213/2005

Raport Badawczy

Research Report

RB/41/2005

Wyznaczanie oceny grupowej przy użyciu metody Dodgsona. Wady i zalety

H. Bury, D. Wagner

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 8373578

fax: (+48) (22) 8372772

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Janusz Kacprzyk

Warszawa 2005

Wyznaczanie oceny grupowej przy użyciu metody Dodgsona. Wady i zalety

Wstęp

Metoda Dodgsona [1, 5] należy do grupy metod, w których ocena grupowa jest wyznaczana na podstawie porównań parami. Wywodzą się one z metody Condorceta [2, 6], której podstawę stanowi intuicyjne założenie, że obiekt który ma być uznany za najlepszy – zgodnie z uporządkowaniami podanymi przez ekspertów – powinien być oceniony przez większość ekspertów jako lepszy od pozostałych. Stosowanie tej metody utrudnia fakt, że zwycięzca w sensie Condorceta nie zawsze istnieje.

W 1874 Charles Dodgson, znany również jako Lewis Carrol, zaproponował sposób wyznaczenia najlepszego obiektu w sytuacji, gdy nie ma zwycięzcy w sensie Condorceta.

W ostatnich latach pojawiło się kilka interesujących publikacji analizujących właściwości metody Dodgsona [3, 4, 7, 8, 9]. Większość z nich wskazuje na istotne niedostatki tej metody. H. Nurmi [7] – powołując się na prace Fishburna [3] i Ratliffa [8] zwrócił uwagę na fakt, że nie spełnia ona tak zwanego warunku jednorodności.

1. Opis metody Dodgsona

Załóżmy, że dany jest zbiór n obiektów $O = \{O_1, \ldots, O_n\}$ oraz grupa K ekspertów. Zadaniem ekspertów jest uporządkowanie zbioru obiektów zgodnie z danym kryterium (zbiorem kryteriów). Zakładamy również, że opinie ekspertów są podawane w skali porządku. Aby nie komplikować rozważań zakładamy, że w uporządkowaniach ekspertów nie występują obiekty równoważne.

Przyjmujemy, że opinia eksperta o numerze k ma postać uporządkowania P^k, w którym obiekt uważany za najlepszy jest umieszczony na pierwszym miejscu zaś na ostatnim miejscu umieszczony jest obiekt uznany za najgorszy.

$$P^k = \{O_{i_1^k}, \ldots, O_{i_n^k}\}, \qquad k = 1, \ldots K,$$

gdzie $O_{i_j^k}$ oznacza obiekt umieszczony przez k-tego eksperta na j-tym miejscu w uporządkowaniu.

Zgodnie z założeniem, że nie występują obiekty równoważne, liczba pozycji w uporządkowaniu jest równa liczbie obiektów.

Niech l_{ij} oznacza liczbę ekspertów, którzy uznali, że obiekt O_i jest lepszy od obiektu O_j. Dla uproszczenia zapisu będziemy stosować oznaczenie $O_i > O_j$. Przy przyjętych założeniach liczba ekspertów, którzy wyrazili przeciwną opinię, to znaczy uznali, że $O_j > O_i$ jest równa $l_{ji} = K - l_{ij}$, i, j = 1,…,n.

Na podstawie współczynników l_{ij} można wyznaczyć macierz rozkładu głosów ekspertów

$$\begin{array}{c|c|c|c|c} & O_1 & O_2 & \cdots & O_n \\ \hline O_1 & - & l_{12} & \cdots & l_{1n} \\ \hline O_2 & l_{21} & - & \cdots & l_{2n} \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline O_n & l_{n1} & l_{n2} & \cdots & - \end{array} \tag{1}$$

Przyjmiemy, nie wchodząc w szczegółowe rozważania, że pojęcie większości ekspertów K_w jest określone następująco

$$K_w = \begin{cases} \dfrac{K+1}{2} & \text{dla K nieparzystych} \\ \dfrac{K}{2} + 1 & \text{dla K parzystych} \end{cases} \tag{2}$$

Definicja 1 [2], [6]

Zwycięzcą w sensie Condorceta jest obiekt O_i, dla którego $l_{ij} \geq K_w$, $j=1,\ldots, n, j \neq i$.

Definicja 2 [1], [5]

Jeżeli nie istnieje zwycięzca w sensie Condorceta, to zwycięzcą w sensie Dodgsona jest obiekt O_i, dla którego liczba zmian jego pozycji w uporządkowaniach podanych przez ekspertów, których należy dokonać, aby ten obiekt został zwycięzcą w sensie Condorceta, jest najmniejsza.

Z reguły przyjmuje się, że jedynie obiekty zajmujące sąsiednie pozycje mogą być przemieszczane [7].

2. Analiza własności metody Dodgsona

Definicja 3 [5, 7]

Ocena grupowa ma własność jednorodności, jeżeli jest ona niezmiennicza względem zwielokrotnienia liczby ekspertów. Będziemy przyjmować, że w takim samym stopniu zwielokrotnieniu ulegają liczby ekspertów, którzy podali te same uporządkowania.

Zakładamy, że K ekspertów podało uporządkowania zbioru n obiektów $O = \{O_1, \ldots, O_n\}$. Ponadto przyjmujemy, że na podstawie podanych uporządkowań można wywnioskować, że zwycięzca w sensie Condorceta nie istnieje a zwycięzcą w sensie Dodgsona jest obiekt O_i. Założymy ponadto, że dla s_i obiektów nie jest spełniony warunek $O_i > O_t$, $t = i_1, \ldots, i_{s_i}$.

Jeżeli przyjmiemy, zgodnie z (2), że K_w oznacza większość ekspertów a l_{it} oznacza liczbę ekspertów, których zdaniem $O_i > O_t$, to liczba zmian opinii ekspertów niezbędna, aby obiekt O_i został zwycięzcą w sensie Condorceta – oznaczymy ją przez d_i – jest równa

$$d_i = s_i K_w - \sum_{t \in J_i} l_{it} \qquad J_i = \{t : l_{it} < K_w\} \quad s_i = \mathrm{card}(J_i) \tag{3}$$

Przykład 1.

Dany jest zbiór 5 obiektów $O = \{O_1, O_2, O_3, O_4, O_5\}$ oraz zbiór 11 uporządkowań $\{P^1,\ldots,P^{11}\}$ podanych dla tych obiektów

Tabela 1

L.ekspertów	Uporządkowanie	
1	P^1:	$O_4 > O_2 > O_3 > O_1 > O_5$
1	P^2:	$O_5 > O_1 > O_4 > O_3 > O_2$
1	P^3:	$O_4 > O_2 > O_3 > O_5 > O_1$
1	P^4:	$O_3 > O_1 > O_4 > O_2 > O_5$
1	P^5:	$O_3 > O_1 > O_4 > O_5 > O_2$
1	P^6:	$O_1 > O_4 > O_2 > O_3 > O_5$
1	P^7:	$O_3 > O_1 > O_4 > O_2 > O_5$
1	P^8:	$O_2 > O_1 > O_5 > O_3 > O_4$
1	P^9:	$O_3 > O_1 > O_4 > O_5 > O_2$
1	P^{10}:	$O_1 > O_4 > O_5 > O_2 > O_3$
1	P^{11}:	$O_2 > O_1 > O_4 > O_3 > O_5$

Liczba ekspertów jest nieparzysta, a zatem, zgodnie z (2), liczba ekspertów stanowiących większość $K_w = \frac{K+1}{2} = \frac{12}{2} = 6$.

Macierz rozkładu głosów ekspertów ma postać

	O_1	O_2	O_3	O_4	O_5	d_i	WB_i
O_1	-	7	5	9	9	$d_1 = 1$	30
O_2	4	-	6	2	7	$d_2 = 2 + 2 + 2 = 6$	19
O_3	6	5	-	5	8	$d_3 = 1 + 1 = 2$	24
O_4	2	9	6	-	9	$d_4 = 4$	26
O_5	2	4	3	2	-	$d_5 = 4 + 2 + 3 + 4 = 13$	11

(4)

Z macierzy (4) bezpośrednio wynika, że nie istnieje zwycięzca w sensie Condorceta. W przedostatniej kolumnie macierzy (4) są podane odległości d_i dla poszczególnych obiektów. Odległość d_i osiąga najmniejszą wartość dla d_1, tzn. dla obiektu O_1; zależności między odległościami d_i, i = 1, ..., 5) są jak następuje: $d_1 < d_3 < d_4 < d_2 < d_5$.

Obiekt O_1 jest zwycięzcą w sensie Dodgsona. Aby obiekt ten został zwycięzcą w sensie Condorceta, należy dokonać zmian opinii jednego z ekspertów np. w uporządkowaniu P^1, P^4, P^5, P^7 lub P^9, tzn. zamienić miejscami obiekty O_3 i O_1.

Warto zauważyć, że obiekt O_i jest również zwycięzcą w sensie Bordy. Kolejność wartości wskaźnika Bordy WB_i jest następująca:

$WB_1 > WB_4 > WB_3 > WB_2 > WB_5$.

2.1. Metoda Dodgsona a warunek jednorodności

Zakładając, że liczba ekspertów K jest parzysta, dla obiektu d_i mamy

$$d_i = s_i\left(\frac{K}{2}+1\right) - \sum_{t\in J_i} l_{it} \qquad J_i = \left\{t : l_{it} < \frac{K}{2}+1\right\} \; s_i = \mathrm{card}(J_i) \tag{5}$$

Aby wykazać, że ocena grupowa tworzona przy użyciu metody Dodgsona nie spełnia warunku jednorodności, rozpatrzymy dowolny obiekt O_j.

Dla obiektu O_j „odległość" d_j jest równa odpowiednio

$$d_j = s_j\left(\frac{K}{2}+1\right) - \sum_{t\in J_j} l_{jt} \qquad J_j = \left\{t : l_{jt} < \frac{K}{2}+1\right\} \; s_j = \mathrm{card}(J_j) \tag{6}$$

Zgodnie z założeniem, że obiekt O_i jest zwycięzcą w sensie Dodgsona przyjmiemy, że

$$d_i < d_j. \tag{6}$$

Jeżeli liczba ekspertów zostanie zwiększona do ρK[1], przy czym – jak podkreślono we Wstępie - zakładamy, że w tej samej skali wzrosną wszystkie liczby l_{it} oraz l_{jt}, to nowe odległości – oznaczymy je odpowiednio przez $d_{i\rho}$, oraz $d_{j\rho}$ – można zapisać następująco:

$$d_{i\rho} = s_i\left(\frac{\rho K}{2}+1\right) - \sum_{t\in J_i} \rho l_{it} \tag{8}$$

$$d_{j\rho} = s_j\left(\frac{\rho K}{2}+1\right) - \sum_{t\in J_j} \rho l_{jt} \tag{9}$$

Można łatwo wykazać, że

$$d_{i\rho} = \rho\left[s_i\left(\frac{K}{2}+1\right) - \sum_{t\in J_i} \rho l_{it}\right] - s_i(\rho-1) = \rho d_i - s_i(\rho-1) = \rho(d_i - s_i) + s_i \tag{10}$$

[1] Zakładamy, że ρ jest liczbą całkowitą.

Podobnie $d_{jp} = \rho(d_j - s_j) + s_j$. (7)

Przeanalizujemy teraz, jakie warunki powinny być spełnione, aby

$$d_{ip} > d_{jp}. \tag{12}$$

Uwzględniając zależności (10) i (11) mamy

$$\rho[(d_i - d_j) + (s_j - s_i)] > (s_j - s_i) \tag{13}$$

Z warunku (7) wynika, że $d_i - d_j < 0$. Należy zatem rozważyć dwa przypadki:

a) $s_j - s_i > 0$ (14)

wtedy

$$\rho\left(\frac{d_i - d_j}{s_j - s_i} + 1\right) > 1. \tag{15}$$

Zatem, aby warunek (15) był spełniony, musi zachodzić nierówność

$$\frac{d_i - d_j}{s_j - s_i} > \frac{1}{\rho} - 1 \tag{16}$$

b) $s_j - s_i < 0$, $s_j \neq s_i$, czyli $s_i - s_j > 0$

wtedy

$$\rho\left(\frac{d_i - d_j}{s_i - s_j} - 1\right) > -1 \tag{8}$$

Przy przyjętych założeniach ta nierówność nie może być spełniona.

Założymy teraz, że liczba ekspertów K jest nieparzysta, wtedy

$$d_i = s_i\left(\frac{K+1}{2}\right) - \sum_{t \in J_i} l_{it} \tag{18}$$

W dalszej analizie należy rozróżnić dwa przypadki:

1. ρK jest liczbą parzystą, co oznacza – przy przyjętych założeniach odnośnie K – że ρ jest liczbą parzystą,

wtedy

$$d_{i\rho} = s_i\left(\frac{\rho K}{2}+1\right)-\sum_{l\in J_i}\rho l_{il} = \rho d_i - s_i(\rho-1) = \rho(d_i - s_i)+s_i \tag{19}$$

Zatem warunki, jakie powinien spełniać współczynnik ρ są takie, jak podaje nierówność (**Błąd! Nie można odnaleźć źródła odwołania.**).

2. ρK jest liczbą nieparzystą, co oznacza – przy przyjętych założeniach odnośnie K – że ρ jest liczbą nieparzystą, wtedy

$$d_{i\rho} = s_i\left(\frac{\rho K+1}{2}\right)-\sum_{l\in J_i}\rho l_{il} = \rho d_i - \frac{s_i}{2}(\rho-1) = \rho\left(d_i - \frac{s_i}{2}\right)+\frac{s_i}{2} \tag{20}$$

Zatem warunek (20) przyjmie postać

$$\rho\left[(d_i - d_j)+\left(\frac{s_j - s_i}{2}\right)\right] > \frac{s_j - s_i}{2} \tag{21}$$

Analiza nierówności (21) wymaga uwzględnienia dwóch przypadków:

a) $s_j - s_i > 0,$ (9)

wtedy $\rho\left(\frac{d_i - d_j}{s_j - s_i}+\frac{1}{2}\right) > \frac{1}{2}$

Aby warunek (9) mógł być spełniony, musi zachodzić nierówność

$$\frac{d_i - d_j}{s_j - s_i} > \frac{1}{2\rho} - \frac{1}{2} \tag{10}$$

b) $s_j - s_i < 0$, $s_j \neq s_i$, czyli $s_i - s_j > 0$, wtedy

$$\rho\left(\frac{d_i - d_j}{s_i - s_j} - \frac{1}{2}\right) > -\frac{1}{2}. \tag{11}$$

Z nierówności (11) bezpośrednio wynika, że przy przyjętych założeniach nierówność ta nie może być spełniona.

Należy podkreślić, że przy $s_i = s_j$ zarówno nierówność (10), jak i (11) nie mogą być spełnione.

Przykład 2

W pracy Ratliffa [8] jest podany następujący przykład przedstawiający uporządkowania 5 obiektów podane przez 70 ekspertów. Dane dotyczące tego przykładu zawiera Tabela 2.

Tabela 2

Liczba ekspertów		Uporządkowanie	
11	(44)	P^1:	$O_1 > O_2 > O_3 > O_4 > O_5$
9	(36)	P^2:	$O_2 > O_1 > O_3 > O_4 > O_5$
14	(56)	P^3:	$O_5 > O_1 > O_4 > O_3 > O_2$
14	(56)	P^4:	$O_2 > O_3 > O_4 > O_5 > O_1$
11	(44)	P^5:	$O_5 > O_4 > O_2 > O_3 > O_1$
11	(44)	P^6:	$O_1 > O_3 > O_4 > O_5 > O_2$

Macierz rozkładu głosów ma w tym przypadku postać

	O_1	O_2	O_3	O_4	O_5
O_1	–	36	45	45	31
O_2	34	–	45	34	34
O_3	25	25	–	45	25
O_4	25	36	25	–	45
O_5	39	36	45	25	–

(12)

Liczba ekspertów stanowiących większość $K_w = \frac{K}{2} + 1 = \frac{70}{2} + 1 = 36$.

Z macierzy (12) bezpośrednio wynika, że nie istnieje zwycięzca w sensie Condorceta.

Odległości d_i ($i = 1, \ldots, 5$) są jak następuje:

$$d_1 = 1 \cdot 36 - 31 = 5$$

$$d_2 = 3 \cdot 36 - 3 \cdot 34 = 6$$

$$d_3 = 3 \cdot 36 - 3 \cdot 25 = 33 \qquad (13)$$

$$d_4 = 2 \cdot 36 - 2 \cdot 25 = 22$$

$$d_5 = 1 \cdot 36 - 25 = 11$$

Mamy zatem $d_1 < d_2 < d_5 < d_4 < d_3$

Zwycięzcą w sensie Dodgsona jest więc obiekt O_1. Aby obiekt ten został zwycięzcą w sensie Condorceta, należy dokonać zmian opinii 5 ekspertów w uporządkowaniu P^3.

Aby obiekt O_2 został zwycięzcą w sensie Condorceta, należy dokonać dwu zmian opinii ekspertów w uporządkowaniu P^1, dwu zmian w uporządkowaniu P^5 oraz dwu zmian w uporządkowaniu P^6, razem 6 zmian.

Jeżeli liczbę ekspertów podających swoje opinie w postaci uporządkowań $P^1, \ldots, P^6$ pomnożymy przez 4 (nowe liczby ekspertów są podane w Tabeli 2 w nawiasach), to macierz rozkładu głosów ekspertów przyjmie postać:

	O_1	O_2	O_3	O_4	O_5
O_1	–	144	180	180	124
O_2	136	–	180	136	136
O_3	100	100	–	180	100
O_4	100	144	100	–	180
O_5	156	144	180	100	–

(14)

Liczba ekspertów stanowiących większość $K_w = \frac{\rho K}{2} + 1 = \frac{70 \cdot 4}{2} + 1 = 141$.

Odległości $d_{i\rho}$ $(i = 1, \ldots, 5)$ są jak następuje:

$$d_{1\rho} = 1 \cdot 141 - 124 = 17 \qquad d_{2\rho} = 3 \cdot 141 - 3 \cdot 136 = 15 \qquad d_{3\rho} = 3 \cdot 141 - 3 \cdot 100 = 123$$

$$d_{4\rho} = 2 \cdot 141 - 2 \cdot 100 = 82 \qquad d_{5\rho} = 1 \cdot 141 - 100 = 41 \tag{28}$$

Mamy zatem $d_{2\rho} < d_{1\rho} < d_{5\rho} < d_{4\rho} < d_{3\rho}$.

Zwycięzcą w sensie Dodgsona jest więc obiekt O_2 a nie obiekt O_1.

W przypadku obiektu O_1 mamy: $s_1 = 1$, $\sum_{t \in J_1} l_{1t} = 31$, $d_1 = 5$, (29)

a dla obiektu O_2 mamy: $s_2 = 3$, $\sum_{t \in J_2} l_{2t} = 102$, $d_2 = 6$.

Liczba ekspertów K = 70 jest liczbą parzystą oraz $s_2 > s_1$, a zatem na podstawie wzoru (**Błąd! Nie można odnaleźć źródła odwołania.**) warunek, aby po zwielokrotnieniu liczby ekspertów o liczbę ρ zwycięzcą w sensie Dodgsona został obiekt O_2 jest następujący:

$$\rho[(5-6)+(3-1)]>(3-1) \tag{15}$$

Przy przyjętych założeniach mamy

$$\frac{5-6}{3-1}=-\frac{1}{2}>-1+\frac{1}{\rho} \tag{16}$$

skąd bezpośrednio wynika, że nierówność (15) jest spełniona jeżeli $\rho > 2$.

Tak więc obiekt O_2 będzie również zwycięzcą w sensie Dodgsona dla $\rho \geq 3$, a nie tylko dla $\rho = 4$, jak podał Ratliff.

Dla $\rho = 3$ macierz rozkładu głosów ekspertów ma postać

	O_1	O_2	O_3	O_4	O_5
O_1	–	108	135	135	93
O_2	102	–	135	102	102
O_3	75	75	–	135	75
O_4	75	108	75	–	135
O_5	117	108	135	75	–

(17)

Liczba ekspertów stanowiących większość $K_w = \frac{\rho K}{2}+1=\frac{70\cdot 3}{2}+1=106$.

Odległości $d_{i\rho}$ (i = 1, ...,5) są jak następuje:

$d_{1\rho} = 1\cdot 106 - 93 = 13$ $d_{2\rho} = 3\cdot 106 - 3\cdot 102 = 12$ $d_{3\rho} = 3\cdot 106 - 3\cdot 75 = 93$

$d_{4\rho} = 2\cdot 106 - 2\cdot 75 = 62$ $d_{5\rho} = 1\cdot 106 - 75 = 31$

Mamy zatem $d_{2\rho} < d_{1\rho} < d_{5\rho} < d_{4\rho} < d_{3\rho}$.

Uwagi końcowe

Metodę Dodgsona można traktować jako naturalne rozszerzenie metody Condorceta, bowiem polega ona na „skorygowaniu" rzeczywistej macierzy rozkładu głosów ekspertów tak, aby zaistniał zwycięzca w sensie Condorceta. Ta korekcja macierzy rozkładu głosów ekspertów narzuca konieczność „poprawienia" rzeczywistych uporządkowań podanych przez ekspertów. Zmiany te są dokonywane w taki sposób, aby odległość (4) wyznaczona dla obiektu, którego pozycja jest zmieniana, była najmniejsza. Aby obiekt, dla którego ta odległość osiąga najmniejszą wartość, był zwycięzcą w sensie Dodgsona, muszą istnieć – w odpowiedniej liczbie – rzeczywiste uporządkowania, w których przemieszczane obiekty zajmują sąsiednie pozycje.

Fakt, że ocena grupowa nie ma własności jednorodności wynika z przyjętej postaci odległości (4). Z drugiej jednak strony należy podkreślić, że jest to metoda nieskomplikowana pozwalajaca wyznaczyć ocenę grupową bez złożonych obliczeń, jak ma to miejsce wna przykład w przypadku mediany Kemeny'ego czy Litvaka. Mimo oczywistych wad można ją więc wprowadzić do katalogu metod stosowanych do wyznaczania oceny grupowej w komputerowych systemach wspomagania pracy grupowej.

Literatura

1.Black D. Theory of Committees and Elections, Cambridge University Press, Cambridge 1958

2.de Condorcet M. Essai sur l'Application de l'Analyse à la Probabilité des Rendues à la Pluralité des Voix. Paris 1785

3.Fishburn P.C. Condorcet social choice functions. SIAM Journal of Applied Mathematics, vol. 33, 1977

4.Klamler C. The Dodgson Ranking and Its Relation to Kemeny's Method and Slater Rule, Social Choice and Welfare, 2002

5.Nurmi H. Comparing voting systems, Dordrecht, D.Reidel, 1987

6.Nurmi H. Voting Paradoxes and How to Deal with Them. Springer Verlag, Berlin-Heidelberg-New York 1999

7.Nurmi H. A comparison of some distance - based choice rules in ranking environments. Report of Department of Political Science, University of Turku, 2004

8.Ratliff T.C. A comparison of Dodgson's method and Kemeny's rule, Social Choice and Welfare, vol.18, 2001

9.Ratliff T.C. A comparison of Dodgson's method and the Borda count, Economic Theory, vol. 20, 2002