ATLANTIS PRESS
Advances in Economics, Business and Management Research, volume 148
Proceedings of the Russian Conference on Digital Economy and Knowledge Management (RuDEcK 2020)

Aggregated Rating Construction as a Collective Choice Problem

Nikitin B.E.*
Voronezh State University of Engineering Technologies
Voronezh, Russia
e-mail: nbe6419@gmail.com

Ivliev M.N.
Voronezh State University of Engineering Technologies
Voronezh, Russia
e-mail: max1m@mail.ru

Bugaev Y.V.
Voronezh State University of Engineering Technologies
Voronezh, Russia
e-mail: y_bugaev52@mail.ru

Kovaleva E.N.
Voronezh State University of Engineering Technologies
Voronezh, Russia

Chikunov S.V.
Voronezh State University of Engineering Technologies
Voronezh, Russia
e-mail: chiksv@rambler.ru

Negoda V.A.
Voronezh State University of Engineering Technologies
Voronezh, Russia

***Abstract* — The article deals with the issues of national aggregated ranking (NAR) of higher education institutions. This case is formulated as a collective choice problem. It is proposed to use voting procedures in small groups that meet the Condorcet principle as aggregation method. The paper applies collective choice rules to the problem under consideration. The alternatives final ordering stability obtained from the Board, Copeland, and Kemeny procedures is illustrated by specific examples. The empirical average was considered as Kemeny median approximate estimate. The example showed Board procedure instability to a slight change in the initial ratings obtained in the rating mechanisms under consideration. Using the voting procedures results described in the paper in small groups on a limited sample containing data from fifteen elements group are presented. The constructed three aggregated rankings proximity degree was estimated using two metrics – the Kendall rank correlation coefficient and the Kemeny distance. Based on a comparative results analysis, it can be concluded that it is appropriate to use collective selection procedures that are well-off in Condorcet when constructing aggregated rating of various organizations, including educational institutions.**

Keywords — rating, voting procedure, Condorcet principle, Kemeny distance, alternative, Board rule, Copeland rule.

I. Introduction

Nowadays ratings are widely used as research tools for various human activity areas. For example, in the higher education system, the method is proposed that allows aggregating the different assessment procedures results for educational institutions [1]. We can see this method implementation in the form of a national aggregate rating (NAR) on the information resource https://best-edu.ru/. Building the NAR took into account the eight ratings that meet the four criteria [2]:

- publicity – full information is in open access;
- stability – there are at least three years;
- mass – evaluated at least 100 organizations;
- frequency – assessment occurs annually.

The Board rule is used as an aggregation procedure [3]. However, this procedure is one of the most manipulated collective selection procedures [4,5]. In the framework of the methodology proposed in [1], the authors suggest using voting procedures in small groups that are Condorcet-rich as different rating results aggregation procedures. These procedures include the Copeland procedure and the Kemeny procedure.

The selection function generated by the corresponding selection mechanism satisfies the Condorcet principle if both the forward and reverse Condorcet conditions are met for it [6]. The direct condition presupposes that an alternative is selected from all paired presentations containing it and the other alternatives selected when the entire options set X is considered. The inverse Condorcet condition says that an alternative selected from the entireset X must also be selected when any pair containing it, is presented from the set alternatives under consideration [6].

II. Mathematical model

The creation of NAR of higher education institutions can be presented as acollective choice problem. Let be $X = \{x_i\}$ – universities set under consideration, $i \in I$, $I = \{1, 2, ..., n\}$. Denote $P = \{P_j\}$ – universities aggregated ratings set, $j = 1, 2, ..., m$. We assume that each mechanism P_j induces reflexive, antisymmetric, transitive, and complete binary relation R_j on the set X. According to [1], each rating mechanism scale P_j is divided into k equivalence classes (for example, quartiles). Objects that fall into l -class rating P_j, receive the corresponding rating $A^j_{l_i}$, $l_i \in \{1,2,...,k\}$. In this case, it is considered that $A^j_{1_i} > A^j_{2_i} > ... > A^j_{k_i}$. If university didn't get into any class by rating P_j, it gets a rating A^j_{k+1}. Thus, each x_i is assigned a corresponding m-dimensional estimatesvector. Forexample, if $m = 8$ and $k = 4$ the evaluation vector x_i, that characterizes a particular alternative may have the following form $-(A^1_{3_i}, A^2_{4_i}, A^3_{5_i}, A^4_{2_i}, A^5_{2_i}, A^6_{1_i}, A^7_{4_i}, A^8_{1_i})$. This means that this university, in particular, was in the first class in the sixth and eighth rankings, and in the third ranking this university wasn't present. Universities that fall into the same rating class P_j, are considered equivalent for it. So, the set X is given $m$linear orders $L = \{R_1, R_2,..., R_m\}$. We need to build the resulting (aggregated) ordering R_* using the rule F: $L^m \to R_*$.

Various voting procedures in small groups can be used to build an aggregate score [4,7]. Here is three procedures description that are considered in this paper as F.

The Board Procedure. According to the Board rule, at the beginning for each x_i is calculated number $b_j(x_i) = Card(L_j(x_i))$, where $L_j(x_i)$ −is alternative x_i lower slice in the binary relation R^2_j, $j = 1,2,...,m$ .In our case:

$$L_j(x_i) = \{x_s \in X \mid A^j_{l_i} > A^j_{l_s}, s \in I, s \neq i\}.$$

Then, the Board score for alternative x_i is defined as the sum for all j, i.e. $b(x_i) = \sum_{j=1}^{m} b_j(x_i)$. Universities final ordering R_* is obtained by ranking relative $x_i \in X$ to the calculated scores $b(x_i)$.

Example 1. Let $X = \{x_1, x_2, x_3, x_4\}$, $m = 5$, $k = 4$ and estimates alternative vectors as follows:

$$x_1 = (A^1_{1_1}, A^2_{1_1}, A^3_{4_1}, A^4_{4_1}, A^5_{4_1}),\ x_2 = (A^1_{2_2}, A^2_{3_2}, A^3_{1_2}, A^4_{2_2}, A^5_{3_2}),$$

$$x_3 = (A^1_{3_3}, A^2_{2_3}, A^3_{2_3}, A^4_{3_3}, A^5_{1_3}),\ x_4 = (A^1_{4_4}, A^2_{4_4}, A^3_{3_4}, A^4_{1_4}, A^5_{2_4}).$$

Thus, five strict linear orders are imposed on the set X :

$$R_1 : x_1 \succ x_2 \succ x_3 \succ x_4,\ R_2 : x_1 \succ x_3 \succ x_2 \succ x_4,$$

$$R_3 : x_2 \succ x_3 \succ x_4 \succ x_1,\ R_4 : x_4 \succ x_2 \succ x_3 \succ x_1,$$

$$R_5 : x_3 \succ x_4 \succ x_2 \succ x_1,$$

Where "$\succ$" symbol means "better". Then, for alternative x_1 we get $L_1(x_1) = \{x_2, x_3, x_4\}$, $L_2(x_1) = \{x_2, x_3, x_4\}$, $L_3(x_1) = L_4(x_1) = L_5(x_1) = \{\varnothing\}$ and the Board score is $b(x_1) = 3 + 3 + 0 + 0 + 0 = 6$. For the rest x_2, x_3, x_4 the Board scores are 9, 9, and 6.As a result, the final aggregated ranking R_* has the following form:

$$(x_2 \sim x_3) \succ (x_1 \sim x_4),$$

where the symbol "$\sim$" means "equivalence".

As it was mentioned above, the Board procedure does not satisfy the Condorcet principle. According to this principle, the Condorcet winner, if there is one, must also be selected according to this rule. However, as the example shows, this principle is violated when using the Board rule. Indeed, the Condorce twinnerin Example 1 is x_2, and the Board procedure puts two alternatives in first place – x_2 and x_3. Inaddition, thefinalordering R_* is unstable in relation to a slight change in the initial alternatives estimates.

Example 2. Now let alternativese valuation vectors x_1 and x_2 have not changed, and the vectors of evaluation x_3 and x_4 are as follows:

$$x_3 = (A^1_{3_3}, A^2_{2_3}, A^3_{2_3}, A^4_{3_3}, A^5_{1_3}),\ x_4 = (A^1_{4_4}, A^2_{4_4}, A^3_{3_4}, A^4_{1_4}, A^5_{2_4}),$$

i.e., third alternative x_3 assessment has deteriorated by one gradation, and the alternative x_4 has improved by one gradation. Corresponding binary relation R_3 takes the form $x_2 \succ x_4 \succ x_3 \succ x_1$. In this case, the Board estimates for the alternatives under consideration are $b(x_2) = 9$, $b(x_3) = 8$, $b(x_4) = 7$. As a result, we get the final ranking R_* in the form of a strict linear order:

$$x_2 \succ x_3 \succ x_4 \succ x_1.$$

Here is another example that demonstrates solution instability obtained by applying the Board procedure.

Example 3. Let us assume that alternatives evaluation vectors x_1 and x_3 have changed in comparison with the original data from Example 1:

$$x_1 = (A^1_{1_1}, A^2_{1_1}, A^3_{4_1}, A^4_{3_1}, A^5_{4_1}),\ x_3 = (A^1_{3_3}, A^2_{2_3}, A^3_{2_3}, A^4_{4_3}, A^5_{1_3}).$$

i. e., the fourth grade x_1 improved by one gradation, but it also worsened for x_3. Then, we have $b(x_1) = 7$, $b(x_2) = 9$, $b(x_3) = 8$, $b(x_4) = 5$. Final ordering R_*:

$$x_2 \succ x_3 \succ x_1 \succ x_4.$$

The Copeland Procedure. The first, second, and third Copeland rules are distinguished [4,5]. According to these rules, at the beginning, a majority relation μ is built on the considered set X:

$$x_s\, \mu\, x_p \leftrightarrow card(\{P_j \in P \mid x_s \overset{P_j}{\succ} x_p\}) > card(\{P_j \in P \mid x_p \overset{P_j}{\succ} x_s\}),$$

where $s, p \in I$, $s \neq p$ and the symbol "$\overset{P_j}{\succ}$" means "better in the ranking P_j". The score $z(x_i)$ is calculated for each $x_i \in X$ .In the first Copeland rule, this estimate is defined as the difference in alternative x_i lower and upper sections capacity in the majority ratio μ:

$$z(x_i) = Card(L(x_i)) - Card(D(x_i)).$$

In the second rule, score $z(x_i)$ is determined by alternative x_i lower cut power in the majority ratio μ, i.e. $z(x_i) = Card(L(x_i))$. In the third rule $z(x_i)$ is determined by alternative x_i upper cut power in the majority relation μ, i.e. $z(x_i) = Card(D(x_i))$. Final ordering R_* is obtained by ranking relative $x_i \in X$ to the calculated scores $z(x_i)$. If first or second rule is applied, the higher the value $z(x_i)$, the higher the alternative position x_i in the final ranking. During the use of the third Copeland rule, the lower the value $z(x_i)$, the higher the alternative position x_i in the final ranking.

Example 4. We take original data from Example 1. Let us build corresponding majority relation matrix (Table 01). Applying first Copeland rule, we get the following estimates $z(x_i): z(x_1) = 0-3 = -3$, $z(x_2) = 3-0 = 3$, $z(x_3) = 2-1 = 1$, $z(x_4) = 1-2 = -1$. Note that $Card(L(x_i)) = \sum_{p=1}^{4} \mu_{ip}$, $Card(D(x_i)) = \sum_{p=1}^{4} \mu_{pi}$, $i \in \{1,2,3,4\}$, $\mu_{ip} \in \{0,1\}$ - are majority relation μ matrix elements.

TABLE I. MAJORITY RELATIONS MATRIX

μ	x_1	x_2	x_3	x_4
x_1	0	0	0	0
x_2	1	0	1	1
x_3	1	0	0	1
x_4	1	0	0	0

Final ordering is $R_*: x_2 \succ x_3 \succ x_4 \succ x_1$. Note that in this example, applying second and third Copeland rules results in a similar final alternatives ranking.

Let us consider applying the Copeland procedure in Example 2 and Example 3 conditions result.

Example 5. Let data meet Example 2 requirements. Then five linear orders are imposed on alternatives set:

$$R_1: x_1 \succ x_2 \succ x_3 \succ x_4,\ R_2: x_1 \succ x_3 \succ x_2 \succ x_4,$$

$$R_3: x_2 \succ x_4 \succ x_3 \succ x_1,\ R_4: x_4 \succ x_2 \succ x_3 \succ x_1,$$

$$R_5: x_3 \succ x_4 \succ x_2 \succ x_1.$$

Corresponding majority ratio matrix μ will be the same as in Example 4(Table 01). Thus, the final alternatives ordering R_* have not changed compared to the previous example.

Example 6. Let original ones be similar to Example 3. Then, in comparison with previous example, binary relations R_1, R_2, R_3 и R_5 will not change, but R_4 will take the following form: $x_4 \succ x_2 \succ x_3 \succ x_1$. As a result of applying the Copeland procedure, the alternatives final ordering R_* will be the same as in Example 4 and Example 5.

The Kemeny Procedure. According to the Kemeny rule, Kemeny median serves as alternatives $x_i \in X$ aggregated ordering R_*, which calculation is reduced to solving discrete optimization problem

$$R_* = \arg(\min_R(\sum_{j=1}^{m} d(R, R_j))),$$

where $d(R, R_j)$ – Kemeny distance between two rankings [8]. [9] presents the author's algorithm for finding the Kemeny median, and [10,11] describes calculating Kemeny median analog examples. Let us consider the examples in which empirical mean (similar to the Kemeny median) is used as Kemeny median approximate estimate.

Example 7. Let source data match Example 1. Let us construct pairwise distances between rankings R_1, R_2, R_3, R_4, R_5 induced on the set X matrix $M = (d(R_i, R_j))$.

TABLE II. PAIRWISE DISTANCES MATRIX

M	R_1	R_2	R_3	R_4	R_5
R_1	0	2	6	10	10
R_2	2	0	8	12	8
R_3	6	8	0	4	4
R_4	10	12	4	0	4
R_5	10	8	4	4	0

In the matrix M element $d(R_i, R_j)$ is equal to non-matching elements number in matrices which correspond to the binaric relations R_i and R_j. For example, $d(R_2, R_4) = 12$. Below we can see binary relations matrices R_2 and R_4:

TABLE III. BINARY RELATION MATRIX R_2

R_2	x_1	x_2	x_3	x_4
x_1	0	1	1	1
x_2	0	0	0	1
x_3	0	1	0	1
x_4	0	0	0	0

TABLE IV. BINARY RELATION MATRIX R_4

R_4	x_1	x_2	x_3	x_4
x_1	0	1	1	1
x_2	0	0	0	1
x_3	0	1	0	1
x_4	0	0	0	0

As we can see, non-matching elements number in these matrices is 12.

Next, calculate for each R_i value $D(R_i) = \sum_{j=1}^{m} d(R_i, R_j)$, $i = 1,2,3,4,5$:

$$D(R_1) = 2+6+10+10 = 28,\ D(R_2) = 2+8+12+8 = 30,$$

$$D(R_3) = 6+8+4+4 = 22,\ D(R_4) = 10+12+4+4 = 30,$$

$$D(R_5) = 10+8+4+4 = 26.$$

The $D(R_i)$ takes the smallest value when R_3. Hence, the final ordering is: $R_* = R_3 : x_2 \succ x_3 \succ x_4 \succ x_1$. Let's consider applying this search procedure in the Example 3 conditions result.

Example 8. Let source data match Example 3 conditions. Recall that in this case, compared to the previous one, evaluation vectors for alternatives x_1 and x_3 fourth components values change by one gradation. Table 05 shows pair wise distances matrix M obtained for this case and calculated values $D(R_i)$ (rightmost column). As can be seen, even in this case, minimum value $D(R_i)$ is reached at R_3, i.e. $R_3 = \arg(\min_{R_i}(\sum_{j=1}^{5} d(R_i, R_j)))$. Consequently, alternatives final ordering R_* has not changed in comparison with the previous example.

TABLE V. PAIRWISE DISTANCES MATRIX

M	R_1	R_2	R_3	R_4	R_5	$D(R_i)$
R_1	0	2	8	10	10	30
R_2	2	0	10	12	8	32
R_3	8	10	0	2	6	26
R_4	10	12	2	0	4	28
R_5	10	8	6	4	0	28

III. RESULTS AND ANALYSIS

The voting procedures discussed above were used in small groups in order to build an aggregated ranking on universities (*Un*) in one of the Russian Federation regions sample. An information resource was used as initial data source [2]. Based on eight ratings evaluation vectors P_1, P_2, …, P_8 were formed for fifteen educational organizations. Below are the scores for eight universities in the region under review.

TABLE VI. UNIVERSITIES (UN) EVALUATIONS VECTOR

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8
Un1	$A^1_{5_1}$	$A^2_{3_1}$	$A^3_{3_1}$	$A^4_{1_1}$	$A^5_{1_1}$	$A^6_{1_1}$	$A^7_{5_1}$	$A^8_{1_1}$
Un2	$A^1_{1_2}$	$A^2_{4_2}$	$A^3_{2_2}$	$A^4_{2_2}$	$A^5_{2_2}$	$A^6_{2_2}$	$A^7_{2_2}$	$A^8_{3_2}$
Un3	$A^1_{3_3}$	$A^2_{4_3}$	$A^3_{5_3}$	$A^4_{2_3}$	$A^5_{2_3}$	$A^6_{1_3}$	$A^7_{4_3}$	$A^8_{1_3}$
Un4	$A^1_{3_4}$	$A^2_{3_4}$	$A^3_{5_4}$	$A^4_{1_4}$	$A^5_{2_4}$	$A^6_{3_4}$	$A^7_{5_4}$	$A^8_{1_4}$
Un5	$A^1_{4_5}$	$A^2_{4_5}$	$A^3_{5_5}$	$A^4_{3_5}$	$A^5_{2_5}$	$A^6_{1_5}$	$A^7_{4_5}$	$A^8_{1_5}$
Un6	$A^1_{2_6}$	$A^2_{4_6}$	$A^3_{4_6}$	$A^4_{2_6}$	$A^5_{2_6}$	$A^6_{3_6}$	$A^7_{4_6}$	$A^8_{3_6}$
Un7	$A^1_{4_1}$	$A^2_{4_7}$	$A^3_{5_7}$	$A^4_{1_7}$	$A^5_{2_7}$	$A^6_{4_7}$	$A^7_{4_7}$	$A^8_{2_7}$
Un8	$A^1_{5_1}$	$A^2_{5_8}$	$A^3_{5_8}$	$A^4_{2_8}$	$A^5_{2_8}$	$A^6_{3_8}$	$A^7_{5_7}$	$A^8_{2_8}$

The eight binary relations corresponding to the estimates given in Table 06 are as follows:

$$R_1 : Un2 \succ Un6 \succ (Un3 \sim Un4) \succ (Un5 \sim Un7) \succ Un8,$$

$$R_2 : (Un1 \sim Un4) \succ (Un2 \sim Un3 \sim Un5 \sim Un6 \sim Un7) \succ Un8,$$

$$R_3 : Un2 \succ Un1 \succ Un6 \succ (Un3 \sim Un4 \sim Un5 \sim Un7 \sim Un8),$$

$$R_4 : (Un1 \sim Un4 \sim Un7) \succ (Un2 \sim Un3 \sim Un6 \sim Un8) \succ Un5,$$

$$R_5 : Un1 \succ (Un2 \sim Un3 \sim Un4 \sim Un5 \sim Un6 \sim Un7 \sim Un8),$$

$$R_6 : (Un1 \sim Un3 \sim Un5) \succ Un2 \succ (Un4 \sim Un6 \sim Un8) \succ Un7,$$

$$R_7 : Un2 \succ (Un3 \sim Un5 \sim Un6 \sim Un7) \succ (Un1 \sim Un8),$$

$$R_8 : (Un1 \sim Un3 \sim Un4 \sim Un5) \succ (Un7 \sim Un8) \succ (Un2 \sim Un6).$$

Eight universities aggregated ordering under consideration, based on the Board procedure, looks like a strictly linear order:

$$R_*^B : Un2 \succ Un1 \succ Un6 \succ Un3 \succ Un5 \succ Un4 \succ Un7 \succ Un8.$$

Applying the Copeland procedure results in a clustered ranking looks like:

$$R_*^C : Un1 \succ Un2 \succ (Un3 \sim Un4 \succ Un6) \succ Un5 \succ Un7 \succ Un8.$$

As an Kemeny median on the set $\{R_1, R_2, R_3, R_4, R_5, R_6, R_7, R_8\}$ approximate estimate, the clustering ranking is used R_2. Pairwise Kemeny distances corresponding matrix is given in Table 07.

Three aggregated rankings found comparative analysis shows, in particular, that the Copeland procedure and Kemeny median analogue use as R_* improves two higher education institutions positions (university 3 and university 6). The improvement is in comparison its positions in the aggregated ranking based on the Board procedure. Three aggregated university rankings proximity degree was evaluated using two metrics: Kendall rank correlation coefficient and the Kemeny distance.

TABLE VII. PAIRWISE DISTANCES MATRIX

M	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
R_1	0	24	16	30	22	26	9	33
R_2	24	0	20	12	15	20	25	15
R_3	16	20	0	24	18	18	17	29
R_4	30	12	24	0	26	30	31	19
R_5	21	15	18	26	0	18	27	25
R_6	26	20	18	30	18	0	10	15
R_7	9	25	17	31	27	10	0	15
R_8	33	15	29	19	25	15	15	0

TABLE VIII. PAIRWISE COMPARISONS MATRIX

	R_*^B	R_*^C	R_2
R_*^B	0	7	21
R_*^C	7	0	14
R_2	21	14	0

Obtained metric values show alternatives close to each other (in particular, rank correlation degree between rankings constructed on Board and Copeland procedures basis is 0.87). Table 8 shows Kemeny pairwise distances values between found aggregate alternatives rankings considered in the work.

IV. CONCLUSION

The article considers the aggregation of various mechanisms for rating alternatives from collective choice theory. It considers perspective results and problems caused by the creation of national aggregated rating methodology of higher education institutions. Given numerical examples showed alternatives final ranking instability obtained on Board procedure basis. Therefore, it is proposed to use voting procedures satisfying the Condorcet principle as aggregation procedures for various mechanisms evaluating alternatives results. This collective choice includes the Copeland and the Kemeny procedures considered in the article. The authors note that collective selection theory models and methods used to build aggregated object ratings can be used in different subject areas [12, 13].

References

[1] V.G. Navodnov, G.N. Motova, O.E. Ryzhakova, “Universities effectiveness Russian Monitoring international ratings and results comparison using the league analysis method”, Educational Issues, no. 3, pp. 130–151, 2019.

[2] Retrieved from: https://best-edu.ru/ratings/nacionalnyj-agregirovannyj-rejting

[3] O.I. Larichev, Theory and decision-making methods, as well as the events chronicle in magic countries: a textbook. Moscow: Logos.

[4] F.T. Aleskerov, E. Kurbanov, “On collective choice rules manipulability degree”, Automation and Telemechanics, Iss. 10, pp. 134–136, 1988.

[5] F.T. Aleskerov, D.S. Karabekyan, R.M. Sanver, V.I. Yakuba, “Known aggregation schemes under multiple choice conditions manipulability assessment”, Econ. Theory Probl., no. 1-2, pp. 37–61, 2009.

[6] M.A. Aizerman, F.T. Aleskerov, Options choice: Theory Fundamentals. Moscow: Nauka, Ch. ed. Phys.-Math. lit., 1990, 240 p.

[7] V.I. Volsky, “Voting procedures in small groups”, Manag. Probl., no. 2, pp. 2–40, 2016.

[8] J. Kemeny, J. Snell, Cybernetic modeling. Some applications., New York, 1963–1970, Per. ingl. B.G. Mirkin, ed. I.B. Gutchina. Moscow: Publ. House “Soviet Radio”, 1972, 192 p.

[9] B.G. Litvak, Expert information. Obtaining and analysis methods. Moscow: Radio and communication, 1982.

[10] A.I. Orlov, Decision Making Theory, Tutorial. Moscow: Publ. house “Mart”, 2004, 656 p.

[11] A.I. Orlov, Econometrics. Moscow: Publ. house “Exam”, 2002, 576 p.

[12] F.T. Aleskerov, E.V. Katayeva, V.V. Pislyakov, V.I. Yakuba, “Scientists contribution assessment by threshold aggregation”, Large syst. manag. works collect., no. 44, pp. 172–189, 2013.

[13] B.E. Nikitin, M.N. Ivliev, L.A. Korobov, “Scientific periodicals rating calculation and analysis”, Voronezh State Univer. of Engineeri. Technol. Vest., vol. 79, no. 4(74), pp. 97–103, 2017.