4$
1	18.24	18.99	18.32	19.48	21.91
2	20.94	20.39	20.63	19.23	24.45
3	21.47	22.98	22.01	23.97	25.83
4	8.73	8.57	8.75	8.87	10.38
5	4.29	3.89	4.06	4.06	4.77
6	9.15	9.47	9.73	9.80	11.55
7	2.39	2.11	2.13	2.16	2.62
8	3.73	3.33	3.37	3.44	4.03
9	0.63	0.71	0.64	0.74	0.74
10	10.42	9.57	10.35	8.24	12.36
MAPE		7.07	3.91	9.65	16.67

ctr

thick rule

the methodology cannot recognize the different intensities of preferences among individual decision makers. Second, the tasks used to collect the empirical data were not decision-making tasks. The experiments involved specifying criteria preferences (ranks), but no selection of an alternative decision on the basis of these ranks. The presence of concrete decision alternatives might influence individual criteria ranks. For example, a student might consider speaking clearly as the most important criteria for instructor evaluation. However, in choosing between two instructors who are both clear speakers, "speaking clearly" may become less important because it does not influence the current instructor's selection decision.

To summarize, an empirical methodology has been presented for calculating aggregate (group) criteria weights on the basis of ordinal ranking of these criteria by several decision makers. Experiments involving university students and faculty were conducted to collect necessary data for developing this methodology. Several aggregation methods have been investigated for two possible cases, depending on whether or not the ranks provided by different individuals correspond to the same set of criteria. For both cases, the best aggregation method is determined on the basis of comparison with the actual aggregate weights.

If all the decision makers rank the same set of criteria, we recommend aggregation method $S1$, which converts individual ranks into individual weights and then calculates the average weight for each criterion. If different individuals rank different subsets of the criteria, the recommended method is $D2$, which converts individual ranks into individual weights and then calculates aggregate weights as averages of individual weights. Potential future extensions include partial or fuzzy rankings, group decision making with weighted voting to reflect different intensities of preference, and aggregation for other rank-weight functions, such as the centroid and the inverse weight models. Another extension is to collect empirical data in a decision-making task setting, aiming to find out whether the same empirical results would be obtained. Finally, theoretical analysis is needed to confirm, explain, and generalize the empirical results.

Acknowledgments

The authors are grateful to King Fahd University of Petroleum and Minerals for supporting this research and also to Mr. Saleh Al-Duwais for help in data collection.

References

1. H. K. Alfares and S. O. Duffuaa, Assigning cardinal weights in multi-criteria decision making based on ordinal ranking, *J. Multi-Crit. Decis. Anal.* (2008), in press.
2. D. Bouyssou, *Aiding Decisions with Multiple Criteria* (Springer, 2002).
3. J.-L. Marichal and M. Roubens, Determination of weights of interacting criteria from a reference set, *Eur. J. Oper. Res.* **124** (2000) 641–650.
4. E. Hinloopen, P. Nijkamp and P. Rietveld, Integration of ordinal and cardinal information in multi-criteria ranking with imperfect compensation, *Eur. J. Oper. Res.* **158** (2004) 317–338.
5. A. Salo and A. Punkka, Rank inclusion in criteria hierarchies, *Eur. J. Oper. Res.* **163** (2005) 338–356.
6. A. Kangas, The risk of decision making with incomplete criteria weight information, *Can. J. Forest Res.* **36** (2006) 195–205.
7. J. R. Doyle, R. H. Green and P. A. Bottomley, Judging relative importance: Direct rating and point allocation are not equivalent, *Organ. Behav. Hum. Decis. Process.* **70** (1997) 65–72.
8. P. A. Bottomley, J. R. Doyle and R. H. Green, Testing the reliability of weight elicitation methods: Direct rating versus point allocation, *J. Market. Res.* **37** (2000) 508–513.

1st Reading

November 6, 2008 16:14 WSPC/173-IJITDM 00317

9. P. A. Bottomley and J. R. Doyle, A comparison of three weight elicitation methods: Good better, and best, *Omega* **29** (2001) 553–560.
10. J. H. P. Paelinck, Qualitative multiple-criteria analysis, environmental protection and multiregional development, *Pap. Reg. Sci.* **36** (1976) 59–74.
11. W. G. Stillwell, D. A. Seaver and W. Edwards, A comparison of weight approximation techniques in multiattribute utility decision making, *Organ. Behav. Hum. Perform.* **28** (1981) 62–77.
12. T. Solymosi and J. Dompi, Method for determining the weights of criteria: The centralized weights, *Eur. J. Oper. Res.* **26** (1985) 35–41.
13. F. H. Barron, Selecting a best multiattribute alternative with partial information about attribute weights, *Acta Psychol.* **80** (1992) 91–103.
14. F. A. Lootsma, *Multi-Criteria Decision Analysis via Ratio and Difference Judgment* (Kluwer Academic Publishers, Dordecht, 1999).
15. F. A. Lootsma and P. W. G. Bots, The assignment of scores for output-based research funding, *J. Multi-Crit. Decis. Anal.* **8** (1999) 44–50.
16. R. Roberts and P. Goodwin, Weight approximations in multi-attribute decision models, *J. Multi-Crit. Decis. Anal.* **11** (2002) 291–303.
17. Z. F. Lansdowne, Ordinal ranking methods for multicriteria decision ranking, *Naval Res. Logist.* **43** (1996) 613–627.
18. Q. Wei, H. Yan, J. Ma and Z. Fan, A compromise weight for multi-criteria group decision making with individual preference, *J. Oper. Res. Soc.* **51** (2000) 625–634.
19. J. Barzilai and F. A. Lootsma, Power relations and group aggregation in the multiplicative AHP and SMART, *J. Multi-Crit. Decis. Anal.* **6** (1997) 155–165.
20. R. Lahdelma and P. Salminen, SMAA-2: Stochastic multicriteria acceptability analysis for group decision making, *Oper. Res.* **49** (2001) 444–454.
21. R. Lahdelma, P. Salminen and J. Hokkanen, Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria, *Eur~~p~~. J. Oper. Res.* **142** (2002) 345–356. del
22. W. J. Hurley and D. U. Lior, Combining expert judgment: On the performance of trimmed mean vote aggregation procedures in the presence of strategic voting, *Eur. J. Oper. Res.* **140** (2002) 142–147.
23. A. Mateos, A. Jiménez and S. Ríos-Insuaa, Monte Carlo simulation techniques for group decision making with incomplete information, *Eur. J. Oper. Res.* **174** (2006) 1842–1864.
24. J. Gonzalez-Pachon and C. Romero, Aggregation of partial ordinal rankings: an interval goal programming approach, *Comput. Oper. Res.* **28** (2001) 827–834. Cap
25. F. A. Lootsma, A model for the relative importance of the criteria in the multiplicative AHP and SMART, *Eur. J. Oper. Res.* **94** (1996) 467–476. space
26. D. Von Winterfeldt and W. Edwards, *Decision Analysis and Behavioral Research* (Cambridge University Press, Cambridge, 1986).
27. Z. Xu and Q. La, Projection method for uncertain multi-attribute decision-making with preference information on alternatives, *Int. J. Inform. Tech. Decis. Making* **3** (2004) 429–434.
28. Z. Xu, An approach to group decision making based on incomplete linguistic preference relations, *Int. J. Inform. Tech. Decis. Making* **4** (2005) 153–160.
29. N. Ahmad, D. Berg and G. R. Simons, The integration of analytical hierarchy process and data envelopment analysis in a multi-criteria decision-making problem, *Int. J. Inform. Tech. Decis. Making* **5** (2006) 263–276.