Bogusław Guzik

SEKWENCYJNE WERSJE METODY HELLWIGA

1. Metoda Hellwiga

Metoda Z.Hellwiga doboru zmiennych objaśniających do liniowego równania regresji polega - jak wiadomo - na znalezieniu takiej kombinacji potencjalnych zmiennych objaśniających, dla której miernik nazwany przez Z.Hellwiga integralną pojemnością informacyjną osiąga wartość największą.

Oznaczmy przez Y zmienną objaśnianą, a przez $X_1,\ldots,X_M$ potencjalne zmienne objaśniające[1]. Wektory wyników obserwacji zmiennej Y, zmiennej X_i $(1 \leqslant i \leqslant M)$ oznaczymy odpowiednio przez y oraz x_i. Niech K będzie dopuszczalną kombinacją potencjalnych zmiennych objaśniających, tzn. taką kombinacją, że jej liczebność jest mniejsza od liczby wyników obserwacji L. Natomiast K niech będzie zbiorem kombinacji dopuszczalnych.Dalej przez kombinację zmiennych rozumieć będziemy zbiór numerów tych zmiennych.

Oznaczmy jeszcze przez H(K) integralną pojemność informacyjną kombinacji K:

$$H(K) = \sum_{i \in K} \frac{r_i^2}{\sum_{j \in K} |r_{ij}|}, \tag{1}$$

gdzie r_i jest współczynnikiem korelacji wektorów y, x_i, natomiast r_{ij} jest współczynnikiem korelacji wektorów x_i, x_j.

Najlepszą kombinacją w zbiorze $\{K\}$ jest kombinacja K_p taka, że

$$H(K_p) = \max_{\{K\}} H(K). \tag{2}$$

---

[1] Jest zrozumiałe, że każda zmienna potencjalna powinna być jakoś uzasadniona: teorią, doświadczeniem, intuicją, przekonaniem itp. Jeśli bowiem zmienne potencjalne są „byle jakie", to i ostateczna lista zmiennych zakwalifikowanych daną metodą do modelu, też może być „byle jaka".

W oryginalnej propozycji Z.Hellwiga sugeruje się procedurę kompletnego przeglądu wszystkich kombinacji dopuszczalnych.

Oznaczmy przez m maksymalny dopuszczalny rozmiar kombinacji zmiennych objaśniających

$$m = \begin{cases} M & \text{jeśli } L > M \\ L-1 & \text{jeśli } L \leqslant M \end{cases} . \tag{3}$$

W procedurze przeglądu kompletnego zbadać trzeba

$$P^m = \binom{M}{1} + \binom{M}{2} + \ldots + \binom{M}{m} \tag{4}$$

kombinacji. W szczególności, jeśli $L > M$ trzeba zbadać

$$P^M = 2^M - 1 \tag{5}$$

kombinacji. Przy dużej liczbie potencjalnych zmiennych objaśniających procedura przeglądu kompletnego jest obliczeniowo bardzo nieefektywna. Np. gdy $L > M$ i $M = 10$, zbadać trzeba 1023 kombinacji, a jeśli $M = 15$ trzeba zbadań aż 32767 kombinacji.

Z tych powodów można proponować inne, sprowadzające się do ograniczania liczby badanych kombinacji, numeryczne realizacje idei Z.Hellwiga.

2. Ograniczanie liczby badanych kombinacji

Podejścia w tym zakresie mogą być różne. Przykładowo

1) merytoryczne lub intuicyjne ograniczenie zbioru kombinacji dopuszczalnych,
2) sformułowanie i rozwiązanie zadania doboru zmiennych jako zadania programowania matematycznego,
3) zastosowanie procedur sekwencyjnych.

Przykładem podejścia pierwszego typu jest

- rozpatrzenie kombinacji o liczebnościach z pewnego ustalonego przedziału, np. z przedziału $\overline{1,5}$ [2],
- forsowanie zespołu zmiennych, które wg opinii badacza powinny znaleźć się w równaniu regresji i potraktowanie tylko pozostałych zmiennych jako „autentycznych" zmiennych potencjalnych[3],
- grupowy dobór zmiennych, w którym zamiast zbioru zmiennych rozpatruje się zbiór potencjalnych grup zmiennych i dobiera się grupy zmiennych[4].

---

[2] Wstępne ograniczanie rozmiaru kombinacji jest o tyle usprawiedliwione, że modele z większą liczbą zmiennych objaśniających na ogół źle się interpretuje. Spotyka się je ponadto dosyć rzadko.

[3] Forsowany zespół zmiennych może zawierać np. (a) zmienne uznane za „ważne", (b) zmienne reprezentujące grupy zmiennych potencjalnych itp. Pewne przykłady forsowania zmiennych omawia E.Nowak w [2], § 1.3.

[4] Na przykład dobierając zmienne do modelu popytu moglibyśmy wyróżnić grupy zmiennych: (dochód, cena dobra), (cena dobra i, popyt na dobro i) ..., (cena dobra n, popyt na dobro n).

Drugie podejście sprowadza się do rozwiązania zadania programowania: znaleźć wartości zmiennych zero-jedynkowych v_i (i=1,...,M) przy których funkcja

$$\sum_{i=1}^{M} \frac{v_i r_i^2}{\sum_{j=1}^{M} v_j |r_{ij}|} \tag{6}$$

osiąga wartość największą.

Jest to zadanie programowania binarnego z nieliniową funkcją celu. Wartość $v_i = 1$ oznacza, że zmienna X_i występuje w kombinacji optymalnej, zaś $v_i = 0$, że X_i nie wchodzi do tej kombinacji.

Postulat by $v_i = 1$, bądź $v_i = 0$ można oczywiście zapisać w postaci warunku

$$v_i(1-v_i) = 0 \quad (i=1,...,M). \tag{7}$$

Zadanie doboru zmiennych (1), (2) jest więc zadaniem szukania maksimum nieliniowej funkcji celu (6) przy nieliniowych warunkach ograniczających (7)[5].

Podejścia trzeciego typu (procedury sekwencyjne) stosowane są na przykład powszechnie przy doborze zmiennych z punktu widzenia wartości współczynnika determinacji. Do tej grupy procedur należą procedury regresji krokowej, regresji stopniowej, selekcji a priori, selekcji a posteriori itd.[6]. Brak jest natomiast opracowań na temat sekwencyjnych realizacji metody Z.Hellwiga. W artykule przedstawimy kilka takich procedur.

3. Procedura DOLØ

Jest to najprostsza (i zarazem „naiwna") procedura sekwencyjna. Polega ona na tym, że najlepszą kombinację jednoelementową uznajemy za kombinację podstawową, badamy wszystkie rozszerzenia tej kombinacji o jeden element i sprawdzamy, czy najlepsze rozszerzenie jest lepsze od kombinacji podstawowej. Jeśli nie, to kończymy obliczenia, jeśli tak to rozszerzamy to najlepsze rozszerzenie o jeden element itd.

Ogólnie procedurę można przedstawić następująco. Oznaczmy

$\mathcal{X}$ - zbiór numerów potencjalnych zmiennych objaśniających,

H(K) - wartość miary Hellwiga dla kombinacji K,

K^* - aktualnie najlepsza kombinacja,

---

[5] Propozycję rozwiązywania zadania doboru zmiennych jako zadania programowania matematycznego zgłosił M.Rocki w [3]. W pracy M.Rockiego proponuje się rozwiązywanie zadania programowania binarnego $\max \sum v_i r_i^2$ przy liniowych ograniczeniach $\sum_i v_j r_{ij} \leqslant d_i$. To sformułowanie opiera się na idei zbliżonej do propozycji Z.Hellwiga, nie jest jednak z nią identyczne.

[6] W procedurach tych postuluje się znalezienie w zbiorze kombinacji spełniających pewne dodatkowe warunki (np. spełniających postulat istotności, jak w metodzie regresji krokowej) kombinacji, dla której współczynnik determinacji jest największy.

$K_l^{\#}$ - najlepsza kombinacja w l-tej iteracji,
m - maksymalny rozmiar kombinacji ($m \leqslant \min M, \{L-1\}$),
M - liczba potencjalnych zmiennych objaśniających.

A) Iteracja zerowa (l=0)

1^o Podstaw $l \Leftarrow 0$.

2^o Znajdź najlepszą kombinację jednoelementową $K_o^{\#}$ i podstaw $K^{\#} \Leftarrow K_o^{\#}$.

B) Iteracja l-ta

3^o Podstaw $l \Leftarrow l+1$. Jeśli liczebność kombinacji $K^{\#}$ jest równa m, to STOP. Jeśli nie, to 4^o.

4^o Zbadaj wszystkie rozszerzenia kombinacji $K_{l-1}^{\#}$ o jeden element $K_{(l)}^{i} = \{K_{l-1}^{\#} \cup i\}$, gdzie $i \in \mathcal{X} \setminus K_{l-1}^{\#}$ i znajdź kombinację $K_l^{\#}$ najlepszą w zbiorze $\{K_{(l)}^{i}\}$.

5^o Jeśli $H(K_l^{\#}) \leqslant H(K_{l-1}^{\#})$, to STOP. W przeciwnym razie podstaw $K^{\#} \Leftarrow K_l^{\#}$ i przejdź do 3^o.

Liczba kombinacji badanych w najgorszym przypadku

Najgorszy, z punktu widzenia liczby badanych kombinacji, przypadek będzie wtedy gdy dochodzi się aż do badania kombinacji o największym dopuszczalnym rozmiarze m. Wówczas w procedurze DOLØ bada się

$$P_o^m = \sum_{t=1}^{M} t - \sum_{t=1}^{M-m} t = \frac{1}{2}\left[M(M+1) - (M-m)(M-m+1)\right]$$

kombinacji. W szczególności, jeśli m = M, to P_o^m równa się

$$P_o^M = \frac{M(M+1)}{2}.$$

Iloraz D_o^M liczby P_o^M do liczby kombinacji badanych w procedurze przeglądu kompletnego $P^M = 2^M - 1$ przyjmuje na przykład wartości

M	5	10	15
D_o^M	0,48	0,054	0,004

Ogólniejsza wersja procedury DOLØ:

A) Iteracja zerowa

1^o $l \Leftarrow 0$.

2^o Ustal kombinację startową $K_o^{\#}$ i podstaw $K^{\#} \Leftarrow K_o^{\#}$.

B) Iteracja l-ta - jak wyżej.

4. Procedura DOL1

Procedura DOLØ może być modyfikowana w wielu kierunkach. Na przykład można uwzględnić

- większą liczbę punktów startowych (w DOLØ był jeden punkt startowy),

- punkty startowe o różnych rozmiarach (w DOLØ punkt startowy miał rozmiar 1).

Oznaczmy

$\mathcal{X}$ - zbiór potencjalnych zmiennych objaśniających,

S - liczba punktów startowych,

$\overline{m_1,m_2}$ - dopuszczalny rozmiar kombinacji startowej ($m_1 \geqslant 1$, $m_2 \leqslant m$),

K_s - s-ta kombinacja startowa (s-ty punkt startowy),

$K_s^{\#}$ - aktualnie najlepsza kombinacja uzyskana z s-tego punktu startowego,

$K_{s,l}^{\#}$ - kombinacja najlepsza po wykonaniu l-tej iteracji dla s-tego punktu startowego,

$K^{\#}$ - aktualnie najlepsza kombinacja z procedury sekwencyjnej.

Ogólnie procedurę można opisać następująco:

- Ustal kombinacje startowe $K_1,\ldots,K_S$.
- Dla każdego punktu startowego K_s zastosuj procedurę DOLØ w jej wersji ogólniejszej i znajdź kombinację najlepszą $K^{\#}(s)$ z tego punktu startowego ($s=1,\ldots,S$).
- Za rozwiązanie zadania przyjmij tę $K^{\#}(s)$, dla której wartość H jest największa.

Przedstawimy pewien wariant tej procedury, który nazwiemy DOL1. Kombinacje startowe ustalane będą na drodze pełnego przeglądu kombinacji o rozmiarach z przedziału $\overline{m_1,m_2}$. W procedurze znajdą się pewne rozwiązania eliminujące ponowne badanie kombinacji (zob. punkt 5° i 9°).

A) Etap wstępny

1° W zbiorze kombinacji o liczebnościach z przedziału $\overline{m_1,m_2}$ znajdź na drodze przeglądu kompletnego S najlepszych kombinacji.

2° Kombinację znajdującą się na s-tym miejscu przyjmij za kombinację startową K_s ($s=1,\ldots,S$).

3° Podstaw $s \Leftarrow 0$, $K^{\#} \Leftarrow K_1$.

B) Etap s-ty (dla s-tego punktu startowego)

4° Podstaw $s \Leftarrow s+1$.

Jeśli $s > S$, to STOP. W przeciwnym razie podstaw $l \Leftarrow 0$, $K_s^{\#} \Leftarrow K_s$, $K_{s,l}^{\#} \Leftarrow K_s$.

5° Jeśli liczebność kombinacji startowej K_s jest równa m_2, to przejdź do 6°, a jeśli jest mniejsza od m_2 to 4°.

6° Podstaw $l \Leftarrow l+1$.

Jeśli liczebność kombinacji $K_{s,l-1}^{\#}$ jest równa m, to 10°, a jeśli jest mniejsza od m, to 7°.

7° Znajdź kombinację $K_{s,l}^{\#}$ najlepszą w zbiorze kombinacji $\{K_{s,l-1}^{\#} \cup i\}$, $i \in \mathcal{X} \setminus K_{s,l-1}^{\#}$.

8° Jeśli $H(K_{s,l}^{\#}) \leqslant H(K_{s,l-1}^{\#})$ przejdź do 10°. W przeciwnym razie podstaw $K_s^{\#} \Leftarrow K_{s,l}^{\#}$ i wykonaj 9°.

9° Jeśli s = 1 przejdź do 6°, a jeśli s > 1, zbadaj czy kombinacja $K^{*}_{s,1}$ jest identyczna z tego samego rozmiaru rozwiązaniem K^{W}_{*} dla jakiegoś etapu wcześniejszego (w < s).

Jeśli $K^{*}_{s,1} = K^{W}_{*}$, zaniechaj obliczeń dla etapu s-tego i przejdź do 4°.

Jeśli $K^{*}_{s,1} \neq K^{1}_{*},\ldots,K^{s-1}_{*}$, przejdź do 6°.

10° Jeśli $H(K^{*}_{s}) \leqslant H(K^{*})$, przejdź do 4°, a jeśli $H(K^{*}_{s}) > H(K^{*})$ podstaw $K^{*} \Leftarrow K^{*}_{s}$ i przejdź do 4°.

Komentarza mogą wymagać działania opisane w punktach 5°, 9°. Ich celem jest zmniejszenie liczby badanych kombinacji.

Pkt. 9°: Kontynuowanie obliczeń, gdy $K^{*}_{s,1}$ jest identyczna z jakimś wcześniejszym rozwiązaniem K^{W}_{*} oznaczałoby powtarzanie obliczeń już przeprowadzonych dla kombinacji K^{W}_{*}, stąd skok do 4° jeśli $K^{*}_{s,1} = K^{W}_{*}$ [7].

Pkt. 5°: Jeśli kombinacja startowa K_s ma rozmiar mniejszy od m_2, to rozszerzając K_s o jeden element (tworząc kombinacje $\{K_s \cup i\}$ $i \in \mathcal{X} \setminus K_s$) rozpatrywać będziemy kombinacje już przebadane w etapie wstępnym, co oczywiście jest zbędne bowiem

- albo natrafimy na kombinację lepszą od K_s, którą może być jednak tylko „wcześniejsza" (już zbadana) kombinacja startowa K_w (w < s),
- albo natrafimy na kombinacje nie lepsze od K_s, co wobec działań z punktu 8° oznaczałoby zaprzestanie badania kombinacji z s-tego punktu startowego.

Liczba kombinacji badanych w najgorszym przypadku w procedurze DOL1 jest sumą dwóch wielkości

- liczby kombinacji badanych w etapie wstępnym

$$P^{D}_{W} = \binom{M}{m_1} + \binom{M}{m_1+1} + \ldots + \binom{M}{m_2},$$

- S krotności liczby kombinacji $P^{D}(1)$ badanych w najgorszym przypadku przy realizacji jednego etapu.

---

[7] W programie obliczeniowym za pomocą którego badano skuteczność procedur sekwencyjnych (zob. dalej paragraf 7), sprawdzanie, czy rozwiązanie z danego punktu startowego s jest identyczne z tego samego rozmiaru rozwiązaniem dla etapów wcześniejszych przebiegało następująco:

1° Niech $K^{*}_{n(r)}$ będzie rozwiązaniem końcowym takiej iteracji z n-tego punktu startowego, że $K^{*}_{n(r)}$ ma rozmiar r, n=1,...,s.

2° Z kombinacją $K^{*}_{n(r)}$ związujemy liczbę $TZ(r,n) = \sum_{j=1}^{M} 10^{j-1}$, gdzie

$$j = \begin{cases} 0 & \text{jeśli zmienna } X_j \text{ nie występuje w } K^{*}_{n(r)} \\ 1 & \text{jeśli zmienna } X_j \text{ występuje w } K^{*}_{n(r)}. \end{cases}$$

(Zauważmy, że liczba TZ(r,n) jest M pozycyjna. Jedynka na j-ej pozycji oznacza, że X_j występuje w kombinacji, natomiast zero na pozycji j-ej oznacza, że X_j nie występuje w kombinacji).

3° Kombinacja $K^{*}_{s(r)}$ jest identyczna z $K^{*}_{w(r)}$, w < s, jeśli TZ(r,s) = TZ(r,w) i nie jest identyczna z $K^{*}_{w(r)}$, jeśli $TZ(r,s) \neq TZ(r,w)$.

Ten najgorszy przypadek będzie wtedy, gdy liczebność kombinacji startowej jest równa m_2 i dochodzi się aż do badania kombinacji o największym dopuszczalnym rozmiarze m. Jest

$$P^D(1) = \sum_{t=1}^{M-m_2} t - \sum_{t=1}^{M-m} t = \frac{1}{2}\left[(M-m_2)(M-m_2+1) - (M-m)(M-m+1)\right].$$

Jeśli na przykład $m_1, m_2 = 1$, $m = M$, to

$$P_W^D(1) = M, \quad P^D(1) = \frac{(M-1)M}{2}.$$

Gdybyśmy wzięli trzy punkty startowe, to w najgorszym przypadku trzeba byłoby więc zbadać $M + \frac{3}{2}M(M-1)$ kombinacji. Dla tej sytuacji iloraz D_1^M liczby kombinacji badanych w najgorszym przypadku do liczby kombinacji P^M badanych w procedurze przeglądu kompletnego miałby wartości

M	5	6	10	15
D_1^M	1,13	0,81	0,14	0,01

W praktyce liczba kombinacji badanych w procedurze DOL1 będzie mniejsza ze względu na działania podane w punktach 5° oraz 9° oraz ze względu na to, że rzadko dochodzi do badania kombinacji o rozmiarze m.

Procedura DOLØ jest szczególnym przypadkiem procedury DOL1, jeśli S=1 a $m_1 = m_2 = 1$.

5. Procedura COF1

Poprzednio prezentowane procedury sekwencyjne były procedurami „od początku", gdyż sprowadzały się do stopniowego rozszerzania kombinacji startowej. Można również proponować procedury „od końca", polegające na stopniowym zawężaniu kombinacji startowej. Odpowiednikiem procedury DOL1 byłaby procedura, którą nazwiemy COF1.

A) Etap wstępny

1° W zbiorze kombinacji o rozmiarach z przedziału $\overline{m_1, m_2}$ $(m_1 \geqslant 1, m_2 \leqslant m)$ znajdź na drodze przeglądu kompletnego S najlepszych kombinacji.

2° Kombinację znajdującą się na s-tym miejscu przyjmij za kombinację startową K_s $(s = 1, \ldots S)$.

3° Podstaw $s \Leftarrow 0$, $K^{\pi} \Leftarrow K_1$.

B) Etap s-ty

4° Podstaw $s \Leftarrow s+1$.

Jeśli $s > S$, to STOP. W przeciwnym razie podstaw $l \Leftarrow 0$, $K_s^{\pi} \Leftarrow K_s$, $K_{s,l}^{\pi} \Leftarrow K_s$ i przejdź do 5°.

5° Jeśli liczebność kombinacji startowej K_s jest większa od m_1, to 4°, jeśli jest równa m_1, to 6°.

6° Podstaw $l \Leftarrow l+1$.

Jeśli liczebność kombinacji $K_{s,l-1}^{\pi}$ jest równa 1, to 10°, a jeśli jest większa od 1, to 7°.

7^o Znajdź kombinację $K^{\#}_{s,1}$ najlepszą w zbiorze kombinacji $\{K^{\#}_{s,1-1} \setminus i\}$, $i \in K^{\#}_{s,1-1}$ będących zawężeniem kombinacji $K^{\#}_{s,1-1}$ o jeden element.

8^o Jeśli $H(K^{\#}_{s,1}) \leqslant H(K^{\#}_{s,1-1})$, przejdź do 10^o. W przeciwnym razie podstaw $K^{\#}_s \Leftarrow K^{\#}_{s,1}$ i wykonaj 9^o.

9^o Jeśli s=1 przejdź do 6^o, a jeśli $s > 1$, zbadaj czy kombinacja $K^{\#}_{s,1}$ jest identyczna z tego samego rozmiaru rozwiązaniem $K^{W}_{\#}$ dla etapu wcześniejszego ($w < s$).

Jeśli $K^{\#}_{s,1} = K^{W}_{\#}$, przejdź do 4^o. Jeśli $K^{\#}_{s,1} \neq K^{1}_{\#}, \ldots, K^{s-1}_{\#}$, przejdź do 6^o.

10^o Jeśli $H(K^{\#}_s) \leqslant H(K^{\#})$, to przejdź do 4^o, a jeśli $H(K^{\#}_s) > H(K^{\#})$, podstaw $K^{\#} \Leftarrow K^{\#}_s$ i przejdź do 4^o.

Liczba kombinacji badanych w najgorszym przypadku jest sumą dwóch wielkości

- liczby kombinacji badanych w etapie wstępnym

$$P^C_W = \binom{M}{m_1} + \binom{M}{m_1+1} + \ldots + \binom{M}{m_2},$$

- S krotności liczby kombinacji $P^C(1)$ badanych w najgorszym przypadku przy realizacji jednego etapu.

Najgorszy przypadek będzie wtedy, gdy liczebność kombinacji startowej jest równa m_1 i stopniowe zawężanie kombinacji startowej doprowadzi do badania kombinacji jednoelementowych. Jest

$$P^C(1) = \sum_{t=1}^{m_1} t - 1 = \frac{m_1(m_1+1)}{2}.$$

Na przykład jeśli m_1, $m_2 = M$ oraz $m = M$, to

$$P^C_W = 1, \quad P^C(1) = \frac{M(M+1)}{2} - 1.$$

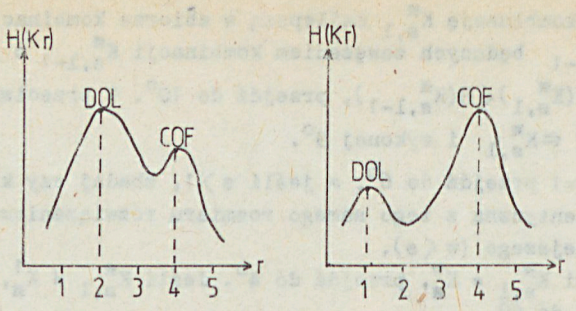
Gdybyśmy w takiej sytuacji wzięli S = 3, to w najgorszym przypadku trzeba byłoby zbadań $\frac{3}{2}M(M+1)-2$ kombinacji. Iloraz C^M_1 liczby kombinacji badanych w najgorszym przypadku do liczby kombinacji P^M badanych w procedurze przeglądu kompletnego byłby następujący

M	5	6	10	15
C^M_1	1,39	0,97	0,159	0,011

Jeśli przyjmiemy S = 1, $m_1 = m_2 = m$, to otrzymamy szczególny przypadek procedury COF1, który nazwać można COFØ, i który jest „lustrzanym" odpowiednikiem procedury DOLØ.

6. Procedura DC (dołączanie i cofanie)

Procedury DOL i COF są procedurami „jednostronnymi" i mogą być niezłe, gdy funkcja $H(K^{\#}_r)$ określająca wartości miary H dla najlepszych kom-



Rys. 1

binacji r-elementowych (r = 1,...,m) jest jednomodalna. Jeśli jednak jest to funkcja wielomodalna procedury DOL oraz COF zastosowane z osobna mogą być bardzo nieskuteczne. Na przykład jeśli jest tak, jak pokazano na rysunku 1, to w pierwszym przypadku procedura COF będzie nieskuteczna, a w drugim przypadku nieskuteczna będzie procedura DOL. Korzystne może być więc połączenie obu procedur (choć spowoduje to wzrost nakładu obliczeniowego). To połączenie nazwiemy procedurą DC.

Algorytm procedury DC

1° Wykonaj procedurę DOL przy dopuszczalnych rozmiarach kombinacji startowej m_1^D, m_2^D oraz liczbie punktów startowych S_D. Niech rozwiązaniem z DOL będzie K_D^{*}.

2° Wykonaj procedurę COF przy dopuszczalnych rozmiarach kombinacji startowej m_1^C, m_2^C i liczbie punktów startowych S_C. Niech rozwiązaniem z procedury COF będzie K_C^{*}.

3° Za rozwiązanie z procedury DC przyjmij

$$K^{*} = \begin{cases} K_D^{*} & \text{jeśli } K_D^{*} \text{ jest niegorsza od } K_C^{*} \\ K_C^{*} & \text{jeśli } K_D^{*} \text{ jest gorsza od } K_C^{*}. \end{cases}$$

Procedurę DC, w której DOL = DOL1, COF = COF1 nazwiemy procedurą DC1.

Liczba kombinacji badanych w najgorszym przypadku jest sumą liczby kombinacji badanych w najgorszym przypadku w procedurze DOL oraz liczby kombinacji badanych w najgorszym przypadku w procedurze COF.

7. Skuteczność procedur sekwencyjnych

Analityczne stwierdzenie, czy przedstawione procedury sekwencyjne są skuteczne w tym sensie, że dostarczają rozwiązania K_{*} identycznego z rozwiązaniem „optymalnym" K_p jest niezmiernie trudne nawet w przypadku najprostszej procedury sekwencyjnej DOLØ. Wyrażenie określające różnicę integralnej pojemności informacyjnej między kombinacją K, a kombinacją $K_n = \{K \cup n\}$, $n \in \mathcal{X} \setminus K$ jest bowiem następujące

$$\Delta = \frac{r_n^2}{\sum_{j \in K} |r_{nj}| + 1} - \sum_{i \in K} \frac{r_i^2}{S_i(S_i + |r_{ni}|)}$$

gdzie

$$S_i = \sum_{j \in K} |r_{ij}|.$$

Wyrażenie to jest dość zawikłane względem r_j, r_{ij}, $i,j \in K_n$.

Z tego powodu badanie skuteczności przedstawionych procedur sekwencyjnych przeprowadzono za pomocą eksperymentu obliczeniowego. Wyniki tego eksperymentu (dokonanego przy użyciu programów obliczeniowych opracowanych przez autora artykułu) przedstawiono w pracy magisterskiej A.Szały i G.Nowaczyk [4].

W eksperymencie realizowano po 500 zadań doboru zmiennych dla każdej rozpatrywanej wersji procedury sekwencyjnej. Lista potencjalnych zmiennych objaśniających dla każdego zadania doboru zmiennych objaśniających były 10 elementowa. Uwzględniono w nim

- trzy warianty rozmiaru RKS kombinacji startowej:
  - 1,1-2,1-3 w przypadku procedury DOL1,
  - 10,9-10,8-10 w przypadku procedury COF1,
- trzy warianty liczby LPS punktów startowych: 1,2,3.

Okazało się, że najmniej skuteczna była procedura COF1. Nieco lepsza od COF1 była procedura DOL1, a najlepsza była procedura DC1. Syntetyczne wyniki uzyskane dla procedury DC1 podano w tabeli 1. Użyto w niej trzech wskaźników

Lider - który określa częstość wskazywania przez procedurę sekwencyjną kombinacji optymalnej,

Podium - który określa częstość wskazywania przez procedurę sekwencyjną kombinacji będącej jedną z trzech najlepszych kombinacji wg przeglądu kompletnego,

N.obl. - który określa nakład obliczeniowy procedury sekwencyjnej, rozumiany jako iloraz liczby kombinacji zbadanych podczas realizacji procedury sekwencyjnej do liczby kombinacji zbadanych przy realizacji przeglądu kompletnego.

Tabela 1

Wskaźniki dla procedury DC1 (DOL1, COF1)

Rozmiar komb. start. RKS		Lider (%)			Podium (%)			N.obl. (%)		
		liczba punktów startowych LPS								
DOL1	COF1	1	2	3	1	2	3	1	2	2
1	10	85,6	90,8	95,4	96,4	98,8	100	7,3	10,2	14,2
1-2	9-10	91,8	95,0	97,2	98,2	99,6	99,8	14,2	16,4	18,7
1-3	8-10	96,8	98,4	99,6	99,4	100	100	36,3	37,7	39,3

Gdybyśmy przyjęli, że dana wersja procedury DC1 jest „dobra" jeśli Lider $\geqslant 95\%$, Podium $\geqslant 99\%$, a wskaźnik nakładu obliczeniowego N.obl. $\leqslant 20\%$, to za „dobre" mogłyby być uznane następujące warianty procedury DC1

$$\text{przy } [RKS_{DOL} = 1,\ RKS_{COF} = 10],\ LPS = 3$$

$$\text{przy } [RKS_{DOL} = \{1\text{-}2\},\ RKS_{COF} = \{9\text{-}10\}],\ LPS = 2$$

$$\text{przy } [RKS_{DOL} = \{1\text{-}3\},\ RKS_{COF} = \{9\text{-}10\}],\ LPS = 3.$$

Wydaje się, że procedurę DC1 można proponować do zastosowania praktycznego. Przy tym - z punktu widzenia nakładu obliczeniowego - korzystniej jest zwiększać liczbę punktów startowych niż rozmiar kombinacji startowej. Zwiększanie rozmiaru kombinacji startowej powoduje bowiem bardzo silny wzrost liczby kombinacji badanych w etapie wstępnym[8].

Literatura

[1] Hellwig Z.: Problem optymalnego wyboru predyktant, Przegląd Statystyczny, 3-4(1969).

[2] Nowak E.: Problemy doboru zmiennych do modelu ekonometrycznego, PWN, 1984.

[3] Rocki M.: O pewnym zadaniu liniowego programowania całkowitoliczbowego jako metodzie doboru zmiennych w modelach ekonometrycznych, Przegląd Statystyczny, 1-2(1982).

[4] Szała A., Nowaczyk G.: Eksperymentalne badanie zgodności procedur sekwencyjnych z metodą Z.Hellwiga doboru zmiennych, praca magisterska, AE Poznań, 1985.

---

[8] W etapie wstępnym stosuje się procedurę kompletnego przeglądu kombinacji o rozmiarach z ustalonego przedziału.