GROUP DECISION MAKING

Abstract

The process of management is based on constant decision making, where the decisions are becoming increasingly complex and multifaceted. In the modern world most decisions are made by groups of people, often from various disciplines. The ability to use multicriteria methods (e.g. AHP/ANP) of supporting decisions is one of the key challenges that modern managers face. That's why the methods supporting decision making are an inherent element of the knowledge of management science. During the current track and organized sessions we will try to considerably expand the knowledge in this area with using AHP/ANP.

Key words: group multicriteria decision making, GMCDM, AIJ, AIP, Anlytic Hierarchy Process, Analytic Network Proces, AHP/ANP

AGGREGATING PAIR-WISE COMPARISONS GIVEN IN SCALES OF DIFFERENT DETAIL DEGREE

V.V.Tsyganok [*]
E-mail: vitaliy.tsyganok@gmail.com

O.V.Andriichuk
E-mail: andreychuck@ukr.net

National Academy of Sciences of Ukraine
The Institute for Information Recording
Kyiv, Ukraine

ABSTRACT

In this paper we suggest an original approach to conducting individual pair comparisons during group decision-making (including AHP/ANP-based decisions). Under this approach every expert is given an opportunity to use the scale, in which degree of detail (number of points) most adequately reflects this expert's competence in the issue under consideration. Before aggregation all separate expert estimates (judgments) are brought to a unified scale, and scales, in which these judgments were built, are assigned respective weights. A respective instrument for pair comparison conduction has been developed, and an experiment has been organized. The experiment statistically proves that as a result of suggested technology usage, there is an increase in degree of correspondence between estimates, input by an expert, and his (her) own notions on examination objects.

Keywords: group decision making, expert judgments, pair-wise comparisons, scales of different detail degree.

1. Introduction

The practice of expert examination conduction (including AHP-based examinations) indicates that there are certain difficulties arising when verbal scales are used for expert examination. Expert/decision-maker is often offered to use only one scale for pair comparisons. Judging from experience, in order to get thorough and undistorted data from an expert, (s)he must be offered to input estimates in a scale, which most adequately corresponds to his/her competence (awareness) level in the issue under consideration. The suggested research resolves the issue of using verbal scales with different degree of detail for each particular pair comparison, in order to ensure maximal credibility of knowledge obtained from an expert (expert information must be thorough and undistorted).

[*] Corresponding author

To ensure thoroughness of information obtained from an expert, we suggest using verbal scales of sufficient degree of detail: the more points the scale includes, the more information an expert can, potentially, input into a DSS, using this particular scale. To avoid information distortion (if an expert is unsure of the degree of dominance between objects in a pair, i.e. (s)he is not competent enough), we suggest giving experts the opportunity to use scales with low degree of detail, or, even, refuse to estimate preferences in certain object pairs. Besides, in our research we also consider an important factor, influencing the level of expert information distortion – quantitative equivalent, corresponding to this or that point of a verbal scale. Correspondence between preference value input by an expert and this expert's notions about ratio of object weights on a pair is an issue of great importance, as it influences the credibility of expert data-based decision-making recommendations.

2. Literature Review

The key studies in the described area include the recent research by (Elliott, 2010), addressing the influence of a chosen quantitative scale upon correspondence between estimation results and expert's own notions. It was demonstrated, that scale selection has considerable impact upon the resulting decision variant estimate. Three quantitative scale types were analyzed, whose point values were assigned to fundamental scale points (Saaty, 2006) of two kinds, i.e. scales with 5 and 9 grades. Quantitative scales under consideration included integer, balanced (Salo & Hamalainen, 1997) and power (Stevens, 1957) scales. In contrast to research, described in the listed publications, in our study we suggest choosing a different scale for each single pair comparison, and not for all pair comparisons.

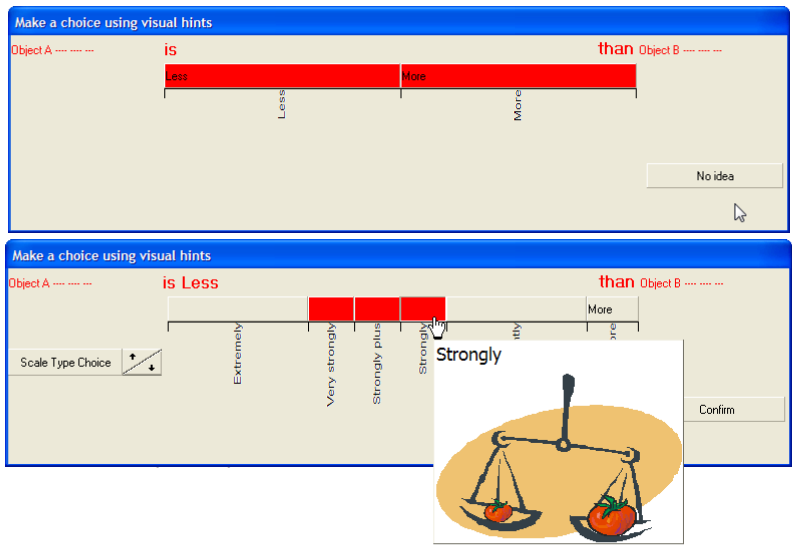
3. Hypotheses/Objectives

The purpose of this study is to prove, that to ensure obtaining of thorough and undistorted expert information on relation between objects (on estimates provided during pair comparisons), an expert should be given an opportunity to use scales with different degree of detail (accuracy). This hypothesis is based on a presumption that in every issue under consideration (and in every pair comparison) an expert has a different level of knowledge/competency/awareness. Each expert's competence level can correspond to a respective estimation scale: the higher the expert's competence is, the more detailed scale (s)he can use to adequately present his(her) knowledge. According to the same principle, an uninformed/incompetent expert should have an opportunity to use a scale with small number of grades (including ordinal scale with only two values – "more" or "less") for pair comparisons, or even refuse to compare objects in a pair because of incompetence. It is understandable that an expert judgment provided in a more detailed scale should be considered more significant than that same judgment provided in a less detailed scale, because in the first case the expert is more confident, and his self-estimated competence in the issue under consideration is higher. Consequently, if during pair comparisons an expert considers objects equal, this judgment can be considered the same as refusal to conduct this particular comparison (inability to evaluate preference of objects in a pair due to doubts/low competency in the issue under consideration). As we see, in verbal scales there is no real need for a grade "equal"/"no preference", because if an expert chooses this value, (s)he might as well "skip" (refuse to estimate) respective preference. Anyway, the choice of "equal" preference value does not introduce any additional information on relation between objects.

Proof (confirmation) of any hypothesis in a weakly structured domain (in which we are conducting our research) is problematic, as there are absolutely no benchmarks to compare results with. That is why, the only way to confirm the hypothesis is an experiment using estimates provided by experts. Such an experiment is described in section 5 of this paper.

4. Research Design/Methodology

During the research a methodology and respective software tools were developed to conduct expert estimation based on the abovementioned approach. In group estimation every expert is offered to provide pair comparisons in verbal scales with different degrees of detail. Each particular pair comparison starts with the scale including only two values («Less» и «More») with an opportunity to refuse to provide the judgment – «No idea» (Figure 1 *a*).



a)

b)

Figure 1 Software tools for gradual estimate precision increase

If ordinal comparison is provided (one of the values «Less» or «More» is selected) the expert is offered to gradually make the estimate more precise, and stop estimation at any stage («Confirm» button on Figure 1 *b*). In the process of this iterative procedure the final estimate is conducted in the scale, which most adequately corresponds to expert's competence in the issue of defining the preference relation between two particular objects. The final estimate may be provided in a scale including 2 to 8 grades.

It should be noted that the developed tool allows an expert to be sure that the quantitative equivalent really corresponds to this or that verbal phrase from estimation scale. Such confidence is achieved through providing user (expert) with interactive graphic tips (hints), allowing him to imagine the approximate relation between objects, and, thus, improve the degree of correspondence between the expert's personal notions and the information (s)he inputs during pair comparisons.

For aggregation of incomplete comparison matrices provided by a group of experts, when different comparisons can be conducted in scales with different accuracy, we suggest using the method known as enumerating all spanning trees with further averaging of priority vectors, calculated based on every tree (Tsyganok, 2010). Before calculation of priority vectors, all pair comparison matrix elements (judgments) are brought to a unified (most detailed) scale. During this process weights of particular judgments (pair comparisons) are taken into consideration. The weights depend on the degree of detail of scales the comparisons were provided in. The scale's degree of detail (informativeness) is calculated according to Hartley's formula for quantity of information: $I = \log_2 N$, where N – number of points in expert estimation scale.

5. Data/Model Analysis

To confirm the hypothesis set forth in section 3 of this paper, an experimental research was conducted with real experts involved. As there are no benchmarks (model values), only the result of individual (not group) expert examination was analyzed. Every experiment participant was offered to select a subject domain, (s)he is competent in, and freely formulate an understandable goal. After that the participant (expert) was offered to formulate 5 to 7 factors making positive impacts upon the formulated goal. It should be noted that, since every expert chooses the subject domain (s)he is familiar with, (s)he must also be aware of contributions of each formulated factor into the goal's achievement.

During the next stage pair comparisons of impacts of formulated factors were conducted. Experts were offered to conduct further comparisons in 3 ways: in the fundamental scale with 5 grades ('Equivalent' (1), 'Moderately' (3), 'Strongly' (5), 'Very strongly' (7), 'Extremely' (9)), in the fundamental scale with nine grades ('Equivalent' (1), 'Weakly or slightly' (2), 'Moderately' (3), 'Moderately plus' (4), 'Strongly' (5), 'Strongly plus' (6), 'Very strongly' (7), 'Very, very strongly' (8), 'Extremely' (9)), and using the suggested tool. In order to minimize the correlation between repeated comparisons of same pairs of objects provided in different ways (every pair was compared three times – each time in a different way) the sequence of pairs presented to an expert for comparison was randomized.

After all pair comparisons were performed (3 pair comparison matrices were filled), 3 priority vectors were calculated. Eigenvector method was used to process matrices, built using the first two approaches, while to define a priority vector based on a matrix including comparisons provided in different scales, the so-called combinatorial (or spanning tree enumeration) method (Tsyganok, 2010) was used (particularly, its modification allowing for usage of different weights for different estimation scales).

At the final stage every experiment participant was offered to rank 3 priority vectors calculated for the factors (s)he formulated. Vectors were displayed as unsigned bar charts in random order. The participant was offered to rank the vectors according to their correspondence to his/her perceptions of quantitative relations between impacts of the formulated factors.

Result obtained by an expert (experiment participant) qualified only if pair comparison matrices satisfied sufficient consistency condition (C.I. value). Statistically credible results were obtained. These results are presented in Table 1. As a result of the research, we can conclude that in most of the analyzed cases, expert estimates obtained using the suggested technology, are more consistent with experts' individual perceptions of examination subject, in comparison to estimates, based on traditional estimation techniques (where fixed number of verbal scale grades is used). Consequently, wide implementation of the suggested pair comparison instrument in decision support technologies (including those using AHP/ANP) seems adequate.

Table 1 Comparative experimental research results

Name of pair comparison technology	Number of participants, who assigned the specified rank to respective technology		
	„1”	„2”	„3”
Fundamental preference scale with 5 grades	10	15	37
Fundamental preference scale with 9 grades	12	33	17
Technology suggested in the paper	41	15	7

6. Limitations

Usage of the suggested tool for pair comparisons may require longer time during expert estimation, and, as a result, more resources, than traditional methods. This may result from the fact that more actions are required from experts during each pair comparison. But in reward we get higher credibility of expert estimates and recommendations to decision makers.

7. Conclusions

As a result of the research, we have suggested an expert estimation mechanism, allowing experts to use scales of different accuracy for each pair comparison. Relevance of the suggested approach is experimentally proven. It has been demonstrated that usage of the respective tool for pair comparisons allows us to improve the degree of correspondence between expert's estimates and his notions of examination subject. This improvement results from the fact that experts use scales, whose accuracy is most consistent with their competency in every issue under consideration.

Implementation of the suggested expert estimation technology in combination with pair comparison matrix aggregation methods (including group methods) improves the credibility of AHP/ANP-based recommendations, given to decision makers.

8. Key References

Elliott, M.A. (2010). Selecting numerical scales for pairwise comparisons. *Reliability Engineering and System Safety, 95*, 750–763.

Saaty, T.L. (2006). *Fundamentals of decision making and priority theory with the analytic hierarchy process*. Pittsburgh, PA: RWS Publications.

Salo, A.A., & Hamalainen, R.P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-criteria Decision Analysis, 6*, 309–319.

Stevens, S.S. (1957). On the psychophysical law. *Psychology Review, 64*, 153–181.

Tsyganok, V.V. (2010). Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method. *Mathematical and Computer Modelling, 52(3-4)*, 538–544.

CONSISTENCY IMPROVEMENT IN COMBINATORY SPANNING TREE ENUMERATION METHOD

Sergey Kadenko
Institute for Information Recording
of the National Academy of Sciences of Ukraine
Kiev, Ukraine
E-mail: sergeykadenko@mail.ru

ABSTRACT

The paper addresses the problem of consistency improvement in group decision-making. The research is done in the context of studies performed by the Laboratory for Decision Support Systems of IIR of NASU (http://dss-lab.org.ua/). Particularly, the paper focuses on the problem of improving the consistency of pair comparison matrices (PCM) in AHP-based group decision support method called “enumeration of all spanning trees” or “combinatory algorithm”. Quite often expert judgments do not satisfy consistency requirements. PCM provided by an individual expert can be inconsistent within itself, while matrices built by several experts in the context of the same decision-making procedure can be mutually inconsistent. Combinatory methods of expert judgment aggregation are designed to utilize the redundancy of expert data most thoroughly. But such aspects as satisfactory PCM consistency level and ways of consistency improvement still need to be studied more carefully. The task, tackled in the current paper, is to study the opportunities for development of a converging consistency improvement procedure, allowing to achieve satisfactory levels of initially inconsistent expert judgments in combinatory aggregation methods.

Keywords: pair comparison matrix, expert judgment consistency, enumeration of all spanning trees, combinatory algorithm

1. Introduction

Expert data-based decision making is used mostly in weakly structured subject domains. In such domains it is problematic to perform quantitative measurements of these or that indicators influencing particular decisions. Moreover, there are no yardstick values which could be used for reference when expert evaluation is performed. For these reasons, expert data can be the only source of information under the abovementioned circumstances. But for these same reasons, expert data is characterized by inconsistency. Inconsistency can be witnessed in both ordinal and cardinal expert judgments. Also, inconsistency characterizes both individual and group expert estimates. Consequently, consistency considerations must be taken into account in every decision-making support method where expert data is used. Particularly, the questions to be addressed are: “how can consistency of expert judgments be measured in a given method?”; “what is the satisfactory expert judgment consistency level when a certain number of objects is evaluated (compared) by a certain number of experts?” and “how can expert judgment consistency level be improved, if necessary?” Or, in other words, “what is the borderline between redundancy of expert information (considered a good feature) and inconsistency of expert information, and how is it crossed?"

This particular paper focuses on consistency considerations in the context of AHP-based combinatory method of enumeration of all spanning trees (see Tsyganok (2010), Mikhailov/Siraj/Keane (2012)). The method fully utilizes the redundancy of expert information in group and individual decision-making, but consistency of expert judgments and its improvement procedures require further study.

2. Literature Review

Combinatory method was first suggested by V.Tsyganok in early 2000-s – respective publication appeared in "Mathematical Modelling" journal in 2010 (Tsyganok (2010)). A few years later a very similar approach was suggested by Siraj\Mikhailov \Keane (2012).
As for consistency in the context of AHP and related methods, the following studies should be mentioned: Iida studied ordinal consistency improvement in AHP through elimination of circular triads (details can be found in his paper from ISAHP 2009). Mikhailov and Siraj, again, studied ordinal consistency improvement (details can be found in their paper from ISAHP 2011). Brunelli and Fedrizzi in their paper from ISAHP 2011 analyzed several consistency indicators in AHP but did not suggest any particular consistency improvement methods. Mikhailov/Siraj/Keane (2012) do not suggest any feedback procedure when PCM are not consistent enough. Saaty (1996) himself does not prescribe any particular methods for consistency improvement if consistency index values are unsatisfactory: in such cases he just recommends to "revise the judgments and reconsider the problem". The nature of consistency indices in AHP (CR, CI, RI) does not provide for particular consistency improvement steps to be taken (particular objects swapped in rankings or PCM, or particular experts to be re-addressed with suggestions to change their judgments in order to improve their consistency). Tsyganok (2010) uses spectral consistency coefficient suggested by Totsenko (1996). Consistency improvement mechanism based on this coefficient is somewhat similar to Delphi approach (experts whose judgments lie outside the "majority" area are asked to change their judgments accordingly), but it is more target-oriented and flexible. Utilization of spectral consistency coefficient as consistency measure in combinatory algorithm allows to tell, which expert must be addressed with a suggestion to change his judgment, and which particular judgment (pair comparison) must be changed. On the other hand, the spectral coefficient is not devoid of certain drawbacks: firstly, it heavily depends on the scale point size, and, secondly, it unites two indicators – dispersion and entropy, which are, in the general case, independent. Consistency improvement procedure used by Tsyganok is not monotonously convergent, although its testing on multiple examples indicates that sufficient consistency level (Totsenko (1996) called it "usability threshold" or T_u) can be achieved. It would be adequate to try developing a monotonously convergent procedure for expert judgment consistency improvement during combinatory algorithm-based aggregation of individual PCM.

3. Hypotheses/Objectives

The problem to be addressed can be formulated as follows. Let us say, a certain number of experts estimates a certain number of objects (or decision variants) according to a certain criterion. Aggregation of individual PCM into a group PCM is performed using the method of enumeration of all spanning trees (or combinatory method). The task is to study the opportunities for devising a monotonously convergent procedure allowing to achieve expert judgment consistency level, which is sufficient for aggregation of individual PCM into a group one. I.e., the procedure must show which expert must be addressed with suggestions to change the judgments and which particular pair comparisons must be changed in the first place to improve the overall consistency level.

4. Research Design/Methodology

The first thing to be kept in mind is that there are several aspects of consistency to be tackled: ordinal versus cardinal consistency and inner versus mutual PCM consistency. The case can be illustrated by an example of an orchestra where each instrument is perfectly tuned to a different pitch. Inner consistency of pair comparisons does not guarantee mutual consistency of individual pair comparison matrices.

Approaches to ordinal consistency improvement within pair comparison matrices were suggested by Mikhailov, particularly, in his paper from ISAHP 2011. If the matrix is ordinally consistent, its elements can be rearranged in such a way that all elements above the principal diagonal are positive. Again, if we need to get a consistent set of pair comparison matrices, built by several experts, the order of alternatives in all experts' rankings needs to be the same, and this condition is the most problematic to fulfill (experts may be reluctant to swap ranks of alternatives), and there is no clear mutual ordinal consistency improvement algorithm. However, there are some rules which can be followed (see Tsyganok/Kadenko (2012)).

If the matrix is ordinally consistent within itself, all the elements lying below the principal diagonal are less than 1 (alternatives can be rearranged in such a way). When such matrices are obtained, spanning tree enumeration method (Tsyganok 2010, Mikhilov/Siraj/Keane 2012) can be launched. As a result, we obtain an aggregate PCM. Based on this matrix alternative weights can be calculated (using eigenvecor or some other method).

But if consistency level is not considered sufficient enough (respective PCM elements are considered "too different"), it is appropriate to shift the respective element of individual pair comparison matrix towards the element of the aggregate matrix, to improve consistency level. The first element to shift (to suggest to an expert for a change) would be the element, which differs from respective aggregate matrix element most significantly. The size of the shift can constitute half of this difference (to ensure better fine-tuning of respective matrix elements).

If we allow the elements of pair comparison matrices to assume values not only from fundamental scale (1/9,…,1/2, 1, 2, …, 9), but all the values within the range, then the consistency improvement procedure can allow to make differences between pair comparison matrices (individual ones and aggregate one) as small as it is required. Thus, we shall have a monotonously converging consistency improvement procedure.

5. Data/Model Analysis

In this section we shall analyze a simple numerical consistency improvement example, where 3 experts estimate 4 objects, and aggregate preference matrix is built using the combinatory method (Table 1).

Table 1. Initial pair comparison matrices

	Expert 1				Expert 2				Expert 3				Aggregate matrix			
A1	1.00	3.00	4.00	8.00	1.00	2.00	4.00	9.00	1.00	2.00	4.00	8.00	1	2.14	4.14	8.26
A2	0.33	1.00	2.00	5.00	0.50	1.00	2.00	4.00	0.50	1.00	2.00	3.00	0.47	1	1.93	3.86
A3	0.25	0.50	1.00	2.00	0.25	0.50	1.00	2.00	0.25	0.50	1.00	2.00	0.24	0.52	1	2.00
A4	0.13	0.20	0.50	1.00	0.11	0.25	0.50	1.00	0.13	0.33	0.50	1.00	0.12	0.26	0.50	1

The largest difference from respective aggregate matrix is witnessed in element a_{24} of the matrix built by Expert 1. Consequently, it is this element that needs to be changed (offered to Expert 1 for change) in the first place. If we permit to use real numeric values, and not only integer ones from fundamental scale, the picture after the 1st iteration will look as follows (Table 2).

Table 2. Consistency improvement: iteration 1 (real values)

	Expert 1				Expert 2				Expert 3				Aggregate matrix			
A1	1.00	3.00	4.00	8.00	1.00	2.00	4.00	9.00	1.00	2.00	4.00	8.00	**1**	**2.17**	**4.14**	**8.20**
A2	0.33	1.00	2.00	**4.43**	0.50	1.00	2.00	4.00	0.50	1.00	2.00	3.00	**0.46**	**1**	**1.91**	**3.79**
A3	0.25	0.50	1.00	2.00	0.25	0.50	1.00	2.00	0.25	0.50	1.00	2.00	**0.24**	**0.52**	**1**	**1.98**
A4	0.13	0.23	0.50	1.00	0.11	0.25	0.50	1.00	0.13	0.33	0.50	1.00	**0.12**	**0.26**	**0.51**	**1**

If real values are not permitted, the picture after the 1st iteration will look as follows (Table 3).

Table 3. Consistency improvement: iteration 1 (integer fundamental scale values)

	Expert 1				Expert 2				Expert 3				Aggregate matrix			
A1	1.00	3.00	4.00	8.00	1.00	2.00	4.00	9.00	1.00	2.00	4.00	8.00	**1**	**2.19**	**4.15**	**8.15**
A2	0.33	1.00	2.00	**4.00**	0.50	1.00	2.00	4.00	0.50	1.00	2.00	3.00	**0.46**	**1**	**1.90**	**3.73**
A3	0.25	0.50	1.00	2.00	0.25	0.50	1.00	2.00	0.25	0.50	1.00	2.00	**0.24**	**0.53**	**1**	**1.97**
A4	0.13	0.23	0.50	1.00	0.11	0.25	0.50	1.00	0.13	0.33	0.50	1.00	**0.12**	**0.27**	**0.51**	**1**

Following the pattern of changing individual pair comparisons, which are most considerably differing from respective aggregate ones at every new iteration, we will get a thoroughly consistent individual matrices and aggregate matrix (Table 4)

Table 4. Thoroughly consistent matrix

1	2	4	8
0.5	1	2	4
0.25	0.5	1	2
0.13	0.25	0.5	1

It should be noted that in case if real values are permitted, the procedure is monotonously convergent.

6. Limitations

In reality experts are not operating with real values, they are only presented verbal scales with respective integer number equivalents. Consequently, in a real expert examination it is, virtually, impossible to achieve ideal consistency of pair comparison matrices (both mutual and inner). The procedure of consistency improvement should, definitely, stop when the absolute value of difference between individual and aggregate pair comparisons start to increase with the new iteration. It should be also noted that prior to launching the described consistency improvement procedure, ordinal consistency of pair comparison matrices should be ensured.

7. Conclusions

Several aspects of pair comparison consistency improvement in combinatorial aggregation method have been analyzed. Based on the analyses, a monotonously convergent consistency improvement procedure has been suggested. The suggested approach has its limitations, namely, in order for procedure to converge, it requires ordinal consistency of individual pair comparison matrices. However, the approach can be utilized as a consistency improvement method in group decision-making (including AHP/ANP-based decision-making).

8. Key References

Kadenko,S., Tsyganok, V. (2012). A Method for Improving the Consistency of Individual Expert Rankings during Their Aggregation. *Journal of Automation and Information Sciences, 44(4), 23-31.*

Saaty, T.L. (1996). Decision Making with Dependence and Feedback: The Analytic Network Process. Pittsburgh, PA: RWS Publications.

Siraj S., Mikhailov L., Keane J. (2012). Enumerating All Spanning Trees for Pairwise Comparisons. *Computers and Operations Research, 39(2), 191-199.*

Totsenko, V.G. (1996). The Agreement Degree of Estimations Set with Regard for Experts' Competence. *Proceedings of the IV International Symposium on the AHP, ISAHP '96.* Simon Fraser University, Burnaby, B.C.,CANADA, 229-241.

Tsyganok, V. (2010). Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method. *Mathematical Modeling, 52(3-4),* 538–544.

GROUP DECISION MAKING WITH THE AHP/ANP – AN OVERVIEW OF APPROACHES TO AGGREGATION OF JUDGMENTS AND PRIORITIES

Anna Florek-Paszkowska (nee Greda)
Cracow University of Economics
Cracow, Poland
E-mail: anna.greda@uek.krakow.pl; greda.anna@gmail.com

Anna Prusak (nee Strada)
Cracow University of Economics
Cracow, Poland
E-mail: anna.strada@uek.krakow.pl; strada@cm-uj.krakow.pl

Piotr Stefanów
Andrzej Frycz Modrzewski Cracow University College
Cracow, Poland
E-mail: piotr@stefanow.net

ABSTRACT

The objective of the present paper is to review knowledge in the area of analytic hierarchy and network processes (AHP/ANP), with respect to group decision-making and aggregating results from many respondents. This knowledge refers to one of the most important aspects of methodology of the AHP/ANP. It reviews two main approaches to aggregating the AHP/ANP results: qualitative (behavioural) and quantitative (mathematical). Qualitative approach includes consensus and voting or compromising, while quantitative approach consists calculating geometric mean of individuals' judgments (aggregating individual judgments – AIJ), and combining results from individual models or parts of a model (aggregating individual priorities – AIP). The authors review these approaches and recommend the matrix of selection of the most appropriate aggregation procedure of the AHP/ANP judgments and priorities dependent on the character of a group.

Keywords: group multicriteria decision making, GMCDM, Analytic Hierarchy Process, Analytic Network Proces, AHP/ANP, aggregating individual judgments, AIJ, aggregating individual priorities, AIP.

1. Introduction

The process of management is based on constant decision making, where decisions are becoming increasingly complex and multifaceted. The ability to use multicriteria methods of supporting decisions is one of the key challenges that modern managers face. The analytic hierarchy and network processes (AHP/ANP) are the most often used multicriteria decision-support techniques in the world, both in science and in practice. With the use of these methods, many different problems have been solved in commercial and governmental bodies. Such decisions, however, are always made in the group settings, involving people from various disciplines. Consequently, development of a solid methodological base for aggregating judgments of multiple players may contribute to the increase of the successful decision making stories and at the same time, have positive impact on various areas of civilisation. Almost every user of the AHP/ANP faces the problem of synthesising the judgments coming from various respondents. The objective of the present paper is to review the existing knowledge on the aggregation of the AHP judgments and priorities obtained by the groups. It has been observed that a large number of publications reporting the AHP/ANP methods in solving various decision-making problems usually „halt" at the stage of the analysis of individual models. The authors of these works do not state how the results have been aggregated. Such information is typically provided in statistical reports. The decision analysis made with the AHP/ANP has its own specific nature. The opinions are expressed using bipolar, pairwise comparisons scale referred as the *Saaty's fundamental scale*. It is based on a comparison of two elements (objects), of which one may be "better", "more important", "more preferred" or "more likely" than the other. The degree of dominance may be indicated from "1" ("the same significance of both elements"), to "9" ("extreme dominance"). While pairwise comparisons are considered to be a very effective measurement tool (Saaty, 2000), in case of group decisions they require special treatment, which is not always average of the judgments. This paper reviews two main approaches to aggregating individual results: qualitative (behavioural) and quantitative (mathematical), as defined by (Goodwin & Wright, 2011).

2. Qualitative approaches to bring together individual judgments

Qualitative (also called behavioural) approaches involve two ways: (1) *consensus*, and (2) *voting or compromising*. According to (Dyer & Forman, 1992) *consensus* is the most attractive way of synthesizing individual opinions, in both constructing the hierarchical model and making judgments, for two reasons. First, the interaction between the group members helps ensure that the relevant information is available to the entire groups. Second, participants feel as "owners" of the decision and make their best efforts to make the whole process successful. Moreover, in some decisions arriving at a consensus may be more important than choosing an alternative. It happens when decision variants do not differ considerably from one another, and the success of the entire decision-making process depends on the subsequent implementation efforts. Consensus can be also reached in non-common objectives context, assuming that certain aspects of a decision problem can be shared. For the above reasons, the process of reaching the consensus in the group environment remains the focus of numerous studies. If consensus cannot be reached on a judgment, the group may *vote or compromise* on an intermediate judgment (Dyer & Forman, 1992). However, voting generates a range of problems. One of them is so called the voting paradox, also known as Condorcet's paradox, which refers to non-transitivity of the group preferences (Goodwin & Wright, 2011).

Perhaps the most widely discussed method of arriving at the consensus is Delphi technique. It was developed in the 1950's by workers of the RAND Corporation, as a procedure to obtain the most consistent agreement within a group of experts. The consensus should be gained through a series of questionnaires with controlled opinion feedback. It consists of several rounds, including: unstructured discussions, consolidating the opinions, producing questionnaires for subsequent rounds where quantitative assessments are required. After each round, the responses are analysed statistically and presented to the panellists for further considerations. In the next rounds, participants have the opportunity to alter prior estimates based on the group feedback. Delphi procedure was initially used for long-term forecasts in science and technology, but later it was applied to a wide range of decision situations (Rowe, Wright, & Bolger, 1991). Yet the method attracted the attention of a wider audience when the first edition of the Linstone & Turoff's book on Delphi appeared in 1975 (Linstone & Turoff, 1975). It has also become a subject of numerous studies and critical reviews (Rowe & Wright, 2011).

There have been several attempts to combine Delphi method with the AHP in solving diverse problems. For example, (Kim, Jang, & Lee, 2013) used Delphi-AHP methods to determine criteria and priorities of recycling of waste electronic and electrical products (WEEE). Nonetheless, the AHP was run as a separate part of the study due to the fact that "it is difficult to determine the priority ranking of each evaluation criterion by brainstorming or using the Delphi method" (p. 945). Similar research structure (first Delphi, then the AHP) was adopted by (Hsu, Lee, & Kreng, 2010; Meesapawong, Rezgui, & Li, 2014; Vidal, Marle, & Bocquet, 2011), who used Delphi to refine a list of factors (criteria) retrieved from the relevant literature, and as a next step applied the AHP method to prioritize these factors (criteria). A different approach was demonstrated by (Tavana, Kennedy, Rappaport, & Ugras, 1993) who incorporated the AHP into Delphi procedure. They asserted that: "the integration of AHP into a Delphi framework enhances the power of AHP by using it in an iterative sequence of individual questioning and anonymous feedback to elicit judgments from a group of individuals who are knowledgeable about issues which are not subject to objective solution" (p. 49). Nonetheless, such a "hybrid" reminds the aggregation of individual judgment (AIJ) procedure, and should be considered a quantitative rather than the qualitative measure.

3. Quantitative methods of aggregating individual judgments and priorities

Quantitative (or mathematical) approach is recommended if it is difficult to obtain the consensus and the group is unwilling to vote or compromise on a judgment. It may be used in two ways: (1) *calculating geometric mean of individuals' judgments*, and (2) *combining results from individual models or parts of a model* (Dyer & Forman, 1992). In the AHP literature, these procedures are known as aggregation of individual judgments (AIJ) and aggregation of individual priorities (AIP) (Forman & Peniwati, 1998). Such classification is based on whether the group acts as a unit or as separate individuals. When individuals abandon their own preferences or values for the good of the group, they should be perceived as "new individual". In this situation, AIJ should be applied, since individual priorities are irrelevant or even non-existent (if individual skips judgments for a cluster or hierarchy). The AIJ is a synergistic aggregation of individual judgments. Thus, the participants should first work together to agree on a common hierarchical model before synthesizing their judgments through agreement on the relative importance of the criteria. For this purpose, the AHP-Delphi hybrid method can be used to obtain the consensus (Tavana et al., 1993). However, those who decided to use AIJ have to consider homogeneity of individual judgments. Thus, the use of AIJ requires additional measure to check the inter-judge homogeneity. The use of such measure, called Compatibility Index (*SI*), has been explained by (von Solms, 2009). In practice, it refers to "similarity" of pairwise comparison matrices derived for different respondents.

Several examples of how to calculate the Compatibility Index can be found in (Prusak & Stefanów, 2014).
When individuals have different values and objectives and act separately, one should be more interested in their individual priorities. For this purpose, the AHP model (or part of the model) is evaluated individually be each respondent and individual priorities are derived. As a next step these priorities are aggregated, and this procedure is referred as the AIP (aggregating individual priorities). According to (Forman & Peniwati, 1998), to combine the results mathematically, geometric mean is required for AIJ (mathematical evidence of this requirement can be found in Aczél & Saaty, 1983), while *either* arithmetic *or* geometric mean for AIP procedure. It was also pointed out that some authors, for example (Ramanathan & Ganesh, 1994) erroneously suggested to use *only* arithmetic mean in the AIP.
Another problem related to mathematical aggregation of the AHP results is how to represent in a satisfactory way people's experience embedded in their judgments (it was called the "soundness of judgment" by (Saaty, 1980). Some researchers have focused on this issue, for example (Beynon, 2005; Tsyganok, Kadenko, & Andriichuk, 2012). In general, non-equivalent importance of individuals in the group require using weighted means of their judgments or priorities (dependent on the procedure) should be used (formulas 1 and 2).

Weighted arithmetic mean (for the sum of all weights equal to 1):

$$\bar{a} = \sum_{i=1}^{n} w_i \cdot a_i = w_1 \cdot a_1 + w_2 \cdot a_2 + ... + w_n \cdot a_n \tag{1}$$

Weighted geometric mean (for the sum of all weights equal to 1):

$$\bar{a}_g = \prod_{i=1}^{n} a_i^{w_i} = a_1^{w_1} \cdot a_2^{w_2} \cdot ... \cdot a_n^{w_n} \tag{2}$$

where: w_i are weights of opinions expressed by the *i*th participant.

According to (Saaty, 1980), examples of factors affecting judgments may include intelligence, years of experience, past record, knowledge, experience in other fields or personal involvement in the decision problem. All these factors should be the basis for determining weights of individual's judgments or priorities. It can be done either as part of the AHP model or through a subsidiary hierarchical structure constructed for evaluation of the respondent. An example of such evaluation hierarchy has been provided by (Saaty & Peniwati, 2007). On the other hand, (Tsyganok et al., 2012) attempted to evidence that „weighting" the experts' competences in group settings does not always make sense. Moreover, such additional analysis may be a problem in a large number of respondents is involved. For this reason, the group size in AHP decisions should be somehow limited. The review of literature demonstrated that many authors do not provide information on the number of experts participating in the AHP survey. Number of participants of the reviewed AHP-Delphi studies varied from 4 (Tavana et al., 1993) to 33 (Meesapawong et al., 2014), yet in the latter case it referred only to the participants of Delphi phase of the study.

4. Conclusions

The paper aimed at reviewing the existing knowledge on the aggregation of the AHP judgments and priorities obtained in the group settings. There are two general approaches of aggregating the AHP results: qualitative and quantitative. Qualitative (also called behavioral) approach includes consensus and voting or compromising, while quantitative (mathematical) method consists of aggregation of individual judgments or priorities by geometric or arithmetic mean. The selection of the appropriate method of synthesizing individual opinions into a common result depends largely on the character of the group: whether it is homogenous or heterogeneous in terms of competencies, and whether it acts in synergy as a unit, or autonomously, as separate individuals. Figure 1 presents the matrix, which can be used in selection of the appropriate aggregation procedure.

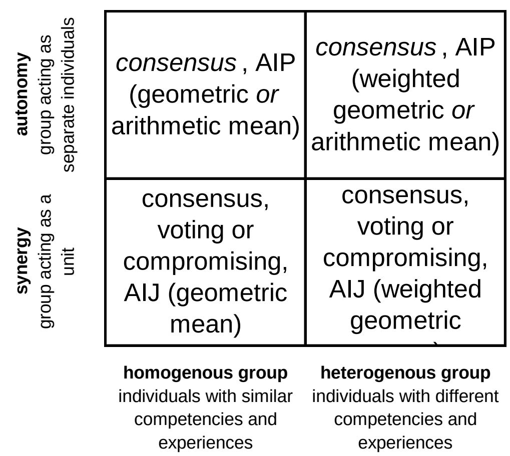


Figure 1. A matrix of selection of the AHP aggregation procedure in the group decisions

The matrix shows four possibilities represented by four quarters. The *bottom-left quarter* refers to the situation when the group is homogenous and acts in synergy. It provides the most favorable conditions for the consensus, and if consensus cannot be reached, for voting or compromising. If it does not work either, the AIJ procedure (aggregation of individual judgment) should be applied using geometric mean. The *bottom right quarter* shows the condition where the group acts as a unit, but participants have different competencies and experiences (or differ in terms of other factors important for a decision). Consensus still can be obtained within such group, while in the AIJ procedure weighted geometric should be used. The *upper quarters* of the matrix represent the environment where decision makers are independent, acting as separate individuals. In our opinion, it is difficult to obtain the consensus in such groups, although Delphi technique does not demand the presence of all participants in one place and time. However, decisions by voting or compromising may be impossible in such settings. Moreover, in case of autonomous decision makers, we are interested in receiving their individual, final priorities of the AHP model, not judgments, and as such it is more practical to use the AIP procedure. To aggregate priorities, either geometric or arithmetic mean can be used, and if the group is heterogeneous (*upper right quarter*), we should use weighted means.

5. References

References will be available in the final version of the paper or the authors will provide them upon request.

6. Key References

Dyer, R. F., & Forman, E. H. (1992). Group decision support with the Analytic Hierarchy Process. *Decision Support Systems*, *8*(2), 99–124.

Forman, E., & Peniwati, K. (1998). Aggregating individual judgments and priorities with the analytic hierarchy process. *European Journal of Operational Research*, *108*(1), 165–169.

Tavana, M., Kennedy, D. T., Rappaport, J., & Ugras, Y. J. (1993). An AHP-Delphi Group Decision Support System Applied to Conflict Resolution in Hiring Decisions. *Journal of Management Systems*, *5*(1), 49–74.