New Metrics to Measure the Quality of the Ranking Results Obtained by the Multi-criteria Decision-Making Methods

Mohammed Chaouki Abounaima, Loubna Lamrini,
Fatima Zahra EL Mazouri, Noureddine EL Makhfi,
Mohammed Talibi Alaoui, and Mohamed Ouzarf

Abstract Nowadays, there is a panoply of multi-criteria decision-making methods which are proposed in the literature to solve the ranking problematic, where each method has its resolution process and has its drawbacks and advantages. These methods aim to rank from best to worst a finite set of alternatives while taking into account a set of conflictual criteria. The purpose of this article is to propose specific metrics that will be useful to measure the quality of the rankings obtained by different methods. Thus, these quality measures can help the decision-maker to choose the best ranking objectively when adopting several methods. To show and prove the importance and relevance of the metrics proposed, a set of twenty-five examples of rankings will be examined. The results of the experiment conclusively show that all the proposed metrics lead to significant and equivalent quality measures.

M. C. Abounaima (✉) · L. Lamrini · F. Z. EL Mazouri · M. Talibi Alaoui · M. Ouzarf
Laboratory of Intelligent Systems and Applications, Faculty of Sciences and Technologies, Sidi Mohammed Ben Abdellah University, Fez, Morocco
e-mail: medchaouki.abounaima@usmba.ac.ma

L. Lamrini
e-mail: loubna.lamrini@usmba.ac.ma

F. Z. EL Mazouri
e-mail: Fatimazahra.elmazouri@usmba.ac.ma

M. Talibi Alaoui
e-mail: mohammed.talibialaoui@usmba.ac.ma

M. Ouzarf
e-mail: mohamed.ouzarf@usmba.ac.ma

N. EL Makhfi
Faculty of Science and Technology Abdelmalek, Essaadi University, Al Hoceima, Morocco
e-mail: n.elmakhfi@gmail.com

1 Introduction

In our recent work [1], we have demonstrated the usefulness of the Pearson Correlation metric to evaluate the quality of the ranking results obtained by different multi-criteria aggregation procedures (MCAP). We intend in this article to propose other possible metrics which take into account the relative importance of the criteria, which generalizes the case of an equal-weighting where all the criteria have the same importance.

The field of multi-criteria decision-making (MCDM) has reached a well advanced and remarkable maturity, which is justified by the considerable number and the abundance of the methods proposed in the literature and the great variety of real applications who have used MCDM methods [2], like industry, economy, energy, social, environmental, even military, etc. An MCDM problem can be summarized by considering, in first, a finite set of alternatives A and set of conflictual criteria F. Then, each alternative is evaluated on all the retained criteria. In the last, the decision-maker (DM) should choose the problematic to resolve and the appropriate resolution method(s). MCDM methods can mainly resolve three issues. The first allows us to rank the alternatives of set A from the best to the worst choice, known as the Ranking Problematic. The second is to sort the set A into established categories, called the Sorting Problematic. Furthermore, the third problem consists in selecting the best alternative, known as the Choice Problematic.

For the same problem, there are several resolution MCAPs, each of which gives a solution. The DM obtains such different results, which all aggregate the different criteria. Given the embarrassment of the choice of solutions obtained by the different MCPAs adopted, the DM surely will have difficulties in selecting the final solution. This work proposes many metrics to measure the quality of each solution. This quality expresses the dependence degree between the MCPA solution and the performance of the alternatives. The final solution to be retained will be that which leads to the highest dependence degree.

The rest of the article is organized as follows: the first section "Overview of multi-criteria analysis methods and comparison between MCDM methods" gives an overview of MCDM approaches and methods. Some work on the comparison of MCDM methods will also be cited in this section. The following section "The proposed approach to measure the quality of rankings" details the proposed approach to assess, on the basis of several parameters, the quality of a ranking. The different tests of the proposed approach and discussion will be presented in the last section "Numerical experimentation and discussion". In the end, a conclusion will be given with the suggestion of some lines of research.

2 Analysis Methods and Comparison Between MCDM Methods

Undoubtedly the reality of decision has always been naturally multi-criteria, where several criteria should be taken into account to find a solution. Unless if by some artisanal transformations, the decision problem becomes mono-criterion, where a single function, called by economists, the objective function to optimize to find the best solution, called the optimal solution.

Unfortunately, it is not always possible to reduce all the functions expressing the criteria into a single function, because of the diversification of the points of view and the consequences of the problem and which can concern all planes of human life: political, military, economic, urban and interurban infrastructure, social, environment, ecology [3]. All these consequences are not expressed and measured directly by the same measurement scale, so they cannot be reduced by a single function. Furthermore, any reduction of the multi-criteria into single criteria, it will be a simplification of the problem, but surely, it will have influences and impacts on the quality and rationality of the final solution.

It is so easy to optimize a problem based on a single criterion. However, for several criteria, each criterion gives its optimal solution. The problem is to find one solution that represents all the solutions obtained by the different criteria. The main objective of MCDM methods is, therefore, to find the solution aggregating all the solutions arising from the different criteria [4].

For more than forty years, the MCDM field has known significant progress as well on the theoretical level as on the application level [5]. Several approaches have thus emerged, each with its advantages and disadvantages. There are currently two major resolution approaches [6]. The first is called the synthesis criterion approach; it consists in transforming the multi-criteria problem into a simple mono-criterion problem. As an example of methods coming under this approach: we cite the Weight Sum Method (WSM) [7], the Goal Programming method [8], the Technique for Order of Preference by Similarity to the ideal solution (TOPSIS) [9], and many other methods. The second approach is known as the outranking approach, where we build a comparison relationship between actions, called the outranking relationship. This last will be used to find the compromise solution depending on the type of problem to be solved: choice, ranking or sorting problematic. There is a panoply of methods that are based on the principle of this approach. We cite the two prevalent methods ELECTRE (Elimination and Choice Translating the REALITY) method [10, 11] and PROMETHEE (Preference Ranking Organization METHod for Enrichment of Evaluations) method [12].

The problem posed in this work is to compare MCDA methods under the same approach. Several authors have approached this question, but for the majority of them, they have tried to compare methods based on their approaches and procedures. For example, we cite the works [13, 14]. Any direct comparison between methods will be subjective and meaningless, as each method has its limitations and advantages. We propose to use metrics to measure the quality of the compromise solutions in

order to help the DM to choose the best result and not the best method because in the MCDM field, there is no best method and bad method.

3   The Proposed Approach and Metrics to Measure the Ranking Quality

For the rest of this section, we need, as the case of any MCDM method, the flowing data:

- $A = \{a_1, ...,a_i, ...,a_n\}$ is the set of n alternatives.
- $F = \{g_1, ..., g_j, ..., g_m\}$ is the family of m criteria ($m \geq 2$) to be maximized.
- $gj(ai)$ is the performance of the alternative ai on the criterion gj evaluated by de DM. The performance $gj(ai)$, also called by judgment, evaluation and preference.
- $W = \{w_1, ..., w_j, ..., w_m\}$ is the weight vector reflecting the relative importance of each criterion.

3.1   The Proposed Approach to Measure the Ranking Quality

To measure the quality of a given ranking solution, which ranks the alternatives from best to worst, we suggest comparing this ranking to all of the rankings induced by each criterion. More precisely, each rank can be associated with a comparison matrix R of the form: if an alternative a is better classified than another alternative b, then this comparison gives $R(a,b) = 1$ and $R(b, a) = 0$. The proposed metrics are used to measure the dependencies that exist between the comparison matrix associated with the ranking solution obtained and the comparison matrices induced by the criteria. For the correlation measure, this dependence must be maximum, and for the other distance metrics, it must be minimum. The advantage of this proposal is that it does not imply any condition on the scales for measuring the performance of alternatives.

The approach proceeds in three steps: In the first step, the comparison matrices R^k induced by the different criteria gk are calculated for all k in $\{1,...,m\}$, with m is the number of criteria considered. In the second step, we calculate the matrix of comparisons R^k associated with the ranking result of the MCDM method or a scenario of the robustness analysis. In the third and last step, all the matrices R^k are compared to the matrix R. We choose for this comparison one of the proposed metrics. Finally, the quality of the ranking P is calculated by the weighted average of these comparisons. The three steps are detailed and explained below.

- Step 1: Computing the comparisons matrix R^k induced by the criterion gk
- The criterion gk, for $k \in \{1,...,m\}$ induces the R^K comparison matrix.
- Let $(R_{ij}^K)_{i, j \in \{1,...,n\}}$ be the comparison matrix. This matrix can be calculated by the following Eq. 1.

$$R_{ij}^k = \begin{cases} 1 \; if \, gk(i) > gk(j) \\ 0 \quad otherwise \end{cases} \tag{1}$$

- The matrix R^K contains only the numbers 0 and 1. The value 1 means that the alternative i is preferred to the alternative j according to the criterion gk.
- Step 2: Evaluation of the comparisons matrix R associate to the ranking P
- The comparisons matrix R, associate to the ranking P, is calculated in the same way for each criterion. The matrix R is given by the following Eq. 3.

$$R_{ij} = \begin{cases} 1 \; if \; "i" \; is \; better \; ranked \; than \; j \; in \; the \; ranking \; P \\ 0 \qquad\qquad\qquad otherwise \end{cases} \tag{2}$$

- The value 1 indicates that the action i is ranked before the action j for the aggregation multi-criteria method used.
- Step 3: Measure of the quality for the ranking P
- The ranking P will be a better choice if this ranking obtained sticks almost, or at least it is near, to all the criteria gk. This comes down to measuring the dependence between the matrix R and each matrix R^k. The experimentation section shows this link of dependence, which exists between the two matrices R and R^k. We propose in this article to measure this dependence by the correlation and distance metrics between the matrices. The quality $Q(P)$ of a ranking P is given by the Eqs. (3) and (4).

Correlation metric:

$$Q(P) = \frac{\sum_{k=1}^{m} w_k \times \text{correlation}(R^k, R)}{\sum_{k=1}^{m} w_k} \tag{3}$$

Distance metric:

$$Q(P) = 1 - \frac{\sum_{k=1}^{m} w_k \times \text{distance}(R^k, R)}{\sum_{k=1}^{m} w_k} \tag{4}$$

where: m is the number of criteria, and w_k is the importance of the criterion gk.

In Eqs. (3) and (4), we take into account the importance of criteria in measuring quality, both for correlation and distance. Insofar as a better correlation on an important criterion, for example, must take advantage and be taken into account in the final calculation of the measurement.

The correlation and distance metrics are explained in the following paragraphs.

Remark The quality of a ranking will be maximum when it goes in the same direction to all the criteria. In this case, the distance metric gives a result that is equal to 0, and which is the best result. That's why we chose "1-*distance* (R^k,R)" so that it becomes a maximization and which will have the same interpretation as the correlation measure.

3.2 The Pearson Correlation Coefficient

The correlation coefficient is a measure of the link between two variables [1]. This coefficient is used to characterize a positive or negative relationship, and it is the symmetrical measure, the closer it is to 1, in absolute value, the link and the dependence between the two variables will be better. For the measure of the correlation between two matrices X and Y, which are for our case comparison matrices translating rankings, we use the following Formula 5.

$$correlation(X, Y) = \frac{\sum_{i=1}^{n} \sum_{i,\, j=1}^{n} (X_{ij} - \bar{X}) \times (Y_{ij} - \bar{Y})}{\sqrt{\sum_{i=1}^{n} \sum_{i,\, j=1}^{n} (X_{ij} - \bar{X})^2} \times \sqrt{\sum_{i=1}^{n} \sum_{i,\, j=1}^{n} (Y_{ij} - \bar{Y})^2}} \tag{5}$$

where: $\bar{X} = \frac{1}{n(n-1)} \times \sum_{\substack{i,\, j=1 \\ i \neq j}}^{n} X_{ij}$ is the empirical average a square matrix X of

order n. n is the number of alternatives of set A.

3.3 The Distance Metrics

For the evaluation of the quality, we use the matrix distance between the two matrices R^k and R. As will be demonstrated in the experimentation section, a division by $n \times (n - 1)$ is added to the distance, in order to obtain a quality result varying between 0 and 1. So for a distance that gives 0 it means that the ranking result is the same as the ranking induced by the criterion gk, it is a better extreme result. And for a distance that gives 1 this means that the result ranking is entirely the opposite of the ranking induced by the criterion gk, it is an extremely lousy result.

All the distances calculated between R and R^k are then aggregated using a weighted average to have the total quality, given by Eq. 4.

All possible distances are summarized in the following Table 1. For each distance, we give it a name, under which it is known in the literature, as well as the formula

Table 1 The proposed distance metrics

Metric name	distance(X,Y)
Kendall metric: d_1	$d_1(X,\, Y) = \frac{1}{n \times (n-1)} \times \sum_{i=1}^{n} \sum_{j=1}^{n} \left| X_{ij} - Y_{ij} \right|$
Euclidean metric: d_2	$d_2(X,\, Y) = \sqrt{\frac{1}{n \times (n-1)} \times \sum_{i=1}^{n} \sum_{j=1}^{n} (X_{ij} - Y_{ij})^2}$
Hölder metric: d_α, with $\alpha \geq 3$	$d_\alpha(X,\, Y) = \sqrt[\alpha]{\frac{1}{n \times (n-1)} \times \sum_{i=1}^{n} \sum_{j=1}^{n} \left| X_{ij} - Y_{ij} \right|^\alpha}$

to compute the distance between any two matrices $X(Xij)1 \leq i,j \leq n$ and $Y(Yij)1 \leq i,j \leq n$, where n is the number of alternatives.

4 Numerical Experimentation and Discussion

To verify that the metrics can correctly measure the quality of the rankings, we examine in this section an experimental example. In this last one, we suggest an MCDM ranking problem with three criteria $F = \{g1,g2,g3\}$, and a set of four alternatives $A = \{A1,A2,A3,A4\}$. In order to have a significant interpretation of the results of this experimental study, we propose three criteria that lead to the same ranking. Besides, we choose the three criteria with the same weighting $w1 = w2 = w3 = 1$. We suppose that the three criteria give the same ranking: $A1 > A2 > A3 > A4$, as shown in Table 2. This ranking means that the alternatives $A1$, $A2$, $A3$, and $A4$ are respectively ranked first, second, third, and fourth.

The comparison matrices $R1, R2$, and $R3$ deduced respectively by the three criteria $g1$, $g2$, and $g3$ are calculated and give the same result of the comparison between the alternatives. These matrices are given in Table 3.

4.1 Numerical Results

For the evaluation of the quality of the classifications obtained by different MCDM methods to be compared, or obtained in robustness analysis, we propose to distinguish three cases of rankings P.

Table 2 The criteria rankings $P1$, $P2$, and $P3$

Alternatives/Rankings	$P1(g1)$	$P2(g2)$	$P3(g3)$
A1	1	1	1
A2	2	2	2
A3	3	3	3
A4	4	4	4

Table 3 Matrices induced by the criteria R^1, R^2, and R^3

	A1	A2	A3	A4
A1	0	1	1	1
A2	0	0	1	1
A3	0	0	0	1
A4	0	0	0	0

We distinguish two extreme cases with other intermediate cases. The first extreme case concerns a ranking coincides with all the rankings induced by the three criteria. The second extreme case concerns the case where the ranking is the opposite of all the rankings induced by the three criteria. Other intermediate cases are envisaged when the ranking obtained is slightly different from the rankings induced by the three criteria.

In this experiment, we compare 25 examples of the most representative P classifications and for which the alternatives alternate their ranks in the P classifications compared to the classifications of the three criteria $P1$, $P2$, and $P3$. Table 4 shows the calculated quality results.

The graph in Fig. 1 shows and compares the variation of the three quality measures obtained in Table 4.

Table 4 Quality measurement by the three metrics for the 25 selected rankings

| Example number	Ranking P	Q_corr(P)	Q_d1(P)	Q_d2(P)
E1	A1 > A2 > A3 > A4	1.0000	1.0000	1.0000
E2	A1 > A2 > A4 > A3	0.7333	0.8333	0.9129
E3	A1 > A3 > A2 > A4	0.7333	0.8333	0.9129
E4	A1 > A3 > A4 > A2	0.4667	0.6667	0.8165
E5	A2 > A1 > A3 > A4	0.7333	0.8333	0.9129
E6	A2 > A1 > A4 > A3	0.4667	0.6667	0.8165
E7	A2 > A3 > A1 > A4	0.4667	0.6667	0.8165
E8	A2 > A3 > A4 > A1	0.2000	0.5000	0.7071
E9	A2 > A4 > A3 > A1	−0.0667	0.3333	0.5774
E10	A3 > A1 > A2 > A4	0.4667	0.6667	0.8165
E11	A3 > A2 > A1 > A4	0.2000	0.5000	0.7071
E12	A3 > A2 > A4 > A1	−0.0667	0.3333	0.5774
E13	A3 > A4 > A1 > A2	−0.0667	0.3333	0.5774
E14	A3 > A4 > A2 > A1	−0.3333	0.1667	0.4082
E15	A4 > A1 > A2 > A3	0.2000	0.5000	0.7071
E16	A4 > A1 > A3 > A2	−0.0667	0.3333	0.5774
E17	A4 > A2 > A1 > A3	−0.0667	0.3333	0.5774
E18	A4 > A2 > A3 > A1	−0.3333	0.1667	0.4082
E19	A4 > A3 > A1 > A2	−0.3333	0.1667	0.4082
E20	A4 > A3 > A2 > A1	−0.60	0.0000	0.0000
E21	A1 > A2 > A3 = A4	0.8704	0.9167	0.9574
E22	A1 > A2 = A3 = A4	0.6202	0.7500	0.8660
E23	A1 = A2 > A3 > A4	0.8704	0.9167	0.9574
E24	A1 = A2 = A3 > A4	0.6202	0.7500	0.8660
E25	A1 = A2 > A3 = A4	0.7454	0.8333	0.9129
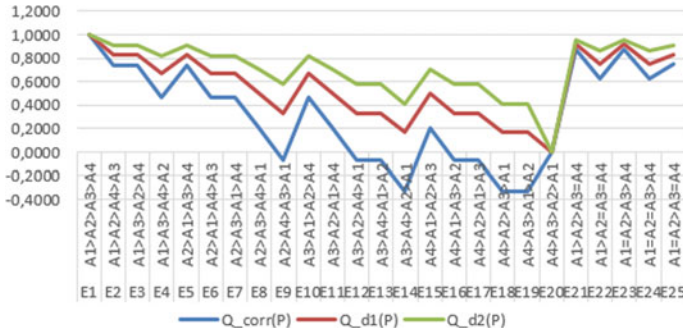
Fig. 1 Graphical representation of quality variations for the three metrics

4.2 Discussion

The graph above clearly justifies that the three metrics vary in the same direction and lead to the same interpretations and analysis of the qualities of the rankings; this allows us to conclude that the three metrics are equivalent and will give the same quality comparison result of the rankings.

Moreover, we notice that the ranking E1: $A1 > A2 > A3 > A4$, see Table 4, gives the maximum quality, which is worth 1. Indeed, this ranking coincides, by hypothesis, with the three rankings induced by the three criteria $g1$, $g2$, and $g3$. However, the ranking E20: $A4 > A3 > A2 > A1$ gives the worst quality, which is 0 for the metrics d1 and d2, and -0.63 for the correlation metric. That is justified by the fact that the ranking E20 is upside down from the three rankings induced by $g1$, $g2$, and $g3$.

For all the other ranking cases, the quality varies between 1 and 0 for the metrics d1 and d2. For the correlation metric, the quality varies between 1 and -0.6. The negative values obtained for the quality show that there is an inverse dependence between the matrices R and R^k [1]. The quality decreases notably when the alternatives change their initial ranks for the E1 ranking.

The same results are obtained with the rankings E21–E22–E23–E24–E25, where some alternatives have equal rank.

In the conclusion of this experiment, the three metrics are equivalent and in perfect coherence. They also have significance for the measurement of the quality of the rankings.

5 Conclusions

The main objective of the paper was to suggest metrics to measure the quality of the rankings. These metrics will be beneficial to compare objectively several MCDM ranking methods.

We have proved and shown in this work that all the proposed metrics give significant results for the measurement of the ranking quality and vary correctly between 1, for the best ranking, and 0 for a bad ranking. Except for correlation, this metric can reach values less than 0, which in this case shows an inverse dependence between the matrices R and R^k. It has also been proven that all the metrics are consistent and equivalent insofar as they lead to the same results of the rankings. The proposed metrics will so serve indifferently as rational and objective tools for the comparison between the ranking results, and can thus help the decision-maker to choose the best ranking of the alternatives.

In our future research, we intend to extend the metrics for the estimation of the quality of the rankings for the case of uncertain criteria as deployed by some MCDM methods such as the famous ELECTRE III method. Other research will focus on quality measurement for sorting and selection problematics.

References

1. Abounaima MC, Mazouri FZE, Lamrini L, Nfissi N, Makhfi NE, Ouzarf M (2020) The pearson correlation coefficient applied to compare multi-criteria methods: case the ranking problematic. In: 2020 1st international conference on innovative research in applied science, engineering and technology (IRASET), pp 1–6. https://doi.org/10.1109/IRASET48871.2020.9092242
2. Aruldoss M, Lakshmi TM, VenkatesanVP (2013) A survey on multi criteria decision making methods and its applications. Am J Inf Syst 1(1). https://doi.org/10.12691/ajis-1-1-5
3. Govindan K, Jepsen MB (2016) Electre: a comprehensive literature review on methodologies and applications. Eur J Oper Res 250(1):1–29. https://doi.org/10.1016/j.ejor.2015.07.019
4. Roy B, Bouyssou D (1993) Aide multicritère méthodes et cas. Economica, Paris
5. Zavadskas EK, Turskis Z, Kildienė S (2014) State of art surveys of overviews on MCDM/MADM methods. Technol Econo Dev Econ 20(1):165–179. https://doi.org/10.3846/20294913.2014.892037
6. El Mazouri FZ, Abounaima MC, Zenkouar K (2019) Data mining combined to the multi-criteria decision analysis for the improvement of road safety: case of France. J Big Data 6(1):5. https://doi.org/10.1186/s40537-018-0165-0
7. Kumar G, Parimala N (2019) A sensitivity analysis on weight sum method MCDM approach for product recommendation. In: Distributed computing and internet technology, pp 185–193. https://doi.org/10.1007/978-3-030-05366-6_15
8. Vivekanandan N, Viswanathan K, Gupta S (2010) Errata to: optimization of cropping pattern using goal programming approach. Opsearch 47(1):104–104. https://doi.org/10.1007/s12597-010-0007-0
9. Behzadian M, Khanmohammadi Otaghsara S, Yazdani M, Ignatius J (2012) A state-of the-art survey of TOPSIS applications. Exp Syst Appl 39(17): 13051–13069. https://doi.org/10.1016/j.eswa.2012.05.056
10. Emovon I, Oghenenyerovwho OS (2020) Application of MCDM method in material selection for optimal design: a review. Res Mater 7:100115. https://doi.org/10.1016/j.rinma.2020.100115
11. Mazouri FZE, Abounaima MC, Zenkouar K, Alaoui AEH (2018) Application of the ELECTRE III Method at the Moroccan Rural electrification program. Int J Electr Comput Eng (IJECE) 8(5):3285–3295. https://doi.org/10.11591/ijece.v8i5.pp3285-3295
12. Behzadian M, Kazemzadeh RB, Albadvi A, Aghdasi M (2010) PROMETHEE: a comprehensive literature review on methodologies and applications. Eur J Oper Res 200(1):198–215. https://doi.org/10.1016/j.ejor.2009.01.021

13. Wang M, Lin S-J, Lo Y-C (2010) The comparison between MAUT and PROMETHEE. In: 2010 IEEE international conference on industrial engineering and engineering management, pp 753–757. https://doi.org/10.1109/IEEM.2010.5675608
14. Sánchez-Lozano JM, García-Cascales MS, Lamata MT (2016) Comparative TOPSIS-ELECTRE TRI methods for optimal sites for photovoltaic solar farms. Case study in Spain. J Clean Prod 127:387–398. https://doi.org/10.1016/j.jclepro.2016.04.005