Materials Research Proceedings 35 (2023) 276-285 https://doi.org/10.21741/9781644902714-33

The ranking-aggregation problem in manufacturing: potential, pitfalls, and good practices

Fiorenzo Franceschini[1,a], Domenico Augusto Maisano[1,b *] and Luca Mastrogiacomo[1,c]

[1] Dept. of Management and Production Engineering (DIGEP), Politecnico di Torino, Italy

[a] fiorenzo.franceschini@polito.it, [b] domenico.maisano@polito.it, [c] luca.mastrogiacomo@polito.it

Keywords: Decision Making, Performance Indicators, Ranking-Aggregation Problem

Abstract. A number of *experts*, who individually rank a set of *objects* based on a certain attribute, and the need to aggregate the resulting (subjective) *rankings* into a *collective judgement*: these are the "ingredients" of the ranking-aggregation problem, which is typical of *social choice*, *psychometrics* and *economics*. This paper shows that the problem has many interesting applications even in *manufacturing* and must be approached with care, in order to avoid misleading results. Through a real-world case study concerning cobot-assisted manual (dis)assembly, the paper illustrates (i) a methodology to tackle the problem in a practical and effective way and (ii) various useful tools (e.g., for estimating the degree of *concordance* among experts, the *consistency* and *robustness* of collective judgment, etc.). The article is addressed to scientists and practitioners in the manufacturing field.

Introduction

Ranking aggregation is an ancient problem with three characteristic elements: (i) a set of *objects* to be prioritised according to a certain *subjective attribute*, (ii) a set of *experts* (equally important or with a hierarchy of importance), who formulate *preference rankings* of the objects of interest, and (iii) a *collective judgement* concerning the objects, resulting from the aggregation of expert rankings through a suitable *aggregation technique* [1-3].

Due to the great generality, disciplinary transversality, and multiplicity of potential applications, the ranking-aggregation problem is of interest to many scientific disciplines and operational contexts, including *manufacturing* [1, 3-6]. Some of the many possible manufacturing applications are:

- *Conceptual design*, regarding the opinions of different designers about alternative design concepts, from the perspective of a specific attribute [7];
- *Production management*, regarding the selection of the most appropriate production system on the basis of productivity, flexibility or another performance attribute [8-9];
- *Quality control*, regarding the prioritization of defects on manufactured parts, aggregating (subjective) expert judgments by visual inspection [10].

The analyst's attention is often directed to the aggregation technique, which can be interpreted as a "black box" transforming *input* data (i.e., experts' rankings and importance hierarchy) into *output* data (i.e., collective judgement and related data) [6]. However, this may lead to overlooking other important aspects that characterize the ranking-aggregation problem, e.g., preliminary assessment of the degree of *concordance* among experts, verification of the *consistency* and *robustness* of output data, etc. The above considerations can turn into the research question: "*What methodological approach should be adopted to address the problem of interest with full awareness?*". Despite the variety of applications to specific cases of interest to individual authors,

Materials Research Proceedings 35 (2023) 276-285 https://doi.org/10.21741/9781644902714-33

the scientific literature in the manufacturing field lacks general guidelines and a collection of good practices for addressing the ranking-aggregation problem.

Aimed at scientists and manufacturing professionals, this work is intended to increase their awareness of the complexity of the ranking-aggregation problem, while providing a set of useful tools to tackle it in a practical and effective manner. Following a pedagogical approach, the description is accompanied by a case-study application concerning cobot-assisted manual (dis)assembly.

Case study

A company reconditions different types of automotive components, mainly starters and alternators. Because of the wide variety of components and the complexity of (dis)assembly and repair operations, the company has been supporting human operators with *collaborative robots*, or simply *cobots*. Cobots are useful for assisting operators in manual operations that require great precision, dexterity and strength [11]. They are extremely versatile for multiple tasks, such as (i) picking up, clamping, handing the tools and parts to be machined/assembled, (ii) supporting dimensional inspection, online quality control, etc., and (iii) guiding less experienced operators, like virtual tutors.

The current market includes a relatively wide range of cobot models, which could be adapted to the operational context of interest. The company management decided to identify the most appropriate cobot model depending on *programming practicality*; in fact, this aspect is crucial in making task preparation faster and easier, while reducing the level of technical skills required by operators [11]. Having previously selected five cobot models from those at the cutting edge of the market, the company relies on the evaluation of a panel of eight experts, including technicians, engineers and external consultants with relatively in-depth and complementary expertise.

Methodology

The flowchart in Fig. 1 summarizes the proposed methodology, which can be divided into three operational phases, illustrated in the following subsections. The multiple feedback loops denote the iterative nature of the proposed procedure, which includes several intermediate verifications, with possible in-progress corrections and adjustments.

Problem definition

First, the specific problem and its characteristics should be identified clearly and unambiguously.

Based on the above case study, a specific ranking-aggregation problem can be formulated: the n = 5 *objects* (o_1 to o_5) are the cobot models that will be evaluated in terms of programming practicality, i.e., the *attribute* of interest, which is inherently *subjective*. The m = 8 *experts* are technicians (e_1 to e_8), engineers and external consultants who formulate their individual preference rankings of the cobot models.

In selecting experts, (at least) two aspects must be taken into account:

1. The greater the number of experts formulating their individual rankings, the higher the statistical relevance of the problem output [12, 13]. Pragmatically, it would be desirable for m to be no less than 5-6, in order for the results of the study to be relevant [6].
2. It may sometimes be appropriate to have a hierarchy of importance of experts, for instance by discriminating those with greater technical expertise. This hierarchy can be constructed in different ways, typically by associating each expert with a *weight* or defining an *importance ranking* [12]. For simplicity, in the case study all experts are considered as equally important.

Next, the type of expert rankings can be determined depending on several factors, such as the *goal* of the problem (e.g., identifying the best/worst object(s), drawing up a complete ranking, etc.), the *data-collection* strategy (e.g., through focus groups, personal telephone/street interviews, online forms, etc.), etc.

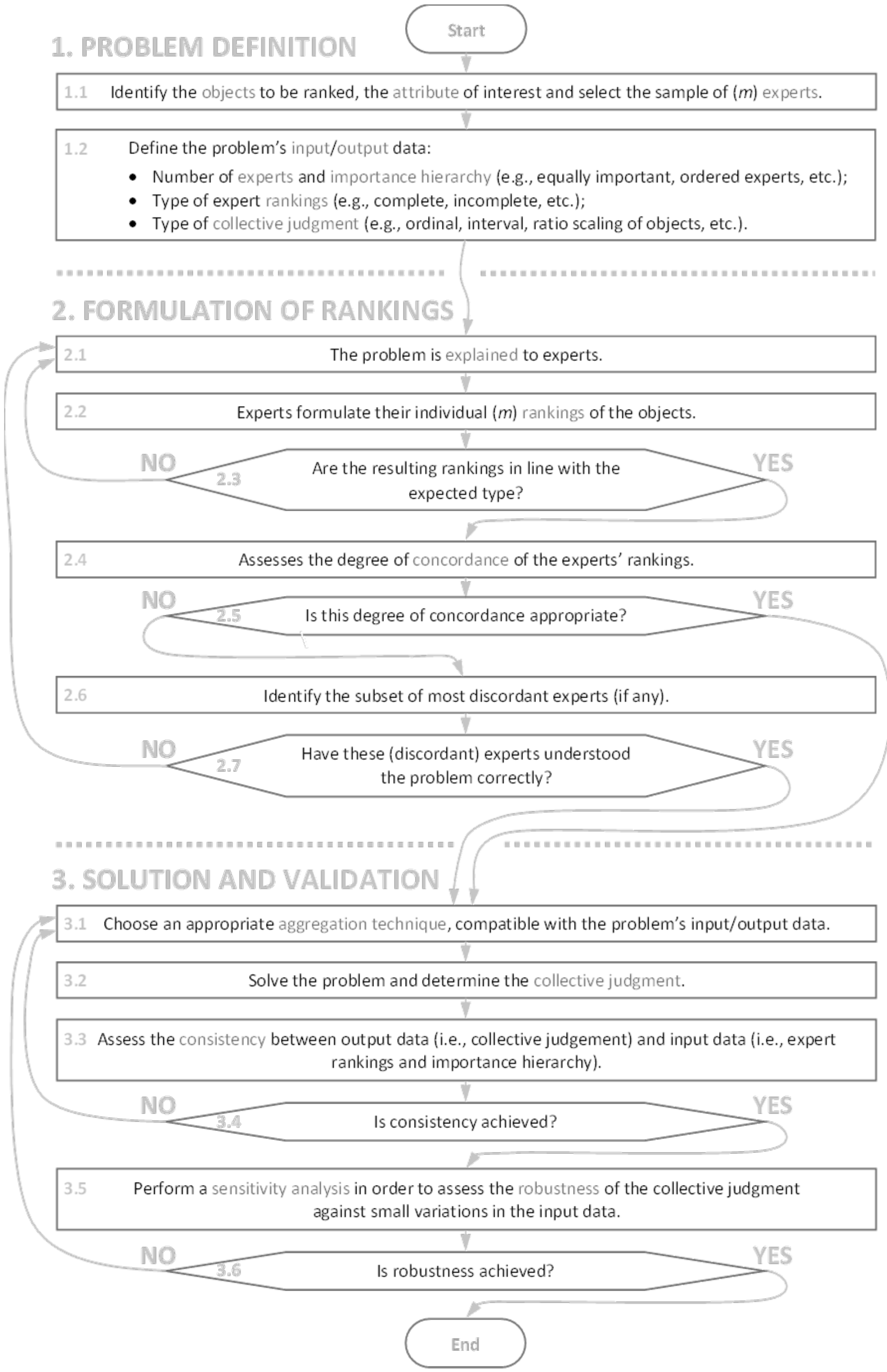

Fig. 1. Flow chart summarizing the proposed operational methodology.

Complete rankings – i.e., rankings in which experts order all objects by linking them with *strict preference* ("$o_i \succ o_j$") or *indifference* relationships ("$o_i \sim o_j$") – represent a classic scenario, although their formulation requires some effort, especially if the number of objects is large [12]. On the other hand, *incomplete* rankings are more "digestible" for experts, because they can take into account possible hesitations or doubts. E.g., incomplete are those rankings in which only a small number of top or bottom objects are included (e.g., the three most/least preferred), or in which the expert decides to omit an object from his/her ranking (e.g., since he/she is not familiar with), or even rankings with *incomparability* relationships between objects ("$o_i \| o_j$") [6]. Given the relatively small number of objects, in the present case experts formulate complete rankings of all five objects.

Subsequently, the type of collective judgment has to be defined depending on the properties that are "desirable" for the specific problem; there is a wide range of possibilities: *rankings*, *scalings* on different scale types (e.g., *interval*, *ratio*), *clusterings*, *scorings*, or collective judgments designating only the winner/loser object, etc. [6]. For the sake of simplicity, in the case study the expected collective judgment is represented by a complete ranking.

Formulation of rankings

This stage begins with a detailed explanation of the problem to experts, who need to understand exactly which objects are to be evaluated, the attribute against which the evaluation is to be made, and how to formulate individual rankings. As seen before, in the case study each expert is required to formulate a complete ranking of all objects; Fig. 2(a) reports the resulting expert rankings, which include relationships of strict preference ("$o_i \succ o_j$") and indifference ("$o_i \sim o_j$") between objects. At this stage, it must be ensured that the expert rankings are formulated consistently with the expected type; if necessary, the formulation must be corrected/revised (see feedback loop from block 2.3 in Fig. 1).

Evaluating the *concordance* among expert rankings is a preliminary check of the plausibility of input data, which is useful to prevent difficulties, such as excessive heterogeneity in the selection of experts, poor understanding of the problem, errors in the formulation of rankings, or other potential obstacles to achieving consensus. The scientific literature includes various statistical indicators, which can be used depending on the problem characteristics [12, 14]. Since the present case is characterized by complete expert rankings with equally-important experts, the Kendall's W and Spearman's ρ can be used [6].

W, known as *coefficient of concordance*, is a *multivariate* statistic that applies at the level of expert rankings and is related to the dispersion of the ranks associated with each object [6, 15]. This measure belongs to [0, 1], with 1 indicating perfect concordance and 0 indicating independence [12]. Returning to the case study, each ranking can be translated into a set of ranks – that is, permutations of the integers {1, 2, 3, 4, 5} – which are then organized into a so-called *rank table*, i.e., a bidirectional matrix of size $m \times n$, with row and column labels designating experts and objects (see Fig. 2(b)). In the case of *tied* objects (i.e., pairs of objects with *indifference* relationships, e.g., "$o_i \sim o_j$"), we conventionally use the *average ranks* that each set of bound objects would occupy if a preference could be expressed [12]; for example, in a ranking where objects o_1 and o_3 are tied for 3rd and 4th place (e.g., see the ranking by e_6 in Fig. 2(a)), the average rank of (3+4)/2 = 3.5 would be assigned to both.

W is defined as:

$$W = \frac{\sum_{j=1}^{n}\left(R_j - \bar{R}\right)^2}{\left[m^2 \cdot n \cdot (n^2-1) - m \cdot \sum_{i=1}^{m} T_i\right]/12}, \quad (1)$$

being

n the number of objects;

m the number of experts;

R_j the column total related to the j-th column of the rank table;

$\bar{R} = m \cdot (n+1)/2$ the average column total (i.e., 24 in the present case);

$T_i = \sum_{k=1}^{g_i}(t_k^3 - t_k)$ a correction factor for ties, in which t_k is the number of tied ranks in the k-th group of tied ranks (where a group is a set of values having constant tied rank) and g_i is the number of groups of ties in the set of ranks (ranging from 1 to n) for expert i.

Experts	(a) Rankings	(b) Rank table o_1	o_2	o_3	o_4	o_5	Row totals	T_i
e_1	$o_3 \succ (o_1 \sim o_5) \succ o_2 \succ o_4$	2.5	4.0	1.0	5.0	2.5	15	6
e_2	$o_5 \succ o_1 \succ (o_2 \sim o_3 \sim o_4)$	2.0	4.0	4.0	4.0	1.0	15	24
e_3	$(o_1 \sim o_3) \succ o_2 \succ (o_4 \sim o_5)$	1.5	3.0	1.5	4.5	4.5	15	12
e_4	$o_1 \succ o_5 \succ o_3 \succ o_4 \succ o_2$	1.0	5.0	3.0	4.0	2.0	15	-
e_5	$o_4 \succ (o_1 \sim o_2 \sim o_5) \succ o_3$	3.0	3.0	5.0	1.0	3.0	15	24
e_6	$o_5 \succ o_4 \succ (o_1 \sim o_3) \succ o_2$	3.5	5.0	3.5	2.0	1.0	15	6
e_7	$o_3 \succ o_5 \succ o_4 \succ o_2 \succ o_1$	5.0	4.0	1.0	3.0	2.0	15	-
e_8	$(o_3 \sim o_5) \succ (o_1 \sim o_2) \succ o_4$	3.5	3.5	1.5	5.0	1.5	15	12
	Col. totals (R_j)	22.0	31.5	21.5	28.5	17.5	120	

Fig. 2. (a) Complete rankings of n = 5 objects, formulated by m = 8 experts; (b) corresponding rank table. T_i is a correction factor for ties (cf. Eq. 1).

With reference to the case study, it is obtained $W = 23.1\%$, denoting a relatively low level of concordance. To further investigate the reasons for this low inter-expert concordance, the *bivariate* perspective of Spearman's correlation coefficient (ρ) related to each possible pair of rankings can be considered. Tab. 1 contains the ρ coefficients between all the possible pairs of expert rankings under consideration [15].

Tab. 1. Spearman's ρ correlation table for the expert rankings in Fig. 2(a). The most pronounced negative correlations (i.e., $\rho < -0.4$) are bolded.

Ranking	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
e_1	1							
e_2	0.287	1						
e_3	0.649	-0.177	1					
e_4	0.564	0.783	0.316	1				
e_5	**-0.918**	0.000	**-0.707**	-0.224	1			
e_6	-0.026	0.574	**-0.649**	0.410	0.344	1		
e_7	0.462	-0.112	-0.158	-0.100	**-0.447**	0.410	1	
e_8	0.865	0.412	0.250	0.369	**-0.825**	0.162	0.632	1

Rather pronounced negative correlations (i.e., $\rho < -0.4$) between certain pairs of expert rankings stand out. Curiously, they (almost) always involve the ranking by e_5, denoting a sort of "countertrend" with respect to the other rankings. Upon brief investigation, it turns out that e_5 expert misunderstood the ranking construction, formulating it in the sense of reverse preference; therefore, the correct ranking should be "$o_3 \succ (o_1 \sim o_2 \sim o_5) \succ o_4$" instead of "$o_4 \succ (o_1 \sim o_2 \sim o_5) \succ o_3$" (see feedback loop from block 2.7 in Fig. 1). After this correction, the W value is significantly higher than before (i.e., $W = 39.6\%$ versus 23.1%). Simultaneously, the relatively large negative ρ values for e_5 are "reabsorbed".

As exemplified, the concordance analysis can be useful in pointing out possible anomalies and "pitfalls" in the formulation of expert rankings [6].

Solution and validation

At this point, it is needed to solve the ranking-aggregation problem utilizing an appropriate aggregation technique and, subsequently, verifying the plausibility of the resulting output. Unfortunately, presenting an exhaustive overview of the state-of-art techniques would require an encyclopaedic analysis. Far from this ambition, Tab. 2 simply recalls some possible aspects to be taken into account while selecting the aggregation technique [6]. For an overview of the aggregation techniques in the "mare magnum" of the scientific literature, we refer the reader to relevant surveys and extensive reviews [2, 6, 16].

Materials Research Proceedings 35 (2023) 276-285 https://doi.org/10.21741/9781644902714-33

Tab. 2. Aspects to consider when selecting a ranking-aggregation technique [6].

(a) Input-data characteristics	(b) Aggregation mechanism	(c) Output-data characteristics
• Problem size: - Number of objects (n); - Number of expert rankings (m). • Type of expert rankings • Type of expert hierarchy	• Rule-based. • Optimization-based; • Distribution-based.	• Designation of a unique winner/loser; • Complete/incomplete *ranking*; • *Classification* in categories; • Collective *scoring* of objects; • Collective *scaling* of objects. …

A relatively simple aggregation technique is applied for the problem of interest: the so-called *Borda Count* (BC), according to which, for each expert ranking, the first object accumulates one point, the second two points, and so on [3, 17]. The collective score of one object can be calculated by cumulating the scores related to each ranking.

The application of the BC technique to the expert rankings (after the correction of the ranking by e_5) leads to the following collective scoring: o_1 = 22.0, o_2 = 31.5, o_3 = 16.5, o_4 = 32.5, o_5 = 17.5 (cf. also Tab. 5(b-i)), from which the collective ranking $o_3 \succ o_5 \succ o_1 \succ o_2 \succ o_4$ is deduced.

Every aggregation technique provides a collective judgment; but how does one know whether it is plausible? Certainly, the rationale of the aggregation technique represents a conceptual guarantee that it is capable of producing reasonable results. However, the aggregation technique that most consistently reflects expert rankings cannot be assessed *ex ante*, but only *ex post* and on a case-by-case basis [4, 18].

Studies have focused on the concept of *consistency of the collective judgment with respect to input data*, defined as "*the ability of a collective judgment to reflect the rankings of experts, while taking the importance hierarchy into account*" (i.e., giving priority to the more important experts) [6].

Among the available tools to assess the degree of consistency of the solution to a certain ranking-aggregation problem, p-indicators are very versatile, as they can be adapted to a variety of contexts, such as those in which expert rankings are (i) not necessarily complete, (ii) equally important, or (iii) characterized by an importance hierarchy [6]. In general, p-indicators can be divided into two families:

- $\boldsymbol{p_j}$, indicators of *local consistency*, which are based on the comparison of each j-th expert's ranking with the collective judgement. A preliminary operation for determining p_j is constructing "*a paired-comparison table*" in which each ranking (i.e., those from experts and that one deduced from the collective judgment) is transformed into sets of paired-comparison relationships (see symbols "$\succ$" and "~" in Tab. 3(a)). Next, a "*consistency table*" – which turns the paired-comparison relationships of each expert into scores, according to the following scoring system is constructed:
  1. *Full consistency*, i.e., identical relationship of strict preference ("$\succ$") or indifference ("~") $\Rightarrow$ score 1;
  2. *Weak consistency*, i.e., consistency with respect to a weak preference relationship only ("$\succ$ or ~" and "$\prec$ or ~", i.e., strict preference or indifference); e.g., when comparing the relationship $o_1 \succ o_2$ with $o_1 \sim o_2 \Rightarrow$ score 0.5;
  3. *Inconsistency* (with respect to both strict and weak preference relationships); e.g., when comparing the relationship $o_1 \succ o_2$ with $o_2 \succ o_1 \Rightarrow$ score 0.

  The conventional assignment of 0.5 points in the case of *weak consistency* is justified by the fact that this is the intermediate case between that of *full consistency* (with score 1) and that of *inconsistency* (with score 0) [6]. The consistency table also reports the sum of the scores (x_j)

obtained by each j-th expert ranking. Tab. 3(b) exemplifies the consistency table related to the case study of interest.

Next, for each j-th expert, the portion of "consistent" paired-comparisons can be calculated as:

$$p_j = \frac{x_j}{\binom{n}{2}} = \frac{x_j}{10}, \tag{2}$$

being:

x_j the total score related to the j-th expert;

$\binom{n}{2} = \frac{n \cdot (n-1)}{2}$ the overall number of paired comparisons (i.e., 10 in the present case).

- ***p***, i.e., indicator of *global consistency*. In the case of equally-important experts, the p_j values are aggregated through the arithmetic average [6]:

$$p = \frac{1}{m} \cdot \sum_{j=1}^{m} p_j, p \in [0,1]. \tag{3}$$

In this specific case, the aggregation technique results into p = 73.8% (see Tab. 3(c)), denoting a relatively good consistency [6, 10].

Tab. 3. (a) Paired-comparison table, (b) consistency table, and (c) p-indicators related to the BC technique.

(a) Paired-comparison table

Paired comparison		Experts e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	Collective judgment
1	o_1, o_2	$\succ$	$\succ$	$\succ$	$\succ$	$\sim$	$\succ$	$\prec$	$\sim$	$\succ$
2	o_1, o_3	$\prec$	$\succ$	$\sim$	$\succ$	$\prec$	$\sim$	$\prec$	$\prec$	$\prec$
3	o_1, o_4	$\succ$	$\succ$	$\succ$	$\succ$	$\succ$	$\prec$	$\prec$	$\succ$	$\succ$
4	o_1, o_5	$\sim$	$\prec$	$\succ$	$\succ$	$\sim$	$\prec$	$\prec$	$\prec$	$\prec$
5	o_2, o_3	$\prec$	$\sim$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$
6	o_2, o_4	$\succ$	$\sim$	$\succ$	$\prec$	$\succ$	$\prec$	$\prec$	$\succ$	$\succ$
7	o_2, o_5	$\prec$	$\prec$	$\succ$	$\prec$	$\sim$	$\prec$	$\prec$	$\prec$	$\prec$
8	o_3, o_4	$\succ$	$\sim$	$\succ$	$\succ$	$\succ$	$\prec$	$\succ$	$\succ$	$\succ$
9	o_3, o_5	$\succ$	$\prec$	$\succ$	$\prec$	$\succ$	$\prec$	$\succ$	$\sim$	$\succ$
10	o_4, o_5	$\prec$	$\prec$	$\sim$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$	$\prec$

(b) Consistency table

	Scores e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	
1	1	1	1	1	0.5	1	0	0.5	
2	1	0	0.5	0	1	0.5	1	1	
3	1	1	1	1	1	0	0	1	
4	0.5	1	0	0	0.5	1	1	1	
5	1	0.5	1	1	1	1	1	1	
6	1	0.5	1	0	1	0	0	1	
7	1	1	0	1	0.5	1	1	1	
8	1	0.5	1	1	1	0	1	1	
9	1	0	1	0	1	0	1	0.5	
10	1	1	0.5	1	1	1	1	1	
x_j	9.5	6.5	7	6	8.5	5.5	7	9	
(c) p-indicators $\boldsymbol{p_j}$	**95%**	**65%**	**70%**	**60%**	**85%**	**55%**	**70%**	**90%**	$\Rightarrow$ ***p* = 73.8%**

The formulation of rankings is often affected by inherent variability, which can "propagate" onto the variability of the output [19]. In general, it may be useful to perform a *sensitivity analysis* to assess the robustness of the solution against small variations in the input data [19]. An example of sensitivity analysis follows.

Tab. 4 contains three sets of expert rankings: the initial one and two additional ones, obtained by applying small distortions (e.g., some "rank-reversal") to the initial one. For each set, the collective scoring/ranking was determined by applying the BC aggregation technique (see results in Tab. 5). Next, the average dispersion in the rank position of individual objects was used as a proxy for the robustness of the resulting collective rankings (see Tab. 5(c)). In this specific case, BC seems to provide a somewhat robust result (i.e., mean standard deviation of 0.44). Thus, no revision of the aggregation technique adopted is necessary (cf., feedback loop from block 3.6 of Fig. 1).

Discussion and general remarks

This paper focused on the ranking-aggregation problem, due to the variety of potential applications in manufacturing. Through a pedagogical approach based on a case study, the paper illustrated a sequential and iterative operational methodology to tackle the problem of interest at multiple levels:

Tab. 4. Set of rankings used for sensitivity analysis.

Experts	(i) Initial set of rankings	(ii) 1st additional set	(iii) 2nd additional set
e_1	$o_3 \succ (o_1 \sim o_5) \succ o_2 \succ o_4$	$(o_3 \sim o_1) \succ o_5 \succ (o_2 \sim o_4)$	$o_5 \succ o_3 \succ o_1 \succ o_4 \succ o_2$
e_2	$o_5 \succ o_1 \succ (o_2 \sim o_3 \sim o_4)$	$o_1 \succ o_5 \succ (o_2 \sim o_3) \succ o_4$	$(o_5 \sim o_1 \sim o_2) \succ o_4 \succ o_3$
e_3	$(o_1 \sim o_3) \succ o_2 \succ (o_4 \sim o_5)$	$(o_1 \sim o_3) \succ o_5 \succ (o_2 \sim o_4)$	$(o_1 \sim o_3 \sim o_2) \succ o_4 \succ o_5$
e_4	$o_1 \succ o_5 \succ o_3 \succ o_4 \succ o_2$	$o_1 \succ (o_5 \sim o_3) \succ o_4 \succ o_2$	$o_5 \succ o_1 \succ (o_3 \sim o_4 \sim o_2)$
e_5	$o_3 \succ (o_1 \sim o_2 \sim o_5) \succ o_4$	$o_3 \succ (o_1 \sim o_2) \succ o_5 \succ o_4$	$o_3 \succ o_2 \succ o_1 \succ (o_5 \sim o_4)$
e_6	$o_5 \succ o_4 \succ (o_1 \sim o_3) \succ o_2$	$o_5 \succ o_4 \succ (o_1 \sim o_2) \succ o_3$	$o_4 \succ o_5 \succ (o_1 \sim o_3 \sim o_2)$
e_7	$o_3 \succ o_5 \succ o_4 \succ o_2 \succ o_1$	$o_5 \succ o_3 \succ o_2 \succ o_4 \succ o_1$	$o_3 \succ o_4 \succ o_5 \succ (o_2 \sim o_1)$
e_8	$(o_3 \sim o_5) \succ (o_1 \sim o_2) \succ o_4$	$(o_3 \sim o_5) \succ o_1 \succ o_2 \succ o_4$	$o_1 \succ (o_3 \sim o_5) \succ (o_2 \sim o_4)$

Tab. 5. Rank tables and collective scorings/rankings resulting from sensitivity analysis.

(i) Initial set of rankings

	(a) Rank table								(b) Collect. scoring	(rank)
	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8		
o_1	2.5	2	1.5	1	3	3.5	5	3.5	22.0	(3.0)
o_2	4	4	3	5	3	5	4	3.5	31.5	(4.0)
o_3	1	4	1.5	3	1	3.5	1	1.5	16.5	(1.0)
o_4	5	4	4.5	4	5	2	3	5	32.5	(5.0)
o_5	2.5	1	4.5	2	3	1	2	1.5	17.5	(2.0)

Collective ranking: $o_3 \succ o_5 \succ o_1 \succ o_2 \succ o_4$

(ii) 1st additional set

	(a) Rank table								(b) Collect. scoring	(rank)
	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8		
o_1	1.5	1	1.5	1	2.5	3.5	5	3	19.0	(3.0)
o_2	4.5	3.5	4.5	5	2.5	3.5	3	4	30.5	(4.0)
o_3	1.5	3.5	1.5	2.5	1	5	2	1.5	18.5	(2.0)
o_4	4.5	5	4.5	4	5	2	4	5	34.0	(5.0)
o_5	3	2	3	2.5	4	1	1	1.5	18.0	(1.0)

Collective ranking: $o_5 \succ o_3 \succ o_1 \succ o_2 \succ o_4$

(c) Rank dispersion

Object	St. dev.
o_1	0.3
o_2	0.3
o_3	0.8
o_4	0.3
o_5	0.6

Mean st.dev. **0.44**

(iii) 2nd additional set

	(a) Rank table								(b) Collect. scoring	(rank)
	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8		
o_1	3	2	2	2	3	4	4.5	1	21.5	(2.5)
o_2	5	2	2	4	2	4	4.5	4.5	28.0	(4.5)
o_3	2	5	2	4	1	4	1	2.5	21.5	(2.5)
o_4	4	4	4	4	4.5	1	2	4.5	28.0	(4.5)
o_5	1	2	5	1	4.5	2	3	2.5	21.0	(1.0)

Collective ranking: $o_5 \succ (o_1 \sim o_3) \succ (o_2 \sim o_4)$

- Checking the plausibility of expert rankings in terms of *concordance*, through *multivariate* and *bivariate* statistical indicators;
- Guiding the aggregation-technique selection, depending on the desired types of input and output data;
- Evaluating the *consistency* and *robustness* of the resulting collective judgment.

Interestingly, the application of the aggregation technique is only one of several steps in the proposed methodology. This study gives greater awareness of the complexity of the ranking-aggregation problem, providing some practical tools for dealing with it in a structured and effective way. The results of this study may be useful for scientists and practitioners in manufacturing, who are facing various kinds of decision-making problems that can be linked to that of ranking aggregation. Although the case study focused on a few specific practical tools (e.g. ρ, W, and p-indicators), the proposed methodology is open to the use of other similar tools.

Regarding the future, it is planned to define an in-depth taxonomy of the aggregation techniques and analytical tools, so as to facilitate their selection for a specific problem of interest.

References

[1] Spohn, W. (2009). A survey of ranking theory. In Degrees of belief (pp. 185-228). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-9198-8_8

[2] Reich, Y. (2010). My method is better!, Research in engineering design 21(3): 137-142. https://doi.org/10.1007/s00163-010-0092-3

[3] Saari D.G. (2011). Decision and elections, Cambridge: Cambridge University Press.

[4] Arrow K.J. (2012). Social choice and individual values, 3rd edn. Yale University Press, New Haven.

[5] Köksalan, M., Wallenius, J., Zionts, S. (2013). An early history of multiple criteria decision making. Journal of Multi-Criteria Decision Analysis, 20(1-2): 87-94. https://doi.org/10.1002/mcda.1481

[6] Franceschini, F., Maisano, D., Mastrogiacomo, L. (2022) Rankings and Decisions in Engineering: Conceptual and Practical Insights. International Series in Operations Research & Management Science Series, Vol. 319, Springer International Publishing, Cham (Switzerland), ISSN: 0884-8289. https://doi.org/10.1007/978-3-030-89865-6

[7] Franceschini, F., Maisano, D. (2019). Design decisions: concordance of designers and effects of the Arrow's theorem on the collective preference ranking. Research in Engineering Design, 30(3), 425-434. https://doi.org/10.1007/s00163-019-00313-9

[8] Chatterjee, P., Chakraborty, S. (2014). Flexible manufacturing system selection using preference ranking methods: A comparative study. International Journal of Industrial Engineering Computations, 5(2), 315-338. https://doi.org/10.5267/j.ijiec.2013.10.002

[9] Qin, Y., Qi, Q., Scott, P.J., Jiang, X. (2020). An additive manufacturing process selection approach based on fuzzy Archimedean weighted power Bonferroni aggregation operators. Robotics and Computer-Integrated Manufacturing, 64, 101926. https://doi.org/10.1016/j.rcim.2019.101926

[10] Franceschini, F., Maisano, D. (2018). Classification of objects into quality categories in the presence of hierarchical decision-making agents. Accreditation and Quality Assurance, 23(1), 5-17. https://doi.org/10.1007/s00769-017-1291-7

[11] Gervasi, R., Mastrogiacomo, L., Maisano, D. A., Antonelli, D., Franceschini, F. (2022). A structured methodology to support human-robot collaboration configuration choice. Production Engineering, 16: 435-451. https://doi.org/10.1007/s11740-021-01088-6

[12] Gibbons, J.D., Chakraborti, S. (2010). Nonparametric statistical inference (5th ed.). CRC Press, Boca Raton, ISBN 978-1420077612. https://doi.org/10.1201/9781439896129

[13] Kendall, M. G. (1962). Ranks and measures. Biometrika, 49(1/2), 133-137. https://doi.org/10.1093/biomet/49.1-2.133

[14] Agresti, A. (2010). Analysis of Ordinal Categorical Data (2nd ed.), New York: John Wiley & Sons. https://doi.org/10.1002/9780470594001

[15] Ross, S.M. (2009). Introduction to probability and statistics for engineers and scientists. Academic Press. https://doi.org/10.1016/B978-0-12-370483-2.00006-0

[16] Figueira, J., Greco, S., Ehrgott, M. (2005) Multiple criteria decision analysis: state of the art surveys. Springer, New York. https://doi.org/10.1007/b100605

[17] Borda, J.C. (1781). Mémoire sur les élections au scrutin, Comptes Rendus de l'Académie des Sciences. Translated by Alfred de Grazia as Mathematical derivation of an election system, Isis, 44:42-51. https://doi.org/10.1086/348187

[18] McComb C., Goucher-Lambert K., Cagan J. (2017), Impossible by design? Fairness, strategy and Arrow's impossibility theorem. Design Science, 3:1-26. https://doi.org/10.1017/dsj.2017.1

[19] Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. Reliability engineering & system safety, 91(10-11), 1109-1125. https://doi.org/10.1016/j.ress.2005.11.014