Defining Relative Weights of Data Sources during Aggregation of Pair-wise Comparisons

© Sergii V. Kadenko
Institute for Information Recording of National Academy of Sciences of Ukraine, Kyiv, Ukraine
seriga2009@gmail.com

Abstract

A method for defining the relative data source weights in the process of aggregation of pair-wise comparisons is suggested. It is shown that in order to define the relative weights of data sources, providing data in the form of pair-wise comparison matrices, it is not enough to consider just a-priori weight estimate, or the so-called objective component of the indicator. It is reasonable to assume that data sources, providing more consistent, compete, and detailed information, should be assigned larger weights when data from several sources is aggregated. The suggested approach allows us to consider consistency, compatibility and completeness of pair-wise comparisons as well as the level of detail of pair-wise comparison scales (in addition to the objective component) while calculating the relative data source weights.

Four conceptual levels of data source weight definition are described (from top to bottom): subject domain, specific problem, specific pair-wise comparison matrix, and specific pair-wise comparison. On the subject domain and the problem levels it is suggested to use any of the previously-developed approaches, including pair-wise comparison-based methods. On the level of a specific pair-wise matrix, provided by a data source, it is suggested to assign it a rating, reflecting its inner consistency and compatibility with pair-wise comparison matrices, provided by other sources. Finally, on the level of each pair-wise comparison, it is suggested assign it a weight, depending on the number of grades in the scale, in which it has been provided.

The suggested approach allows us to improve the pair-wise comparison aggregation methods, particularly, combinatorial method (sometimes called enumeration of all spanning trees). Obtained experimental results confirm that all different aspects of data source credibility should be taken into consideration during content-analysis-based research and expert examinations, particularly when completeness, consistency and compatibility levels of estimates vary across data sources, and scales with different numbers of grades are used. On the whole, obtained results allow us to improve the existing pair-wise comparison aggregation methods and increase the credibility of results of content-analysis-based research and group expert examinations, using pair-wise comparisons.

Keywords: content analysis, relative data source weight, decision-making support, pair-wise comparison matrix, pair-wise comparison scale, expert estimation, consistency, compatibility.

1 Introduction: why pair-wise comparisons?

Information space is characterized by a multitude of factors, both tangible and intangible. In [1] it is stated that information operations are influenced by many solely quantitative (for instance, socio-psychological) criteria, factors, and parameters. Their formal mathematical and analytical description is a challenging task. Authors also stress the impossibility of development and implementation of some universal methodology for modeling of information operations, first of all, because of weak formalization of related factors and concepts.

In [2] it is shown that information security in general and information operations in particular represent a weakly structured subject domain. Decision-making support technologies prove to be an effective analytical tool in other weakly-structured domains, so they, definitely, should be among the technological means for analysis and modeling of information operations. Beside that, in [1] it is stressed that information security strategy is an important component of the national security strategy. Tsyganok et al. in [3] show that decision support technologies are a powerful instrument of strategic planning, especially, in weakly structured domains, where other approaches are not as effective.

In any weakly-structured domain, that has to be analytically described, we often come to the point when a set of objects or factors has to be measured or compared. Pair-wise comparisons are a flexible way of representation of data on a set of objects that cannot be measured or described by absolute quantitative values. Any measurement, by the way, is a pair-wise comparison, in a sense, because the measured object is actually compared to some benchmark value (meter, pound, second, byte, etc.). However, if there are no benchmarks to compare the objects with, the only way to obtain at least some quantifiable data on these objects is to compare them among themselves.

In the contexts of information security, information policy, and information operations, the necessity to deal with pair-wise comparisons might arise in several situations.

1) Strategic planning (more general case). As information security, informational policy, and information operations represent weakly structured domains, decision-making support technologies prove useful as strategic planning tools in these areas. Strategic planning process using expert data is described in [3] and includes such steps

as: formulation of the main strategic goal by the decision-maker (DM), selection of experts, decomposition (by experts) of the main goal into sub-goals (criteria, factors) down to the level of projects, that can be directly influenced by the DM (building of a hierarchy of criteria), evaluation of relative impacts of criteria by experts, definition of the strategy as the most rational distribution of resources among projects. Under such a scenario, most of the data comes from experts, so they are the primary information source. Pair-wise comparisons come into play when experts evaluate relative impacts of criteria (factors) in the hierarchy.

2) Building some rating or ranking of a set of objects (or factors) based on comparing them among themselves (concepts, topics, product brands, political parties, candidates etc) during some informational-analytical research (more particular case). Just like in the previous case, the data may come from experts, however, it may also come from analysts, that monitor information sources (online publications, meta-search results, blogs, or even comments in social networks). For example, if an online publication says something like "Brand A has much stronger standing at the market than Brand B" or "Candidate X is following close behind Candidate Y, according to the polls", such statements do provide the basis for pair-wise comparisons in some scale. Ordinal pair-wise comparison scales include only two values (such as "better" or "worse", "more" or "less" etc). Cardinal pair-wise comparison scales include particular quantitative values. For instance, a popular scale used by Saaty [4] includes the following values: 1 – no preference, 2 – weak preference, 3 – moderate preference, 4 – moderate plus, 5 – strong preference, 6 – strong plus, 7 – very strong preference, 8 – very strong plus, 9 – extreme, as well as reciprocal values. The result of a session of pair-wise comparisons is a pair-wise comparison matrix (PCM), where each element shows ordinal or/and cardinal relation between the two respective objects.

Besides, pair-wise comparisons can be also derived from the frequencies, with which compared objects (again, product brands, political parties, electoral candidates etc) are mentioned in data sources.

If we need to build a unified rating or ranking of a set of objects or factors (that is a common task in both above-mentioned cases) and the data on these objects comes from several data sources, this data needs to be aggregated. However, before aggregation it is necessary to define the relative importance of these data sources, i.e., which of them should be assigned larger weights. Some weights can be assigned to the data sources by experts or analysts a priori, but there are also other important components of relative weights of data sources. It would be reasonable to assume, that the more complete, consistent, and detailed the data is, the more credible the data source is and the greater weight should be assigned to it. In this paper we are going to address all the listed components of data source weight in greater detail.

2 Literature overview; existing approaches

When it comes to decision-making based on pair-wise comparisons (coming from several data sources), Tom Saaty (author of analytic hierarchy/network process (AHP/ANP) [4]) and his followers can, in a way, be called monopolists. During the last few decades this scientific school suggested a set of approaches to definition of expert competence indicators that can and should be used during aggregation pair-wise comparisons coming from several different sources. In [4] Saaty suggests building a hierarchy of such factors as "skills", "experience", "accomplishments", "persuasion", "efforts" etc and calculate expert competence as an aggregate estimate of the expert according to these criteria (again, based on eigenvectors of PCM, as prescribed by AHP method). If the DM is unable to evaluate experts according to these criteria, then the tasks of self-evaluation and mutual evaluation can be delegated to the experts themselves. At this phase Saaty considers data sources to be equally competent. Ramanathan and Ganesh [5] instead suggest compiling a matrix of alternative weights, obtained as eigenvectors of individual PCM $W = (\vec{w}_1,...,\vec{w}_m)$ (where m is the number of experts) and calculate relative weights of experts based on the equation $\lambda\vec{x} = W\vec{x}$. Another approach is suggested by Yang et. al in [6]. The authors use a kind of combination of two methods, AHP and TOPSIS (described, for instance, by Hwang and Yoon in [7]), and propose to calculate weights of expert judgments based on their distance from the best ones and the worst ones in the set ("ideal alternatives"). These approaches can be relatively easily extrapolated to the cases when data sources are not experts. Online or printed data sources, for instance, can be similarly evaluated by a set of criteria, and assigned respective relative weights.

Speaking of Ukrainian academic schools, we can mention Totsenko [8] who, in a way, summarized the earlier achievements of soviet researchers. Totsenko suggested calculating relative competence of experts as product of self-estimate and the weighted sum of mutual estimate and objective estimate (see formula (1) below). Gnatienko and Snytiuk [9] set forth two approaches for defining the expert's competence: 1) based on the questionnaire with questions of different types; 2) based on PCMs, provided by the experts themselves. PCM is approximated by a "polyhedron" of weight coefficients. The more its volume is, the less consistent the PCM is, and the lesser weight should be assigned to the respective expert. In the case, when data sources are not experts, we cannot apply the concept of self-estimate (and mutual estimate as well) or talk about completion of questionnaires. In other aspects, the approaches specified in this passage can be extrapolated to the case when experts are not the primary data source. Rather promising attempts to apply decision support methods for processing of data that does not come from experts, but is derived from content-monitoring results have been made by Andriichuk in [10].

The brief literature overview, provided above, indicates, that every existing approach to calculation of weights of data sources takes into account just one particular aspect of data source weight: objective estimate ([4,9]), mutual or self-estimate ([5]), mutual agreement (compatibility) ([6]), or inner agreement (consistency) ([9]) of data. Consequently, there is a need to develop a method for calculation of data source weights, that would take all the listed aspects into consideration: self-estimate (if applicable), mutual estimate (if applicable), objective a priori estimate of the weights, as well as consistency and compatibility of data. Beside that, it would be reasonable to assume, that the data source weight should depend on the degree of detail, with which the data is presented (the more detailed the data is the greater weight should be assigned to its source). In other words, the data source weight should reflect both a priori estimate and a posteriori estimate (the latter depending on the quality of information, coming from the source). Further we will try to take all the listed aspects into consideration.

3 Problem statement and solution idea

Data source weight calculation problem can be formulated as follows. Let us assume, that at some phase of strategic planning procedure, described in [3], or in the course of some independent content research m data sources are used to compile a rating of n objects. The data from these sources is represented in the form of m PCMs: $\{A^{(k)}; k=1..m\} = \{a_{ij}^{k} : k=1..m; i,j=1..n\}$. ***We should find*** the relative weights of data sources that would most thoroughly reflect the quality of quality of information coming from these sources (in view of aspects listed in previous sections of this paper). These weights are to be further used for aggregation of data, coming from all the sources: $c_l : l=[1..m]: \sum_{l=1}^{m} c_l = 1$.

Solution idea. Until recently the author of this paper (being the advocate of expert data-based decision support methods) would have preferred to use the approaches suggested by Totsenko [8] to define data source weights. In an expert group the key components of l-th member's competence c_l are (as we mentioned above) self-estimate, mutual estimate, and objective component ($c_{l,self}, c_{l,obj}, c_{l,mutual}$, respectively):

$$c_l = c_{l,self}(x_1 c_{l,obj} + x_2 c_{l,mutual}) \quad (1)$$

In (1) x_1, x_2 are the relative weights of objective and mutual estimates. In the case when data sources are not experts, only the objective component applies.

The results of recent research (such as improvement of combinatorial pair-wise comparison aggregation method [11] and development of a multi-scale approach to data representation [3]) of the academic school the author belongs to call for revision of existing approaches to definition of data source weights based on the requirements, specified in the previous section of this paper.

If in the course of some analytic research data from several sources is used and represented in the form of PCM, then the weights of these sources can be considered at several conceptual or contextual levels (from more general to more specific): 1) level of overall credibility of the data source within the subject domain in general; 2) level of the specific analytic research; 3) level of a specific series of pair-wise comparisons (resulting in completion of one PCM); 4) level of a specific pair-wise comparison, presented in some selected scale. Let us address each of the levels in particular.

1) Data source weight at the level of the *subject domain in general* is reflected by the components, presented in formula (1): self-estimate, mutual estimate, and objective estimate. In the case when the data source is a member of an expert group, all three components apply, if the data source is not an expert, just the objective component remains. It can be set a priori by the analyst (knowledge engineer) who conducts monitoring and research of online or printed data sources.

2) Data source weight in the context of *specific analytic research or examination* can be defined based in keywords, describing the main goal of the research. The principles of calculation of data source weights in terms of basic keywords (BKW) (again, based on the three listed components) were set forth by Totsenko in [8]. Totsenko (similarly to Saaty) suggested building a hierarchy of keywords. The non-normalized weight of the *j*-th data source with respect to *h*-th keyword, located at *i*-th level of the hierarchy k_{jih}^{*} should be calculated as follows.

$$k_{jih}^{*} = v_{jih}(x_1 o_{jih} + x_2 w_{jih}) \quad (2)$$

In (2) x_1, x_2 – are, respectively, the relative weights of objective and mutual estimates of expert group members' competences (if data comes from experts), while $v_{jih}, o_{jih}, w_{jih}$ are, respectively, self-estimate, objective, and mutual estimates of experts' competence according to the given keyword. If experts are not involved and data comes from online or printed sources, then only the objective component remains. If experts are involved then self-estimates, mutual estimates and objective estimates of individual experts' competences, as well as relative weights of components x_1, x_2 can be defined using the methods, listed in the introductory section of this paper.

3) Relative weight of a data source, providing the basis for a *specific PCM* can be defined using the approach described in [12]. According to the approach, the data source should be considered more credible, if the PCM is more consistent within itself and, at the same time, compatible with the respective PCM, based on data from other sources.

At this point we should remind that weights are assigned to data sources in order to ensure more "balanced" procedure of aggregation of PCMs, provided by different sources. Combinatorial method of PCM aggregation ([11, 13]) (sometimes called enumeration of all spanning trees) proves to be one of the most effective data aggregation methods. Within this method, in order to exploit the redundancy of information most thoroughly, we decompose PCM, provided by each given data source, into basic pair-wise comparison sets (that can be represented by connected graphs called spanning trees). Each spanning tree is used to build (reconstruct) an ideally consistent PCM (ICPCM) [11, 12]. Each of these ICPCM is assigned a rating, reflecting the weigh of the data source at the level of the respective *basic (informatively meaningful) set of pair-wise comparisons* of objects [12].

Weight or "rating" of a certain ICPCM depends on its proximity to original PCM, provided by data sources. It is reasonable to assume, that the more distant from original PCM the ICPCM is, the lower its rating should be (because this "distance" or "difference" is the indicator of inconsistency and incompatibility of the PCM). In [12] it is suggested to calculate ICPCM rating as shown in formulas (3) and (3a).

$$R_{kql} = \frac{c_k c_l}{\ln(\sum_{u,v} |a_{uv}^{kq} - a_{uv}^{l}| + e)} \quad (3)$$

$$R_{kql} = \frac{c_k c_l}{\log_2(\sum_{u,v} |a_{uv}^{kq} - a_{uv}^{l}| + 2)} \quad (3a)$$

In formulas (3) and (3a) R_{kql} is the rating (weight) of the ICPCM, "reconstructed" from the q-th informatively-significant set of pair-wise comparisons, taken from original PCM, provided by k-th data source, that is compared to the original PCM, provided by the l-th data source; a_{uv}^{kq} and a_{uv}^{l} are the respective elements of q-th ICPCM from k-th source and the original PCM from the l-th source; c_k, c_l are a priori estimates of data source weights (representing the first two of the above-mentioned conceptual levels). Logarithm operator is used to avoid large differences between PCM ratings and ensure that the rating fall within the same order of magnitude.

In [12] only the case of additive pair-wise comparisons is addressed. That is, the transitivity condition is formulated as $\forall i, j = 1..n : a_{ij} = w_i - w_j = (w_i - w_k) + (w_k - w_j) = a_{ik} + a_{kj}$, where n is the general number of compared objects, w_i, w_j are the relative weights of objects with the respective numbers. For multiplicative pair-wise comparisons (when the transitivity (consistency) condition of PCM is formulated as $\forall i, j = 1..n : a_{ij} = \frac{w_i}{w_j} = \frac{w_i}{w_k} \times \frac{w_k}{w_j} = a_{ik} \times a_{kj}$) we can suggest alternative formulas for calculation of ICPCM ratings, where sum operation is replaced by product operation:

$$R_{kql} = \frac{c_k c_l}{\ln(\prod_{u,v} \max(\frac{a_{uv}^{kq}}{a_{uv}^{l}}; \frac{a_{uv}^{l}}{a_{uv}^{kq}}) + e - 1)} \quad (4)$$

$$R_{kql} = \frac{c_k c_l}{\log_2(\prod_{u,v} \max(\frac{a_{uv}^{kq}}{a_{uv}^{l}}; \frac{a_{uv}^{l}}{a_{uv}^{kq}}) + 1)} \quad (4a)$$

The denominators in relations (3), (3a), (4), (4a) reflect the differences between the respective PCMs. The greater the difference, the less consistent (and compatible) the data source is. ICPCM, based on data from the given source are compared to original PCM, provided by both this same source and other sources ($k, l = 1..m$). This allows us to consider both consistency and compatibility of the estimates. Based on the look of formulas (3), (3a), (4), (4a), we can make the following conclusion. If the DM (research organizer) needs the results of analytic research to reflect the differences in the data source weights most thoroughly, then (s)he should choose the multiplicative "version" of the formula to calculate the PCM rating, and set smaller logarithm base.

4) At the level of a *specific pair-wise comparison* the weight of the data source should depend on the degree of detail of the scale, in which the comparison is provided. The data source, providing the comparison value in the more detailed scale, should be considered more credible in regard to the respective pair of compared objects. In order to take this component of data source weight into account, the authors of [3] suggest assigning to every pair-wise comparison, provided in some given scale, a coefficient, proportional to the amount of information, "contained" by the scale. The amount of information in the scale of N grades is calculated based on Hartley's formula ([14]): $S_N = I = \log_2 N$.

Consequently, the rating (or weight) of a basic set of pair-wise comparisons of n objects (described above) can be calculated as the geometric mean of ratings of its elements. Thus, the average non-normalized weight of q-th basic pair-wise comparison set, based on PCM, provided by k-th data source, will amount to:

$$s^{kq} = \left(\prod_{u=1}^{n-1} \log_2 N_u^{(kq)}\right)^{\frac{1}{n-1}} \quad (5)$$

According to Caley's theorem on trees ([15]), the total number of basic pair-wise comparison sets, derived from a complete PCM of n objects is n^{n-2}. If the matrix is incomplete (comparisons of some objects are not provided by the data source), then the number of such basic sets (spanning trees) is $T \in (0,..,n^{n-2})$. In formula (5) the range of q is $\{q = 1..T\}$. Consequently, when it comes to the whole PCM, provided by data source number k, its rating (or weight) in terms of the degree of detail or scales, the matrix elements are provided in, can be calculated as follows.

$$s^{k} = \left(\prod_{\substack{u,v=1 \\ v>u}}^{n} \log_2 N_{uv}^{(k)}\right)^{\frac{2}{n(n-1)}} \quad (6)$$

By definition, the matrix is reciprocally symmetrical, so we can consider only the elements lying above the principal diagonal of the matrix ($v > u$).

All the listed components of relative weights of data sources should be taken into consideration during aggregation of pair-wise comparisons, coming from these sources. So far we have addressed only separate conceptual aspects of the problem solution. In the next section we are going to complete the problem solution process and present a method for data source weight calculation that should be used for aggregation of data from different sources.

4 Application of the approach for improvement of combinatorial method of pair-wise comparison aggregation

Based on considerations, set forth in the previous sections, we can introduce certain improvements into combinatorial method of pair-wise comparison aggregation, described in [11, 12], which will allow us to make the results of its application more compliant with the actual level of data sources' credibility levels.

Combinatorial method is used in the process of strategic planning ([3]) when experts are estimating the relative impacts of factors that influence the main goal of the strategic plan. Beside that, it can be used to compile a rating of a set of objects based on data from several sources, expressed in the form of complete or incomplete PCMs. We have chosen combinatorial approach from among various pair-wise comparison aggregation methods, because it allows us to use redundant data from different sources most thoroughly, and because of advantages of the approach (such as stability and others), demonstrated in [13].

We should remind that in combinatorial method basic pair-wise comparison sets (so-called spanning trees) are selected from PCMs, provided by each data source. Normalized weights (in Saaty's terms, priorities) of compared objects can be calculated based on each of these sets: $\{w_j^q; j = 1..n; q = 1..T\}$. If we neglect the differences in the relative credibility of data sources and calculate aggregate values of weights of n objects ($w_j^{aggregate}; j = 1..n$) based on PCMs, provided by m sources using the geometric mean formula, we will get the following expression.

$$w_j^{aggregate} = \left(\prod_{q=1}^{T} w_j^q\right)^{\frac{1}{T}};\ j = 1..n;\ T \in [1..mn^{n-2}] \quad (7)$$

In order to consider consistency and compatibility of PCMs during aggregation of object weight values, it is reasonable to assign each ICPCM (and the respective weight vector), derived from the individual PCM, provided by a specific data source, a rating (as explained in the previous section). In order to introduce a rating, reflecting both consistency of an individual PCM and compatibility of PCMs, provided by different sources, it is suggested to produce *m* copies of every ICPCM. As a result, we will get *T**=*mT* ICPCMs ($m \le T^* \le m^2 n^{n-2}$). Beside the level of agreement of estimates, the rating should depend on the weights of scales pair-wise comparisons are provided in. (4-th conceptual level in section 3 of this paper) and a priori estimate of relative weight of the data source (levels 1 and 2). In view of these considerations, we suggest the following formula for calculation of the rating of each single ICPCM (and the respective priority vector) from among all *T**=*mT* ICPCMs.

$$R_{kql} = \frac{c_k c_l s^{kq} s^l}{\ln(\sum_{u,v} \left| a_{uv}^{kq} - a_{uv}^l \right| + e)} \quad (8)$$

$$R_{kql} = \frac{c_k c_l s^{kq} s^l}{\ln(\prod_{u,v} \max(\frac{a_{uv}^{kq}}{a_{uv}^l}; \frac{a_{uv}^l}{a_{uv}^{kq}}) + e - 1)} \quad (8a)$$

In formulas (8) and (8a) k,l are the numbers of data sources ($k,l = 1..m$) that provide PCMs being compared; c_k, c_l are weight coefficients, representing the levels of subject domain and particular analytic research; again, we should stress that k and l can be equal or different, i.e. ICPCM, based on original PCM, provided by data source number *k*, can be compared with original PCM, provided by this source (inner agreement or consistency), and with PCMs, provided by other sources (mutual agreement or compatibility); q is the number of ICPCM “copy” ($q = 1..mT_k$); s^{kq} is the average relative weight of scales, in which pair-wise comparisons from the given spanning tree were provided (formula (5)); s^l is the average weight of scales, in which the respective PCM by data source number *l* was provided (formula (6)). If we incorporate ICPCM ratings into formula (7), we get the following expression for aggregated object weights.

$$w_j^{aggregate} = \prod_{k,l=1}^{m} \left(\prod_{q^k=1}^{T_k} (w_j^{(kq_k l)})^{\frac{R_{kq_k l}}{\sum_{u,p,v} R_{upv}}}\right);\ j = 1..n \quad (9)$$

5 Experimental research and numeric results

The modified combinatorial method of pair-wise comparison aggregation was tested on multiple numeric examples, using the prototype application in MS Excel. The experiment indicated that aggregation results, that incorporated the weights of data sources, significantly differed from results that did not incorporate these weights, although the same PCMs were used. This fact confirms the necessity to take into consideration all the aspects of relative importance of data sources when aggregating pair-wise comparisons, because in this case the final result reflects the credibility of these sources more adequately.

Naturally, if all data sources provide data in the same scale, and all comparisons are ideally consistent, then the respective conceptual levels will never come into play (as we can see from the previous sections of this paper). However, in the real world the situation is, usually, quite the opposite.

In this section we are going to consider a numeric example, which illustrates the differences between priority values, derived from pair-wise comparisons, respectively, with and without taking of data source weights into consideration. Let us say, 3 experts (E_1, E_2, E_3) compare 4 alternatives (A_1, A_2, A_3, A_4), and use integer scales with numbers of grades from 2 to 9 for pair-wise comparisons. In order to make the example more illustrative, let us assume that a priori estimates of experts’ relative competence levels are equal ($c_1 = c_2 = c_3 = 1$, if normalized – $c_1 = c_2 = c_3 = 1/3$). In this example we will use additive PCM rating function (formula (8)).

Ordinal PCMs are shown in Table 1. Let ordinal estimates be perfectly consistent, i.e. alternatives can be sorted in such a way, that all pair-wise comparisons above the principal diagonal of every PCM exceed 1. Let us assume the 1st expert is unable to compare the 1st alternative with the 3rd one. This means that the respective PCM is incomplete, while all spanning trees, including the respective element from the 1st expert's PCM, are informatively insignificant, and, consequently, the respective multipliers in formula (9) equal 1.

Table 1 Ordinal pair-wise comparison matrices, provided by three experts

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	>	*	>	1	>	>	>	1	>	>	>
A_2		1	>	>		1	>	>		1	>	>
A_3			1	>			1	>			1	>
A_4				1				1				1

Table 2 features the total numbers of grades in the scales, selected by the 3 experts for input of pair-wise comparisons. Table 3 contains the numbers of specific grades in the respective scales. Table 4 contains PCMs of the 3 experts, brought to the most detailed scale with 9 grades (see section 1 of this paper) according to the rules, proposed in [3]. The unified value (brought to the more detailed scale from the less detailed one) does not always coincide with exact grade value. Pair-wise comparisons, skipped by the expert, are replaced by 1 («no preference»).

Table 2 Number of grades in scales, selected by experts for input of comparisons

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	9	*	7	1	3	4	5	1	9	9	8
A_2		1	6	5		1	6	7		1	3	9
A_3			1	4			1	8			1	7
A_4				1				1				1

Table 3 Numbers of specific grades in the scales, selected by experts

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	2	*	7	1	3	4	5	1	2	4	8
A_2		1	3	4		1	2	4		1	2	4
A_3			1	2			1	2			1	3
A_4				1				1				1

Table 4 Pair-wise comparison values, brought to the most detailed scale

	E_1				E_2				E_3			
	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4	A_1	A_2	A_3	A_4
A_1	1	2	1	8 5/6	1	7 1/2	8 1/6	8 1/2	1	2	4	9
A_2	1/2	1	3 8/9	6 1/2	1/7	1	2 2/7	4 5/6	1/2	1	3 1/2	4
A_3	1	1/4	1	2 5/6	1/8	3/7	1	2	1/4	2/7	1	3 1/2
A_4	1/8	1/6	1/3	1	1/8	1/5	1/2	1	1/9	1/4	2/7	1

Based on matrices from table 4, 48 ICPCM are built ($mn^{n-2} = 3 \times 4^2 = 48$). In order to compare each of these ICPCMs with original PCMs, 3 copies of eeach ICPCM are made (according to the number of experts). That is, the general number of ICPCMs being processed (and, respectively, the number of multipliers in formula (9)) is $T = m^2 n^{n-2} = 144$. Ratings of 144 matrices are calculated according to formula (8) (if the spanning tree lacks the pair-wise comparison, skipped by the 1st expert, then the rating of the respective matrix equals 0). After that, the ratings are normalized by their sum (see power index in formula (9)). At the same time, alternative weight vectors (priorities) are calculated based on every ICPCM. Any basic set of pair-wise comparisons (for instance, the last row of an ICPCM) can be used as non-normalized weight vector. Finally, non-normalized aggregate priorities are calculated according to formula (9) and normalized.

Normalized priority values, calculated according to formula (9) and according to formula (7) (i.e. not taking expert competence into consideration) are shown in table 5.

Table 5 Priority vectors, incorporating and not incorporating expert competence values

	W_1	W_2	W_3	W_4
Competence taken into consideration (formula (9))	0.533	0.264	0.133	0.071
Competence not taken into consideration (formula (7))	0.547	0.269	0.129	0.054
Difference (%)	2.6	1.9	3.0	23.9

As we can see, in spite of perfect ordinal and fairly high cardinal agreement of original data, and in spite of usage of additive (and not multiplicative) expression for ICPCM rating calculation, the results obtained using simple and weighted average mean are different. The greatest difference is witnessed between the smallest priority values.

The example confirms that if the level of agreement, completeness, and detail of initial data from a given source is low, then the final aggregation results will be significantly influenced by the source's relative weight.

We should also stress, that if data sources are experts, then feedback can be organized to improve completeness, consistency, and compatibility of the data. Feedback with experts can be based on spectral approach, suggested in [16] and further elaborated in [17]. However, if data sources are not experts (humans), then feedback cannot be organized, and data source weight components, reflecting the above-listed aspects, will play far greater role than in the cases when experts are involved.

6 Conclusions

It has been shown that during aggregation of data from different sources it is necessary to take their relative weights into account. Moreover, in order to calculate the relative weight of a data source it is not enough to just estimate it a priori. The weight of a data source should incorporate completeness, agreement, and degree of detail of information, provided by the source. A method for calculating the relative weights of data sources has been suggested. Beside a priori estimates, the data source weight coefficient incorporates the degree of detail of scales, in which the data is provided, as well as consistency and compatibility of the data. The method has been tested on a large number of numeric examples, where initial data was input in the form of individual PCMs and aggregated. Obtained experimental results illustrate significant differences between aggregate priority values, respectively, incorporating and not incorporating data source weights. The method's key advantage is its universal nature: it allows us to take all the above-mentioned aspects of data source credibility into account, not just some single aspect. The original data source weight calculation mechanism provides the basis for improvement of pair-wise comparison aggregation methods and for increasing the credibility of results of their application.

This paper is prepared as part of project #F73/23558 "Development of Decision-making Support Methods and Means for Detection of Information Operations". The project won the contest #F73 for grant support of scientific research projects held by The State Fund for Fundamental Research of Ukraine and Belarusian Republican Foundation for Fundamental Research.

References

1. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія – К., Інтертехнологія, 2009 – 164 с.
2. Kadenko S.V. Prospects and Potential of Expert Decision-making Support Techniques Implementation in Information Security Area / S.V.Kadenko // CEUR Workshop Proceedings – 2016 – pp. 8-14.
3. Tsyganok V.V. Using Different Pair-wise Comparison Scales for Developing Industrial Strategies / V.V.Tsyganok, S.V.Kadenko, O.V.Andriichuk // Int. J. Management and Decision Making. – 2015. – Vol. 14, No 3. – pp. 224–250.
4. Saaty, T.L., 1994. Fundamentals of Decision Making and Priority Theory with The Analytic Hierarchy Process. / T.L.Saaty; RWS Publications, Pittsburgh PA, 204220.
5. Ramanathan, R. Group Preference Aggregation Methods Employed in AHP: An Evaluation and Intrinsic Process for Deriving Members' Weightages / R.Ramanathan, L.S.Ganesh// European Journal of Operational Research 79 – 1994, pp. 249-265.
6. Yang Q, Du P-a, Wang Y, Liang B. A rough set approach for determining weights of decision makers in group decision making. PLoS ONE 12(2) – 2017: e0172679. doi:10.1371
7. Hwang, C. L. Multiple attribute decision making: methods and applications : a state-of-the-art survey / C.L. Hwang, K. Yoon ; Berlin ; New York : Springer-Verlag, 1981 – 259 p.
8. Totsenko, V. G. Determination of Relative Competence of Group Members in the Subject under Discussion on Group Decision Making / V.G.Totsenko // JAutomatInfScien.v34.i4.30 – 2002. DOI: 10.1615
9. Гнатієнко Г. М. Експертні технології прийняття рішень / Г.М.Гнатієнко, В.Є.Снитюк. – К.: ТОВ "Маклаут", 2008 – 444 с.
10. Андрійчук О.В. Побудова баз знань систем підтримки прийняття рішень при виявленні інформаційних операцій / О.В.Андрійчук, Д.В.Ланде // Реєстрація, зберігання і обробка даних. Щорічна підсумкова наукова конференція – Київ, 2017 – с. 101-103.
11. Циганок В.В. Комбінаторний алгоритм парних порівнянь зі зворотним зв'язком з експертом / В.В.Циганок // Реєстрація, зберігання і обробка даних. – 2000. – Т.2, №2. – с.92-102.
12. Циганок В. В. Метод обчислення ваг альтернатив на основі результатів парних порівнянь, проведених групою експертів / В.В.Циганок // Реєстрація, зберігання і обробка даних. – 2008. – Т.10, №2. – с. 121-127.
13. Tsyganok, V. Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method / V.Tsyganok // Mathematical and Computer Modeling, 52(3-4) – 2010 – pp. 538–544.
14. Hartley R.V.L. Transmission of information. / R.V.L.Hartley // Bell System Technical J. 1928 – 7, pp. 535-63.

15. Cayley, A.A Theorem on Trees. / A.Cayley // Quarterly Journal of Mathematics 1889 – Volume 23, pp. 376-378.
16. Totsenko V.G., Tsyganok V.V. Method of paired comparisons using feedback with expert. Journal of Automation and Information Sciences. – 1999. – Vol.31, No9. – P.86-97.
17. Olenko Andriy & Tsyganok Vitaliy Double Entropy Inter-Rater Agreement Indices. Applied Psychological Measurement. – 2016. – v.40(1). – P.37-55.