

## Wykorzystanie gradacyjnej analizy danych do klasyfikacji podregionów pod względem struktury agrarnej

---

**Streszczenie.** *W artykule podjęto problem dotyczący pomiaru niepodobieństwa struktur. Okazało się, że zaproponowana w nim miara, zbudowana na zasadzie analogii ze współczynnikiem Giniego, pozwala wychwycić subtelności, na które nie są czule powszechnie stosowane w literaturze miary oparte na metrykach. W opracowaniu przedstawiono sposób wizualizacji struktur przy wykorzystaniu map nadreprezentacji oraz sposoby grupowania obiektów za pomocą gradacyjnej analizy danych. Otrzymane wyniki grupowania przedstawiono na tle wyników uzyskanych innymi metodami.*

**Słowa kluczowe:** podobieństwo struktur, mapa nadreprezentacji, grupowanie obiektów, gradacyjna analiza danych.

---

Znalezienie podobieństwa obiektów opisanych przez zestaw cech — w celu ich uporządkowania liniowego — jest zagadnieniem często rozważanym przez analityków. W literaturze można znaleźć wiele propozycji dotyczących budowy wskaźników podobieństwa ze względu na struktury (Walesiak, 1984; Strahl, 1985; Kukuła, 2010). Wybór określonej miary podobieństwa jest zatem sprawą dyskusyjną ze względu na rozbieżności w wynikach uzyskanych różnymi metodami. Wydaje się też niemożliwym znalezienie jednej najlepszej metody do osiągnięcia tego celu. Możliwy natomiast może być wybór metody bardziej skutecznej, gdy badacz weźmie pod uwagę narzucone przez siebie dodatkowe warunki, jakie taka miara powinna spełniać. Zazwyczaj warunki te zależą od przedmiotu badania, w sytuacji gdy badacz zechce wyodrębnić charakterystyczne dla tego badania szczegóły. Celem artykułu jest opis zaadoptowania wybranych metod pomiaru podobieństwa obiektów, ich grupowania oraz wizualizacji na potrzeby klasyfikacji podregionów pod względem struktury agrarnej.

### *PROBLEMY ZWIĄZANE Z POMIAREM PODOBIEŃSTWA STRUKTUR*

Problem pomiaru podobieństwa struktur bywa niejednokrotnie rozwiązywany z wykorzystaniem tzw. metryki Minkowskiego. Okazuje się jednak, że miara ta w niektórych przypadkach może nie brać pod uwagę dość istotnych szczegółów,

które można zauważyć w sposób intuicyjny. W celu wyjaśnienia posłużymy się tutaj teoretycznym przykładem.

Załóżmy, że chcemy porównać 3 różne podregiony o identycznym obszarze użytków rolnych (271625 ha). Niech te różnice lub zmiany dotyczą struktury ilościowej gospodarstw pod względem powierzchni użytków rolnych oraz, co za tym idzie, zmiany struktury powierzchni gospodarstw rolnych. Szczegółowe dane zawiera tabl. 1.

**TABL. 1. PRZYKŁAD TEORETYCZNY WEDŁUG PODREGIONÓW 1—3**

Grupy obszarowe	P1				P2			
	liczba gospodarstw	średni obszar	struktura		liczba gospodarstw	średni obszar	struktura	
			ilościowa	po-wierzchni			ilościowa	po-wierzchni
<b>R a z e m</b> .....	<b>50000</b>	<b>5,43</b>	<b>100,00</b>	<b>100,00</b>	<b>37051</b>	<b>7,33</b>	<b>100,00</b>	<b>100,00</b>
Do 1 ha .....	11500	0,75	23,00	3,18	3705	0,75	10,00	1,02
1— 5 .....	18500	3,00	37,00	20,43	13709	3,00	37,00	15,14
5—10 .....	12000	7,50	24,00	33,13	8892	7,50	24,00	24,55
10—15 .....	4500	12,50	9,00	20,71	3335	12,49	9,00	15,34
Powyżej 15 ha .....	3500	17,50	7,00	22,55	7410	16,11	20,00	43,95

(dok.)

Grupy obszarowe	P3			
	liczba gospodarstw	średni obszar	struktura	
			ilościowa	po-wierzchni
<b>R a z e m</b> .....	<b>49500</b>	<b>5,49</b>	<b>100,00</b>	<b>100,00</b>
Do 1 ha .....	4950	0,75	10,00	1,37
1— 5 .....	24750	2,50	50,00	22,78
5—10 .....	11880	7,50	24,00	32,80
10—15 .....	4455	12,51	9,00	20,52
Powyżej 15 ha .....	3465	17,66	7,00	22,53

Ź r ó d ł o: obliczenia własne na podstawie danych umownych.

Jeżeli przyjrzymy się zmianom w strukturze ilościowej podregionów P1 i P2 oraz podregionów P1 i P3, intuicyjnie wydaje się, że odległość między podregionami P1 i P2 powinna być większa niż podregionami P1 i P3. Ma na to wpływ kilka czynników:

- liczba gospodarstw ogółem w P2 (37051) jest znacząco mniejsza niż w P1 (50000), czego nie można zauważyć przy porównaniu P3 (49500) z P1. Ma to bezpośredni związek z bardziej znaczącymi zmianami w strukturze ilościowej dla P2 w porównaniu z P1 (większa liczba gospodarstw o powierzchni powyżej 15 ha);
- średnia powierzchnia gospodarstwa w P2 (7,33 ha) znacząco różni się od analogicznej średniej dotyczącej P1 (5,43 ha), natomiast o widocznej różnicy pomiędzy P3 a P1 można mówić tylko w przypadku średniej powierzchni gospodarstw w grupach obszarowych od 1 ha do 5 ha;

- znacząca różnica przy porównaniu  $P1$  z  $P2$  wystąpiła również w przypadku struktury ziemi, szczególnie gospodarstw o powierzchni powyżej 15 ha. Podobnych różnic nie można zauważyć porównując  $P1$  z  $P3$ .

Okazuje się, że odległość np. euklidesowa jest nieczuła na tego rodzaju subtelności. Dzieje się tak z resztą dla dowolnego  $p$  w metryce Minkowskiego (1).

$$d_{yx} = \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (1)$$

Na przykład dla struktury ilościowej (tabl. 1) przy  $p=2$ :

$$d_{P2-P1} = d_{P3-P1} = 0,18385$$

Wydaje się więc uzasadnione zastosowanie czasami takiej miary, która takie subtelności weźmie pod uwagę. Są one o tyle ważne, że wychwytyują zmianę struktury w pożądanym kierunku, np. zgodnym ze wspólną polityką rolną.

Aby przybliżyć omawiany problem, przypomnijmy niektóre postulowane przez praktyków wymagania stawiane wskaźnikom mierzącym koncentrację dobra w posiadaniu obiektów. Koncentracja jest przecież niczym innym, jak niepodobieństwem pomiędzy strukturą mającą rozkład egalitarny (czyli zakładający, że każdy obiekt posiada jednakową ilość dobra) a faktyczną strukturą dobra w posiadaniu obiektów. Podstawowe postulaty stawiane wskaźnikom koncentracji przez praktyków to:

1. Wskaźnik koncentracji  $\varphi$  przyjmuje wartość „0”, jeżeli dobro rozdzielone jest równomiernie pomiędzy wszystkie obiekty, czyli:

$$\varphi(\mathbf{e}) = 0 \quad \text{dla } \mathbf{e} = \left( \frac{1}{n}, \dots, \frac{1}{n} \right)$$

2. Wartości wskaźnika  $\varphi$  są zgodne z zasadą transferów (*principle of transfers*), która głosi, że przekazanie przez „biedniejszy” obiekt dowolnej części swoich zasobów „bogatszemu” zawsze pociąga za sobą wzrost wartości wskaźnika, czyli:

$$\varphi(\mathbf{x}_{ij}) \leq \varphi(\mathbf{x}_{ij,\varepsilon})$$

gdzie:

$$\mathbf{x}_{ij} = (x_1, \dots, x_i, \dots, x_j, \dots, x_n),$$

$$\mathbf{x}_{ij,\varepsilon} = (x_1, \dots, x_i - \varepsilon, \dots, x_j + \varepsilon, \dots, x_n).$$

3. Wartości wskaźnika  $\varphi$  są zgodne z zasadą wrażliwości transferu (*transfer sensitivity axiom*), czyli transfer zasobów z obiektu „biednego” do „bogatego” powoduje wzrost wartości wskaźnika, przy czym wzrost wartości wskaźnika przy stałej wielkości transferu jest tym większy, im „bogatszy” jest obiekt, do którego dokonano transferu:

$$\bigwedge_{k>j>i} \varphi(\mathbf{x}_{ij,\varepsilon}) \leq \varphi(\mathbf{x}_{ik,\varepsilon})$$

gdzie:

$$\mathbf{x}_{ij,\varepsilon} = (x_1, \dots, x_i - \varepsilon, \dots, x_j + \varepsilon, \dots, x_k, \dots, x_n),$$

$$\mathbf{x}_{ik,\varepsilon} = (x_1, \dots, x_i - \varepsilon, \dots, x_j, \dots, x_k + \varepsilon, \dots, x_n).$$

4. Wskaźnik przyjmuje maksymalną wartość w przypadku, gdy całość zasobów jest w posiadaniu jednego obiektu:

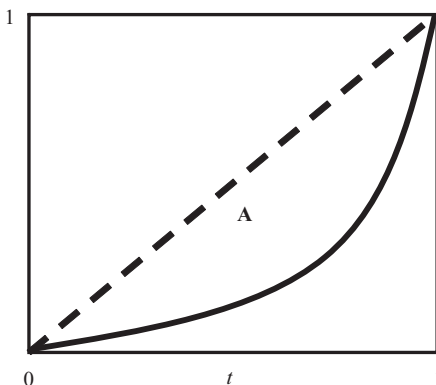
$$\varphi(\mathbf{s}) = 1 \quad \text{dla } \mathbf{s} = (0, \dots, 0, 1)$$

Jednym z najczęściej stosowanych wskaźników koncentracji jest współczynnik Giniego (Gini, 1914; Glasser, 1962):

$$G = 2 \cdot P_A = 1 - 2 \cdot \int_0^1 f(t) dt \quad (2)$$

czyli podwojone pole powierzchni pomiędzy przekątną kwadratu reprezentującą rozkład egalitarny a krzywą Lorentza (Gastwirth, 1971; Arnold, 1987) (wykr. 1).

Wykr. 1. KRZYWA LORENTZA — RYSUNEK POGLĄDOWY



Źródło: opracowanie własne.

W przypadku dyskretnym — aby spełnić postulat (4) w zakresie przyjęcia maksymalnej wartości równej 1 — używa się często unormowanej postaci współczynnika  $G$ :

$$G^* = \frac{n}{n-1} \cdot G \quad (3)$$

Ogólnie wskaźnik koncentracji można więc określić wykorzystując pojęcie odległości  $d$ :

$$\varphi(\mathbf{x}) = \frac{d(\mathbf{e}, \mathbf{x})}{d(\mathbf{e}, \mathbf{s})} \quad (4)$$

o ile odległość  $d$  spełnia warunki (2) i (3).

Wskaźnik koncentracji jest więc szczególną miarą niepodobieństwa, ponieważ mierzy niepodobieństwo pomiędzy strukturą egalitarną i faktyczną strukturą dobra w posiadaniu obiektów. Gdybyśmy założyli, że interesuje nas wskaźnik niepodobieństwa  $d$ , który ma podobne własności jak wskaźnik koncentracji, to jego pożądane własności można zapisać następująco:

$$1) \ d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}) \geq 0,$$

$$\text{gdzie: } \mathbf{y} = (y_1, \dots, y_n), \ \mathbf{x} = (x_1, \dots, x_n), \ x_i \geq 0, \ \sum_{i=1}^n x_i = 1, \ y_i \geq 0, \ \sum_{i=1}^n y_i = 1,$$

$$2) \ d(\mathbf{x}, \mathbf{x}) = 0,$$

3) miara niepodobieństwa (odległości) zmienia się zgodnie z przyjętą we wskaźnikach koncentracji „zasadą transferów” i „wrażliwością transferów”:

$$\bigwedge_{\varepsilon > 0} \bigwedge_{n \geq k > j > i \geq 1} 0 < d(\mathbf{x}, \mathbf{x}_{ij, \varepsilon}) < d(\mathbf{x}, \mathbf{x}_{ik, \varepsilon})$$

gdzie:

$$\mathbf{x} = (x_1, \dots, x_i, \dots, x_j, \dots, x_k, \dots, x_n),$$

$$\mathbf{x}_{ij, \varepsilon} = (x_1, \dots, x_i - \varepsilon, \dots, x_j + \varepsilon, \dots, x_k, \dots, x_n),$$

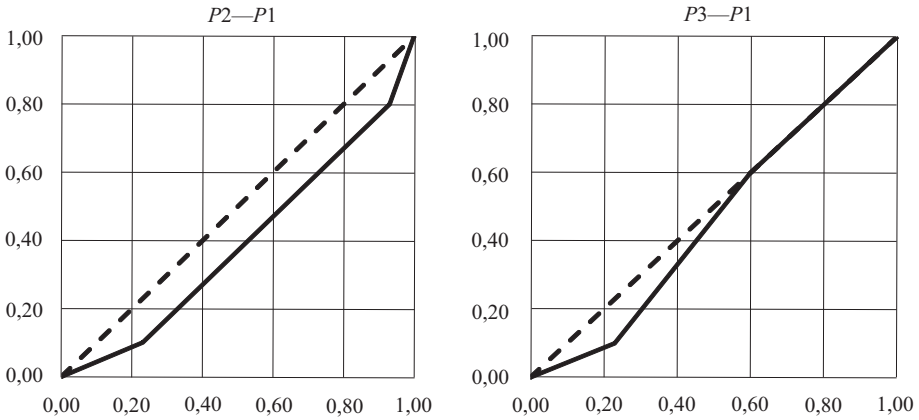
$$\mathbf{x}_{ik, \varepsilon} = (x_1, \dots, x_i - \varepsilon, \dots, x_j, \dots, x_k + \varepsilon, \dots, x_n) \text{ oznaczają struktury.}$$

Na zasadzie analogii z krzywą Lorentza niepodobieństwo struktury  $\mathbf{y}$  do struktury  $\mathbf{x}$  możemy przedstawić w postaci łamanej łączącej pewne punkty (Binderman i in., 2014). Punkty te mają współrzędne będące w tym przypadku kolejnymi skumulowanymi wartościami wskaźników struktur (np. strukturami ilościowymi gospodarstw w porównywanych podregionach):

$$(0; 0), (x_1, y_1), (x_1 + x_2; y_1 + y_2), \dots, (x_1 + \dots + x_n; y_1 + \dots + y_n) \quad (5)$$

Porównanie  $P2$  z  $P1$  i  $P3$  z  $P1$  pokazuje wykr. 2.

Wykr. 2. ŁAMANA STRUKTUR SKUMULOWANYCH



Źródło: opracowanie własne na podstawie danych z tabl. 1.

Można przyjąć, że łamana ta jest wykresem pewnej ciągłej funkcji:  $C_{[y:x]} : [0; 1] \rightarrow [0; 1]$ .

Miarą niepodobieństwa struktury  $y$  do struktury  $x$  również na zasadzie analogii (tym razem ze współczynnikiem Giniego) jest miara  $|ar|$ , gdzie:

$$ar(y : x) = ar(C_{[y:x]}) = 1 - 2 \int_0^1 C_{[y:x]}(t) dt \quad (6)$$

Tym samym mierząc odległości  $P2$  od  $P1$  i  $P3$  od  $P1$  z przykładu teoretycznego przy pomocy miary  $|ar|$  możemy wychwycić subtelnosci, na które nieczuła była metryka Minkowskiego (1):

$$|ar(P2 : P1)| = 0,22 \quad \text{podczas gdy} \quad |ar(P3 : P1)| = 0,08$$

a więc  $|ar(P2 : P1)| > |ar(P3 : P1)|$

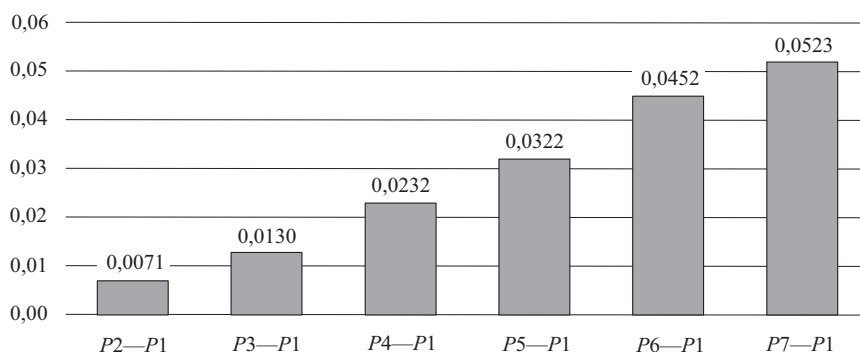
Tabl. 2 i wykr. 3 pokazują wielkości miary  $|ar|$  dotyczące przykładowych zmian np. w strukturze powierzchni (pogrubione wartości w tabl. 2), porównujących odległości innych niż poprzednio podregionów  $P2, \dots, P7$  od  $P1$  — zgodnych z „zasadą transferów” i „wrażliwością transferów”.

**TABL. 2. STRUKTURA MIARY ODLEGŁOŚCI  $|ar|$  I STOPIEŃ WRAŻLIWOŚCI TRANSFERU**

Wyszczególnienie	P1	P2	P3	P4	P5	P6	P7
<b>R a z e m</b> .....	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>	<b>100,00</b>
Do 1 ha .....	3,18	<b>0,18</b>	3,18	<b>0,18</b>	3,18	3,18	<b>0,18</b>
1— 5 .....	20,43	<b>23,43</b>	20,43	20,43	<b>17,43</b>	<b>17,43</b>	20,43
5—10 .....	33,13	33,13	33,13	<b>36,13</b>	33,13	33,13	33,13
10—15 .....	20,71	20,71	<b>17,71</b>	20,71	<b>23,71</b>	20,71	20,71
Powyżej 15 ha .....	22,55	22,55	<b>25,55</b>	22,55	22,55	<b>25,55</b>	<b>25,55</b>
$d(P1; Pi)$ .....	<b>0,0000</b>	<b>0,0424</b>	<b>0,0424</b>	<b>0,0424</b>	<b>0,0424</b>	<b>0,0424</b>	<b>0,0424</b>
$ar(Pi; P1)$ .....	<b>0,0000</b>	<b>0,0071</b>	<b>0,0130</b>	<b>0,0232</b>	<b>0,0322</b>	<b>0,0452</b>	<b>0,0523</b>

Źródło: opracowanie własne na podstawie danych teoretycznych.

**Wykr. 3. MIARY ODLEGŁOŚCI  $|ar|$  I STOPIEŃ WRAŻLIWOŚCI TRANSFERU**



Źródło: opracowanie własne na podstawie danych z tabl. 2.

### WIZUALIZACJA STRUKTUR PRZY WYKORZYSTANIU MAP NADREPREZENTACJI

Problem związany z wizualizacją danych pod względem stopnia odległości (niepodobieństwa) obiektu od ustalonego wzorca można rozwiązać wykorzystując mapy nadreprezentacji (Ciok i in., 2004; Borkowski, Szczesny, 2005; Binderman i in., 2012). Aby przybliżyć pojęcie mapy nadreprezentacji, niech jako przykład posłużą rzeczywiste dane dotyczące podregionów ostrołęckiego (26), bielskiego (44) i bytomskiego (45). Dane dotyczące liczby i struktur ilościowych gospodarstw zawierają odpowiednio tabl. 3 i 4.

**TABL. 3. LICZBA GOSPODARSTW ROLNYCH WEDŁUG GRUP  
OBSZAROWYCH POWIERZCHNI UŻYTKÓW ROLNYCH DLA TRZECH PODREGIONÓW**

Podregiony	(0, 1]	(1, 5]	(5, 10]	(10, 15]	(15, ~)	Razem	Struktura
<b>O g ó l e m</b> .....	<b>38330</b>	<b>44906</b>	<b>23771</b>	<b>12689</b>	<b>13098</b>	<b>132794</b>	<b>100,00</b>
26 .....	8548	27512	21855	12053	12205	82173	61,88
44 .....	27402	14668	1020	301	385	43776	32,97
45 .....	2380	2726	896	335	508	6845	5,15

Źródło: opracowanie własne na podstawie danych Powszechnego Spisu Rolnego 2010 (<http://stat.gov.pl/>).

**TABL. 4. STRUKTURA ILOŚCIOWA GOSPODARSTW ROLNYCH WEDŁUG GRUP OBSZAROWYCH POWIERZCHNI UŻYTKÓW ROLNYCH DLA TRZECH PODREGIONÓW**

Podregiony	(0, 1]	(1, 5]	(5, 10]	(10, 15]	(15, ~)	Razem
<b>O g ó ł e m</b> .....	<b>28,86</b>	<b>33,82</b>	<b>17,90</b>	<b>9,56</b>	<b>9,86</b>	<b>100,00</b>
26 .....	10,40	33,48	26,60	14,67	14,85	100,00
44 .....	62,60	33,51	2,33	0,69	0,88	100,00
45 .....	34,77	39,82	13,09	4,89	7,42	100,00

Ź r ó d ł o: opracowanie własne na podstawie danych z tabl. 3.

Wzorcem, z którym zostaną porównane struktury gospodarstw będzie odpowiednia wartość dla wzorca strukturalnego. Należy zatem uzmysłowić sobie, że wzorzec strukturalny w zależności od podejścia można rozumieć dwojako:

- 1) bez uwzględniania wielkości obiektów (średnia z odpowiednich wartości trzech wskaźników strukturalnych dla każdego rodzaju gospodarstwa) — wzorzec strukturalny 1;
- 2) z uwzględnieniem wielkości obiektów (stosunek liczby gospodarstw ogółem według grup obszarowych użytków rolnych do liczby gospodarstw ogółem) — wzorzec strukturalny 2.

Obie te średnie z oczywistych względów nie będą ze sobą tożsame (tabl. 5).

**TABL. 5. WZORCE STRUKTURALNE GOSPODARSTW ROLNYCH WEDŁUG POWIERZCHNI UŻYTKÓW ROLNYCH**

Wyszczególnienie	(0, 1]	(1, 5]	(5, 10]	(10, 15]	(15, ~)	Razem
Wzorzec strukturalny 1 .....	35,92	35,60	14,01	6,75	7,72	100,00
Wzorzec strukturalny 2 .....	28,86	33,82	17,90	9,56	9,86	100,00

Ź r ó d ł o: opracowanie własne na podstawie danych z tabl. 3 i 4.

Po wyznaczeniu wartości średnich można obliczyć tzw. „wskaźniki nadreprezentacji” (iloraz poszczególnych składowych struktur i odpowiednich składowych danego wzorca strukturalnego). W zależności od przyjętego sposobu wyznaczenia średniej będą one miały różne wartości:

1. Wskaźniki nadreprezentacji względem wzorca strukturalnego 1 — czyli bez uwzględnienia wielkości obiektów — zawiera tabl. 6.

**TABL. 6. WSKAŹNIKI NADREPREZENTACJI 1**

Podregiony	(0, 1]	(1, 5]	(5, 10]	(10, 15]	(15, ~)
26 .....	0,2896	0,9404	1,8990	2,1731	1,9245
44 .....	1,7425	0,9411	0,1664	0,1019	0,1140
45 .....	0,9679	1,1185	0,9346	0,7251	0,9616

Ź r ó d ł o: opracowanie własne na podstawie danych z tabl. 4 i 5.



2. Wskaźniki nadreprezentacji względem wzorca strukturalnego 2 — czyli z uwzględnieniem wielkości obiektów — zawiera tabl. 7.

TABL. 7. WSKAŹNIKI NADREPREZENTACJI 2

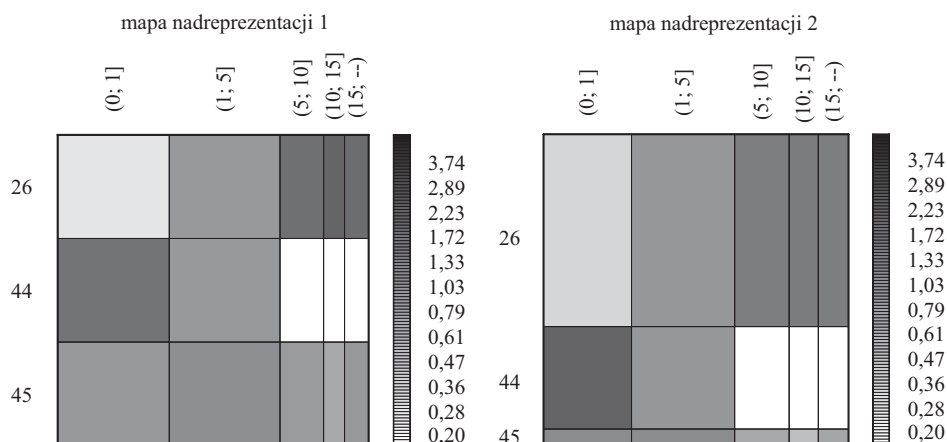
Podregiony	(0, 1]	(1, 5]	(5, 10]	(10, 15]	(15, ~)
26 .....	0,3604	0,9901	1,4858	1,5350	1,5059
44 .....	2,1686	0,9909	0,1302	0,0720	0,0892
45 .....	1,2046	1,1777	0,7312	0,5122	0,7524

Źródło: jak przy wykr. 6.

Wyznaczone współczynniki nadreprezentacji pozwalają narysować tzw. mapę nadreprezentacji (wykr. 4). Mapa nadreprezentacji jest kwadratem o bokach równych 1, gdzie w wierszach znajdują się (w tym przypadku) rozpatrywane trzy podregiony, a w kolumnach — grupy obszarowe gospodarstw. Kolory pokazują nadreprezentację (skrajnie czarny) lub niedoreprezentację (skrajnie biały). W zależności od sposobów wyznaczania średniej mapa nadreprezentacji będzie wyglądać inaczej:

- mapa nadreprezentacji 1 (bez uwzględniania wielkości obiektów) ma wiersze równej wysokości, a szerokością kolumn są tu wzorcowe wartości strukturalne poszczególnych rodzajów gospodarstw, pochodzące z wzorca strukturalnego 1;
- mapa nadreprezentacji 2 (z uwzględnieniem wielkości obiektów) ma wiersze różnej wysokości. Wysokość ta określona jest przez procentowy udział liczby gospodarstw dla każdego z podregionów w liczbie gospodarstw ogółem. Szerokością kolumn są tu wzorcowe wskaźniki strukturalne poszczególnych rodzajów gospodarstw, pochodzące z wzorca strukturalnego 2.

Wykr. 4. MAPY NADREPREZENTACJI PRZYKŁAD



Źródło: opracowanie własne na podstawie programu GRADESTAT.

Zarówno omawiana krzywa koncentracji, jak i miara niepodobieństwa  $|ar|$  oraz wizualizacja danych przy pomocy map nadreprezentacji to podstawowe pojęcia związane z tzw. gradacyjną analizą danych (GAD), o której można szerzej przeczytać w pracach Szczesnego (2002) i Kowalczyk i in. (2004).

W bardziej formalny sposób mapę nadreprezentacji dla macierzy danych  $\mathbf{P}$  o nieujemnych elementach można przedstawić jako gęstość na kwadracie jednostkowym  $([0; 1] \cdot [0; 1])$ . Bez straty ogólności można założyć, że:

$$\mathbf{P} = \{p_{ij} : i = 1, \dots, m, j = 1, \dots, n\}, \quad \text{gdzie } \sum p_{ij} = 1, p_{ij} \geq 0$$

Macierzy  $\mathbf{P}$  można przypisać pewien rozkład ciągły  $\mathbf{P}^*$ , wyznaczony przez gęstość opisaną w kwadracie jednostkowym według wzoru:

$$h^*(u, v) = \frac{p_{ij}}{p_{i+} \cdot p_{+j}}, \quad (u, v) \in \mathbf{R}_{ij}, \quad \mathbf{R}_{ij} = [(S_{i-1}, S_i) \cdot (T_{j-1}, T_j)] \quad (7)$$

gdzie:  $i = 1, \dots, m; \quad j = 1, \dots, n; \quad p_{i+} = \sum_{j=1}^n p_{ij}; \quad p_{+j} = \sum_{i=1}^m p_{ij}$

$$S_i = p_{1+} + \dots + p_{i+}; \quad T_j = p_{+1} + \dots + p_{+j}; \quad S_0 = T_0 = 0$$

Ilustracja graficzna tej gęstości nazywa się mapą nadreprezentacji macierzy danych  $\mathbf{P}$ .

### GRUPOWANIE OBIEKTÓW NARZĘDZIAMI

Podstawowym zagadnieniem w ramach GAD jest badanie zróżnicowania wierszy i kolumn oraz dążenie do takiego ich uporządkowania w macierzy danych, aby osiągnąć (widoczny na mapie nadreprezentacji) maksymalny kontrast pomiędzy skrajnymi wierszami lub skrajnymi kolumnami.

Cel ten realizuje algorytm GCA (*Grade Correspondence Analysis*). Porządkuje on wiersze lub kolumny macierzy  $\mathbf{P}$  w taki sposób, aby zmaksymalizować pewien wskaźnik zależności. Wskaźnik ten obliczany jest dla  $\mathbf{P}^*$ , który przyporządkowany jest macierzy danych  $\mathbf{P}$ . Jako wskaźnik zależności wykorzystuje się w tym przypadku współczynnik korelacji *tau-Kendalla* lub *rho-Spearmana*

(Ząbkowski, Szczesny, 2012). Wskaźniki te można potraktować jako miary zróżnicowania wierszy (analogicznie kolumn) macierzy danych  $\mathbf{P}$ :

$$\tau(\mathbf{P}) = 2 \cdot \sum_{t=2}^m \sum_{s=1}^{t-1} [(p_{t+} \cdot p_{s+}) \cdot ar(t:s)] \quad (8)$$

$$\rho^*(\mathbf{P}) = 6 \cdot \sum_{t=2}^m \sum_{s=1}^{t-1} [(S_t^* - S_s^*) \cdot (p_{t+} \cdot p_{s+}) \cdot ar(t:s)] \quad (9)$$

gdzie  $ar(t:s)$  oznacza wartość wskaźnika zróżnicowania wektorów:

$$\left( \frac{p_{t1}}{p_{t+}}, \frac{p_{t2}}{p_{t+}}, \dots, \frac{p_{tm}}{p_{t+}} \right) \text{ oraz } \left( \frac{p_{s1}}{p_{s+}}, \frac{p_{s2}}{p_{s+}}, \dots, \frac{p_{sn}}{p_{s+}} \right)$$

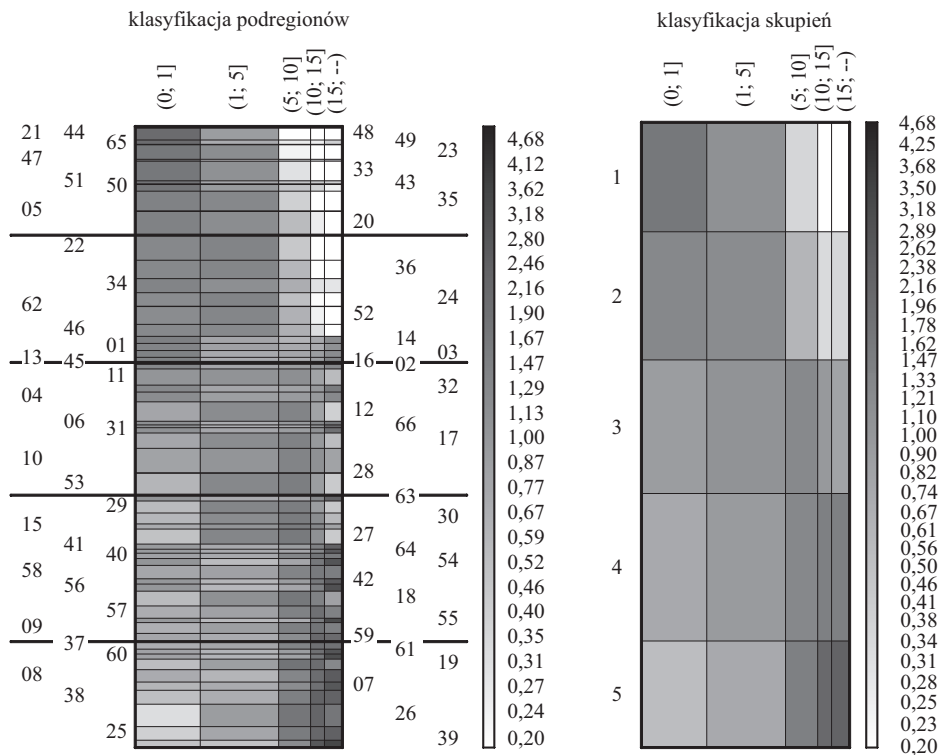
oraz  $S_i^* = \frac{S_i + S_{i-1}}{2}$  dla  $S_i = p_{1+} + \dots + p_{i+}$ ,  $i = 1, \dots, m$ .

Na przedstawianych wskaźnikach budowane są skupienia w taki sposób, aby zmaksymalizować zróżnicowanie, przy czym to występujące pomiędzy dwoma skupieniami rozumie się jako zróżnicowanie pomiędzy dwoma obiektami utworzonymi z tych skupień jako sumy obiektów wchodzących w ich skład.

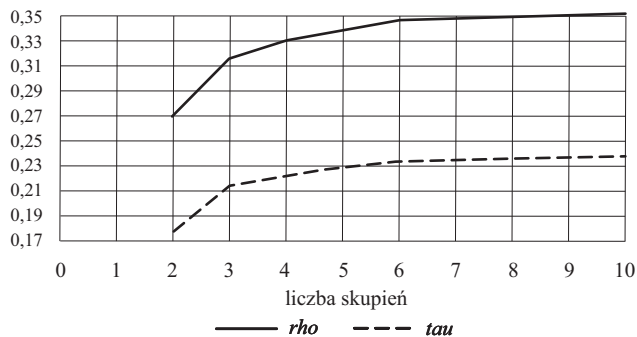
Na wyk. 5 pokazano strukturę liczby gospodarstw rolnych według grup obszarowych powierzchni użytków rolnych dla 66 podregionów w Polsce. Dane pochodzą z Powszechnego Spisu Rolnego 2010. Rysunek po lewej zawiera wizualizację podregionów z jednoczesnym podziałem na 5 skupień. Kolory przedstawiają nadreprezentację i niedoreprezentację względem wzorca strukturalnego wyznaczonego sposobem 2. Wizualizacja uwzględnia wielkość obiektów, a więc poszczególne wiersze nie mają równej wysokości. Po prawej stronie wizualizacji podregionów przedstawiono uporządkowanie wyodrębnionych 5 klas. Podziału dokonano przy pomocy programu GRADESTAT, w którym jest zaimplementowany algorytm GCCA (*Grade Correspondence Cluster Analysis*) (Kowalczyk i in., 2004; Szczesny, 2002). Oba rysunki z formalnego punktu widzenia są kwadratami, ale dla potrzeb lepszego odbioru zostały odpowiednio zeskalowane.

Decyzję o podziale podregionów na 5 skupień podjęto na podstawie wyznaczonych miar zróżnicowania pomiędzy skupieniami  $\rho$  i  $\tau$  dla różnej liczby wyznaczanych klas. Wydaje się, że znaczący wzrost wskaźników zróżnicowania między skupieniami nastąpił w przypadku podziału obiektów na 5 skupień. Dalsze zwiększanie liczby skupień nie powodowało znaczącego wzrostu wielkości obydwu wskaźników, choć przedmiotem dyskusji może być odpowiedź na pytanie, czy podział 6-klasowy nie byłby lepszy (wykr. 6).

**Wykr. 5. KLASYFIKACJA I PODZIAŁ PODREGIONÓW ZE WZGLĘDU NA STRUKTURĘ ILOŚCIOWĄ GOSPODARSTW UŻYTKÓW ROLNYCH (z uwzględnieniem wielkości obiektów)**



**Wykr. 6. ZMIANY WIELKOŚCI WSKAŹNIKÓW ZRÓŻNICOWANIA MIĘDZY SKUPIENIAMI W ZALEŻNOŚCI OD LICZBY UTWORZONYCH SKUPIEŃ**

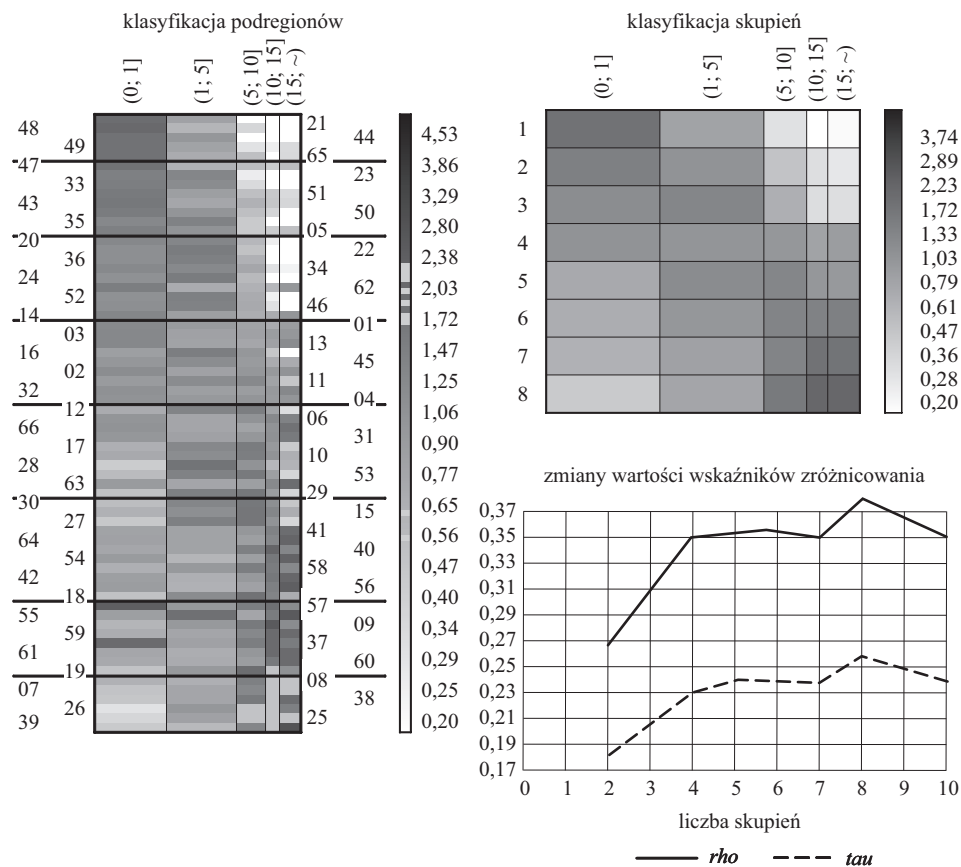


Ź r ó d ł o: opracowanie własne.

W uzupełnieniu przeprowadzonych rozważań należy podkreślić, że w przypadku gdy nie uwzględnimy wielkości obiektów, ze zrozumiących względów wizuali-

zacja podregionów, ich podział na skupienia, a także wizualizacja samych skupień będzie przedstawiała się inaczej. Podregiony zagregowano tu w skupienia w postaci środka ciężkości (jako średnia w skupieniu), co jest nietypowym podejściem jeżeli chodzi o zwyczajową procedurę w GAD. Podział obiektów na 8 skupień wyjaśnia na wyk. 7 przebieg zmian wartości wskaźników zróżnicowania pomiędzy skupieniami ( $\rho$  i  $\tau$ ). Wykres zmian wielkości wskaźników nie jest w tym przypadku standardowy ze względu na sposób agregacji obiektów w skupienia.

**Wykr. 7. KLASYFIKACJA I PODZIAŁ PODREGIONÓW ZE WZGLĘDU NA STRUKTURĘ ILOŚCIOWĄ GOSPODARSTW ROLNYCH WEDŁUG GRUP OBSZAROWYCH UŻYTKÓW ROLNYCH (bez uwzględnienia wielkości obiektów)**



Źródło: jak przy wyk. 4.

### WYNIKI BADAŃ — PUNKTY ODNIESIENIA

W celu sprawdzenia stabilności otrzymanych metodami GAD podziałów na skupienia w opracowaniu (ze względu na jednolitość danych) zdecydowano się na porównanie ich z wynikami uzyskanymi w artykule pt. *Klasyfikacja podre-*

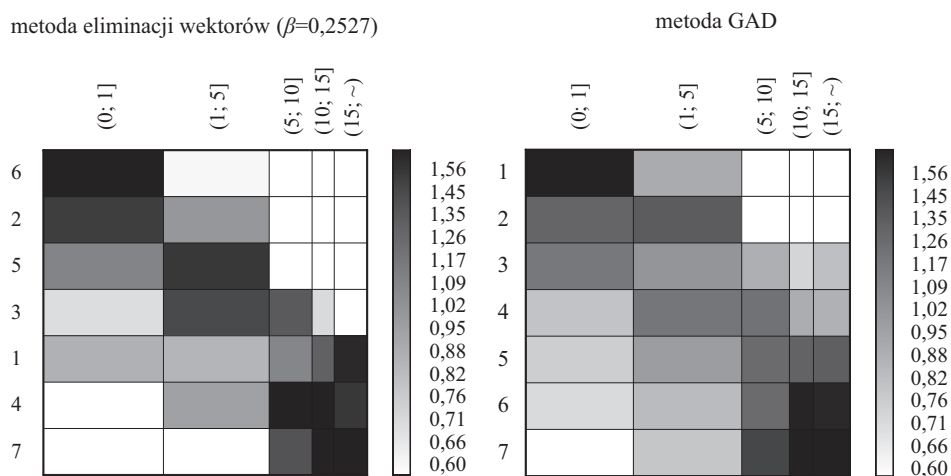
gionów pod względem podobieństwa struktury agrarnej (Bożek, 2013). W pracy tej do podziału na skupienia zastosowano 2 metody:

- eliminacji wektorów według różnych wartości progowych zróżnicowania  $\beta$ ,
- zbiorów rozmytych.

Chcąc zapewnić jak najwyższy stopień porównywalności wyników otrzymanych metodami GAD z wynikami uzyskanymi metodami opisanymi wcześniej, zdecydowano się na identyczne podziały podregionów w sensie liczby skupień. Porównanie wyników należy tu rozumieć jako porównanie wielkości wskaźników zróżnicowania  $\rho$  i  $\tau$  pomiędzy uzyskanymi skupieniami. Okazuje się, że w zależności od podejścia do problemu (bez uwzględnienia wielkości obiektów i z uwzględnieniem wielkości obiektów) można zauważyć pewne prawidłowości.

Na wykr. 8 i 9 pokazano przykładowe porównanie podziału uzyskanego metodą GAD z podziałem uzyskanym metodą eliminacji wektorów według wartości progowej zróżnicowania  $\beta=0,2527$ . Podział obiektów metodą eliminacji wektorów doprowadził tutaj do uzyskania siedmiu skupień (dwa z nich były jednoelementowe) i taką samą liczbę skupień utworzono metodą GAD. Na wykr. 8, ze względu na podejście do problemu, które nie uwzględniało wielkości obiektów, dokonano agregacji podregionów w utworzonych skupieniach w postaci jego środka ciężkości. Uporządkowanie skupień (które potraktowano jako osobne obiekty) przy pomocy algorytmu GCA pozwoliło na zmaksymalizowanie stopnia zróżnicowania pomiędzy nimi. Stało się tak zarówno w przypadku skupień uzyskanych metodą eliminacji wektorów ( $\beta=0,2527$ ), jak i skupień uzyskanych metodami GAD. Przy takim podejściu podział uzyskany metodą eliminacji wektorów pozwolił uzyskać wyższy stopień zróżnicowania między utworzonymi skupieniami ( $\rho=0,4513$  i  $\tau=0,3127$ ) niż w przypadku 7 skupień utworzonych metodami GAD ( $\rho=0,3504$  i  $\tau=0,2371$ ).

**Wykr. 8. KLASYFIKACJA SKUPIEŃ UTWORZONYCH METODĄ ELIMINACJI WEKTORÓW ORAZ METODĄ GAD (bez uwzględnienia wielkości obiektów)**

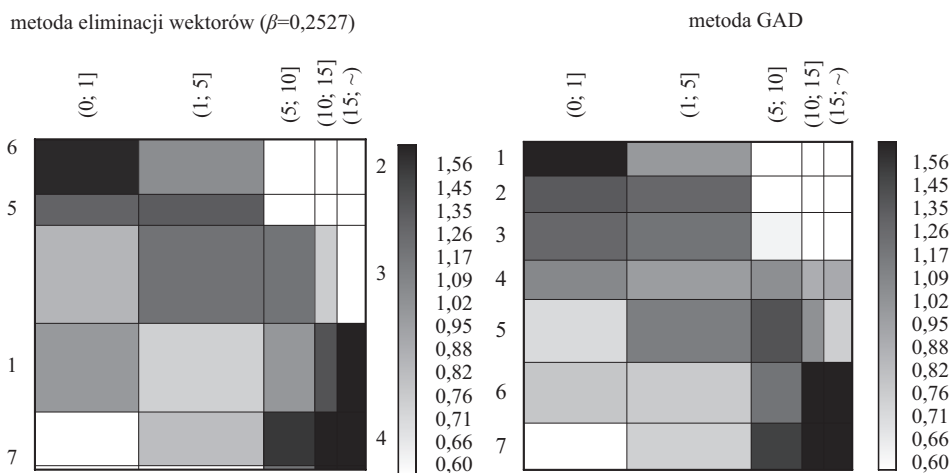


Źródło: jak przy wykr. 4.

O lepszym zgrupowaniu obiektów (na korzyść metody eliminacji wektorów) zapewniającym wyższą wartość wskaźnika zróżnicowania pomiędzy skupieniami świadczy również widoczny na mapie (wykr. 7) większy kontrast pomiędzy skrajnymi wierszami i kolumnami.

Na wykr. 9 przedstawiono skupienia z uwzględnieniem wielkości zgrupowanych w nich obiektów. Z 7 skupień utworzonych metodą eliminacji wektorów ( $\beta=0,2527$ ) utworzono również oddzielne obiekty, ale tym razem należy rozumieć je jako sumy obiektów wchodzących w ich skład. W tym układzie stopień zróżnicowania między utworzonymi metodą eliminacji wektorów skupieniami był niższy ( $\rho=0,3038$  i  $\tau=0,2053$ ) niż w przypadku 7 skupień utworzonych metodą GAD ( $\rho=0,3474$  i  $\tau=0,2353$ ).

**Wykr. 9. KLASYFIKACJA SKUPIEŃ UTWORZONYCH METODĄ ELIMINACJI WEKTORÓW ORAZ METODĄ GAD (z uwzględnieniem wielkości obiektów)**

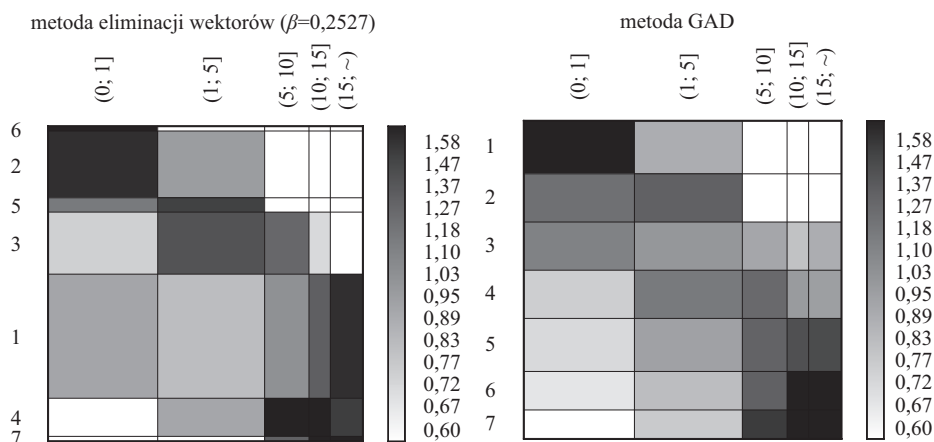


Źródło: jak przy wykr. 4.

O lepszym zgrupowaniu obiektów (na korzyść metody GAD) zapewniającym wyższą wartość wskaźnika zróżnicowania pomiędzy skupieniami (dla GAD) świadczy również widoczny na mapie (wykr. 9) większy kontrast pomiędzy skrajnymi wierszami i kolumnami (choć nie tak wyrazisty w tym przypadku, jak poprzednio). Zwraca też uwagę fakt, że grupowanie obiektów metodą GAD (z uwzględnieniem wielkości tych obiektów) zapewnia większą regularność, jeżeli chodzi o wielkość samych skupień. Nie jest to jednak zaskoczeniem, gdyż algorytm GCA porządkujący obiekty w skupienia bierze pod uwagę także wielkość samych skupień. W przypadku metody eliminacji wektorów skupienia 6 i 7 były bardzo małe (zawierały tylko po 1 obiekcie) i dlatego ich wielkość (szerokość wiersza) jest na mapie praktycznie niewidoczna.

Można jeszcze sklasyfikować utworzone skupienia nie uwzględniając wielkości obiektów, ale uwzględniając liczbę zgrupowanych w nich obiektów. Tym razem również wskaźniki różnicowania pomiędzy utworzonymi 7 skupieniami były większe w przypadku skupień uzyskanych metodą GAD ( $\rho=0,3421$  i  $\tau=0,2317$ ) niż w przypadku metody eliminacji wektorów dla  $\beta=0,2527$  ( $\rho=0,2918$  i  $\tau=0,1980$ ). Wynik ten również jest czymś naturalnym ze względu na to, że algorytm GCA tworząc skupienia optymalizuje wskaźniki różnicowania również pod względem liczby zgrupowanych w nich obiektów.

**Wykr. 10. KLASYFIKACJA SKUPIEŃ UTWORZONYCH METODĄ ELIMINACJI WEKTORÓW ORAZ METODĄ GAD (bez uwzględnienia wielkości obiektów, ale z uwzględnieniem liczebności w skupieniach)**

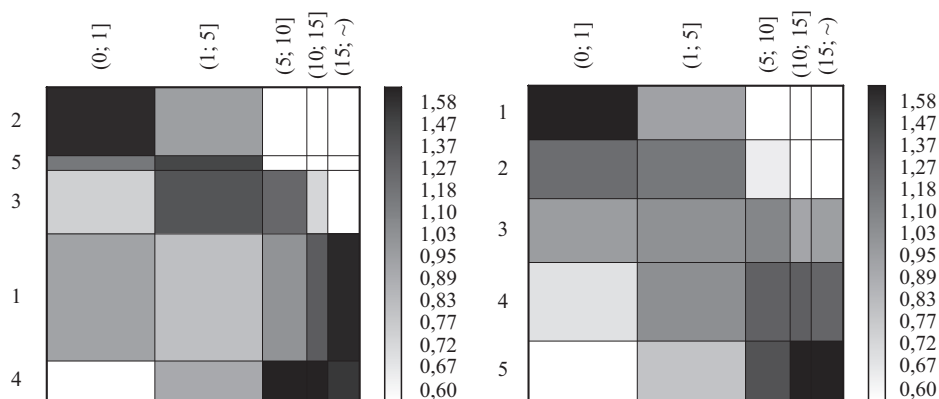


Źródło: jak przy wykr. 4.

Szerokość wierszy na mapie nadreprezentacji dotyczącej skupień uzyskanych metodą GAD (wykr. 10) wskazuje również na większą regularność biorąc pod uwagę liczebność obiektów zgrupowanych w poszczególnych skupieniach. Jeżeli nawet widoczne na mapie nadreprezentacji jednoelementowe skupienia 6 i 7 (uzyskane metodą eliminacji wektorów) uznamy za elementy odstające i z analizy usuniemy zawarte w nich podregiony (48 — katowicki i 49 — rybnicki), wielkości wskaźników  $\rho$  i  $\tau$  dla pozostałych w tym wypadku 5 skupień (wykr. 11) również będą przemawiały na korzyść metody GAD. Z pięciu nowo utworzonych metodą GAD skupień z uwzględnieniem liczebności w skupieniu, ale bez podregionów 48 i 49, otrzymano wskaźnik różnicowania między skupieniami  $\rho=0,3209$  oraz  $\tau=0,2168$ , podczas gdy wskaźnik różnicowania dla 5 skupień uzyskanych metodą eliminacji wektorów (po eliminacji skupienia 6 i 7) wyniósł —  $\rho=0,2734$ , a —  $\tau=0,1851$ .



**Wykr. 11. KLASYFIKACJA SKUPIEŃ UTWORZONYCH METODĄ ELIMINACJI WEKTORÓW Z POMIĘCIEM SKUPIENIA 6 I 7 ORAZ METODĄ GAD Z POMIĘCIEM OBIEKTU 48 I 49 (bez uwzględnienia wielkości obiektów, ale z uwzględnieniem liczebności w skupieniach)**



Źródło: jak przy wykr. 4.

Jeżeli nie uwzględnimy wielkości obiektów i ich liczebności w skupieniach, ale odrzucimy — tak jak poprzednio — jednoelementowe skupienia 6 i 7 uzyskane metodą eliminacji wektorów oraz utworzymy 5 skupień metodą GAD (bez podregionów 48 i 49), wielkości wskaźników  $\rho$  i  $\tau$  w przypadku metody GAD dadzą również korzystniejszy wskaźnik zróżnicowania między skupieniami (GAD —  $\rho=0,3436$ ;  $\tau=0,2324$  oraz metoda eliminacji wektorów —  $\rho=0,3323$ ;  $\tau=0,2260$ ).

W celu porównia skupień utworzonych metodą GAD ze skupieniami uzyskanymi innymi metodami zaprezentowanymi w pracy Bożka utworzono odpowiednią liczbę skupień metodą GAD oraz wyznaczono dla porównywanych metod wskaźniki zróżnicowania między skupieniami  $\rho$  i  $\tau$ . Wyniki porównań skupień utworzonych metodą eliminacji wektorów według wartości progowej zróżnicowania  $\beta$  z identyczną liczbą skupień utworzoną metodą GAD zawiera tabl. 8.

**TABL. 8. PORÓWNIANIE SKUPIEŃ UTWORZONYCH METODĄ ELIMINACJI WEKTORÓW WEDŁUG WARTOŚCI PROGOWEJ ZRÓŻNICOWANIA I METODĄ GAD**

Wyszczególnienie	Metoda eliminacji wektorów według wartości progowej zróżnicowania $\beta$ (z klasami jednoelementowymi)			Metoda GAD		Liczba skupień
	$\beta$	$\rho$	$\tau$	$\rho$	$\tau$	
NIE .....	0,2527	<b>0,4513</b>	<b>0,3127</b>	0,3504	0,2371	7
TAK .....	0,2527	0,3038	0,2053	<b>0,3474</b>	<b>0,2353</b>	7
NIE .....	0,2400	<b>0,3546</b>	<b>0,2399</b>	0,3535	0,2396	6
TAK .....	0,2400	0,3228	0,2190	<b>0,3456</b>	<b>0,2337</b>	6

**TABL. 8. PORÓWNANIE SKUPIEŃ UTWORZONYCH METODĄ ELIMINACJI WEKTORÓW WEDŁUG WARTOŚCI PROGOWEJ ZRÓŻNICOWANIA I METODĄ GAD (dok.)**

Wyszczególnienie	Metoda eliminacji wektorów według wartości progowej zróżnicowania $\beta$ (z klasami jednoelementowymi)			Metoda GAD		Liczba skupień
	$\beta$	$\rho$	$\tau$	$\rho$	$\tau$	
NIE .....	0,2200	<b>0,3796</b>	<b>0,2586</b>	0,3676	0,2498	9
TAK .....	0,2200	0,3274	0,2223	<b>0,3508</b>	<b>0,2374</b>	9
NIE .....	0,2000	<b>0,3756</b>	<b>0,2552</b>	0,3514	0,2383	10
TAK .....	0,2000	0,3309	0,2237	<b>0,3512</b>	<b>0,2377</b>	10

Źródło: obliczenia własne.

Analizując wartości wyznaczonych wskaźników zróżnicowania między skupieniami ( $\rho$  i  $\tau$ ) można zauważyć, że przemawiają one na korzyść metody eliminacji wektorów w przypadku, gdy skupienia nie uwzględniały wielkości obiektów wchodzących w ich skład. W przypadku uwzględniania w utworzonych skupieniach wielkości obiektów wskaźniki zróżnicowania osiągały wyższą wartość w obliczeniach metodą GAD.

Wyniki porównań skupień utworzonych metodą zbiorów rozmytych (Bożek, 2013) z identyczną liczbą skupień utworzoną metodą GAD zawierają tabl. 9 i 10. Dodatkowo w tablicach tych zestawienia wielkości wskaźników zróżnicowania dotyczą liczby skupień uwzględniających skupienia jednoelementowe, jak i bez skupień jednoelementowych. W tym drugim przypadku, aby zachować pewien stopień porównywalności przy tworzeniu skupień metodą GAD, wyeliminowano ze zbioru danych te podregiony, które w metodzie zbiorów rozmytych tworzyły skupienia jednoelementowe.

**TABL. 9. PORÓWNANIE SKUPIEŃ UTWORZONYCH METODĄ ZBIORÓW ROZMYTYCH (5 KLAS) I METODĄ GAD**

Wyszczególnienie	Metoda zbiorów rozmytych — 5-klasowa			Metoda GAD		Liczba skupień
	Klasy jednoelementowe	$\rho$	$\tau$	$\rho$	$\tau$	
NIE .....	TAK	0,3261	0,2197	<b>0,3514</b>	<b>0,2383</b>	10
TAK .....	TAK	0,3339	0,2256	<b>0,3512</b>	<b>0,2377</b>	10
NIE .....	NIE	0,3247	0,2210	<b>0,3437</b>	<b>0,2331</b>	5
TAK .....	NIE	0,3257	0,2211	<b>0,3354</b>	<b>0,2276</b>	5

Źródło: obliczenia własne.

**TABL. 10. PORÓWNANIE SKUPIEŃ UTWORZONYCH METODĄ ZBIORÓW ROZMYTYCH (6 KLAS) I METODĄ GAD**

Wyszczególnienie	Metoda zbiorów rozmytych — 6-klasowa			Metoda GAD		Liczba skupień
	Klasy jednoelementowe	$\rho$	$\tau$	$\rho$	$\tau$	
NIE .....	TAK	0,2977	0,2022	<b>0,3676</b>	<b>0,2498</b>	9
TAK .....	TAK	0,3346	0,2258	<b>0,3508</b>	<b>0,2374</b>	9
NIE .....	NIE	0,3332	0,2255	<b>0,3526</b>	<b>0,2387</b>	6
TAK .....	NIE	0,3376	0,2278	<b>0,3495</b>	<b>0,2364</b>	6

Źródło: obliczenia własne.

W tym przypadku wskaźniki różnicowania między skupieniami były za każdym razem wyższe dla skupień utworzonych metodą GAD (pogrubione wartości w tabl. 9 i 10).

## Wnioski

Przedstawione w opracowaniu wyniki są tylko swego rodzaju koncepcją podejścia do problemu podziału obiektów na skupienia, choć zastosowanie metod GAD (ich adaptacja) w grupowaniu obiektów może w niektórych przypadkach mieć uzasadnienie praktyczne i jest wygodne z uwagi na łatwą i przejrzystą wizualizację otrzymanych wyników. Koncepcja mierzenia odległości między dwoma obiektami przy pomocy wartości bezwzględnej funkcji miary  $ar$  w sposób wartościujący zasadę dotyczącą wrażliwości transferów także wydaje się być czasami uzasadniona, choć może być to w niektórych wypadkach przedmiotem dyskusji.

Porównanie wyników uzyskanych przez Bożka z wynikami uzyskanymi metodą GAD miało w tym przypadku charakter poznawczy, a nie wartościujący. Były tu zastosowane dwie metody oceny różnicowania obiektów i utworzonych skupień, a próba porównania wyników (w formie liczbowej) ograniczała się tylko do jednej z nich (GAD). Porównanie różnicowania między skupieniami polegało na wyznaczaniu miar  $\rho$  i  $\tau$ , co jest typowym podejściem dla metody GAD. Gdy jednak różnicowanie między skupieniami zmierzymy, np. przy pomocy metryki Minkowskiego, to niezależnie od tego, czy uwzględnimy wielkość obiektów w skupieniach czy też nie, różnicowanie pomiędzy skupieniami będzie niższe dla metody GAD.

Wartości funkcji  $ar$ , na których opiera się proponowana miara odległości mogą być także dobrym punktem wyjścia do utworzenia rankingu podregionów pod względem struktury agrarnej. Traktując wartości funkcji  $ar$  jako miarę niepodobieństwa danego podregionu do pewnego wzorca (którym może być np. wzorzec strukturalny utworzony ze średnich wartości wskaźników strukturalnych) uzyskamy dodatkowo informację o „kierunku” tego niepodobieństwa. Porządkowanie podregionów według tak wyznaczonych miar niepodobieństwa od wartości dodatnich do wartości ujemnych może mieć w tym podejściu logiczne uzasadnienie — podregion jest tym wyżej w rankingu, im większa jest przewaga strukturalna gospodarstw o większym areale w stosunku do średniej struktury. W dalszym etapie, po uszeregowaniu podregionów, można wyznaczyć odpowiadające im wartości  $S_i^*$  (wzór 9), które mogą być odpowiednikiem zmiennej syntetycznej służącej jako podstawa do podziału na skupienia według zaproponowanej przez Kukułę (2014) metody podziału obiektów na klasy na podstawie wskaźnika syntetycznego. W tym przypadku dobierając odpowiednio zaproponowany parametr podziału  $k$  można wartościować grupy w zależności od większego „nasilenia” gospodarstw z większą powierzchnią agrarną.

---

dr inż. Grzegorz Koszela — SGGW

## LITERATURA

- Arnold B. C. (1987), *Majorization and the Lorenz Order: A Brief Introduction*, Lecture Notes in Statistics 43, Springer-Verlag, Berlin.
- Binderman Z., Borkowski B., Szczesny W., Shachmurove Y. (2012), *Zmiany struktury eksportu produktów rolnych w wybranych krajach UE w okresie 1980—2010*, „Metody Ilościowe w Badaniach Ekonomicznych”, Tom 13, nr 1, Wydawnictwo SGGW, Warszawa.
- Binderman Z., Koszela G., Szczesny W. (2014), *Zmiany w strukturze gospodarstw rolnych w krajach Unii Europejskiej w latach 2003—2010 (aspekty metodyczne)*, „Problemy rolnictwa światowego”, Tom 14, nr 3, Wydawnictwo SGGW, Warszawa.
- Borkowski B., Szczesny W. (2005), *Metody wizualizacji danych wielowymiarowych jako narzędzie syntezy informacji*, Roczniki Naukowe SERIA, Tom VII, z. 5.
- Bożek J. (2013), *Klasyfikacja podregionów pod względem podobieństwa struktury agrarnej*, „Wiadomości Statystyczne”, nr 9, GUS.
- Ciok A., Kowalczyk T., Pleszczyńska E., Szczesny W. (1995), *Algorithms of grade correspondence-cluster analysis*, The Coll. Papers on Theoretical and Applied Computer Science.
- Gastwirth J. L. (1971), *A general definition of the Lorenz curve*, „Econometrica”, Vol. 39, No. 6: s. 1037—1039.
- Gini C. (1914), *Sulla misura della concentrazione e della variabilità dei caratteri*, Atti del R. Istituto Veneto di Scienze Lettere ed Arti.
- Glasser G. J. (1962), *Variance formulas for the mean difference and coefficient of concentration*, „Journal of the American Statistical Association”, No. 57.
- Kowalczyk T., Pleszczyńska E., Ruland F. eds. (2004), *Grade Models and Methods of Data Analysis. With applications for the Analysis of Data Population*, Studies in Fuzziness and Soft Computing, Vol. 151, Springer Verlag, Berlin-Heidelberg-New York.
- Kukuła K. (red.) (2010), *Statystyczne studium struktury agrarnej w Polsce*, PWN, Warszawa.
- Kukuła K. (2014), *Budowa rankingu województw ze względu na wyposażenie techniczne rolnictwa w Polsce*, „Wiadomości Statystyczne”, nr 7, GUS.
- Strahl D. (1985), *Podobieństwo struktur ekonomicznych*, PN AE, nr 281, Wrocław.
- Szczesny W. (2002), *Grade correspondence analysis applied to contingency tables and questionnaire data*, Intelligent Data Analysis, Vol. 6 (1), IOS Press, Amsterdam.
- Walesiak M. (1984), *Pojęcie, klasyfikacja i wskaźniki podobieństwa struktur gospodarczych*, „Prace Naukowe AE we Wrocławiu”, nr 285, Wrocław.
- Ząbkowski T., Szczesny W. (2012), *Badanie atrakcyjności oferty dostępu do Internetu za pomocą analizy gradacyjnej*, „Metody Ilościowe w Badaniach Ekonomicznych”, Tom XIII/3, Wydawnictwo SGGW, Warszawa.

**Summary.** *The article deals with the problem of measuring the dissimilarity structures. It turned out that the proposed measure, built by analogy with the Gini coefficient, can capture the subtleties, to which are not sensitive measure based on the metrics, commonly used in the literature. This paper presents a way to visualize structures using maps over-representation and ways of grouping objects by gradation data analysis. The groupings results are shown in comparison with the results obtained by other methods.*

**Keywords:** similarity of structures, map the over-representation, grouping objects, gradation data analysis.

**Резюме.** В статье была предпринята проблема измерения несходства структур. Оказалось, что предлагаемая мера, разработанная по принципу аналогии с коэффициентом Джини, позволяет обнаружить тонкости, на которые не чувствительны обычно используемые меры основанные на метриках. В разработке был представлен способ визуализации структур с использованием карт перепредставленности, а также способы группировки объектов с помощью градационного анализа данных. Полученные результаты группировки представляются на фоне результатов полученных другими методами.

**Ключевые слова:** сходство структур, карта перепредставленности, группировка объектов, градационный анализ данных.