Podobieństwo wielkości (skali) oraz kształtu (formy) w złożonych badaniach strukturalnych

mgr Marek Walesiak

Akademia Ekonomiczna we Wrocławiu

W pracy [5] autor zasygnalizował potrzebę badania struktur złożonych (opisanych zespołem cech absolutnych i (lub) relatywnych) przedstawianych w postaci blokowej macierzy obserwacji na R strukturach o zapisie (1).

$$\begin{bmatrix} q^1_{11} & q^1_{12} & \cdots & q^1_{1n} \\ \vdots & \vdots & & \vdots \\ q^1_{r1} & q^1_{r2} & \cdots & q^1_{rn} \\ \hline q^2_{11} & q^2_{12} & \cdots & q^2_{1n} \\ \vdots & \vdots & & \vdots \\ q^2_{r1} & q^2_{r2} & \cdots & q^2_{rn} \\ \hline \vdots & \vdots & & \vdots \\ \hline q^R_{11} & q^R_{12} & \cdots & q^R_{1n} \\ \vdots & \vdots & & \vdots \\ q^R_{rl} & q^R_{r2} & \cdots & q^R_{rn} \end{bmatrix} \qquad (1)$$

gdzie:

$i, j = 1, 2, ..., R$;
R — liczba badanych struktur;
r — liczba elementów struktury;
k — numer składnika struktury ($k = 1, 2, ..., r$);
l — kolejny numer cechy ($l = 1, 2, ..., n$);
n — liczba cech;
$q^i_{k,l}$ — wartość k-tego składnika l-tej cechy pomierzonej dla i-tej struktury.

W sytuacji, gdy struktury złożone opisane są zespołem cech absolutnych i każdy element blokowej macierzy obserwacji $q^i_{k,l}$ jest $\geqslant 0$, to do badania podobieństwa wielkości (skali) oraz kształtu (formy) struktur można wykorzystać następujące przekształcenia cech:

$$q^i_{k,l} := \frac{q^i_{k,l}}{\sum_{i=1}^{R} \sum_{k=1}^{r} q^i_{k,l}} \qquad l \in S \qquad (2)$$

$$q^i_{k,l} := \frac{q^i_{k,l}}{\sum_{k=1}^{r} q^i_{k,l}} \qquad l \in S \qquad (3)$$

Zakłada się, że wszystkie cechy mają charakter stymulant [1]. Na destymulantach dokonuje się przekształceń, przykładowo poprzez obliczenie odwrotności ich realizacji lub według wzoru:

$$c - q^i_{k,l} \text{ (gdzie } c \text{ oznacza stałą} \geqslant \max_i \max_k q^i_{k,l})$$

Prowadzenie badań polegających na wydzieleniu grup struktur (obiektów badania) podobnych pod względem skali (wielkości) wymaga stosowania przekształcenia (2), natomiast wyodrębnienie podzbiorów obiektów podobnych co do kształtu (formy) — podstawienia (3).

[1]) Określenia stymulanty i destymulanty podaje Z. Hellwig w pracy [2].

Dwie struktury nazywamy podobnymi pod względem skali (wielkości), jeżeli posiadają zbliżony poziom wartości cech je opisujących.

Podobny kształt (formę) mają takie struktury, które posiadają zbliżone proporcje wartości składowych cech opisujących badane struktury.

W związku z tym, że struktury złożone mogą być opisane zespołem cech absolutnych i (lub) relatywnych [2]) powyższe przekształcenia cech nie mogą być zastosowane do badania podobieństwa skali (wielkości) oraz kształtu (formy) tychże struktur. Wynika to z faktu, że obserwacje na cechach relatywnych nie są sumowalne (nie są połączone zwykłą operacją dodawania).

Przedstawiona w formie zapisu (1) blokowa macierz obserwacji na R strukturach nie może być bezpośrednio poddana badaniom strukturalnym, lecz wymaga pewnych przekształceń, które obecnie omówimy.

Realizacje cechy l-tej pomierzone na wszystkich obiektach badania R możemy przedstawić graficznie na osi liczbowej:

$$-3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3$$

W zależności od tego czy cecha jest stymulantą (S) czy też destymulantą (D) możemy dokonać przesunięć miejsca zerowego na osi liczbowej tak, aby zastosowane formuły normalizacyjne dawały wartości cech zawarte w przedziale $\langle 0, 1 \rangle$.

W przypadku, gdy cecha jest stymulantą miejsce zerowe przesuwamy do jej wartości będącej $\min_i \min_k q^i_{k,l}$ (przy czym wartość ta musi być mniejsza od zera), czyli następuje jednostronne przesunięcie na osi liczbowej każdej wartości cechy in minus o wielkość równą $\min_i \min_k q^i_{k,l}$

Dla cechy będącej destymulantą punkt zerowy przesuwamy do jej wartości równej $\max_i \max_k q^i_{k,l}$ (przy czym wartość ta jest większa od zera). Podobnie jak wyżej następuje jednostronne przesunięcie wszystkich wartości cechy in minus o wielkość równą $\max_i \max_k q^i_{k,l}$

W sytuacji, gdy dla stymulanty wartość cechy równa $\min_i \min_k q^i_{k,l}$ jest $\geqslant 0$, a dla destymulanty $\max_i \max_k q^i_{k,l} \leqslant 0$, to realizacje cech pozostają bez zmian i w tej formie mogą być poddane normalizacji.

Nasze rozważania w tym miejscu znajdą oparcie na przykładzie. Niech wektory realizacji cechy l-tej w dwóch strukturach będą następujące (w pierwszym przypadku cecha jest stymulantą, w drugim destymulantą):

$$[-5 \quad 0 \quad -2] \text{ i } [2 \quad 7 \quad 4]$$

Dla stymulanty przekształcone wektory obserwacji cechy l-tej przyjmują postać:

$$[0 \quad 5 \quad 3] \text{ i } [7 \quad 12 \quad 9]$$

Dla destymulanty transformowane realizacje cechy l-tej określają wektory:

$$[-12 \quad -7 \quad -9] \text{ i } [-5 \quad 0 \quad -3]$$

Po dokonaniu takiego przekształcenia blokowa macierz obserwacji na R strukturach przybiera postać (4). Stanowi ona etap wyjściowy do zastosowania formuł normalizacyjnych umożliwiających badanie podobieństwa wielkości (skali) oraz kształtu (formy) struktur.

[2]) W sytuacji, gdy struktury opisane są cechami absolutnymi i relatywnymi stosujemy przekształcenia określone formułami (5) i (6) lub (7) i (8) ze względu na jednolitość prowadzonego badania.

$$
\left[\begin{array}{cccc}
p_{11}^1 & p_{12}^1 & \cdots & p_{1n}^1 \\
\vdots & \vdots & & \vdots \\
p_{r1}^1 & p_{r2}^1 & \cdots & p_{rn}^1 \\
\hline
p_{11}^2 & p_{12}^2 & \cdots & p_{1n}^2 \\
\vdots & \vdots & & \vdots \\
p_{r1}^2 & p_{r2}^2 & \cdots & p_{rn}^2 \\
\hline
\vdots & \vdots & & \vdots \\
\hline
p_{11}^R & p_{12}^R & \cdots & p_{1n}^R \\
\vdots & \vdots & & \vdots \\
p_{r1}^R & p_{r2}^R & \cdots & p_{rn}^R
\end{array}\right] \tag{4}
$$

gdzie:

$p_{k,l}^{i}$ — transformowana wartość k-tego składnika l-tej cechy pomierzonej dla i-tej struktury, pozostałe oznaczenia bez zmian.

Prowadzenie badań polegających na wydzieleniu grup struktur podobnych pod względem skali (wielkości) wymaga stosowania następujących formuł rachunku normalizacyjnego:

$$p_{k,l}^{i} := \frac{p_{k,l}^{i}}{\max\limits_{i}\max\limits_{k} p_{k,l}^{i}}, \quad l \varepsilon S \tag{5}$$

$$p_{k,l}^{i} := \frac{p_{k,l}^{i}}{\min\limits_{i}\min\limits_{k} p_{k,l}^{i}}, \quad l \varepsilon D \tag{6}$$

Badania prowadzące do wyodrębnienia grup struktur podobnych pod względem kształtu (formy) wymagają zastosowania następujących przekształceń:

$$p_{k,l}^{i} := \frac{p_{k,l}^{i}}{\max\limits_{k} p_{k,l}^{i}}, \quad l \varepsilon S \tag{7}$$

$$p_{k,l}^{i} := \frac{p_{k,l}^{i}}{\min\limits_{k} p_{k,l}^{i}}, \quad l \varepsilon D \tag{8}$$

Przekształcenia (5), (6), (7) i (8) powodują unormowanie wartości cech w przedziale $<0,1>$.

Często w blokowej macierzy obserwacji na strukturach nie występują pewne elementy oznaczane symbolem (—). W przypadku struktur gałęziowych dotyczy to gałęzi przemysłu metalurgicznego, która ze względu na swoją specyfikę nie występuje we wszystkich województwach. Konieczność uwzględnienia tej sytuacji powoduje, że po normalizacji wpisujemy w miejsce znaku (—) wartość 0, co oznacza najgorszy wariant dla danej l-tej cechy. Jest to zabieg nieco sztuczny, ale konieczny.

Kontynuując nasz przykład badania podobieństwa wielkości (skali) struktur znormalizowane wartości cechy l-tej przedstawiają wektory:

dla stymulanty

$$\left[0 \ \frac{5}{12} \ \frac{3}{12}\right] \quad \text{i} \quad \left[\frac{7}{12} \ 1 \ \frac{9}{12}\right]$$

dla destymulanty

$$\left[1 \ \frac{7}{12} \ \frac{9}{12}\right] \quad \text{i} \quad \left[\frac{5}{12} \ 0 \ \frac{3}{12}\right]$$

Z kolei badając podobieństwo kształtu (formy) struktur przekształcone wartości cechy l-tej obrazują wektory:

dla stymulanty

$$\left[0 \ 1 \ \frac{3}{5}\right] \quad \text{i} \quad \left[\frac{7}{12} \ 1 \ \frac{9}{12}\right]$$

dla destymulanty

$$\left[1 \ \frac{7}{12} \ \frac{9}{12}\right] \quad \text{i} \quad \left[1 \ 0 \ \frac{3}{5}\right]$$

Realizacja zakresów badań wymienionych w tytule pracy wymaga obliczenia $\frac{1}{2}$ z R (R — 1) „odległości" pomiędzy znormalizowanymi macierzami obserwacji na strukturach [według formuł (5), (6) lub (7), (8)] i utworzenia z nich odpowiedniej macierzy symetrycznej [P] o wymiarach R $\times$ R.

$$[P] = \begin{bmatrix} 0 \ P_{1,2} \ldots P_{1,R} \\ P_{2,1} \ 0 \ldots P_{2,R} \\ \hline P_{R,1} \ P_{R,2} \ldots 0 \end{bmatrix} \tag{9}$$

Powyższa macierz stanowi podstawę wykorzystania znanych procedur taksonomicznych dla wyodrębnienia grup struktur podobnych ze względu na rozpatrywane zjawisko ekonomiczne.

Jako wskaźnik podobieństwa pomiędzy dwiema strukturami autor proponuje miarę określoną następująco [5] [3]):

$$P_{ij}^{*} = \frac{\sum\limits_{l=1}^{n}\sum\limits_{k=1}^{r} d_{ij,k}^{l}}{r \cdot n - b} \tag{10}$$

gdzie:

$$d_{ij,k}^{l} = \begin{cases} \dfrac{p_{k,l}^{i}:}{p_{k,l}^{j}:} \iff p_{k,l}^{i}: \leqslant p_{k,l}^{j}: \\[2ex] \dfrac{p_{k,l}^{j}:}{p_{k,l}^{i}:} \iff p_{k,l}^{i}: > p_{k,l}^{j}: \end{cases}$$

b — liczba składowych macierzy obserwacji nie występujących jednocześnie w strukturze obiektu i oraz j;

pozostałe oznaczenia bez zmian.

W sytuacji, gdy $p_{k,l}^{i} := p_{k,l}^{j} := 0$ przyjmujemy, że $d_{ij,k}^{l} = 0$.

W praktycznym zastosowaniu wykorzystuje się wskaźnik braku podobieństwa o postaci:

$$P_{ij} = 1 - P_{ij}^{*} = 1 - \frac{\sum\limits_{l=1}^{n}\sum\limits_{k=1}^{r} d_{ij,k}^{l}}{r \cdot n - b} \tag{11}$$

Miara ta jest unormowana w przedziale $<0, 1>$ i spełnia inne postulaty stawiane miarom podobieństwa [4]. Interpretacja miary jest następująca: jeżeli $P_{ij} \to 0$ oznacza to nikłe zróżnicowanie badanych struktur, natomiast w przypadku, gdy $P_{ij} \to 1$, to sygnalizuje nam istotne różnice pomiędzy badanymi strukturami.

Przykładem zastosowania prezentowanej propozycji będzie próba określenia podobieństwa (skali) oraz kształtu (formy) podstawowych relacji ekonomicznych w przemyśle województw makroregionu południowo-zachodniego w ujęciu gałęziowym.

Każde województwo makroregionu południowo-zachodniego ($i = 1, 2, ..., 7$) scharakteryzowane zostało przez zbiór cech ($l = 1, 2, 3, 4$) pomierzonych dla grup gałęzi przemysłu ($k = 1, 2, ..., 9$).

[3]) Inne miary można znaleźć w tej samej pracy [5].

W skład makroregionu południowo-zachodniego wchodzą województwa: gorzowskie, jeleniogórskie, legnickie, leszczyńskie, wałbrzyskie, wrocławskie, zielonogórskie.

Uwzględnione zostały następujące cechy reprezentujące podstawowe relacje ekonomiczne w przemyśle:

- Wydajność pracy na 1 robotnika w tys. zł (mierzona produkcją globalną w cenach porównywalnych z 1 I 1979 r.),
- Techniczne uzbrojenie pracy robotników w tys. zł (mierzone wartością brutto środków trwałych produkcyjnych według stanu w dniu 31 XII przypadających na 1 robotnika grupy przemysłowej i rozwojowej),
- Energetyczne uzbrojenie pracy robotników w tys. kWh (mierzone ilością zużytej energii elektrycznej w przeliczeniu na 1 robotnika grupy przemysłowej i rozwojowej),
- Produktywność środków trwałych produkcyjnych w zł (mierzona wartością produkcji globalnej na 1000 zł wartości brutto środków trwałych według stanu w dniu 31 XII).

Cechy zostały pomierzone dla następujących grup gałęzi przemysłu:

1. przemysł paliwowo-energetyczny,
2. przemysł metalurgiczny,
3. przemysł elektromaszynowy,
4. przemysł chemiczny,
5. przemysł mineralny,
6. przemysł drzewno-papierniczy,
7. przemysł lekki,
8. przemysł spożywczy,
9. pozostałe gałęzie przemysłu.

Wymienione wyżej cechy, pomierzone za rok 1980 [3] umożliwiły konstrukcję blokowej macierzy informacji o wymiarach $7 \times 9 \times 4$ (R x r x n), która stanowiła podstawę stosowania omówionej procedury normalizacyjnej dla wyodrębnienia grup województw podobnych ze względu na wielkość (skalę) oraz kształt (formę) podstawowych relacji ekonomicznych w przemyśle w ujęciu gałęziowym.

W związku z tym, że przyjęte w badaniu cechy mają charakter stymulant, więc dla badania podobieństwa wielkości (skali) stosujemy przekształcenia określone wzorem (5), zaś badając podobieństwo kształtu (formy) — formułę (7).

Tak przedstawione znormalizowane macierze obserwacji na strukturach stanowiły podstawę wykorzystania wskaźnika braku podobieństwa struktur o postaci (11).

Stosując współczynnik braku podobieństwa do struktur, w których normalizację cech stosowano według formuły (5), czyli badając podobieństwo skali (wielkości) otrzymujemy następującą macierz [P]:

$$
[\mathbf{P}] = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
0 & 0,276 & 0,401 & 0,411 & 0,381 & 0,373 & 0,307 \\
0,276 & 0 & 0,418 & 0,400 & 0,316 & 0,402 & 0,329 \\
0,401 & 0,418 & 0 & 0,410 & 0,387 & 0,274 & 0,401 \\
0,411 & 0,400 & 0,410 & 0 & 0,458 & 0,396 & 0,281 \\
0,381 & 0,316 & 0,387 & 0,458 & 0 & 0,318 & 0,396 \\
0,373 & 0,402 & 0,274 & 0,396 & 0,318 & 0 & 0,401 \\
0,307 & 0,329 & 0,401 & 0,281 & 0,396 & 0,401 & 0
\end{array}
$$

Badając podobieństwo kształtu (formy) tychże struktur [stosując przekształcenie (7)] otrzymano następującą macierz braku podobieństwa między strukturami:

$$
[\mathbf{P}] = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
0 & 0,465 & 0,499 & 0,379 & 0,526 & 0,474 & 0,352 \\
0,465 & 0 & 0,437 & 0,490 & 0,513 & 0,477 & 0,498 \\
0,499 & 0,437 & 0 & 0,490 & 0,402 & 0,359 & 0,498 \\
0,379 & 0,490 & 0,490 & 0 & 0,604 & 0,492 & 0,380 \\
0,526 & 0,513 & 0,402 & 0,604 & 0 & 0,482 & 0,569 \\
0,474 & 0,477 & 0,359 & 0,492 & 0,482 & 0 & 0,507 \\
0,352 & 0,498 & 0,498 & 0,380 & 0,569 & 0,507 & 0
\end{array}
$$

W pracy zastosowano algorytm taksonomiczny grupowania obiektów (struktur) podobnych zaproponowany przez autorów pracy [1]. Jako krytyczny poziom braku podobieństwa przyjęto arbitralnie wartość 0,400. Wysoka wartość krytyczna poziomu braku podobieństwa spowodowana jest stosunkowo dużym zróżnicowaniem wielkości (skali) oraz kształtu (formy) relacji ekonomicznych w przemyśle województw makroregionu południowo-zachodniego w ujęciu gałęziowym.

Ostatecznie badając podobieństwo wielkości (skali) podstawowych relacji ekonomicznych w przemyśle województw makroregionu południowo-zachodniego w ujęciu gałęziowym otrzymano następujące grupy:

Grupa 4-elementowa: gorzowskie (1), jeleniogórskie (2), wałbrzyskie (5), zielonogórskie (7).

Grupa 2-elementowa: legnickie (3), wrocławskie (6).

Grupa 1-elementowa: leszczyńskie (4).

Zastosowany algorytm przy badaniu podobieństwa kształtu (formy) tychże struktur złożonych pozwolił wyodrębnić następujące grupy:

Grupa 3-elementowa: gorzowskie (1), leszczyńskie (4), zielonogórskie (7).

Grupa 2-elementowa: legnickie (3), wrocławskie (6).

Grupy 1-elementowe: 1. jeleniogórskie (2),
2. wałbrzyskie (5).

Przytoczony przykład empiryczny wskazuje na praktyczną użyteczność badania struktur złożonych w sensie podobieństwa wielkości (skali) oraz kształtu (formy). Wyciąganie głębszych wniosków praktycznych na podstawie takiego materiału empirycznego wymaga znakomitego rozeznania w przedmiocie badania.

Porównanie dwóch form badania podobieństwa struktur złożonych nasuwa kilka istotnych wniosków.

Wniosek 1

Struktury i oraz j posiadają identyczny kształt (formę) jeśli istnieją związki wprost proporcjonalne pomiędzy wektorami realizacji cech opisujących badane struktury, tzn.:

$$
\begin{aligned}
p_{k,1}^{i} &= a_1 p_{k,1}^{j} \\
p_{k,2}^{i} &= a_2 p_{k,2}^{j} \\
&\vdots \\
p_{k,n}^{i} &= a_n p_{k,n}^{j}
\end{aligned}
$$

gdzie:

$a_1, a_2, .., a_n$ — współczynniki proporcjonalności,
$(a_1, a_2, ..., a_n) \varepsilon (0 + \infty)$,
pozostałe oznaczenia bez zmian.

Wniosek 2

Dwie struktury są identyczne ze względu na kształt (formę) oraz wielkość (skalę) wtedy i tylko wtedy, gdy macierze obserwacji na nich są równe (tzn. posiadają te same wymiary oraz ich odpowiednie elementy są sobie równe).

LITERATURA

1. Chomątowski S., Sokołowski A., Taksonomia struktur. „Przegląd Statystyczny" nr 2/1978.
2. Hellwig Z., Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. „Przegląd Statystyczny" nr 4/1968.
3. Struktura branżowa i przestrzenna przemysłu 1975 i 1980. GUS. Warszawa 1982.
4. Walesiak M., Metoda oceny podobieństwa struktur (na przykładzie struktury gałęziowej zatrudnienia w przemyśle uspołecznionym województw Polski w roku 1980), „Wiadomości Statystyczne" nr 10/1982.
5. Walesiak M., Pojęcie, klasyfikacja i wskaźniki podobieństwa struktur gospodarczych. Zeszyty Naukowe AE we Wrocławiu (zgłoszone do druku).