Consistency analysis in quality classification problems with multiple rank-ordered agents

Consistency analysis in quality classification problems with multiple rank-ordered agents

Fiorenzo Franceschini[1] and Domenico Maisano[2]

[1] *fiorenzo.franceschini@polito.it* [2] *domenico.maisano@polito.it*
Politecnico di Torino, DIGEP (Department of Management and Production Engineering),
Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

Abstract

A relatively diffused quality decision problem is that of classifying some objects of interest into predetermined nominal categories. This problem is particularly interesting in the case: (i) multiple agents perform local classifications of an object, to be fused into a global classification, (ii) there is more than one object to be classified, and (iii) agents may have different positions of power, expressed in the form of an importance rank-ordering. Due to the specificity of the problem, the scientific literature encompasses a relatively small number of data fusion techniques.

For the fusion to be effective, the global classifications of the objects should be consistent with the agents' local classifications and their importance rank-ordering, which represent the input data.

The aim of this paper is to propose a set of indicators, which allow to check the degree of *consistency* between the global classification and the input data, from several perspectives, e.g., that of individual agents, individual objects, agents' importance rank-ordering, etc.. These indicators are independent from the fusion technique in use and applicable to a wide variety of practical contexts, such as problems in which some of the local classifications are uncertain or incomplete.

The proposed indicators are simple, intuitive and practical for comparing the results obtained through different techniques. The description therein is supported by several practical examples.

Keywords: Quality classification problem, Quality inspection, Decision making, Surface-defect classification, Nominal scale, Rank-ordered agents, Fusion technique, Consistency, Agent agreement.

Introduction

When using nominal scales, it is often required to classify some *objects* of interest (o_1, o_2, o_3, etc.) into predetermined scale categories (c_1, c_2, c_3, etc.). By an "object" we will consider a specific feature/attribute of an entity observed; for example, a morphological characteristic of biological species (such as skin and eye colour) or the marital status of individuals (e.g., single, married, divorced, widowed, civil union, etc.). This operation – referred to as *quality classification* problem (Léger and Martel, 2002; Zopounidis and Doumpos, 2002; Bashkansky and Gadrich, 2008; Van Wieringen and De Mast, 2008) – is not trivial, when objective and incontrovertible rules for driving it are not available (Duffuaa and Khan, 2005; Bashkansky et al. 2007; Franceschini et al. 2007; See, 2012). For example, the classification of surface defects in Manufacturing or that of reference

materials in Materials Science are *subjective*, as they may change from subject to subject (Mevik and Næs 2002; Franceschini, Galetto, and Varetto, 2004; Mandroli and al. 2006).

The quality classification problem can be formulated in several forms. A popular formulation is that in which: (i) each category is defined *a priori* and characterized by one or more typical objects, also known as *reference objects* or *prototypes*; (ii) a set of criteria are used for comparing the object of interest with the prototype(s) of each quality category; (iii) the most plausible category for the object of interest is the one minimizing a suitable dissimilarity measure (Van Wieringen and De Mast, 2008; Agresti, 2013; Creswell, 2013; Bress, 2016; Steiner et al. 2016).

In this paper, we consider a different formulation of the problem – formalized in a recent paper by the authors (Franceschini and Maisano, 2016) – in which we assume that multiple *decision-making agents* have to classify several objects of interest into plausible scale categories. By a "decision-making agent" we will consider any of a wide variety of subjects; examples could be human beings, individual criteria in a multi-criteria decision process, etc.. Precisely, (i) each agent performs a (subjective) *local* classification of an object, selecting the most plausible category (e.g., agent d_1 classifies object o_1 into category c_2, d_2 classifies it into c_3, etc.), and (ii) the agents' local classifications are fused into a *global* (or *fused*) one. Other important features of this formulation are that:

1. *Each agent may express just a single preference in favour of the category(ies) that he/she considers to be most plausible*. This feature makes the assignment process easier for agents: selecting a single category is more practical than formulating, for instance, a preference ordering of the categories (e.g., $(c_1 \sim c_2) > c_3 > \ldots$), as these quality categories are often mutually exclusive and/or inconsistent with each other; e.g., considering the problem of classifying biological species into the categories *bacteria*, *protozoa*, *chromista*, *plantae*, *fungi*, *animalia*, it seems unreasonable to formulate preference orderings of the categories.
2. *In the case of hesitation, one agent may (i) refrain from expressing his/her preference or (ii) fractionalize his/her (single) preference between two or more (tied) categories* (e.g., agent d_1 may classify object o_1 into the two categories c_2 and c_3, or even decide not to classify it). In this way, agents are not forced to dubious (local) classifications in uncertain situations.
3. *There is a hierarchy of importance of agents, expressed through a linear ordering*, like $d_1 > (d_2 \sim d_3) > \ldots$, where symbols ">" and "~" depict the "strict preference" and "indifference" relationship respectively (Nederpelt and Kamareddine, 2004). This feature makes the problem more general, since agents should not necessarily be equi-important. Also, the formulation of the agents' importance hierarchy through a rank-ordering is easier (for the analysts) than that through a set of weights (Cook, 2006). In fact, although the literature provides several techniques for guiding weight quantification – for example, the AHP procedure (Saaty, 1980; Ramanathan and Ganesh, 1994) or the method proposed in Wang et al. (2014) – they are often neglected in

practice, probably because of their complexity or the (strong) hypotheses behind their use. As a result, weights are not rarely assigned in arbitrary and questionable ways.

4. *It is assumed that the global classification of an (i-th) object should necessarily consist of one-and-only-one category*, as this is the ultimate goal of the quality classification problem.

Despite its relative simplicity and practicality, the afore-described quality classification problem has been little studied and the state of the art essentially includes two fusion techniques: the first one simply uses the *mode* operator while the second, recently proposed by Franceschini and Maisano (2016), is based on the idea that the winning quality category is the one that reaches a threshold first, during a gradual voting process based on the agents' importance rank-ordering. We will return to these techniques later.

The afore-mentioned fusion techniques, and maybe those that will be proposed in the future, certainly have their *pro* and *contra*; e.g., the first technique is certainly simpler but also rougher than the second one. An interesting question is: *For a generic quality classification problem with rank-ordered agents, how could we identify the best fusion technique?* We are aware that it is a very tricky question, since (i) it is difficult to pinpoint the concept of "best" fusion technique and (ii) the "true" solution to a generic problem is not known *a* priori (Zopounidis and Doumpos, 2002; Figueira et al., 2005; Cook, 2006). Nevertheless, the performance of different fusion techniques may be assessed, at least roughly, according to various aspects, such as:

- The ability to produce a classification that is consistent with the input data;
- The adaptability to a variety of input data (e.g., tied or incomplete local classifications, in the case of agents' hesitation);
- Computational complexity.

Among these aspects, the one concerning the consistency of the classification is particularly important and can be decomposed into two dimensions:

A) *Type-A consistency, i.e., the ability of a solution to reflect the agents' local classifications*;

B) *Type-B consistency, i.e., the ability of a solution to reflect the agents' importance hierarchy*, based on the idea that the more important agents should have a predominant influence on the solution.

The goal of this paper is to provide a practical set of indicators to quantify the consistency of global solutions, taking into account both the above dimensions. A not-so-dissimilar set of indicators was proposed for a different decision-making problem, concerning the fusion of multi-agent preference orderings into a single consensus ordering (Franceschini and Maisano, 2015).

The consistency verification can be performed at different aggregation levels (e.g., at the level of individual objects, individual agents, etc.). Also, some of the proposed indicators allow to depict the so-called level of *agent agreement* (Viera and Garrett, 2005).

The remainder of the paper is organized into four sections. The section "Related work" contains a

brief literature review of the quality classification problems in the field of multi-criteria decision aid. The section "Description of the indicators" presents a detailed description of the proposed indicators, focusing on their construction, aggregation, and practical use. The description is supported by a realistic example in the manufacturing field. The section "Active use of the proposed indicators" reverses the perspective, interpreting the proposed indicators not only as *passive* tools, to check the solution provided by a certain fusion technique, but also as *active* tools to identify the most plausible solution to a quality classification problem. Finally, the section "Discussion" summarizes the original contributions of this paper, highlighting its practical implications, limitations and suggestions for future research.

Related work

The scientific literature encompasses a variety of techniques for supporting the classification of objects into nominal scale categories. These techniques generally depend on (i) the specific formulation of the classification problem, (ii) the initial data available, and (iii) the requirements related to the solution (Yevseyeva, 2007). The majority of these techniques have been developed in the area of *multi-criteria decision aid* (MCDA) and almost exclusively apply to classification problems based on the use of sets of prototypes for the categories.

For instance, the one proposed by Perny (1998), denominated *multi-criteria filtering* (MCF), is based on the *concordance* and *non-discordance* principles that have been first used on the ELECTRE methods (Roy, 1968). The method proposed by Goletsis et al. (2004), denominated *gMCDA classifier*, implements a similar scheme, with less control parameters to be adjusted. Another similar MCDA classification method – denominated *PROAFTN* (Belacel, 2000) – enables to determine the fuzzy indifference relations by generalising the concordance and discordance indices used in the ELECTRE III method.

Despite the abundance of MCDA classification techniques, the literature includes little research on other aspects, such as (i) the analysis of the interdependencies of the control parameters, (ii) their statistical validation, (iii) comparisons and applications of different techniques MCDA methods over the same datasets, and (iv) the establishment of links between these techniques and those coming from related disciplines, such as Pattern Recognition, Machine Learning, Data Mining, etc. (Witten and Frank, 2005).

A "rare bird" is represented by the tool developed by Brasil Fihlo et al., (2009), which allows to compare the effectiveness of different classification techniques, based on a customized genetic algorithm to calibrate their control parameters automatically, under some different sets of prototypes. Unfortunately, this tool cannot be applied to techniques designed for the classification problem formalized by Franceschini and Maisano (2016), i.e., in which (i) multiple agents perform local classifications of an object, to be fused into a global classification, (ii) there is more than one

object to be classified, and (iii) there is an importance rank-ordering of agents.

For this particular classification problem, the different state-of-the-art classification (or fusion) techniques are only two:

- The first one is quite trivial since the "winning" category is selected through the *mode* operator, which identifies the category with the largest number of preferences (even fractionalized). In the example in Tab. 1, the mode corresponds to c_2 as this category collects a number of preferences (i.e., 2) higher than the other ones. We note that the preference by agent d_3 is fractionalized with respect to the two (tied) categories c_1 and c_3 (which obtain 0.5 preferences each). The total number of agents participating in this classification process is 4, since – among the 5 initial agents (i.e., d_1 to d_5) – d_4 is unable to classify the object of interest; the total score, obtained cumulating the agents' preferences, is obviously 4. We remark that the selection technique based on the mode ignores the agents' importance hierarchy, i.e., $d_1 > (d_2 \sim d_3) > d_4 > d_5$ in the case exemplified.

Tab. 1. (a) Hypothetical multi-agent classification of an object (o_i), considering four nominal categories (c_1 to c_4). (b) Selection of the most plausible category through the mode of the agents' local classifications (i.e., c_2).

	(a) Classification of an object					(b) Agents' preferences			
	d_1	d_2	d_3	d_4	d_5	c_1	c_2	c_3	c_4
1. Agents' local classifications	c_1 (1)	c_2 (1)	c_1, c_3 (0.5, 0.5)	- N/A	c_2 (1)	1.5	2	0.5	0
2. Agents' importance rank-ordering: $d_1 > (d_2 \sim d_3) > d_4 > d_5$									

- The second fusion technique, recently proposed by Franceschini and Maisano (2016), is based on the following idea: the winning category is the one that reaches a threshold (t) first, during a gradual voting process based on the agents' importance rank-ordering. Regarding the problem in Tab. 1(a), the voting process is organized into four turns, which correspond to the "blocks" of agents with indifferent importance (i.e., turn 1 including d_1, turn 2 including d_2 and d_3, due to their indifferent importance, turn 3 including d_4, and turn 4 including d_5). In each turn, the categories gradually cumulate a score corresponding to the preferences expressed by the agents. We note that turn 3 is "dull" as agent d_4 refrains from classifying the object of interest.

  According to this fusion technique, c_1 is selected, as its cumulative score reaches t first (i.e., in turn 2). The parameter t was set to 1.5; for details, see (Franceschini and Maisano, 2016). This fusion technique favours c_1 as this category is the preferred one, according to the more important agents (see Tab. 2).

  The indicators presented in the remainder of the paper represent an important novelty for the state of the art, as they allow to check the consistency of the solution obtained by a specific fusion technique and/or to compare the solutions obtained by different techniques.

Tab. 2. Selection of the most plausible category applying the fusion technique by Franceschini and Maisano (2016) to the problem in Tab. 1(a). The selection results in c_1.

	(a) Agents' turn-by-turn preferences				(b) Cumulative score			
	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
Turn 1 (d_1)	1	0	0	0	1	0	0	0
Turn 2 ($d_2 \sim d_3$)	0.5	1	0.5	0	1.5	1	0.5	0
Turn 3 (d_4)	-	-	-	-	1.5	1	0.5	0
Turn 4 (d_5)	0	1	0	1	1.5	2	0.5	0

$t = (1+J_i/2)/2 = 1.5$, where $J_i = 4$ is the number of voting agents, excluding those unable to classify the object (o_i) of interest, i.e., d_3 in this case. Details on the rationale behind this formula are presented in (Franceschini and Maisano, 2016).

Description of the indicators

This section illustrates the proposed indicators in detail. In order to make the description effective and clarify the notation in use, let us consider a realistic problem concerning the classification of defects on the surface of hot-rolled steel plates by visual inspection. In this context, the correct quality recognition and classification can provide effective information for product optimization. We hypothesize that $I = 5$ defects (i.e. the o_i objects of the problem, being $i = 1, 2, \ldots, I$) should be classified by $J = 4$ operators (i.e., the d_j agents of the problem, being $j = 1, 2, \ldots, J$) into $K = 3$ nominal categories (c_k, being $k = 1, 2, \ldots, K$), corresponding to some of the most frequent defect types for the manufacturing process of interest (for details, see (Ai and Xu, 2013)):

(c_1) *crack*, i.e, longitudinal or transverse fractures, which are along or perpendicular to the rolling direction respectively;

(c_2) *scale*, i.e., defects in the form of fish scales, strips or dots;

(c_3) *chap*, i.e., defects shaped like a turtleback or a network of cracks.

Operators/agents are divided into three classes of competence (i.e., I, II, and III, in decreasing order), based on a combination of two attributes: (i) their work experience (e.g. number of years of service), and (ii) their level of professional qualification (e.g., worker, team leader, foreman, etc.). A team of experts selected these two attributes, as they may significantly influence the accuracy of the response while being relatively easy to evaluate. The resulting agents' importance rank-ordering is $d_1 > (d_2 \sim d_4) > d_3$.

Each agent performs a local classification of each of the five objects (see Tab. 3(a)). Through some fusion technique (no matter what), these *local* classifications are merged into *global* ones, which represent the solution to the problem (see Tab. 3(b)).

Tab. 3. (a) Classification problem in which 5 objects (o_i) are classified by 4 agents (d_j) into 3 quality categories (c_k). (b) Possible solution to the problem, obtained through no-matter-what fusion technique.

	(a) Input data					(b) Solution
Agents		d_1	d_2	d_3	d_4	Global classification
Competence class		I	II	III	II	concerning o_i
1. Local classifications concerning o_i	o_1	c_1	c_2	c_2	c_1, c_3	c_1
	o_2	c_2	c_2, c_3	c_2	-	c_2
	o_3	c_3	c_2	c_1, c_2, c_3	c_3	c_2
	o_4	c_1, c_2	c_2	c_1	c_1, c_3	c_1
	o_5	c_3	c_2, c_3	c_2, c_3	-	c_3
2. Agents' importance rank-ordering: $d_1 > (d_2 \sim d_4) > d_3$.						

In the case of hesitation, agents are not forced to formulate dubious classifications; in fact, they may (i) refrain from performing some of the local classifications (e.g., see those characterized by the symbol "-") or (ii) make multiple assignments (e.g., the d_4's local classification of o_1 is "c_1, c_3").

The rest of the following section is divided into three subsections, which describe respectively:

- The proposed indicators to assess the type-A consistency of the solution;
- The proposed indicators to assess the type-B consistency of the solution;
- Aggregation of (some of) the previous typologies of indicators into an overall consistency indicator.

Type-A consistency (level-2 title)

The type-A consistency can be assessed at the level of (i) an individual *i*-th object, and (ii) the totality of the objects, as described in the following two subsections.

Level of an individual object: $p_{i\bullet}^{A}$ indicators (level-3 title)

For each *i*-th object (o_i), the local quality classification by each *j*-th agent (d_j) is transformed into a score (A_{ij}), representing the type-A consistency with respect to the global classification. Tab. 4 reports the possible scores, hereafter denominated as A_{ij}-scores.

Tab. 4. Scores used for evaluating the type-A consistency of a local classification with respect to the global one.

Case	Example		A_{ij}-score
	Local classif.	Global classif.	
1. The category in the global classification and the single one in the local classification are coincident.	c_1	c_1	1
2. The category in the global classification is included among the (l) multiple categories in the local classification.	c_1, c_2, c_3	c_1	$1/l$ (i.e., 1/3 in the case exemplified)
3. The category in the global classification is different from the category(ies) in the local classification.	c_2, c_3	c_1	0
4. The agent is unable to classify the object of interest.	-	c_1	N/A

These scores can be interpreted as relative frequencies of the (single) category in the global classification, with respect to the category(ies) in the local classification: the score is 1 in the case of one-to-one correspondence, $1/l$ in the case of partial matching among the (l) multiple categories, 0 in the case of no matching, and N/A (not applicable) in the case one agent is unable to classify the object of interest.

Next, for each *i*-th object, it is possible to construct a vector of the A_{ij}-scores related to the agents' involved in the classification problem. For example, Tab. 5 reports the vector of the A_{ij}-scores concerning o_1, for the problem in Tab. 3, and the average score ($p_{1\bullet}^{A} = 37.5\%$) related to the object o_1. Considering a generic *i*-th object, $p_{i\bullet}^{A}$ is defined as:

$$p_{i\bullet}^{A} = \frac{\sum_{j} A_{ij}}{J_i}, \tag{1}$$

where J_i is the number of agents able to classify o_i (e.g., 4 for o_1).

The indicator $p_{i\bullet}^{A}$ is a real number included between 0 and 1: the higher the value, the greater the type-A consistency. The superscript "*A*" indicates that this indicator represents the type-A consistency, while the subscript "*i*•" denotes that it is determined by considering all of the usable A_{ij}-scores for the *i*-th object, i.e., those different from "N/A" (see Tab. 4).

Tab. 5. Vector of the scores representing the type-A consistency for the quality classification of o_1, in the problem in Tab. 3.

Agents	d_1	d_2	d_3	d_4
Local classifications	c_1	c_2	c_2	c_1, c_3
A_{1j}-scores	1	0	0	0.5
$p_{1\bullet}^{A} = \left(\sum_{j} A_{1j}\right) \Big/ J_1 = 1.5 / 4 = 37.5\%$				

The global classification is c_1;
The agents' importance rank-ordering is $d_1 > (d_2 \sim d_4) > d_3$;
$J_1 = 4$ is the total number of agents involved in the classification process of o_1;
$p_{1\bullet}^{A}$ depicts the type-A consistency related to the classification of o_1.

Among the possible quality categories in the global classification, the one maximizing $p_{i\bullet}^{A}$ obviously corresponds to the mode: being $p_{i\bullet}^{A}$ proportional to the sum of the relative frequencies of the categories (i.e., the A_{ij}-scores), it will be maximized by the category with the maximum relative frequency, i.e., the mode itself. We remark that $p_{i\bullet}^{A} = 1$ denotes the ideal case of full agreement among agents, in which the local classifications are all coincident with the global one.

The $p_{i\bullet}^{A}$ construction can be extended to the totality of the objects. Tab. 6 illustrates the A_{ij}-scores concerning the solution to the problem exemplified (see Tab. 3(b)).

Tab. 6. Table of the indicators of type-A consistency related to the problem in Tab. 3, at the level of the individual objects and the totality of the objects.

	A_{ij}-scores						
o_i	d_1	d_2	d_3	d_4	J_i	$\sum_{j} A_{ij}$	$p_{i\bullet}^{A}$
o_1	1	0	0	0.5	4	1.5	37.5%
o_2	1	0.5	1	N/A	3	2.5	83.3%
o_3	0	1	1/3	0	4	1.3	32.5%
o_4	0.5	0	1	0.5	4	2.0	50.0%
o_5	1	0.5	0.5	N/A	3	2.0	66.7%
I_j	5	5	5	3	$\sum_{j} I_j = \sum_{i} J_i = 18$		
$\sum_{i} A_{ij}$	3.5	2.0	2.8	1.0		$\sum_{i}\sum_{j} A_{ij} = 9.3$	
$p_{\bullet j}^{A}$	70.0%	40.0%	56.7%	33.3%			$p^{A} = 51.9\%$

J_i number of agents able to classify the *i*-th object;
I_j number of objects that the *j*-th agent is able to classify;
A_{ij} score expressing the type-A consistency between the global classification and the local one by the *j*-th agent, with respect to the *i*-th object;
$p_{i\bullet}^{A}$, $p_{\bullet j}^{A}$ and p^{A} are defined in Eqs. 1, 3 and 2 respectively.

Level of the totality of the objects: p^A and $p^A_{\bullet j}$ indicators (level-3 title)

An overall indicator of the type-A consistency, at the level of the totality of the objects, is defined as the weighted sum of the $p^A_{i\bullet}$ values with respect to the relevant J_i values, i.e., the number of agents able to classify o_i:

$$p^A = \frac{\sum_i p^A_{i\bullet} \cdot J_i}{\sum_i J_i} = \frac{\sum_i \sum_j A_{ij}}{\sum_i J_i}. \quad (2)$$

Eq. 2 shows that p^A can also be interpreted as the average of the A_{ij}-scores (excluding the "N/A" contributions). Among the possible quality categories in the solution to a classification problem, those maximizing p^A correspond to the mode values of the local classifications. A proof is that, since the mode relating to an individual i-th object maximizes $\sum_j A_{ij}$, the mode values of the totality of the objects will maximize $\sum_i \sum_j A_{ij}$, which is proportional to p^A (see Eq. 2).

Let us now define another type-A consistency indicator (i.e., $p^A_{\bullet j}$), which is somehow akin to $p^A_{i\bullet}$, being defined as the average of the A_{ij}-scores related to the local classifications by a certain j-th agent:

$$p^A_{\bullet j} = \frac{\sum_i A_{ij}}{I_J}, \quad (3)$$

where the subscript "•j" denotes that this indicator is determined considering all of the usable A_{ij}-scores related to the (j-th) agent of the problem, i.e., those different from "N/A".

Likewise $p^A_{i\bullet}$, $p^A_{\bullet j}$ is a real number $\in [0, 1]$. In the above example, we note that – among the 4 agents of interest – those formulating the more consistent local classifications are d_1 and d_3, with $p^A_{\bullet 1} = 70.0\%$ and $p^A_{\bullet 3} = 56.7\%$ respectively (see the bottom of Tab. 6).

Returning to the overall indicator p^A, it can also be interpreted as a weighted sum of the $p^A_{\bullet j}$ values with respect to I_j, i.e., the number of objects that the j-th agent is able to classify:

$$p^A = \frac{\sum_j p^A_{\bullet j} \cdot I_j}{\sum_j I_j}. \quad (4)$$

We remark that, since $p^A_{i\bullet}$, $p^A_{\bullet j}$, and p^A depict the type-A consistency, they do not take into account the agents' importance rank-ordering. This limitation can be overcome by introducing the complementary indicators described in the next subsection.

Type-B consistency (level-2 title)

Likewise type-A consistency, the type-B one can be assessed at the level of (i) an individual *i*-th object and (ii) the totality of the objects, as described in the following subsections.

Level of an individual object: $p_{i\bullet}^{B}$ indicators (level-3 title)

The $p_{i\bullet}^{B}$ indicators rely on the idea that the more important agents should have a predominant influence on the determination of the solution, resulting in higher A_{ij}-scores. Following this idea, it seems reasonable to compare the importance rank-ordering of the agents (i.e., $d_1 > (d_2 \sim d_4) > d_3$, in the case exemplified) with the ordering based on the relevant A_{ij}-scores, for an *i*-th object of interest. E.g., considering o_1, we have the following ordering: $(A_{11}=1) > (A_{14}=0.5) > (A_{12}=0) \sim (A_{13}=0)$, therefore the corresponding agents' ordering will be $d_1 > d_4 > (d_2 \sim d_3)$. For simplicity, the agents' ordering resulting from the comparison of the A_{ij}-scores of an *i*-th object will be hereafter denominated as "A_{ij}-ordering".

The comparison between the agents' importance rank-ordering and the A_{ij}-ordering is carried out in two steps:

1. *Decomposition of the two orderings into a number of paired-comparison relationships* (e.g., in the case $(d_1 > d_2)$, the paired comparison (d_1, d_2) will result in the relationships ">". There are 4 types of possible relationships, characterized by the following symbols:

   ">" Strict preference in favour of the first agent in the paired comparison;

   "<" Strict preference in favour of the second agent in the paired comparison;

   "~" Indifference between the two agents;

   "N/A" not applicable, as the paired comparison is not defined in the ordering of interest.

   Tab.7 shows the paired-comparison relationships obtained from the A_{ij}-ordering of o_1 and the agents' importance rank-ordering.

Tab. 7. Paired-comparison relationships obtained from the A_{ij}-ordering of o_1 and the agents' importance rank-ordering.

Ordering	Paired comparison relationships					
	(d_1, d_2)	(d_1, d_3)	(d_1, d_4)	(d_2, d_3)	(d_2, d_4)	(d_3, d_4)
A_{ij}-ordering of o_1: $d_1 > d_4 > (d_2 \sim d_3)$	>	>	>	~	<	<
Agents' importance rank-ordering: $d_1 > (d_2 \sim d_4) > d_3$	>	>	>	>	~	<

2. *Comparison of the paired-comparison relationships and assignment of B_{il}-scores*, according to the conventions in Tab.8. The subscript "*il*" indicates that a B_{il}-score is associated with each *i*-th object and each *l*-th paired comparison; precisely, for each *i*-th object, *l* is a natural number $\in [1, L_i]$, L_i being the number of usable paired-comparison relationships (i.e., different from "N/A"). Tab.8 also contains a definition of the four cases of *full consistency*, *weak consistency*, *inconsistency* and *incomparability*, which are associated with four different possible scores.

The conventional choice of assigning 0.5 points in the case of weak consistency is justified by the fact that this is the intermediate case between that one of full consistency (with score 1) and that of inconsistency (with score 0). We are aware that this choice, although reasonable, is inevitably arbitrary; however, we will return to this point later, showing that the proposed indicators are relatively robust with respect to small variations in these scores (see the section "Sensitivity analysis", in the appendix).

Tab. 8. Conventional scores used for evaluating the type-B consistency, when comparing the paired-comparison relationships resulting from the A_{ij}-orderings (or $p^{A}_{\bullet j}$-ordering) with those resulting from the agents' importance rank-ordering.

Case	Score
1. *Full consistency*, i.e., the two paired-comparison relationships are identical (both ">" or "<" or "~").	1
2. *Weak consistency*, i.e., one of the two paired-comparison relationships is of indifference "~", while the other one is of strict preference (">" or "<"). In other words, the two paired-comparison relationships are consistent with respect to the weak-preference relationship "≤" or "≥"; e.g., when comparing the relationship $d_1 > d_2$ with $d_1 \sim d_2$.	0.5
3. *Inconsistency*, i.e., the two paired-comparison relationships are of opposite strict preference (">" and "<", or "<" and ">"); e.g., when comparing the relationship $d_1 > d_2$ with $d_1 < d_2$.	0
4. *Incomparability*, i.e., at least one of the two paired-comparison relationships is "N/A".	N/A

The type-B consistency vector in Tab.9 contains the B_{il}-scores resulting from the comparison of the paired-comparison relationships in Tab.7.

Tab.9 Vector of the B_{il}-scores of o_1, based on the paired-comparison relationship in Tab.7.

o_i	(d_1, d_2)	(d_1, d_3)	(d_1, d_4)	(d_2, d_3)	(d_2, d_4)	(d_3, d_4)		
o_1	1	1	1	0.5	0.5	1	$\sum_l B_{1l} = 5$	$p^{B}_{1\bullet} = 83.3\%$

Considering a generic *i*-th object, we define the aggregated indicator of type-B consistency:

$$p^{B}_{i\bullet} = \frac{\sum_l B_{il}}{L_i}, \tag{5}$$

where

the subscript "$i\bullet$" denotes that all the usable B_{il}-scores related to o_i are averaged;

$\sum_l B_{il}$ is the total score related to the object o_i;

L_i is the number of usable paired-comparison relationships, i.e.:

$$L_i = \binom{J_i}{2} = J_i \cdot (J_i - 1)/2, \tag{6}$$

J_i being the number of agents able to classify o_i. Returning to the A_{ij}-ordering relating to o_i, exemplified in Tab.7, $J_1 = 4$ and therefore $L_1 = 6$.

Similarly to other existing indicators – e.g., the Kendall's tau (τ) and the Spearman's rho (ρ) – $p^{B}_{i\bullet}$ can be interpreted as an indicator of correlation between pairs of orderings (Kendall, 1970;

Spearman, 1904; Montgomery, 2013). Curiously, in the case of orderings with strict-preference relationships only, τ and $p^B_{1\bullet}$ are linearly related, as shown below:

$$\tau = \frac{\text{no. of concordant pairs} - \text{no. of discordant pairs}}{\text{total no. of pair combinations}} = \frac{\sum_l B_{il} - \left(L_i - \sum_l B_{il} \right)}{L_i} = 2 \cdot p^B_{i\bullet} - 1 . \qquad (7)$$

In this sense, $p^B_{i\bullet}$ can be considered as a variant of τ. The decision to adopt $p^B_{i\bullet}$ is motivated by two reasons:

1. its range – i.e., [0, 1] – is compatible with that of the so-far-defined indicators;
2. it can be easily calculated even if the orderings (i) are incomplete or (ii) contain some indifference relationships.

Level of the totality of the objects: the p^B indicator (level-3 title)

The calculation of the type-B consistency vector and the relevant $p^B_{i\bullet}$ value can obviously be extended to the totality of the objects. Tab.10 and Tab.11 respectively show the paired-comparison relationships and the consistency table (i.e., the table collecting the consistency vectors) relating to the totality of the objects, for the problem exemplified.

Tab.10 shows the paired-comparison relationships obtained from (i) the A_{ij}-orderings of the objects of interest, (ii) the $p^A_{\bullet j}$-ordering (the practical use of this other ordering will be clarified later), and (iii) the agents' importance rank-ordering. Regarding the A_{ij}-orderings, we remark that the number (L_i) of usable paired-comparison relationships can vary from object to object (see the last column). For example, for o_2 and o_5, $L_2 = L_5 = \binom{3}{2} = 3$, since agent d_4 does not formulate any local classification of these objects.

Tab.11 shows that the local/global classifications concerning o_1 denote a relatively high type-B consistency ($p^B_{1\bullet} = 83.3\%$), while this consistency is significantly lower for the other objects.

Tab.10. Paired-comparison relationships obtained from the A_{ij}-orderings of the individual objects, the $p^A_{\bullet j}$-ordering and the agents' importance rank-ordering.

Ordering(s)			Paired-comparison relationships						
			(d_1, d_2)	(d_1, d_3)	(d_1, d_4)	(d_2, d_3)	(d_2, d_4)	(d_3, d_4)	L_i
A_{ij}-orderings:	o_1	$d_1 > d_4 > (d_2 \sim d_3)$	>	>	>	~	<	<	6
	o_2	$(d_1 \sim d_3) > d_2$	>	~	N/A	<	N/A	N/A	3
	o_3	$d_2 > d_3 > (d_1 \sim d_4)$	<	<	~	>	>	>	6
	o_4	$d_3 > (d_1 \sim d_4) > d_2$	>	<	~	<	<	>	6
	o_5	$d_1 > (d_2 \sim d_3)$	>	>	N/A	~	N/A	N/A	3
$p^A_{\bullet j}$-ordering: $d_1 > d_3 > d_2 > d_4$			>	>	>	<	>	>	$L_\bullet$=6
Agents' importance rank-ordering: $d_1 > (d_2 \sim d_4) > d_3$			>	>	>	>	~	<	L=6

L_i is the number of usable paired-comparison relationships (i.e., different from "N/A") obtained from the A_{ij}-ordering related to o_i;
$L_\bullet$ is the number of usable paired-comparison relationships obtained from the $p^A_{\bullet j}$-ordering;
L is the number of usable paired-comparison relationships obtained from the agents' importance rank-ordering.

Tab.11. Consistency table relating to the paired-comparison relationships in Tab.7.

Types of scores		(d_1, d_2)	(d_1, d_3)	(d_1, d_4)	(d_2, d_3)	(d_2, d_4)	(d_3, d_4)	L_i	$\sum_l B_{il}$	$p^B_{i\bullet}$
	o_1	1	1	1	0.5	0.5	1	6	5	83.3%
	o_2	1	0.5	N/A	0	N/A	N/A	3	1.5	50.0%
B_{il}-scores	o_3	0	0	0.5	1	0.5	0	6	2	33.3%
	o_4	1	0	0.5	0	0.5	0	6	2	33.3%
	o_5	1	1	N/A	0.5	N/A	N/A	6	2.5	41.7%
$B_{\bullet l}$-scores		1	1	1	0	0.5	0	$L_\bullet = 6$	$\sum_l B_{\bullet l} = 3.5$	p^B=58.3%

B_{il}-scores are constructed by comparing the paired-comparison relationships obtained from the A_{ij}-ordering with those obtained from the agents' importance rank-ordering;

$B_{\bullet l}$-scores are constructed by comparing the paired-comparison relationships obtained from the $p^A_{\bullet j}$-ordering with those obtained from the agents' importance rank-ordering;

$L_\bullet$ is the number of usable paired-comparison relationships obtained from the $p^A_{\bullet j}$-ordering.

The information given by the $p^B_{i\bullet}$ indicators is quite fragmented, as it is based on data concerning individual objects. A more general indicator of type-B consistency can be constructed based on the comparison between the agents' importance rank-ordering and the $p^A_{\bullet j}$-ordering; in fact, $p^A_{\bullet j}$ indicators provide a type-A consistency evaluation based on the totality of the objects, not just one. In the case exemplified, the $p^A_{\bullet j}$-ordering is ($p^A_{\bullet 1}$=70.0%) > ($p^A_{\bullet 3}$=33.3%) > ($p^A_{\bullet 2}$=56.7%) > ($p^A_{\bullet 4}$=40.0%), therefore the resulting agents' ordering is $d_1 > d_3 > d_2 > d_4$ (see Tab. 6 and Tab.10). For simplicity, the agents' ordering resulting from the comparison of the $p^A_{\bullet j}$-scores will be hereafter denominated as "$p^A_{\bullet j}$-ordering". We can define a synthetic indicator (p^B), which, although being analogous to the afore-described $p^B_{i\bullet}$ indicators, has a richer information content:

$$p^B = \frac{\sum_l B_{\bullet l}}{L_\bullet}, \tag{8}$$

where

$B_{\bullet l}$ is the type-B consistency score, constructed by comparing the usable paired-comparison relationships obtained from the $p^A_{\bullet j}$-ordering with those obtained from the agents' importance rank-ordering, according to the conventions in Tab. 8;

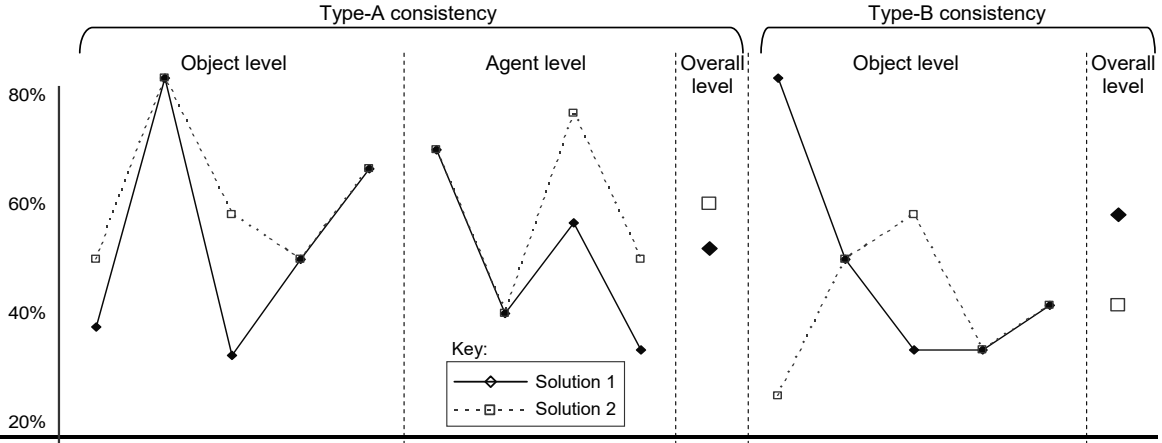
$L_\bullet$ is the number of usable paired-comparison relationship, obtained from the $p^A_{\bullet j}$-ordering.

The absence of subscript indicates that p^B takes into account the totality of the agents and the objects of interest (through the $p^A_{\bullet j}$ indicators). The information given by p^B is certainly more exhaustive than that given by the individual $p^B_{i\bullet}$ indicators.

Aggregation of p^A and p^B (level-2 title)

Among the indicators defined in Sects. 2.1 and 2.2, those with greater information content are p^A and p^B. These indicators are complementary as they respectively depict the overall type-A and type-B consistency.

The indicators p^A, p^B and those used for evaluating the consistency at a local level, i.e., $p^A_{i\bullet}$, $p^A_{\bullet j}$, and $p^B_{i\bullet}$, can be used to make a structured comparison between different solutions to the same quality classification problem. For example, Fig. 1 contains a comparison between two possible solutions to the problem exemplified, i.e., the so-far-examined solution (*solution 1*: c_1, c_2, c_2, c_1, c_3), and the one corresponding to the mode of the local classifications (*solution 2*: c_2, c_2, c_3, c_1, c_3).



Solution	$p^A_{1\bullet}$	$p^A_{2\bullet}$	$p^A_{3\bullet}$	$p^A_{4\bullet}$	$p^A_{5\bullet}$	$p^A_{\bullet 1}$	$p^A_{\bullet 2}$	$p^A_{\bullet 3}$	$p^A_{\bullet 4}$	p^A	$p^B_{1\bullet}$	$p^B_{2\bullet}$	$p^B_{3\bullet}$	$p^B_{4\bullet}$	$p^B_{5\bullet}$	p^B
1. c_1, c_2, c_3, c_1, c_3	37.5%	83.3%	32.5%	50.0%	66.7%	70.0%	40.0%	56.7%	33.3%	51.9%	83.3%	50.0%	33.3%	66.7%	41.7%	58.3%
2. c_2, c_2, c_3, c_1, c_3	50.0%	83.3%	58.3%	50.0%	66.7%	70.0%	40.0%	76.7%	50.0%	60.2%	25.0%	50.0%	58.3%	66.7%	41.7%	41.7%

Fig. 1. Structured comparison between two possible solutions to the problem exemplified, on the basis of the proposed indicators.

We note that the two solutions appear quite similar in terms of type-A consistency, as proved by the relatively close p^A values (i.e. 51.9% for solution 1 and 60.2% for solution 2). Regarding type-B consistency, solution 1 seems better than solution 2 (i.e., p^B values are 58.3% and 41.7% respectively).

It is worth remarking that the so-far-discussed indicators provide a consistency evaluation for a specific quality classification problem, not in absolute terms. E.g., in problems characterized by a very high level of *agent agreement*, global classifications are more likely to be consistent with the agents' local classifications and the relevant p^A values tend to be higher. For example, a solution with $p^A = 79\%$, in a classification problem with a relatively low agent agreement is not necessarily less consistent than a solution with $p^A = 80\%$, in a problem where the agent agreement is much higher.

Based on the above considerations, to obtain a more "absolute" information on the (type-A and type-B) consistency of a certain solution, it seems reasonable to compare the p^A and p^B indicators

with the highest values that they could achieve (i.e., $p^{A,\ Max}$ and $p^{B,\ Max}$) for a specific quality classification problem:

$$\begin{aligned} p^{A,\,Max} &= \underset{K^I \text{solutions}}{Max}\left(p^A\right) \\ p^{B,\,Max} &= \underset{K^I \text{solutions}}{Max}\left(p^B\right), \end{aligned} \tag{9}$$

in which K^I is the total number of possible solutions to the quality classification problem, K being the number of possible categories in which an object can be classified, and I being the total number of objects to be classified.

Finding $p^{A,\ Max}$ (and the corresponding solution(s)) is relatively simple, as – being p^A defined as the (weighted) sum of the $p_{i\bullet}^A$ values related to the individual objects (see Eq. 2) – it can be maximized by independently maximizing the individual $p_{i\bullet}^A$ values, through the mode of the local classifications. Returning to the above-exemplified problem, the solution maximizing p^A is therefore coincident with the solution 2 in Fig. 1, with a corresponding $p^{A,\ Max} = 60.2\%$. The relatively low value of this indicator denotes a relatively poor level of agreement in the agents' local classifications. In this sense, $p^{A,\ Max}$ can be used as a rough indicator of agent agreement, for the problem of interest.

The scientific literature encompasses other indicators aimed at evaluating agent agreement, such as the Cohen's *Kappa* and its variants (Agresti, 2013; Cohen, 1960; 1968). Even though these indicators are relatively simple and intuitive, they have some limitations, e.g., they are considered as overly conservative measure of agreement or they are based on the very unrealistic scenario, that – when not completely certain – agents simply guess; for details see (Strijbos, 2006; Pontius Jr and Millones, 2011). Also, the application of these indicators to the problem of interest can be complicated in the case local classifications are uncertain (e.g., with multiple tied categories) or incomplete.

Finding $p^{B,\ Max}$ is more complicated than finding $p^{A,\ Max}$. Among the possible solutions, it is not immediate to determine the one(s) maximizing p^B, as the problem cannot be decomposed at the level of individual objects. On the contrary, it can be shown that the solution maximizing the individual $p_{i\bullet}^B$ values is not necessarily that one maximizing p^B; see the proof in the section "Note on the p^B maximization", in the appendix. For the above reasons, a way to determine $p^{B,\ Max}$ is (i) generating all the possible (K^I) solutions to the classification problem and (ii) selecting that one(s) with the highest p^B value. Finding a more efficient way of determining the solutions with $p^{B,\ Max}$ is an open problem.

For the classification problem exemplified, ten out of the $K^I = 3^5 = 243$ possible solutions have $p^B = p^{B,\ Max} = 91.7\%$ (see Tab.12).

Despite the above ten solutions all maximize the type-B consistency, they do not necessarily perform well in terms of type-A consistency, as we can see from the generally low p^A values (see the penultimate column of Tab.12).

Tab.12. List of the 10 solutions to the classification problem exemplified, which maximize p^B.

No.	Solution					p^A	$p^B=p^{B,Max}$
	o_1	o_2	o_3	o_4	o_5		
1	c_1	c_1	c_1	c_2	c_1	18.5%	91.7%
2	c_1	c_1	c_2	c_3	c_3	29.6%	91.7%
3	c_1	c_1	c_3	c_2	c_3	40.7%	91.7%
4	c_1	c_2	c_2	c_3	c_1	32.4%	91.7%
5	c_1	c_2	c_3	c_2	c_1	43.5%	91.7%
6	c_1	c_2	c_3	c_2	c_2	49.1%	91.7%
7	c_3	c_2	c_3	c_2	c_3	54.6%	91.7%
8	c_3	3	c_3	c_3	c_3	26.9%	91.7%
9	c_3	c_2	c_3	c_2	c_3	43.5%	91.7%
10	c_3	c_2	c_3	c_2	c_3	49.1%	91.7%

To ease the comparison between the p^A and p^B values related to a certain solution and the corresponding $p^{A,Max}$ and $p^{B,Max}$ values, we can define the following two *normalized* indicators:

$$\begin{aligned} p^{A,Norm} &= p^A \big/ p^{A,\mathrm{Max}} \\ p^{B,Norm} &= p^B \big/ p^{B,\mathrm{Max}} \end{aligned} . \tag{10}$$

Obviously, these indicators reach the maximum value (i.e., 1) for solutions with $p^A = p^{A,Max}$ and $p^B = p^{B,Max}$ respectively. $p^{A,Norm}$ and $p^{B,Norm}$ make the type-A and type-B consistency evaluations more homogeneous and comparable with each other; in fact, although both p^A and $p^B \in [0, 1]$, we have noticed that the p^A values related to the solutions to a generic classification problem generally tend to be lower than the p^B values (see for example the distributions in Fig. 2), and therefore the direct comparison of these indicators would be inappropriate.

An alternative possible approach for normalizing p^A and p^B would be to replace them with their percentile number in the relevant distributions (e.g., for a p^A value in the 76th percentile of the distribution, $p^{A,Norm}$ would be 0.76). However, this normalization would "downgrade" the *cardinal* scales of p^A and p^B to *ordinal* ones, where the "distance" between values is not taken into account (Stevens, 1946).

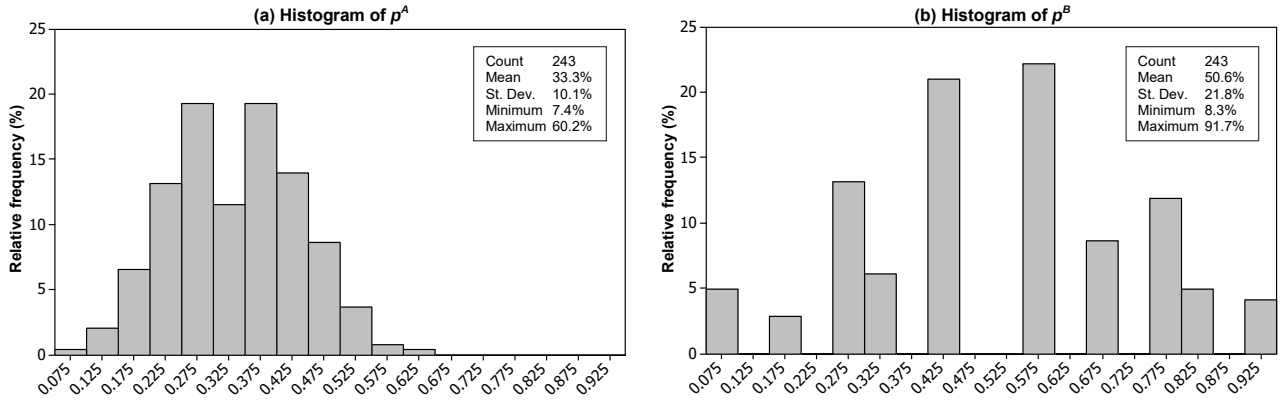


Fig. 2. Distributions of the p^A and p^B values related to the ($K^I = 3^5 = 243$) possible solutions to the classification problem exemplified.

It is interesting to notice that the $p^{A,\,Norm}$ and $p^{B,\,Norm}$ values are generally uncorrelated; see the map in Fig. 3, representing the "positioning" of the possible (K^I) solutions to the classification problem of interest. This lack of correlation corroborates the hypothesis that p^A and p^B (or $p^{A,\,Norm}$ and $p^{B,\,Norm}$) are complementary indicators. It was empirically checked that these considerations can be extended to other classification problems.

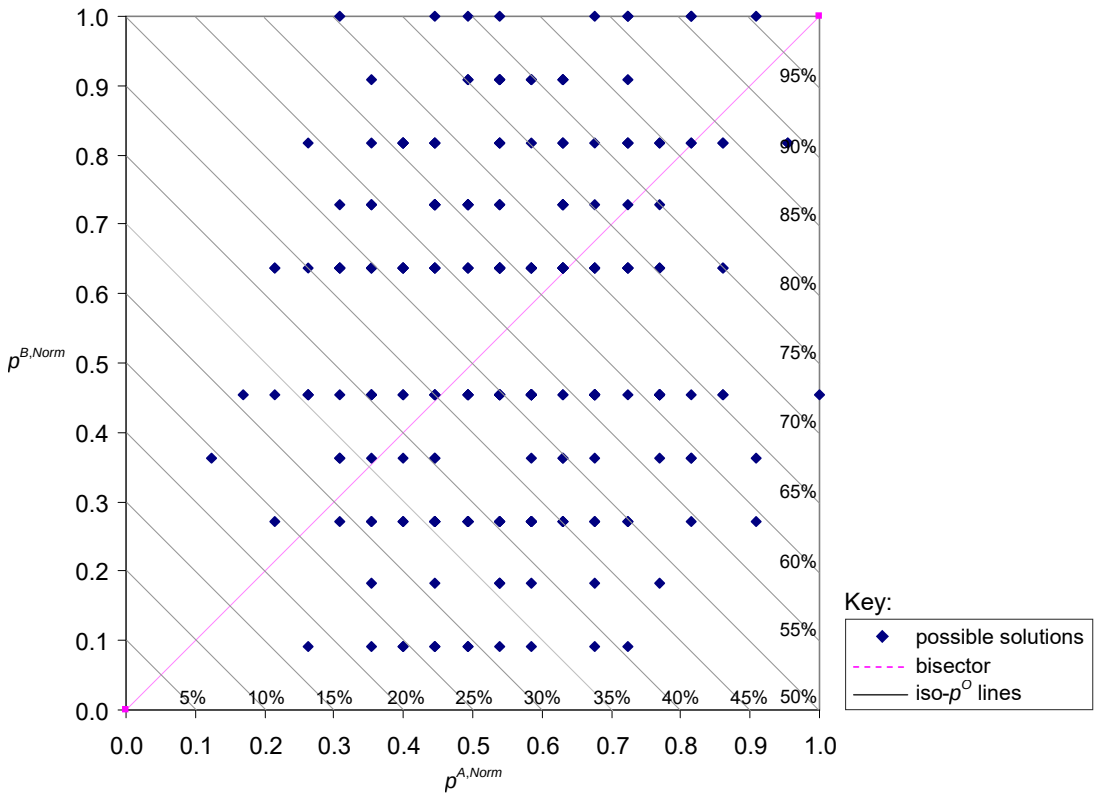


Fig. 3. $p^{A,Norm}$ – $p^{B,Norm}$ map concerning the quality classification problem in Tab. 2. The iso-p^O lines are perpendicular to the bisector of the first quadrant; the corresponding numerical values of p^O are reported along the bottom edge and the right edge of the map.

A synthetic indicator of overall consistency (p^O) can be obtained by averaging $p^{A,\,Norm}$ and $p^{B,\,Norm}$:

$$p^O = \frac{p^{A,Norm} + p^{B,Norm}}{2}. \tag{11}$$

The synthesis by the average value allows condensing the results of the consistency analysis into a single number. Although this choice seems practical and reasonable to us, we are aware that an analogous synthesis can be obtained in other ways, e.g., using the *min()* operator.

Considering the solution in Tab. 3(b), $p^{A,\,Norm} = 95.4\%$ and $p^{B,\,Norm} = 81.8\%$, therefore $p^O = 88.6\%$. This indicator, which $\in [0, 1]$, provides a synthetic estimate of the type-A and type-B consistency of the solution examined.

Having introduced a relatively large number of indicators at different aggregation levels, let us now focus on the tree diagram in Fig. 4, which summarizes their construction and practical use. The input data (i.e., agents' local classifications and importance rank-ordering) and output data (i.e., global classifications of the objects) of the quality classification problem are positioned at the

bottom of the diagram, while the overall consistency indicator (p^O) is positioned at the top vertex. The (sub-)indicators for assessing the type-A and type-B consistency are positioned on the left and right-hand side respectively. Interestingly, the diagram shows a certain asymmetry: while the type-A consistency indicators ignore the agents' importance rank-ordering, the type-B consistency indicators combine this information with that of other (sub-)indicators of type-A consistency (e.g., the $p^A_{\bullet j}$ values).

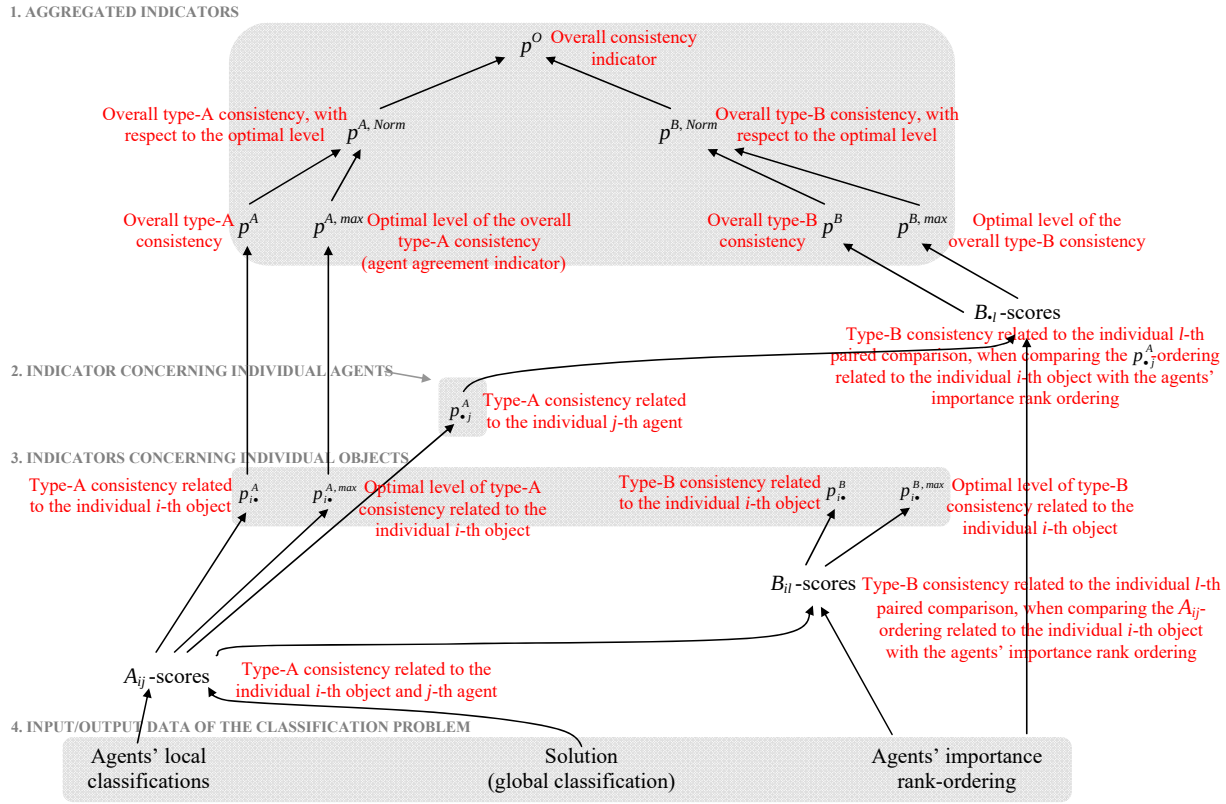


Fig. 4. Tree diagram of the proposed set of indicators, synthesizing their construction and practical use.

Active use of the proposed indicators

The indicators presented in Sects. 2.1, 2.2 and 2.3 have so far been interpreted as practical (*passive*) tools for assessing the consistency of the solutions to a specific quality classification problem. This paradigm can be reversed, interpreting these indicators as *active* tools for determining the optimal solution(s). The adjective "optimal" is related to the conventions adopted in the definition of the indicators. In other words, (some of) them can be used to define an *objective function*, to be maximized for determining the better solution(s) in terms of type-A and type-B consistency. The proposed indicators may therefore become the basic elements of a novel fusion technique for determining the solution to a specific quality classification problem, so as to enrich process diagnostic capabilities.

A way of defining such an objective function is through the synthetic indicator p^O. For the purpose of example, Tab.13 includes the top-30 solutions in terms of p^O, for the problem exemplified in the

section “Description of the indicators”. Interestingly, the afore-examined solution (with $p^O = 74.9\%$) is classified in 23rd position of the p^O-ranking, quite distanced from the optimal solution (with $p^O = p^{O,\,Max} = 95.4\%$); see also the graphical representation in Fig. 5. We are aware that this ranking depends on some potentially questionable assumptions in the construction of the indicators in use, such as: (i) definition of p^A and p^B, (ii) normalization mechanism for obtaining $p^{A,\,Norm}$ and $p^{B,\,Norm}$, and (iii) aggregation mechanism for determining p^O. Nevertheless, it seems reasonable to assume that the better solutions (in terms of type-A and type-B consistency) are those with relatively high p^O values. The section “Sensitivity analysis” (in the appendix) contains a sensitivity analysis, showing that p^O is relatively robust with respect to small variations in the B_{il} and $B_{\bullet l}$-scores, used for evaluating the type-B consistency.

Fig. 5 contains a $p^{A,\,Norm}$–$p^{B,\,Norm}$ map of the top-30 solutions in Tab. 13.

Let us notice that the one corresponding to the mode of the local classifications is ranked in 30th position. Although this solution maximizes the type-A consistency (obviously, $p^{A,\,Norm} = 1$), it totally neglects the agents’ importance hierarchy; $p^{B,\,Norm}$ is therefore relatively low (i.e., 45.5%), penalizing the resulting p^O (i.e., 72.7%).

Tab.13. List of the top-30 solutions, in terms of p^O, for the classification problem in Tab. 3(a).

Solution no.	Rank position	Global classifications					p^A	$p^{A,\,Norm}$	p^B	$p^{B,\,Norm}$	p^O
		o_1	o_2	o_3	o_4	o_5					
1[a, b]	1	c_1	c_2	c_3	c_2	c_3	54.6%	90.8%	91.7%	100.0%	95.4%
2[b]	2	c_1	c_2	c_3	c_2	c_2	49.1%	81.5%	91.7%	100.0%	90.8%
3[b]	2	c_3	c_2	c_3	c_2	c_3	49.1%	81.5%	91.7%	100.0%	90.8%
4	4	c_1	c_2	c_3	c_1	c_3	57.4%	95.4%	75.0%	81.8%	88.6%
5	4	c_2	c_2	c_3	c_2	c_3	57.4%	95.4%	75.0%	81.8%	88.6%
6[b]	6	c_1	c_2	c_3	c_2	c_1	43.5%	72.3%	91.7%	100.0%	86.2%
7[b]	6	c_1	c_3	c_3	c_2	c_3	43.5%	72.3%	91.7%	100.0%	86.2%
8	8	c_1	c_2	c_3	c_1	c_2	51.9%	86.2%	75.0%	81.8%	84.0%
9	8	c_3	c_2	c_3	c_1	c_3	51.9%	86.2%	75.0%	81.8%	84.0%
10[b]	10	c_1	c_1	c_3	c_2	c_3	40.7%	67.7%	91.7%	100.0%	83.8%
11	11	c_1	c_2	c_2	c_2	c_3	49.1%	81.5%	75.0%	81.8%	81.7%
12	11	c_1	c_2	c_3	c_3	c_3	49.1%	81.5%	75.0%	81.8%	81.7%
13	13	c_3	c_2	c_3	c_2	c_2	43.5%	72.3%	83.3%	90.9%	81.6%
14	14	c_1	c_2	c_3	c_1	c_1	46.3%	76.9%	75.0%	81.8%	79.4%
15	14	c_1	c_3	c_3	c_1	c_3	46.3%	76.9%	75.0%	81.8%	79.4%
16	16	c_1	c_1	c_3	c_1	c_3	43.5%	72.3%	75.0%	81.8%	77.1%
17	16	c_1	c_2	c_1	c_2	c_3	43.5%	72.3%	75.0%	81.8%	77.1%
18	16	c_1	c_2	c_2	c_3	c_3	43.5%	72.3%	75.0%	81.8%	77.1%
19	19	c_1	c_3	c_3	c_2	c_2	38.0%	63.1%	83.3%	90.9%	77.0%
20	19	c_3	c_2	c_3	c_2	c_1	38.0%	63.1%	83.3%	90.9%	77.0%
21	19	c_3	c_3	c_3	c_2	c_3	38.0%	63.1%	83.3%	90.9%	77.0%
22[b]	22	c_1	c_2	c_2	c_3	c_1	32.4%	53.8%	91.7%	100.0%	76.9%
23[c]	23	c_2	c_2	c_3	c_3	c_3	51.9%	86.2%	58.3%	63.6%	74.9%
24	23	c_1	c_2	c_2	c_1	c_3	51.9%	86.2%	58.3%	63.6%	74.9%
25	25	c_2	c_2	c_3	c_2	c_1	46.3%	76.9%	66.7%	72.7%	74.8%
26	25	c_1	c_3	c_2	c_1	c_3	40.7%	67.7%	75.0%	81.8%	74.8%
27	27	c_1	c_1	c_3	c_2	c_2	35.2%	58.5%	83.3%	90.9%	74.7%
28	27	c_3	c_1	c_3	c_2	c_3	35.2%	58.5%	83.3%	90.9%	74.7%
29[b]	29	c_1	c_1	c_2	c_3	c_3	29.6%	49.2%	91.7%	100.0%	74.6%
30[d]	30	c_2	c_2	c_3	c_1	c_3	60.2%	100.0%	41.7%	45.5%	72.7%

[a] Optimal solution, which is coincidentally the same solution resulting from the application of the fusion technique proposed by Franceschini and Maisano (2016);
[b] Solutions maximizing p^B;
[c] Solution examined in the section “Description of the indicators” (see Tab. 3(b));
[d] Solution maximizing p^A, resulting from the application of the *mode* operator.

Coincidentally, the optimal solution is the same solution obtained through the fusion technique by Franceschini and Maisano[14] (briefly recalled in the section "Related work"). This coincidence is quite interesting and may somehow be interpreted as a test of *convergent validity* of the optimization approach proposed in this section (Campell, 1946).

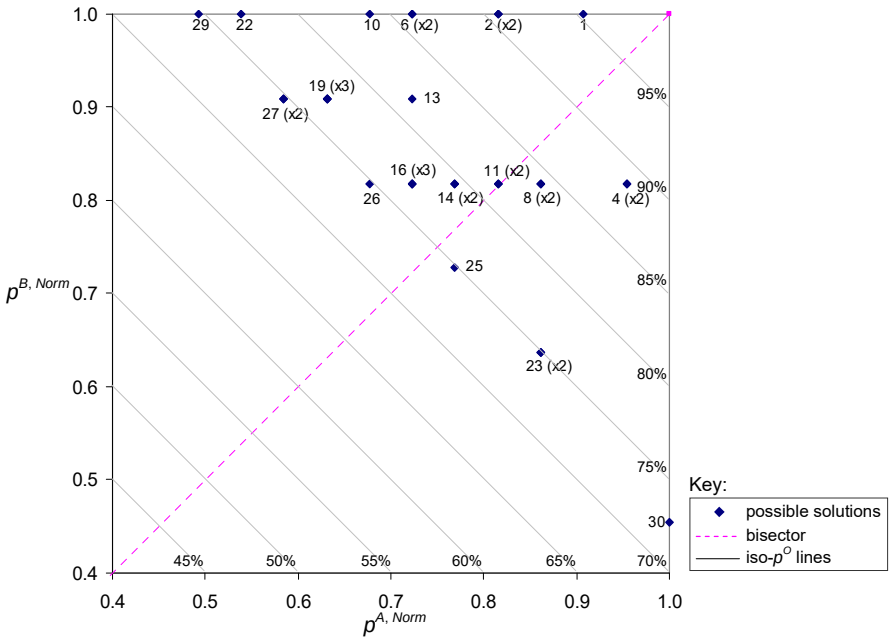


Fig. 5. Detail of the $p^{A,Norm}$– $p^{B,Norm}$ map in Fig. 3, focussing on the top 30 solutions in terms of p^O. The relevant rank-positions (i.e., from 1 to 30, in descending order) are reported in the map; the numbers in brackets (i.e., "x*n*") indicate that *n* different solutions have the same positioning. The iso-p^O lines are perpendicular to the bisector of the first quadrant; the corresponding numerical values of p^O are reported along the bottom edge and the right edge of the map.

Discussion

This paper proposed a set of relatively simple and intuitive indicators for assessing the (type-A and type-B) consistency of the solution(s) to a multi-object quality classification problem with multiple rank-ordered agents. Here follows a synthetic list of the indicators:

- $p^A_{i\bullet}$ and $p^A_{\bullet j}$ can be used to check the type-A consistency at the level of individual objects and individual agents, respectively;
- $p^B_{i\bullet}$ can be used to check the type-B consistency at the level of individual objects;
- p^A and p^B provide an overall assessment on the type-A and type-B consistency respectively. We empirically showed that these indicators are complementary;
- $p^{A, Max}$ provides an indication of the maximum achievable level of type-A consistency, for a specific quality classification problem, and can also be interpreted as a rough measure of the level of agent agreement;
- $p^{B, Max}$ provides an indication of the maximum achievable level of type-B consistency, for a specific quality classification problem;
- $p^{A, Norm}$ and $p^{B, Norm}$ are obtained by normalizing p^A and p^B. This normalization is necessary

because the direct comparison of p^A and p^B would produce biased results.

- Finally, p^O is an overall consistency indicator, obtained by averaging $p^{A,\ Norm}$ and $p^{B,\ Norm}$; this indicator is relatively robust with respect to small variations in the B_{il} and $B_{\bullet l}$-scores.

Let us now focus the attention on the (twofold) practical role of p^O:

1. p^O can be interpreted as a *passive* tool, which provides an overall indication of the level of consistency of a solution to a specific classification problem. E.g., considering the problem exemplified, the solution proposed in Tab. 3(b) (with $p^O = 88.6\%$) seems significantly more consistent than that obtained by applying the mode operator (i.e., the solution ranked in 30th position in Tab.13, with $p^O = 72.7\%$).
2. p^O can be interpreted as an *active* tool, precisely an *objective function* to be maximized for identifying the optimal solution(s).

Although the proposed indicators are simple, intuitive and practical, their construction may present some limitations, such as:

- The determination of some of them (e.g., $p^{B,\ Max}$ and $p^{O,\ Max}$) can be computationally burdensome, since it requires the analysis of the totality of the possible solutions to a certain classification problem. To overcome this problem, we have developed an *ad hoc* software application (available on request), which automatically generates all the possible solutions and determines the relevant indicators.
- The normalization and aggregation mechanism of the indicators is based on potentially questionable assumptions. Nevertheless, a sensitivity analysis showed the robustness of p^O to small variations of the input data.
- The calculation of the indicators for assessing the type-B consistency is based on the comparison of two orderings. This conventional choice is potentially debatable.

In this study we focussed the attention on quality classification problems in which the solution includes one-and-only-one quality category for each object. With relatively modest changes, the proposed indicators could be adapted to more general problems, in which the solution may be indeterminate or include multiple quality categories (e.g., in the case of hesitation by a relatively large portion of the agents).

Future research will aim at analyzing the proposed way of determining the optimal solution(s), from the viewpoint of some popular axioms borrowed from the social choice theory (Arrow and Rayanaud, 1986). Also, we plan to develop a real-use application of the proposed indicators in a bigger scale.

References

Agresti, A. (2013) Categorical data analysis. John Wiley & Sons.

Ai, Y., Xu, K. (2013) Application of bandelet transform to surface defect recognition of hot rolled steel plates, International Conference on Machine Vision Applications (MVA2013), May 20-23, Kyoto, Japan.

Arrow, K.J., Rayanaud, H. (1986) Social choice and multicriterion decision-making, Cambridge: MIT Press.
Bashkansky E., Dror S., Ravid R. (2007), Effectiveness of a Product Quality Classifier. Quality Engineering, 19(3): 235-244.
Bashkansky E., Gadrich T. (2008), Evaluating Quality Measured on a Ternary Ordinal Scale. Quality and Reliability Engineering International, 24: 957-971.
Belacel N. (2000) Multicriteria assignment method PROAFTN: Methodology and medical applications. European Journal of Operational Research, 125: 175-183.
Brasil Filho, A.T., Pinheiro, P.R., Coelho, A.L., Costa, N.C. (2009). Comparison of two prototype-based multicriteria classification methods. IEEE Symposium on Computational intelligence in multi-criteria decision-making, IEEE MCDM 2009, 30th March-2nd April, pp. 133-140.
Campell, D.T., Fiske, D.W. (1959) Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56:81-105.
Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement, 20(1): 37-46.
Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin, 70(4): 213-220.
Cook, W.D. (2006). Distance-based and ad hoc consensus models in ordinal preference ranking. European Journal of Operational Research, 172(2): 369-385.
Creswell, J.W. (2013). Research design: Qualitative, quantitative, and mixed methods approaches (fourth edition). Sage publications, Thousand Oaks (California).
Duffuaa, S.O, Khan M. (2005), "Impact of inspection errors on the performance measures of a general repeat inspection plan", International Journal of Production Research, Vol. 43, No. 23, pp. 4945-4967.
Figueira, J., Greco, S., Ehrgott, M. (2005) Multiple criteria decision analysis: state of the art surveys. Springer, New York.
Franceschini, F., Galetto, M., Maisano, D. (2007) Management by Measurement: Designing Key Indicators and Performance Measurement Systems. Springer, Berlin.
Franceschini, F., Galetto, M., Varetto M. (2004), Qualitative Ordinal Scales: The Concept of Ordinal Range. Quality Engineering, 16(4): 515-524.
Franceschini, F., Maisano, D. (2015) Checking the consistency of the solution in ordinal semi-democratic decision-making problems. Omega, 57(B): 188-195.
Franceschini, F., Maisano, D. (2016) Classification into nominal categories in the presence of multiple rank-ordered agents. Submitted to Fuzzy Sets and Systems. A draft version is available at https://www.dropbox.com/s/kvfzgqtvxhnnrf9/NominalScales.pdf?dl=0.
Goletsis, Y., Papaloukas, C., Fotiadis, D., Likas, A., Michalis, L. (2004) Automated ischemic beat classification using genetic algorithms and multicriteria decision analysis. IEEE Transactions on Biomedical Engineering, 51(10): 1117-1725.
Kendall, M.G. (1970). Rank Correlation Methods, 4th edition, C. Griffin and Company, London.
Léger, J., Martel, J. M. (2002) A multicriteria assignment procedure for a nominal sorting problematic. European Journal of Operational Research, 138(2), 349-364.
Mandroli, S. S., Shrivastava, A. K., Ding, Y. (2006), "A survey of inspection strategies and sensor distribution in discrete-part manufacturing processes". IIE Transactions, 38(4): 309-328.
Mevik B., Næs T. (2002). Strategies for Classification When Classes Arise from a Continuum. Quality Engineering, 15(1): 113-126.
Montgomery, D. C. (2013), Introduction to Statistical Quality Control: A Modern Introduction, 7th Edition, John Wiley and Sons, Singapore.
Nederpelt, R., Kamareddine, F. (2004). Logical Reasoning: A First Course. King's College Publications, London.
Perny, P. (1998) Multicriteria filtering methods based on concordance/nondiscordance principles, Annals of Operational Research, 80: 137-167.
Pontius Jr, R.G., Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing, 32(15): 4407-4429.
Ramanathan, R., Ganesh, L.S. (1994) Group preference aggregation methods employed in AHP: An evaluation and an intrinsic process for deriving members' weightages. European Journal of Operational Research, 79(2): 249-265.
Roy, B. (1968) Classement et choix en presence de points de vue multiples: La methode ELECTRE," Revue Francaise d'Informatique et de Recherche Operationnelle, 8: 57-75.
Saaty, T.L. (1980) The Analytic Hierarchy Process: Planning, Priority and Allocation, New York: McGraw-Hill.
See, J. E. (2012), Visual Inspection: A Review of the Literature, Sandia Report, SAND2012-8590, Sandia

National Laboratories, Albuquerque, NM.
Spearman C. (1904). The proof and measurement of association between two things. American Journal of Psychology 15(1): 72–101.
Steiner S. H., Lu Y., Mackay R. J. (2016). Assessing binary measurement systems and inspection protocols utilizing follow-up data. Quality Engineering, 28(3): 329-336.
Stevens, S.S. (1946) On the Theory of Scales of Measurement. Science, 103(2684): 677-680.
Strijbos, J., Martens, R., Prins, F., Jochems, W. (2006). Content analysis: What are they talking about?. Computers & Education, 46(1): 29-48.
Van Wieringen, W.N., De Mast, J. (2008) Measurement system analysis for binary data. Technometrics, 50(4): 468-478.
Viera, A.J., Garrett, J.M. (2005) Understanding interobserver agreement: the kappa statistic. Family Medicine, 37(5): 360-363.
Wang, B., Liang, J., Qian, Y. (2014) Determining decision makers' weights in group ranking: a granular computing method. Forthcoming to International Journal of Machine Learning and Cybernetics, DOI: 10.1007/s13042-014-0278-5.
Witten, I. H., Frank, E. (2005) Data Mining: Practical Machine Learning Tools and Techniques (2nd edition). Morgan Kaufmann, San Francisco.
Yevseyeva, I. (2007) Solving classification problems with multicriteria decision aid approaches. Ph.D. dissertation, University of Jyväskylä, Jyväskylä, Finland.
Zopounidis, C., Doumpos, M. (2002) Multicriteria classification and sorting methods: A literature review. European Journal of Operational Research, 138(2), 229-246.

Appendix

Note on the p^B maximization (level-2 title)

Sects. 2.1.2 and 2.3 showed that, for a specific quality classification problem, the determination of the maximum possible value of p^A (i.e., $p^{A, Max}$) is relatively simple: this value is associated with the solution(s) that maximize the $p_{i\bullet}^A$ values related to the individual objects. The p^A maximization problem can be therefore decomposed on an object-by-object basis.

The determination of the maximum possible value of p^B (i.e., $p^{B, Max}$) is more complicated as the solution(s) that maximize the $p_{i\bullet}^B$ values related to the individual objects are not necessarily those with $p^B = p^{B, Max}$.

Let us now provide a demonstration of this statement, based on a counter-example contradicting the statement that the solutions maximizing the individual $p_{i\bullet}^B$ values are also those with $p^B = p^{B, Max}$.

We consider a specific classification problem where $I = 6$ objects (i.e. o_1 to o_6) should be classified by $J = 10$ agents (i.e., d_1 to d_{10}) into $K = 4$ nominal categories (i.e., c_1 to c_4). The agents' importance rank-ordering is: $(d_1 \sim d_7) > (d_2 \sim d_4) > (d_3 \sim d_5 \sim d_{10}) > (d_6 \sim d_8 \sim d_9)$. Each agent performs a local classification for each objects, as shown in Tab. A.1.

Tab. A.1. Classification problem in which 6 objects (o_i) are classified by 10 agents (d_j) into 4 categories (c_k).

Agents		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
Local classifications concerning o_i	o_1	c_1	c_1	c_2	c_1, c_3	c_1	c_1, c_2	c_1	c_3	c_1	c_1
	o_2	c_2	c_2, c_4	c_2	-	c_2	c_2	c_2	c_1	c_2, c_3	c_2
	o_3	c_4	c_4	c_1, c_2, c_4	c_4	c_1, c_4	c_1, c_4	c_4	-	c_2	c_4
	o_4	c_1, c_2	c_1	c_1	c_1	c_1	c_2	c_1, c_2	-	c_1, c_2	c_2
	o_5	c_4	c_2, c_4	c_2, c_4	-	c_2, c_3, c_4	c_4	c_4	c_2	c_4	c_4
	o_6	c_1, c_3	c_1	c_1, c_4	c_1	c_1	c_4	c_1	c_4	c_2	c_1, c_4
Agents' importance rank-ordering: $(d_1 \sim d_7) > (d_2 \sim d_4) > (d_3 \sim d_5 \sim d_{10}) > (d_6 \sim d_8 \sim d_9)$.											

Analyzing the $K^I = 4096$ possible solutions to this problem, we respectively selected (i) those maximizing the individual $p_{i\bullet}^B$ values and (ii) those maximizing p^B (see Tab. A.2). It can be noted that the (three) solutions maximizing the individual $p_{i\bullet}^B$ values all have p^B values lower than $p^{B,\,Max}$.

Tab. A.2. List of the three solutions maximizing the individual $p_{i\bullet}^B$ values (on a object-by-object basis) and that one maximizing p^B, for the classification problem in Tab. A.1.

Agents	o_1	o_2	o_3	o_4	o_5	o_6	$p_{1\bullet}^B$	$p_{2\bullet}^B$	$p_{3\bullet}^B$	$p_{4\bullet}^B$	$p_{5\bullet}^B$	$p_{6\bullet}^B$	p^B
Solutions maximizing the individual $p_{i\bullet}^B$ values	c_1	c_4	c_4	c_1	c_1	c_3	65.6%	62.2%	68.9%	66.7%	58.9%	66.7%	73.3%
	c_1	c_4	c_4	c_1	c_3	c_3	65.6%	62.2%	68.9%	66.7%	58.9%	66.7%	73.3%
	c_1	c_4	c_4	c_1	c_4	c_3	65.6%	62.2%	68.9%	66.7%	58.9%	66.7%	76.7%
Solution maximizing p^B	c_1	c_2	c_1	c_1	c_4	c_1	65.6%	52.2%	43.3%	66.7%	58.9%	65.6%	$p^{B,\,Max} = 85.6\%$

This result represents a proof that the solution(s) maximizing the $p_{i\bullet}^B$ values related to the individual objects are not necessarily those with $p^B = p^{B,\,Max}$.

Sensitivity analysis (level-2 title)

This section analyzes the robustness of the proposed indicators with respect to small variations in the B_{il} and $B_{\bullet l}$-scores, used for evaluating the type-B consistency. While the assignment of these scores seems adequate in the case of (i) *full consistency* (score 1), (ii) *inconsistency* (score 0), and (iii) *incomparability* (N/A), it is somehow arbitrary in the case of *weak consistency* (score 0.5) – see the relevant definitions in Tab.8. We now propose an empirical sensitivity analysis aimed at showing how small variations in the weak-consistency score may affect (some of) the proposed indicators. Specifically, considering the classification problem in the section "Description of the indicators", we analysed the variations in the indicators associated with the ($K^I = 243$) possible solutions, when using the three different sets of B_{il} and $B_{\bullet l}$-scores in Tab. A.3 – i.e., set (a), (b), and (c). We note that set (b) is the same set proposed in Tab.8 and used in the previous examples.

Tab. A.3. Sets of B_{il} and $B_{\bullet l}$-scores used in the sensitivity analysis.

	Scores		
	Set (a)	Set (b)	Set (c)
1. *Full consistency*	1	1	1
2. *Weak consistency*	0.25	0.50	0.75
3. *Inconsistency*	0	0	0
4. *Incomparability*	N/A	N/A	N/A

Considering the 30 solutions in Tab.13 (i.e., the top-30 solutions of the problem exemplified in the section "Description of the model", obtained adopting set (b)), we recalculated the corresponding indicators, especially $p^{A,\,Norm}$, $p^{B,\,Norm}$ and p^O. Then we determined the corresponding rank position in terms of p^O (i.e., a number included between 1 and 243). It can be seen that the resulting changes in the p^O values are marginal, e.g., the top-7 solutions are identical for each of the three sets (see also the diagram in Fig. A.1). Although we are aware that it is not a rigorous proof, the sensitivity

analysis revealed a certain robustness of p^O.

Tab. A.4. Results of the robustness analysis: p^O-ranking related to the 30 solutions in Tab.13, when using each of the three sets of B_{il} and $B_{.l}$-scores in Tab. A.3.

Solution No.	Global classifications o_1	o_2	o_3	o_4	o_5	$p^{A,\,Norm}$	Set (a) $p^{B,\,Norm}$	p^O	rank pos.	Set (b) $p^{B,\,Norm}$	p^O	rank pos.	Set (c) $p^{B,\,Norm}$	p^O	rank pos.
1	c_1	c_2	c_3	c_2	c_3	90.8%	100.0%	95.4%	1	100.0%	95.4%	1	100.0%	95.4%	1
2	c_1	c_2	c_3	c_2	c_2	81.5%	100.0%	90.8%	2	100.0%	90.8%	2	100.0%	90.8%	2
3	c_3	c_2	c_3	c_2	c_3	81.5%	100.0%	90.8%	2	100.0%	90.8%	2	100.0%	90.8%	2
4	c_1	c_2	c_3	c_1	c_3	95.4%	81.0%	88.2%	4	81.8%	88.6%	4	82.6%	89.0%	4
5	c_2	c_2	c_3	c_2	c_3	95.4%	81.0%	88.2%	4	81.8%	88.6%	4	82.6%	89.0%	4
6	c_1	c_2	c_3	c_2	c_1	72.3%	100.0%	86.2%	6	100.0%	86.2%	6	100.0%	86.2%	6
7	c_1	c_3	c_3	c_2	c_3	72.3%	100.0%	86.2%	6	100.0%	86.2%	6	100.0%	86.2%	6
8	c_1	c_2	c_3	c_1	c_2	86.2%	81.0%	83.6%	9	81.8%	84.0%	8	82.6%	84.4%	8
9	c_3	c_2	c_3	c_1	c_3	86.2%	81.0%	83.6%	9	81.8%	84.0%	8	82.6%	84.4%	8
10	c_1	c_1	c_3	c_2	c_3	67.7%	100.0%	83.8%	8	100.0%	83.8%	10	100.0%	83.8%	11
11	c_1	c_2	c_2	c_2	c_3	81.5%	81.0%	81.2%	11	81.8%	81.7%	11	82.6%	82.1%	12
12	c_1	c_2	c_3	c_3	c_3	81.5%	81.0%	81.2%	11	81.8%	81.7%	11	82.6%	82.1%	12
13	c_3	c_2	c_3	c_2	c_2	72.3%	85.7%	79.0%	13	90.9%	81.6%	13	95.7%	84.0%	10
14	c_1	c_2	c_3	c_1	c_1	76.9%	81.0%	78.9%	14	81.8%	79.4%	14	82.6%	79.8%	14
15	c_1	c_3	c_3	c_1	c_3	76.9%	81.0%	78.9%	14	81.8%	79.4%	14	82.6%	79.8%	14
16	c_1	c_1	c_3	c_1	c_3	72.3%	81.0%	76.6%	17	81.8%	77.1%	16	82.6%	77.5%	20
17	c_1	c_2	c_1	c_2	c_3	72.3%	81.0%	76.6%	17	81.8%	77.1%	16	82.6%	77.5%	20
18	c_1	c_2	c_2	c_3	c_3	72.3%	81.0%	76.6%	17	81.8%	77.1%	16	82.6%	77.5%	20
19	c_1	c_3	c_3	c_2	c_2	63.1%	85.7%	74.4%	21	90.9%	77.0%	19	95.7%	79.4%	16
20	c_3	c_2	c_3	c_2	c_1	63.1%	85.7%	74.4%	21	90.9%	77.0%	19	95.7%	79.4%	16
21	c_3	c_3	c_3	c_2	c_3	63.1%	85.7%	74.4%	21	90.9%	77.0%	19	95.7%	79.4%	16
22	c_1	c_2	c_2	c_3	c_1	53.8%	100.0%	76.9%	16	100.0%	76.9%	22	100.0%	76.9%	25
23	c_2	c_2	c_3	c_3	c_3	86.2%	61.9%	74.0%	25	63.6%	74.9%	23	65.2%	75.7%	26
24	c_1	c_2	c_2	c_1	c_3	86.2%	61.9%	74.0%	26	63.6%	74.9%	24	65.2%	75.7%	27
25	c_2	c_2	c_3	c_2	c_1	76.9%	66.7%	71.8%	33	72.7%	74.8%	25	78.3%	77.6%	19
26	c_1	c_3	c_2	c_1	c_3	67.7%	81.0%	74.3%	24	81.8%	74.8%	26	82.6%	75.2%	29
27	c_1	c_1	c_3	c_2	c_2	58.5%	85.7%	72.1%	28	90.9%	74.7%	27	95.7%	77.1%	23
28	c_3	c_1	c_3	c_2	c_3	58.5%	85.7%	72.1%	28	90.9%	74.7%	27	95.7%	77.1%	23
29	c_1	c_1	c_2	c_3	c_3	49.2%	100.0%	74.6%	20	100.0%	74.6%	29	100.0%	74.6%	34
30	c_2	c_2	c_3	c_1	c_3	100.0%	42.9%	71.4%	34	45.5%	72.7%	30	47.8%	73.9%	35

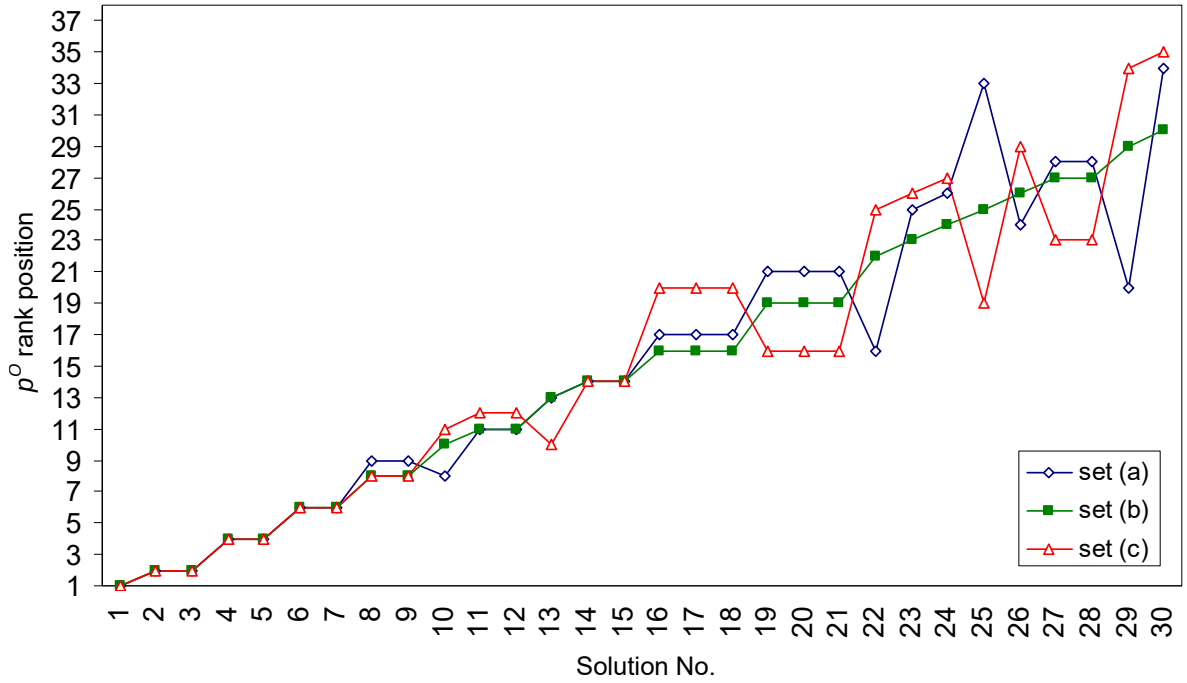


Fig. A.1. Graphical representation of the p^O-ranking related to the 30 solutions in Tab.13, when using each of the three sets of B_{il} and $B_{.l}$-scores in Tab. A.3.