STABILITY OF RANKING-DEPENDENT PAIR-WISE COMPARISON PATTERNS IN THE ANALYTIC HIERARCHY PROCESS

Author 1: *Sergii Kadenko*[1]
Author 2. *Oleh Andriichuk*[2]
Author 3. *Vitalii Tsyganok*[3]

Highlights

- The paper contains an outline of several ranking-dependent pair-wise comparison patterns.
- Such patterns and respective matrices can be both complete and incomplete.
- So, we try to find the specific conditions under which these patterns can be compared in terms of stability to expert errors.
- We also suggest a simulation algorithm, allowing to experimentally compare stability of several ranking-dependent comparison patterns which can be used in the Analytic Hierarchy Process.

ABSTRACT

The paper is dedicated to ranking-dependent pair-wise comparison patterns which can be used in the Analytic Hierarchy Process. Ordinal information on compared objects can be used to improve the quality of expert data during estimation and, potentially, reduce the number of comparisons that the experts need to perform. In our research we focus on several ranking-dependent approaches – Best Worst method, Best-Second Best (Top 2) method, and the original "queues" method. The first two comparison patterns (and respective methods) are incomplete, while the third is a complete one. We determine conditions under which these three methods can be compared in terms of stability to expert errors. We also develop an algorithm, according to which a simulation-type experiment

[1] Sergii Kadenko, Ph D, senior researcher, Department of Intelligent Decision Support Technologies, Institute for Information Recording of the National Academy of Sciences of Ukraine, M. Shpak str., 2, Kyiv, 03113, Ukraine, e-mail: seriga2009@gmail.com (ORCID: 0000-0001-7191-5636).

[2] Oleh Andriichuk, Ph D, senior researcher, Department of Intelligent Decision Support Technologies, Institute for Information Recording of the National Academy of Sciences of Ukraine, M. Shpak str., 2, Kyiv, 03113, Ukraine, e-mail: andreychuck@ukr.net (ORCID: 0000-0003-2569-2026).

[3] Vitalii Tsyganok, Doctor of Engineering, Head of Department of Intelligent Decision Support Technologies, Institute for Information Recording of the National Academy of Sciences of Ukraine, M. Shpak str., 2, Kyiv, 03113, Ukraine, e-mail: Vitaliy.Tsyganok@gmail.com (ORCID: 0000-0002-0821-4877).

should be organized. Subsequent experimental study will allow us to compare the three listed approaches according to their stability.

Keywords: expert estimate, incomplete pair-wise comparison matrix, ranking.

1. Introduction and literature review

Our present research addresses cognitive and algorithmic aspects of knowledge transfer and decision support. Particularly, research of specific pair-wise comparison (PC) patterns in the Analytic Hierarchy Process (AHP) and other related methods is intended to improve the quality of expert data during estimation and, potentially, reduce the number of comparisons that the experts need to perform.

In order to obtain relative wights of n objects in AHP, we need to perform from $n-1$ to $n(n-1)/2$ PC. As the number of compared objects grows, the task becomes more labor-intensive. In order to simplify the process and reduce the computational complexity of AHP, while preserving sufficient levels of credibility and redundancy of expert data, different incomplete PC methods have been developed. Incomplete PC methods have been covered in multiple publications (such as (Wedley, 2009), (Szádoczki, Bozóki & Tekile, 2022)). PC methods, taking rough ranking of compared alternatives into consideration represent a separate group among them. A series of recent publications (including (Andriichuk & Kadenko, 2022) addresses these particular methods.

While comparisons of incomplete PC methods in terms of stability to expert errors have been outlined in several studies (such as (Szádoczki, Bozóki & Tekile, 2022)), a separate comparative analysis of ranking-dependent PC methods has not been conducted yet.

Algorithms for such comparative studies of PC methods as to their stability have been developed quite a long time ago. As AHP and related methods are mostly applied in weakly structured subject domains, influenced by multiple intangible factors, where there are no benchmark values, and engaging actual experts in such experiments requires considerable resources, it makes sense to focus on simulation-based modeling of expert estimation process. Therefore, experimental research of a PC method's stability might include the following conceptual phases:

1. Generate a set of relative wights of compared objects $(w_1, \dots, w_n)$, i.e., priority vector
2. Generate a PC matrix (PCM) of judgments based on these relative object weights $\{a_{ij} = \frac{w_i}{w_j}; i,j = 1..n\}$
3. Generate an incomplete PCM based on the complete one
4. Fluctuate the generated PCM (for instance, $\{a'_{ij} = a_{ij} \pm \varepsilon; i,j = 1..n\}$)
5. Calculate priorities $(w'_1, \dots, w'_n)$ using some given method (eigenvector, least squares, spanning tree enumeration, geometric mean) based on the fluctuated PCM
6. Calculate (ordinal and/or cardinal) difference between initial and calculated priority vectors Δ. This difference can be represented by any of the indicators listed in (Szádoczki et al, 2023).

Maximum deviation of the weight vector, obtained based on the fluctuated PCM, from the initial vector, may be used as the indicator of a method's stability to expert errors. According to this indicator different PC methods are compared with each other.

We should note that in (Szádoczki et al, 2023) PC methods, using different amounts of ordinal information (Best Worst, Best Random, 3(-quasi)-regular-graphs, others), have been compared in terms of both ordinal and cardinal stability. The issues of legitimacy of such "mixed" comparisons became relevant before the article's publication (Szádoczki et al, 2023) and has remained the subject of discussion ever since.

2. Objectives

Our main objective is to compare the three chosen ranking-dependent PC methods: Best Worst (Rezaei, 2015), Top 2 (Best-Second Best) (Szádoczki et al, 2023), and the original "queues" method (Andriichuk, Kadenko, & Tsyganok, 2024). Additionally, we would like to validate the "queues" method by comparing it with other ranking-dependent ones. However, before conducting a simulation-type comparison experiment, described above, we need to define the conditions, under which the three listed methods can be compared among themselves (i.e., conditions allowing us to consider the results of the methods' experimental comparison according to stability criterion credible enough).

3. Methodology

Let us start with a quick overview of the methods under consideration.
If the rough ranking of compared objects is available, then the Best Worst method suggests comparing all the objects only to the best and the worst one in the set (according to the ranking). Respectively, the Top 2 method suggests comparing all the objects only with the best and the second-best ones in the ranking. PC structures of Best Worst and Top 2 methods can be easily illustrated by graphs. The looks of such graphs for the case when 7 objects are compared is shown on Fig. 1. Sizes of nodes indicate the ranks of respective objects.

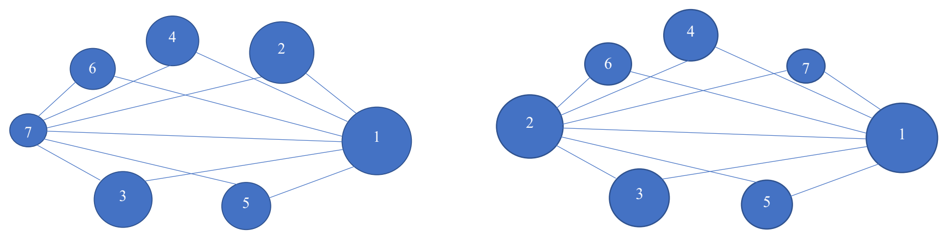

Fig. 1 Graphs of PC structures for 7 objects: Best Worst method (left) and Top 2 method (right)

In both Best Worst and Top 2 methods each object is compared to only two objects in the ranking (either the best and the worst or the best and the second best). Therefore, both methods are incomplete ones (not all PCM cells contain independent PC values), and the number of comparisons amounts to $(n-1)+\big((n-1)-1\big)=2n-3$ PC. If the objects in the ranking are numbered from best to worst, then the respective incomplete PCM for Best Worst (left) and Top 2 (right) methods look as follows:

$$\begin{pmatrix} 1 & a_{12} & \cdots & \cdots & a_{1n} \\ & 1 & \cdots & & \\ & & 1 & \cdots & \cdots \\ & & & 1 & a_{n-1,n} \\ & & & & 1 \end{pmatrix}\begin{pmatrix} 1 & a_{12} & \cdots & \cdots & a_{1n} \\ & 1 & a_{23} & & a_{2n} \\ & & 1 & \cdots & \cdots \\ & & & 1 & \cdots \\ & & & & 1 \end{pmatrix}$$

"Queues" method envisions a certain sequence of PC, defined by the rough ranking of the objects. Let's assume, the objects are numbered according to their rank order: $a_1 > a_2 > \cdots > a_n$, where a_i is the object with rank and number i, $i = \overline{1, n}$ and n is the total number of objects. In this case the sequence of comparisons (i.e., of object pairs presented to the expert), ensuring the highest relevance and consistency of the results (Andriichuk, Kadenko, & Tsyganok, 2024) is as follows: queue 1: (a_1, a_n) (ranks differ by $(n-1)$); queue 2: (a_1, a_{n-1}) or (a_2, a_n) (ranks differ by $(n-2)$); queue 3: (a_1, a_{n-2}) or (a_2, a_{n-1}) or (a_3, a_n) (ranks differ by $(n-3)$); … ; queue $(n-1)$: (a_1, a_2) or (a_2, a_3) or … or (a_{n-1}, a_n) (ranks differ by 1). The respective PCM is filled as shown on Fig. 2.

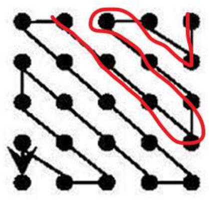

Fig. 2 The pattern of PCM filling in the "queues" method

In (Andriichuk, Kadenko, & Tsyganok, 2024) it is shown that it is not necessary to perform all PC in "queues" method. All PC within the same queue are "equally important", so we can confine ourselves to PC which are required for obtaining a connected PC structure (graph).

If n is an odd number, then in order for the PC graph to become connected it is enough to perform $\left[\frac{n}{2}\right]$ or $(n-1)/2$ PC queues + 1 PC from the next queue, i.e. queue number $((n-1)/2)+1$. This is the PC of objects $(a_1, a_{\frac{n-1}{2}+1})$, i.e. the 1st and the "central" element in the ranking. Likewise, it can be the PC of the "central" element and the last one $(a_{\frac{n-1}{2}+1}, a_n)$, as all PC within a queue are equivalent, and objects can be numbered from best to worst or vice versa. The point is to include the "central" ranking element into the PC structure, so that it becomes connected.

If n is an even number, then there will be two "central" elements in the ranking. Therefore, in order to obtain a connected comparison structure, we need to perform $\left(\frac{n}{2}\right) - 1$ PC queues + 2 PC from queue number $(n/2)$: $(a_1, a_{\frac{n}{2}})$, $(a_{\frac{n}{2}+1}, a_n)$. An example of obtaining a connected comparison graph for 7 objects is shown on Fig. 3.

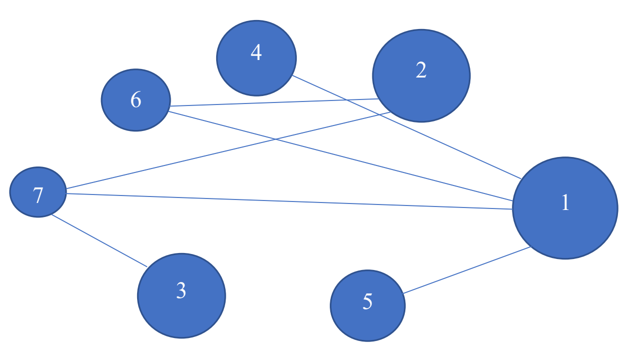

Fig. 3 A connected set of PC of 7 objects in "queues" method

In this example queue 1 includes a single PC (a_1, a_7), queue 2 – PC $\{(a_1, a_6), (a_2, a_7)\}$, queue 3 – PC $\{(a_1, a_5), (a_2, a_6), (a_3, a_7)\}$, queue 4 – PC $\{(a_1, a_4), (a_2, a_5), (a_3, a_6), (a_4, a_7)\}$ and so on. Connectivity is achieved after PC (a_1, a_4) or, alternatively, (a_4, a_7).
By definition, the number of the queue i equals the total number of PC in it. Therefore, the total number of PC in queues from 1 to i equals the sum of an arithmetic progression: $N_i = i(i+1)/2$. That is, in order to obtain a connected set of PC of n objects, compared using the "queues" method, we need to perform the following number of comparisons:

$$N_{conn} = \begin{cases} \left(\frac{n-1}{4}\left(\frac{n-1}{2}+1\right)\right) + 1, n \text{ is odd}, n > 1 \\ \left(\frac{n}{4}\left(\frac{n}{2}-1\right)\right) + 2, n \text{ is even}, n > 2 \end{cases} \quad (1)$$

In order to compare n objects among themselves using either Best Worst or Top 2 method, we need to perform $(2n-3)$ PC (as explained earlier). The "queues" method, in its turn, is a complete one. So, to place the listed 3 methods (Best Worst, Top 2, "queues") in equal conditions and be able to compare their stability to expert errors, we have to ensure that the number of PC in all 3 methods is equal. Therefore, the condition

$$N_{conn} \leq (2n-3) \quad (2)$$

should be fulfilled.

$$\left(\frac{n-1}{4}\left(\frac{n-1}{2}+1\right)\right) + 1 \leq (2n-3), n \text{ is odd}, n > 1$$

$$\left(\frac{n}{4}\left(\frac{n}{2}-1\right)\right) + 2 \leq (2n-3), n \text{ is even}, n > 2$$

Having solved these square inequalities, we can deduce that condition (2) is fulfilled when $3 \leq n \leq 14$.

4. Results Analysis

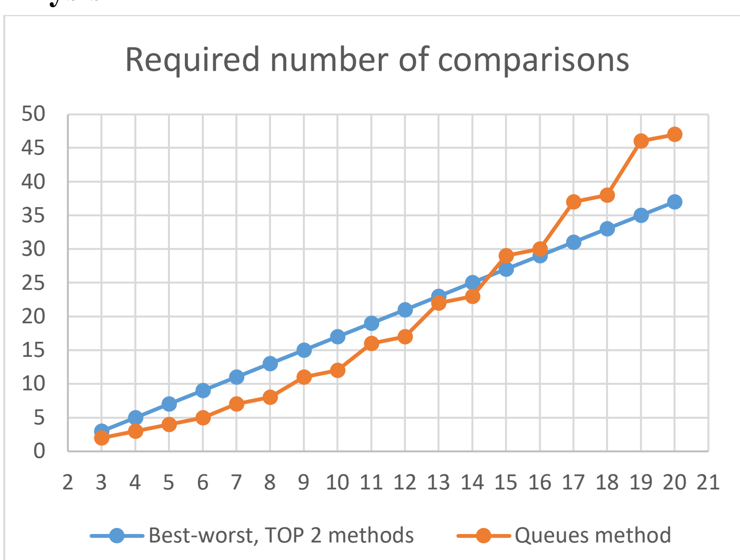

Fig. 4 The number of PC, required for priority vector calculation in Best Worst, Top 2 and "queues" methods

Minimum numbers of PC, required for calculation of priorities using Best Worst, Top 2, and "queues" methods, as functions of n are shown on Fig. 4. For Best Worst and Top 2 methods this dependence function is a linear one ($N = O(n)$), i.e. its graph is a straight line. For "queues" method, this dependence function is a square one ($N_{conn} = O(n^2)$). Its graphs for odd and even n are parabolic. It means that when the number of objects lies within the limits $3 \leq n \leq 14$, in order to compare the 3 methods, we should generate $(2n - 3)$ PC in "queues" method, according to the sequence suggested by the queues. That is, after the "connectivity number" of comparisons N_{conn} is achieved, we should keep generating PC until their number reaches $(2n - 3)$.
If $n > 14$, then condition (2) in "queues" method cannot be fulfilled. In this case we should start generating PC in "queues" method with a basic PC set, suggested in (Andriichuk, Kadenko, & Tsyganok, 2024), that is a bi-partite spanning tree graph, in which a_1 (the best object) is compared to all objects from the 2nd half of the ranking $\left(a_{\left[\frac{n}{2}\right]+1}, \ldots, a_n\right)$, while a_n (the worst object) is compared to all objects from the 1st half of the ranking $(a_1, \ldots, a_{\left[\frac{n}{2}\right]})$. Diameter of this graph equals 3 and does not depend on n, while its edges represent comparisons from first queues. These properties of the graph ensure stability of this initial PC structure to expert errors. An example of such a spanning tree graph for 7 objects is shown on Fig. 5.

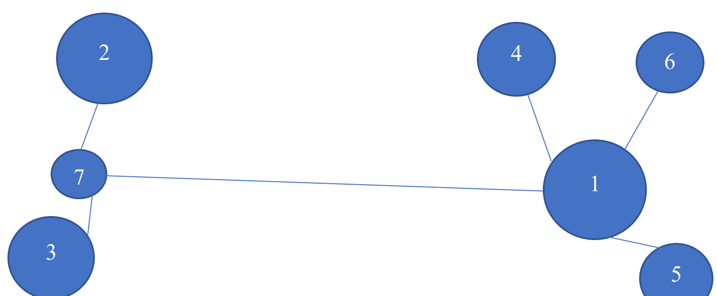

Fig. 5 Spanning tree graph, representing the minimum connected PC set for 7 objects, built using "queues" method

In fact, the only difference between graphs shown on Fig. 5 and Fig. 3 is the edge corresponding to PC (a_2, a_6), which makes PC set on Fig. 3 redundant and is missing on Fig. 5.
Once this initial connected set of $(n - 1)$ PC is built, we should add PC to the set, following the order of queues, until the number of PC reaches $(2n - 3)$. A basic flow chart of PC generation algorithm for "queues" method can be found in the Appendix.
Once the necessary number of PC is generated, we can compare the "queues" method with Best Worst and Top 2 methods, following the procedure outlined in the introduction (steps 3 to 6).
Before moving on to conclusions, we would like to focus on PCM fluctuation process, intended to simulate expert errors. Fluctuation can be additive or multiplicative and fluctuated PCM elements can be rounded to the nearest grade of Saaty's fundamental scale. Analysis of results obtained in (Andriichuk, Kadenko, & Tsyganok, 2024) indicates that adequacy of expert estimates is significantly influenced by the order in which pairs of

objects are presented to the expert for comparison. Particularly, PC belonging to queues with smaller numbers are more credible than PC from queues with larger numbers. Benefits of “queues” method include higher adequacy of priorities and better consistency (CR) of judgments (Andriichuk, Kadenko, & Tsyganok, 2024). Therefore, in our view, it is irrelevant to apply simple additive or multiplicative fluctuations to PCM during the future comparison of the three methods. That is, while fluctuating PCM elements (step 4 of the simulation procedure set forth in the Introduction section), we need to take specific queues these elements belong to into account. Based on our previous studies, we assume that comparisons from queues with smaller numbers are less prone to errors than comparisons from subsequent queues. During simulation process this can be achieved through modification of ordinary additive or multiplicative fluctuations in such a way that fluctuation of a particular PCM element depends on the number of the queue the respective comparison belongs to. Modified additive (3) and multiplicative (4) fluctuation formulas might represent a convex combination of ordinary deviation and ranking-dependent deviation as shown below:

$$a'_{ij} = a_{ij} \pm \left(\alpha\varepsilon + (1-\alpha)\frac{\varepsilon}{n-k}\right); i,j = 1..n; 0 < \alpha < 1; \qquad (3)$$

$$a'_{ij} = a_{ij}(1 \pm (\alpha\delta + (1-\alpha)\ \delta/(n-k)); i,j = 1..n; 0 < \delta < 1;\ 0 < \alpha < 1; \qquad (4)$$

In (3) and (4) k is the number of the queue the respective PC belongs to. By definition of the queues the number of the queue is determined by the distance between the respective compared objects in the ranking: $k = n - |i-j|$. So, formulas (3) and (4) can be rewritten as follows:

$$a'_{ij} = a_{ij} \pm \left(\alpha\varepsilon + (1-\alpha)\frac{\varepsilon}{|i-j|}\right); i,j = 1..n; 0 < \alpha < 1; \qquad (3a)$$

$$a'_{ij} = a_{ij}(1 \pm (\alpha\delta + (1-\alpha)\ \delta/|i-j|); i,j = 1..n; 0 < \delta < 1;\ 0 < \alpha < 1; \qquad (4a)$$

5. Conclusions

In the present paper we have prepared the background for conducting an experimental comparative study of stability of several ranking-dependent PC methods (Best Worst, Top 2, “queues” method) to expert errors. We have defined the conditions under which such an experiment can be performed and its results can be deemed valid. We have also suggested modifications of a typical simulation-type experiment required to compare the three listed methods in terms of stability. As all listed ranking-dependent PC patterns can be used in AHP, we consider the current research and subsequent comparative experimental study an important contribution to AHP theory and methodology, as well as to analysis of cognitive and algorithmic aspects of decision making in general.

6. Limitations

Apart from heuristic nature of expert estimation process, which always calls for additional validation, we would like to stress the psycho-physiological limitations of human mind. In previous simulation-type studies similar to the current one (Szádoczki, Bozóki & Tekile, 2022) dimensionality of analyzed PCM ranged from $n = 4$ to $n = 24$. In our research we have found a kind of a threshold dimensionality value $n = 14$. At the same time, an average expert is capable of ranking and rating only $n = (7 \pm 2)$ objects (with some degree of credibility). Therefore, we should note that actual estimation problems are, usually, limited to these dimensionalities.

7. Key References

Andriichuk, O. & Kadenko, S. (2022). Ranking-dependent Decision Support Methods for Weakly-structured Domains. *CEUR. Workshop Proceedings, Vol. 3503*, 32-41.

Andriichuk, O., Kadenko, S., & Tsyganok, V. (2024). Significance of the order of pair-wise comparisons in Analytic Hierarchy Process: an experimental study. *Journal of Multi-Criteria Decision Analysis, 31(3-4),* e1830. https://doi.org/10.1002/mcda.183.

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega, 53*, 49–57. https://doi.org/10.1016/j.omega.2014.11.009.

Szádoczki, Z., Bozóki, S., Juhász, P. et al. (2023) Incomplete pairwise comparison matrices based on graphs with average degree approximately 3. *Ann Oper Res 326, 783–807*. https://doi.org/10.1007/s10479-022-04819-9.

Szádoczki, Z., Bozóki, S., & Tekile H. (2022). Filling in pattern designs for incomplete pairwise comparison matrices: (quasi-)regular graphs with minimal diameter. *Omega, 107*, doi: 10.1016/j.omega.2021.102557.

Wedley, W. C. (2009). Fewer Comparisons – Efficiency via Sufficient Redundancy. *Proceedings of the 10th International Symposium on the Analytic Hierarchy/Network Process. Multi-criteria Decision Making, Pitsburg, Pensilvania, USA, July 30 – August 2, 2009.* Available at: http://www.isahp.org/2009Proceedings/Final_Papers/94_Wedley_Fewer_Comparisons_REV_FIN.pdf. Accessed on Oct 30, 2024.

8. Appendix

Flow-chart of the PC generation algorithm enabling us to compare "queues" method with Best Worst and Top 2 methods