Lietuvos matematikos rinkinys
Lietuvos matematikų draugijos darbai, ser. B
57 t., 2016, 37–42

ISSN 0132-2818
www.mii.lt/LMR/

Balsavimo teorijos metodų statistinė lyginamoji analizė

Aleksandras Krylovas, Natalja Kosareva

Vilniaus Gedimino technikos universitetas, Fundamentinių mokslų fakultetas
Saulėtekio al. 11, LT-10223 Vilnius
E. paštas: aleksandras.krylovas@vgtu.lt, natalja.kosareva@vgtu.lt

Santrauka. Straipsnyje Monte Carlo eksperimentais palyginti 8 kandidatų rūšiavimo algoritmai. 6 iš jų sudaryti balsavimo teorijos pagrindu ir 2 – Kemeny medianos pagrindu. Didžiausius teisingų sprendimų skaičiaus vidurkius ir mažiausius nebaigtų balsavimo procedūrų skaičiaus vidurkius parodė Kemeny medianos pagrindu grindžiami rūšiavimo algoritmai.

Raktiniai žodžiai: balsavimo teorijos metodai, Monte Carlo eksperimentai, Kemeny mediana.

1 Įvadas

Sprendžiant grupinio sprendimo priėmimo uždavinius, moksliniuose šaltiniuose randame daug skirtingų metodų. Kai kuriuos iš šių metodų palyginsime tarpusavyje Monte Carlo eksperimentų pagalba. Apsiribosime klasikiniais prioritetinio balsavimo teorijos metodais [2], dažnai taikomais grupinio vertinimo uždaviniams spręsti. Taip pat įtrauksime 2 metodus, pasiūlytus šio straipsnio autorių [4], kai pirmenybės nustatomos konstruojant ekspertų nuomonių Kemeny medianą [3]. Tarkime, kad turime N ekspertų vertinimo rezultatus vertinant n objektų: $x_1, x_2, \ldots, x_n$.

Ekspertai	x_1	x_2	$\ldots$	x_n
1	$r_1^{(1)}$	$r_2^{(1)}$	$\ldots$	$r_n^{(1)}$
2	$r_1^{(2)}$	$r_2^{(2)}$	$\ldots$	$r_n^{(2)}$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
N	$r_1^{(N)}$	$r_2^{(N)}$	$\ldots$	$r_n^{(N)}$

Čia $r_j^{(i)}$ yra j-jo objekto rangas, kurį nustatė i-asis ekspertas. Mūsų tikslas – išranguoti n objektų taip, kad šis prioritetų nustatymas geriausiai atitiktų ekspertų išreikštą nuomonę.

2 Balsavimo teorijos metodai

Nagrinėsime tokius 5 kandidatų 10 išdėstymų (rikiuočių):

$$x_3 \succ x_1 \succ x_2 \succ x_5 \succ x_4, \qquad x_4 \succ x_3 \succ x_2 \succ x_1 \succ x_5,$$
$$x_5 \succ x_4 \succ x_3 \succ x_2 \succ x_1, \qquad x_4 \succ x_5 \succ x_3 \succ x_2 \succ x_1,$$

$$\begin{array}{ll} x_3 \succ x_4 \succ x_5 \succ x_2 \succ x_1, & x_4 \succ x_1 \succ x_5 \succ x_3 \succ x_2, \\ x_5 \succ x_4 \succ x_3 \succ x_2 \succ x_1, & x_3 \succ x_4 \succ x_1 \succ x_2 \succ x_5, \\ x_5 \succ x_3 \succ x_1 \succ x_4 \succ x_2, & x_3 \succ x_1 \succ x_4 \succ x_5 \succ x_2. \end{array} \tag{1}$$

Trumpai pristatysime prioritetinių balsavimų teorijos metodus [2], kurių autoriai atsispindi metodų pavadinimuose. Visi šie metodai buvo pasiūlyti dar XVIII amžiuje. Analizei atlikti patogu priskirti kiekvienam iš 5 kandidatų rangą 1 už paskutinę vietą, 2 – už priešpaskutinę ir t. t. Paskutiniame stulpelyje surašytos rangų sumos:

$$\begin{array}{cccccccccc|c} 4 & 2 & 1 & 1 & 1 & 4 & 1 & 3 & 3 & 4 & 24 \\ 3 & 3 & 2 & 2 & 2 & 1 & 2 & 2 & 1 & 1 & 19 \\ 5 & 4 & 3 & 3 & 5 & 2 & 3 & 5 & 4 & 5 & 39 \\ 1 & 5 & 4 & 5 & 4 & 5 & 4 & 4 & 2 & 3 & 37 \\ 2 & 1 & 5 & 4 & 3 & 3 & 5 & 1 & 5 & 2 & 31 \end{array} \tag{2}$$

Paprastas ir dažnai taikomas metodas, vadinamas Borda metodu – paskelbti nugalėtoju (lyderiu) daugiausiai taškų surinkusį kandidatą, o pralaimėjusiu (autsaideriu) – mažiausiai. Šiuo atveju nugalėtojas Borda metodo atžvilgiu būtų trečiasis kandidatas, o pralaimėjusiu – antrasis.

Sudarykime dar vieną lentelę, kurioje surašykime kiekvieno kandidato užimtų vietų skaičius:

Kandidatas	1 vieta	2 vieta	3 vieta	4 vieta	5 vieta
1 kand.	0	3	2	1	4
2 kand.	0	0	2	5	3
3 kand.	4	2	3	1	0
4 kand.	3	4	1	1	1
5 kand.	3	1	2	2	2

Lyderį ir autsaiderį galima nustatyti pagal pirmąsias (daugumos metodas) ir paskutiniąsias (mažumos metodas) vietas. Taigi nugalėtojas daugumos metodu bus trečiasis kandidatas, o autsaideris yra nevienintelis (pirmasis arba antrasis). Mažumos metodas nustato tą patį lyderį – trečiąjį kandidatą, o autsaideris – pirmasis.

Prioritetinių balsavimų teorijoje yra žinomas Condorcet principas – lyginti kiekvieną kandidatą su kiekvienu. Skaičiuosime kiek kartų pirmasis kandidatas turėjo pranašumą (laimėjo dvikovų) prieš kitus:

$x_1 \succ x_2$	$x_1 \succ x_3$	$x_1 \succ x_4$	$x_1 \succ x_5$
5	1	3	5

Matome, kad pirmasis kandidatas nėra stipresnis Condorset principo prasme už bet kurį kitą kandidatą.

Analogiškai apskaičiuokime trečiojo kandidato dvikovų laimėjimus:

$x_3 \succ x_1$	$x_3 \succ x_2$	$x_3 \succ x_4$	$x_3 \succ x_5$
9	10	5	5

1 lentelė. Kandidatų rūšiavimo algoritmai, sudaryti pagal balsavimo teorijos princpus.

Metodas	Algoritmas
$C_\uparrow$	Condorset metodu nustatomas *autsaideris* ir jis pašalinamas iš kandidatų sąrašo. Vėl nustatomas *autsaideris* iš likusių kandidatų ir t. t.
$C_\downarrow$	Condorset metodu nustatomas *lyderis* ir jis pašalinamas iš kandidatų sąrašo. Vėl nustatomas *lyderis* iš likusių kandidatų ir t. t.
D	Daugumos metodu nustatomas *lyderis* ir jis pašalinamas iš kandidatų sąrašo. Vėl nustatomas *lyderis* iš likusių kandidatų ir t. t.
M	Mažumos metodu nustatomas *autsaideris* ir jis pašalinamas iš kandidatų sąrašo. Vėl nustatomas *autsaideris* iš likusių kandidatų ir t. t.
$B_\uparrow$	Borda metodu nustatomas *autsaideris* ir jis pašalinamas iš kandidatų sąrašo. Vėl nustatomas *autsaideris* iš likusių kandidatų ir t. t.
$B_\downarrow$	Borda metodu nustatomas *lyderis* ir jis pašalinamas iš kandidatų sąrašo. Vėl nustatomas *lyderis* iš likusių kandidatų ir t. t.

Taigi trečiasis kandidatas laimėjo prieš pirmąjį ir antrąjį ir pagal Condorset metodą. Nugalėtojas šiuo atveju yra ketvirtasis kandidatas, o autsaiderių yra du – pirmasis ir antrasis kandidatai. Pastebėkime, kad „geriausias“ kolektyvinės atrankos metodas lyderiui arba autsaideriui nustatyti neegzistuoja [1].

Nagrinėsime 6 rūšiavimo algoritmus, kurių aprašymai pateikti 1-oje lentelėje, sudarytus Condorset, Borda, daugumos ir mažumos metodų pagrindu [2].

Visais atvejais algoritmas nutraukia veiklą, kai autsaideris arba lyderis nustatomas nevienareikšmiškai.

3 Kemeny medianos metodai

Be pateiktų rūšiavimo algoritmų nagrinėsime dar du, sudarytus Kemeny medianų pagrindu [4]. Kiekvieną rikiuotę

$$X^{(r)}: \quad x_{i_1}^{(r)} \succ x_{i_2}^{(r)} \succ \cdots \succ x_{i_n}^{(r)}$$

atitinka kvadratinė matrica

$$A^{(r)} = \left(a_{ij}^{(r)}\right)_{n\times n}, \qquad a_{i_j}^{(r)} = \begin{cases} 1, & x_i^{(r)} \succ x_j^{(r)}. \\ 0, & x_i^{(r)} \preceq x_j^{(r)}. \end{cases}$$

Pavyzdžiui, rikiuotę $x_5 \succ x_4 \succ x_3 \succ x_2 \succ x_1$ atitinka matrica

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Atstumas (Kemeny prasme) tarp rikiuočių $X^{(r_1)}$ ir $X^{(r_2)}$ apibrėžiamas taip:

$$\rho_K\left(X^{(r_1)}, X^{(r_2)}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_{ij}^{(r_1)} - a_{ij}^{(r_2)}\right|.$$

Rikiuočių rinkinio $\{X^{(1)}, X^{(2)}, \dots, X^{(r)}\}$ (Kemeny) mediana vadinama tokia rikiuotė M_K, kad

$$M_K = \arg\min_M \sum_{i=1}^{n} \rho_K\bigl(X^{(i)}, M\bigr)$$

nagrinėjant visas įmanomas rikiuotes M. Taigi konstruodami rikiuočių rinkinio medianą turime dar vieną kandidatų rūšiavimo algoritmą (žymėsime K).

Apibrėžkime kitaip atstumą tarp rikiuočių $X^{(r_1)}$ ir $X^{(r_2)}$ traktuodami atitinkamas perstatas, kaip vektorius:

$$\rho_P\bigl(X^{(r_1)}, X^{(r_2)}\bigr) = \sum_{i=1}^{n} \bigl|x_i^{(r_1)} - x_i^{(r_2)}\bigr|.$$

Pavyzdžiui,

$$\rho_P(x_1 \succ x_2 \succ x_3, x_2 \succ x_3 \succ x_1) = |1-2| + |2-3| + |3-1| = 4.$$

Analogiškai apibrėžiame medianą M_P ir turime dar vieną rūšiavimo algoritmą (žymėsime P):

$$M_P = \arg\min_M \sum_{i=1}^{n} \rho_P\bigl(X^{(i)}, M\bigr).$$

4 Skaičiavimo eksperimentai

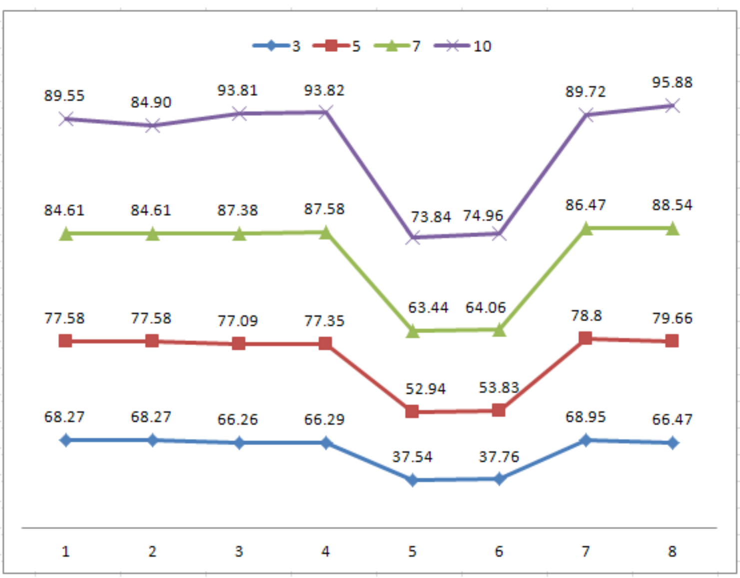
Nagrinėsime atvejus, kai r ekspertų (rinkėjų) rikiuoja 3, 4 arba 5 kandidatus ir šio rikiavimo (balsavimo) rezultatas apdorojamas kiekvienu algoritmu $C_\uparrow$, $C_\downarrow$, D, M, $B_\uparrow$, $B_\downarrow$, K, P. Rikiuočių rinkinys (balsavimo rezultatas) generuojamas taip. Kiekvienas ekspertas su tikimybe p, lygia 0.3, 0.5, 0.6 arba 0.7, gali pasirinkti vienintelę fiksuotą rikiuotę, kurią traktuojame, kaip „teisingą“, arba atitinkamai su tikimybe 0.7, 0.5, 0.4, 0.3 bet kurią „neteisingą“ rikiuotę (jų yra atitinkamai 5, 23 arba 119, priklausomai nuo kandidatų skaičiaus) ir visų „neteisingų“ rikiuočių pasirodymo tikimybės yra vienodos.

Kiekvieno algoritmo taikymo rezultatas gali būti T – rūšiavimo rezultatas sutapo su „teisinga“ rikiuote, K – rezultatas nesutapo su „teisinga“ rikiuote, N – rūšiavimo atlikti nepavyko. Lentelėje pateikti Monte Carlo eksperimentų rezultatai, kai kiekvienas atsitiktinis balsavimas buvo generuojamas kompiuteriu 1000 kartų ir atitinkamame lentelės langelyje surašytas reikšmių T, K, N stulpelis. Kito stulpelio langelyje surašomas kito kompiuterinio balsavimo rezultatas. Kiekvienas 1000 balsavimų eksperimentas buvo pakartotas 10 kartų.

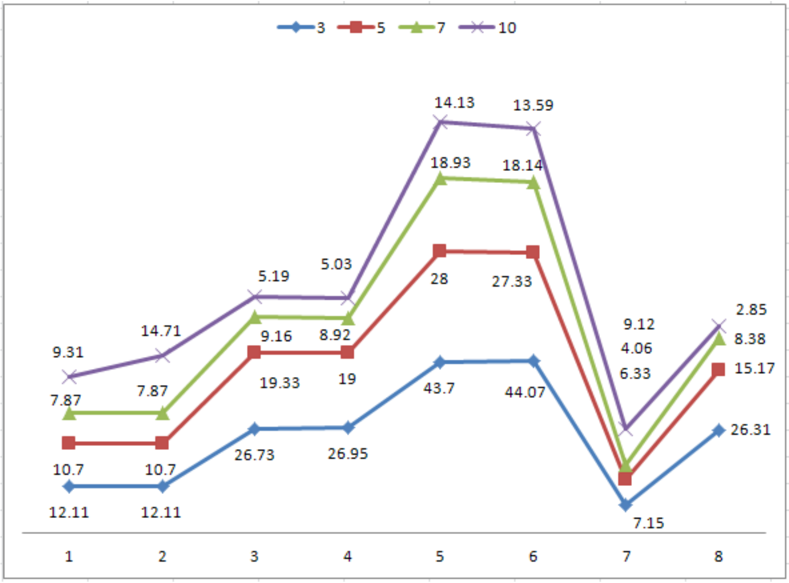
2-oje lentelėje parodyti eksperimentų rezultatai, kai nagrinėjami 5 kandidatai, 10 ekspertų, $p = 0.6$. 1 ir 2 paveiksluose pavaizduoti atitinkamai teisingų sprendimų ir nebaigtų procedūrų skaičiaus vidurkiai, kai nagrinėjami 5 kandidatai, 3, 5, 7, 10 ekspertų, $p = 0.6$. Monte Carlo statistinio eksperimento rezultatai parodė, kad Kemeny medianos pagrindu sudaryti rūšiavimo metodai K (7) ir P (8) daugeliu atvejų yra pranašesni už balsavimo teorijos metodus. Jie daugeliu atveju parodė aukščiausią teisingų sprendimų skaičiais vidurkį ir/arba mažiausią nebaigtų balsavimo procedūrų skaičiaus vidurkį.

2 lentelė. Eksperimentų rezultatai 8 rūšiavimo algoritmams: 5 kandidatai, 10 ekspertų, $p = 0.6$.

1	$C_\uparrow$	T	749	762	735	731	738	749	786	757	769	758
		K	27	34	27	32	36	40	21	31	34	33
		N	224	204	238	237	226	211	193	212	197	209
2	$C_\downarrow$	T	677	668	642	657	659	666	686	665	681	665
		K	12	13	11	16	9	8	4	8	10	13
		N	311	319	347	327	332	326	310	327	309	322
3	D	T	829	830	825	812	818	833	848	836	837	828
		K	32	26	25	35	37	28	34	26	27	36
		N	139	144	150	153	145	139	118	138	136	136
4	M	T	842	843	823	828	814	828	865	833	833	832
		K	20	31	28	29	42	29	22	30	34	20
		N	138	126	149	143	144	143	113	137	133	148
5	$B_\uparrow$	T	579	596	557	562	559	571	601	574	600	543
		K	190	180	195	187	195	176	182	200	162	203
		N	231	224	248	251	246	253	217	226	238	254
6	$B_\downarrow$	T	577	565	578	549	568	579	609	588	600	578
		K	183	176	192	187	176	189	163	186	178	192
		N	240	259	230	264	256	232	228	226	222	230
7	K	T	756	762	739	738	742	753	788	757	772	760
		K	30	36	30	35	37	42	23	39	37	36
		N	214	202	231	227	221	205	189	204	191	204
8	P	T	874	883	866	862	857	881	888	871	879	869
		K	46	38	42	54	53	42	45	40	48	50
		N	80	79	92	84	90	77	67	89	73	81



1 pav. Teisingų sprendimų skaičiaus vidurkiai: 5 kandidatai, 3, 5, 7, 10 ekspertų, $p = 0.6$.



2 pav. Nebaigtų procedūrų skaičiaus vidurkiai: 5 kandidatai, 3, 5, 7, 10 ekspertų, $p = 0.6$.

Literatūra

[1] K.J. Arrow. *Social Choice and Individual Values*. J. Wiley/Chapman & Hall, New Haven, New York, London, 1951.

[2] P.C. Fishburn. A comparative analysis of group decision methods. *Behav. Sci.*, **16**(6):538–544, 1971.

[3] J.G. Kemeny and J. Laurie Snell. *Mathematical Models in the Social Sciences*. New York, 1963.

[4] A. Krylovas, E.K. Zavadskas, N. Kosareva and S. Dadelo. New KEMIRA method for determining criteria priority and weights in solving MCDM problem. *Int. J. Info. Tech. Dec. Mak.*, **13**(6):1119–1133, 2014.

SUMMARY

Statistical benchmarking of voting theory methods
A. Krylovas, N. Kosareva

In the article 8 candidates sorting algorithms were compared by Monte Carlo experiments. 6 of them are constructed on the basis of voting theory and 2 – on the basis of Kemeny median. The largest number of correct decisions averages and the least average number of unfinished voting procedures were shown by the median-based sorting algorithms.

Keywords: voting theory methods, Monte Carlo experiments, Kemeny median.