

Danuta Strahl

KLASYFIKACJA POZYCYJNA W ANALIZACH DYNAMICZNYCH

1. Wstęp

Wśród wielu zadań i celów analiz ekonomicznych znajduje się ocena zmian tendencji rozwoju i wzajemnych relacji badanych obiektów. W badaniach regionalnych zachodzi z kolei potrzeba łączenia kryteriów oceny uwzględniających wymiar dynamiczny oraz przestrzenny. Wielowymiarowa analiza danych zawiera zdecydowanie więcej propozycji dla badań opartych wyłącznie na danych przekrojowych lub czasowych, a znacznie mniej opartych na danych przekrojowo-czasowych czy też przestrzenno-czasowych (por. [1; 6; 7; 8; 12]). Do cennych narzędzi opisu i identyfikacji szczególnych właściwości zjawisk ekonomicznych należą metody klasyfikacji. Metody te są licznie prezentowane w literaturze przedmiotu, ale na ogół ograniczają się do badań statycznych. Jest też kilka propozycji pokazujących możliwości wykorzystania metod klasyfikacji w badaniach dynamicznych (por. [5; 9; 11]). Zasadniczym celem tego artykułu jest opracowanie procedury klasyfikacji obiektów z wykorzystaniem statystyk pozycyjnych w badaniach dynamicznych.

2. Podstawy formalne

Dany jest zbiór obiektów $P = \{P_1, P_2, \dots, P_n, \dots, P_N\}$ opisany zbiorem m zmiennych, oznaczonych symbolami $X = \{X_1, X_2, \dots, X_m\}$. Zakładamy, że w zbiorze zmiennych znajdują się wyłącznie zmienne o charakterze stymulant (por. [10; 11; 12]). Kiedy pojawiają się zmienne o charakterze destymulant lub nominant, należy stosować znane formuły przekształceń ich na stymulanty (por. [10, 11, 12]).

Wartości cech obserwowane są w momentach czasowych $t = 1, 2, \dots, T$. Każdy obiekt zatem może być opisany za pomocą macierzy o postaci:

$$\mathbf{P}_k^t = \begin{bmatrix} x_{k_1}^1 & x_{k_2}^1 & \dots & x_{k_m}^1 \\ x_{k_1}^2 & x_{k_2}^2 & \dots & x_{k_m}^2 \\ \vdots & \vdots & & \vdots \\ x_{k_1}^t & x_{k_2}^t & \dots & x_{k_m}^t \\ \vdots & \vdots & & \vdots \\ x_{k_1}^T & x_{k_2}^T & \dots & x_{k_m}^T \end{bmatrix}_{T \times m}, \quad j = 1, 2, \dots, m, \quad k = 1, 2, \dots, K, \quad t = 1, 2, \dots, T, \quad (1)$$

gdzie: $x_{k_j}^t$ – wartość j -tej zmiennej ($j = 1, 2, \dots, m$) w k -tym obiekcie badania ($k = 1, 2, \dots, K$) w $t = 1, 2, \dots, T$ momencie obserwacji.

Zbiór obiektów $P = \{P_1, P_2, \dots, P_K\}$ może być opisany macierzą blokową o postaci:

$$\mathbf{P} = \begin{bmatrix} x_{11}^1 & x_{12}^1 & & x_{1m}^1 \\ x_{11}^2 & x_{12}^2 & & x_{1m}^2 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & x_{1j}^t & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{11}^T & x_{12}^T & \dots & x_{1m}^T \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & x_{k_j}^t & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{k_1}^T & x_{k_2}^T & & x_{k_m}^T \\ \vdots & \vdots & & \vdots \\ x_{K_1}^1 & x_{K_2}^1 & \dots & x_{K_m}^1 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & x_{K_j}^t & \vdots \\ x_{K_1}^T & x_{K_2}^T & \dots & x_{K_m}^T \end{bmatrix}_{K \times m \times T}, \quad (2)$$

gdzie: $x_{k_j}^t$ – wartość j -tej cechy w k -tym obiekcie w t -tym momencie obserwacji ($t = 1, 2, \dots, T$).

3. Klasyfikacja pozycyjna z medianą

Etap I procedury klasyfikacji

Klasyfikację obiektów $P = \{P_1, P_2, \dots, P_K\}$ przeprowadzamy dla każdego momentu obserwacji $t = 1, 2, \dots, T$ według następujących kroków.

1) Dla każdej zmiennej X dla każdego momentu $t = 1, 2, \dots, T$ wyznaczamy medianę według jednego ze wzorów (por. [2]):

$$Me(X_j) = \frac{x_{kj}^{i=K/2} + x_k^{i=K/2+1}}{2} \text{ dla parzystej liczby badanych obiektów,} \quad (3)$$

$$Me(X_j) = \frac{x_{kj}^{i=K \cdot m/2} + x_k^{i=K \cdot m/2+1}}{2} \text{ dla nieparzystej liczby badanych obiektów.} \quad (4)$$

2) Powstaje wektor median o postaci:

$$[Me(X_1) \ Me(X_2) \ \dots \ Me(X_m)]. \quad (5)$$

3) Klasyfikacja obiektów w momentach $t = 1, 2, \dots, T$.

Proponowana procedura klasyfikacji uwzględnia dwa przypadki. W przypadku pierwszym algorytm klasyfikacji prowadzi do budowy dla każdego momentu $t = 1, 2, \dots, T$ $(m + 1)$ klas oznaczonych symbolem S_g , gdzie $g = 1, 2, \dots, G$ ($G = m + 1$), gdy zbiory opisane są za pomocą m zmiennych. W przypadku drugim algorytm klasyfikacji prowadzi do budowy 2^m (czyli $G = 2^m$) klas możliwych kombinacji z m zmiennych dla każdego momentu $t = 1, 2, \dots, T$.

Przypadek pierwszy

Krok 1. Do klasy S_1^t (dla $t = 1, 2, \dots, T$) wchodzi obiekty ze zbioru P , których wartości wszystkich zmiennych X_j^t , czyli m zmiennych, są wyższe (korzystniejsze) od zadanej statystyki pozycyjnej lub jej równe. W naszych rozważaniach przyjmiemy, że statystyką tą jest mediana (Me). Stąd:

$$\bigwedge_j x_{kj}^t \geq Me(X_j^t), \quad (6)$$

gdzie: $k = 1, 2, \dots, K$, $j = 1, 2, \dots, m$, $t = 1, 2, \dots, T$.

Krok 2. Do klasy S_2^t wchodzi obiekty ze zbioru P , których wartości tylko $(m - 1)$ zmiennych spełniają warunek:

$$x_{kj}^t \geq Me(X_j^t). \quad (7)$$

Krok m . Do klasy S_g^t ($g = m$) wchodziły obiekty ze zbioru P , których tylko wartość jednej zmiennej X_j^t ze zbioru X spełnia warunek (7).

Krok $(m + 1)$. Do klasy S_{g+1}^t ($g = m + 1$) wchodziły obiekty P , których wartość x_{kj}^t żadnej zmiennej X_j^t nie spełnia warunku (7).

Przypadek drugi

Krok 1. Klasę S_1^t (dla $t = 1, 2, \dots, T$) tworzą te obiekty, których wartości wszystkich m zmiennych X_j^t spełniają warunek:

$$x_{kj}^t \geq Me(X_j^t), \text{ gdzie: } j = 1, 2, \dots, m. \quad (8)$$

Krok 2. Klasę S_2^t tworzą te obiekty, których wartości jedynie $(m - 1)$ zmiennych tworzących jedną z kombinacji $\binom{m}{m-1}$ zmiennych spełniają warunek (8).

Krok 3. Klasę trzecią S_3^t tworzą te obiekty, których wartości zmiennych kolejnej kombinacji $(m - 1)$ -elementowej spełniają warunek (8).

Krok 4. Po wyczerpaniu kombinacji $(m - 1)$ -elementowych tworzymy klasy dla kombinacji $(m - 2)$ -elementowych i stawiamy warunek (8).

Krok 2^m . Klasę S_g^t ($g = 2^m$) tworzymy z obiektów, dla których wartości x_{kj}^t wszystkich zmiennych X_j^t nie spełniają warunku (8).

Jak widać, oba przypadki mają wyraźnie odmienne założenia klasyfikacyjne. W przypadku pierwszym przypisujemy identyczne znaczenie wszystkim zmiennym, rozróżniając jedynie klasy obiektów poprzez liczbę zmiennych spełniających zadane warunki. Natomiast w drugim przypadku rozróżniamy grupy obiektów poprzez identyfikację specyfikacji zmiennych spełniających zadane warunki klasyfikacji.

Procedurę klasyfikacji powtarzamy dla każdego momentu $t = 1, 2, \dots, T$.

4. Klasyfikacja dynamiczna z dominantą

Etap II procedury klasyfikacji

1) Dla każdego obiektu P_k ($k = 1, 2, \dots, K$) budujemy wektor, którego elementami są częstości przynależności danego obiektu do poszczególnych klas określonych w I etapie klasyfikacji w badanych momentach $t = 1, 2, \dots, T$.

2) W wyniku tego pomiaru powstaje macierz, odpowiednio o wymiarach: $(m+1) \times K$ dla przypadku pierwszego oraz $2^m \times K$ dla przypadku drugiego o postaci:

$$\left[a_{gk} \right]_{(m+1) \times K} \text{ lub } \left[a_{gk} \right]_{2^m \times K}, \quad (9)$$

gdzie: a_{gk} – częstość występowania k -tego regionu ($k = 1, 2, \dots, K$) w t momentach badania ($t = 1, 2, \dots, T$) w g -tej klasie ($g = 1, 2, \dots, (m+1)$ lub $g = 1, 2, \dots, 2^m$).

3) Dla każdego obiektu P_k wyznaczamy dominantę z wektora: $\left[a_{gk} \right]_{1 \times g}$.

4) Klasy dynamiczne (a więc zawierające obiekty obserwowane we wszystkich okresach $t = 1, 2, \dots, T$) tworzymy w następujących krokach.

Przypadek pierwszy

Krok 1. Do klasy D_1 wchodzi obiekty ze zbioru P , dla których dominanta wartości przynależności obiektów do klas utworzonych w I etapie klasyfikacji w momentach $t = 1, 2, \dots, T$ znalazła się w klasie pierwszej.

Krok 2. Do klasy D_2 wchodzi obiekty ze zbioru P , dla których dominanta wartości przynależności obiektów do klas utworzonych w I etapie klasyfikacji w momentach $t = 1, 2, \dots, T$ znalazła się w klasie drugiej.

Krok m . Do klasy D_g (dla $g = m$) wchodzi obiekty ze zbioru P , dla których dominanta wartości przynależności obiektów do klas utworzonych w I etapie klasyfikacji w momentach $t = 1, 2, \dots, T$ znalazła się w klasie m -tej.

Krok $(m+1)$. Do klasy D_g (dla $g = m+1$) wchodzi obiekty ze zbioru P , dla których dominanta wartości przynależności obiektów do klas utworzonych w I etapie klasyfikacji w momentach $t = 1, 2, \dots, T$ znalazła się w klasie $m+1$.

Przypadek drugi

Krok 1. Klasę D_1 tworzą te obiekty ze zbioru P , dla których dominanta wartości przynależności obiektów do klas utworzonych w I etapie klasyfikacji w momentach $t = 1, 2, \dots, T$ znalazła się w klasie pierwszej.

Krok 2. Klasę D_2 tworzą te obiekty ze zbioru P , dla których dominanta wartości przynależności obiektów do klas utworzonych w I etapie klasyfikacji w momentach $t = 1, 2, \dots, T$ znalazła się w klasie drugiej.

Krok 3. Klasę trzecią D_3 tworzą te obiekty ze zbioru P , dla których dominanta wartości przynależności obiektów do klas utworzonych w I etapie klasyfikacji w momentach $t = 1, 2, \dots, T$ znalazła się w klasie trzeciej.

Po wyczerpaniu kombinacji $(m - 1)$ -elementowych tworzymy klasy dla kombinacji $(m - 2)$ -elementowych.

Krok 2^m. Klasę D_g ($g = 2^m$) tworzą te obiekty ze zbioru P , dla których dominanta wartości przynależności obiektów do klas utworzonych w I etapie klasyfikacji w momentach $t = 1, 2, \dots, T$ znalazła się w klasie 2^m .

Pozostają jeszcze do omówienia przypadki szczególne. Jak wiadomo, może na przykład zaistnieć sytuacja, w której brakuje dominanty. Jeżeli zatem częstość występowania obiektów w określonych klasach jest identyczna, przypisujemy dany obiekt do klasy z ostatniego okresu badania.

5. Ilustracja proponowanej procedury

Obiektami badania będą regiony szereblu NUTS-2 w Wielkiej Brytanii, stąd: $P = 1, 2, \dots, 37$. Obiekty te zostały opisane następującymi cechami:

X_1 – udział pracujących z wyższym wykształceniem w ogólnej liczbie pracujących w regionie,

X_2 – kapitał ludzki w nauce i technologii (*HRST*) jako odsetek aktywnych zawodowo,

X_3 – udział ludności w wieku 25-64 lata uczestniczącej w kształceniu ustawicznym w regionie.

Wartości cech były obserwowane w pięciu latach, tj. w 2001, 2002, 2003, 2004, 2005, stąd $t = 1, 2, 3, 4, 5$.

Zgodnie z procedurą klasyfikacji (według przypadku drugiego) utworzono osiem klas obiektów:

- Klasa I obejmuje obiekty – regiony, dla których wartości wszystkich zmiennych X_1, X_2, X_3 są wyższe od mediany.
- Klasa II obejmuje obiekty – regiony, dla których wartości cech X_1 i X_2 są wyższe od mediany, a wartości cechy X_3 są niższe od mediany.
- Klasa III obejmuje obiekty – regiony, dla których wartości cech X_1 i X_3 są wyższe od mediany, a wartości cechy X_2 są niższe od mediany.
- Klasa IV obejmuje obiekty – regiony, dla których wartości cech X_2 i X_3 są wyższe od mediany, a wartości cechy X_1 są niższe od mediany.
- Klasa V obejmuje obiekty – regiony, dla których wartość cechy X_1 jest wyższa od mediany, a wartości dwóch cech, tj. X_2 oraz X_3 , są niższe od mediany.
- Klasa VI obejmuje obiekty – regiony, dla których wartość cechy X_2 jest wyższa od mediany, a wartości dwóch cech, tj. X_2 oraz X_3 , są niższe od mediany.
- Klasa VII obejmuje obiekty – regiony, dla których wartość cechy X_3 jest wyższa od mediany, a wartości dwóch cech, tj. X_1 oraz X_2 , są niższe od mediany.

- Klasa VIII obejmuje obiekty – regiony, dla których wartości wszystkich cech są niższe od mediany.

Tabela 1. Częstość przynależności badanych regionów do ośmiu klas w latach 2001-2005

Obiekt	1	2	3	4	5	6	7	8	Klasa w dynamicznej klasyfikacji
1. Tees Valley and Durham						5			6
2. Northumberland, Tyne and Wear	1	2				2			6*
3. Cumbria				1		1	1	2	8
4. Cheshire	1	1		2		1			4
5. Greater Manchester	1	1	1		2				5
6. Lancashire				2		3			6
7. Merseyside		1	1		3				5
8. East Riding and North Lincolnshire							1	4	8
9. North Yorkshire			4		1				3
10. South Yorkshire			1		1		2	1	7
11. West Yorkshire			4		1				3
12. Derbyshire and Nottinghamshire				1		4			6
13. Leicestershire, Rutland and Northants				2		3			6
14. Lincolnshire				1		3		1	6
15. Herefordshire, Worcestershire and Warks		1		2		2			6*
16. Shropshire and Staffordshire		1		3		1			4
17. West Midlands	1	2				2			2*
18. East Anglia			1	1	2			1	5
19. Bedfordshire, Hertfordshire	5								1
20. Essex	3				2				1
21. Inner London			5						3
22. Outer London			5						3
23. Berkshire, Bucks and Oxfordshire	5								1
24. Surrey, East and West Sussex	1		4						3
25. Hampshire and Isle of Wight	2			3					4
26. Kent			2		3				5
27. Gloucestershire, Wiltshire and North Somerset	5								1
28. Dorset and Somerset	2			1			2		4*
29. Cornwall and Isles of Scilly				1			1	3	8
30. Devon	1					1	3		7
31. West Wales and The Valleys						2		3	8
32. East Wales	2	1			1	1			1
33. North Eastern Scotland							1	4	8
34. Eastern Scotland			5						3
35. South Western Scotland		2			3				5
36. Highlands and Islands					1		2	2	8*
37. Northern Ireland								5	8

* Klasa z ostatniego okresu badania.

Źródło: obliczenia własne na podstawie danych Eurostatu.

Ze względu na ograniczone ramy artykułu nie podano wyników klasyfikacji dla każdego roku, a w tab. 1 podano tylko częstość występowania każdego regionu w poszczególnych klasach w każdym badanym roku.

Klasyfikacja dynamiczna (za lata 2002-2005) przynosi podział regionów brytyjskich ze względu na przyjęte do analizy cechy, który podano w tab. 2.

Tabela 2. Klasyfikacja regionów Wielkiej Brytanii w ujęciu dynamicznym

Klasa/liczba regionów	$X_1 \geq Me$	$X_2 \geq Me$	$X_3 \geq Me$	Regiony
1/5	+	+	+	1. Bedfordshire, Hertfordshire 2. Essex 3. Berkshire, Bucks and Oxfordshire 4. Gloucestershire, Wiltshire and North Somerset 5. East Wales
2/1	+	+	-	1. West Midlands
3/6	+	-	+	1. North Yorkshire 2. West Yorkshire 3. Inner London 4. Outer London 5. Surrey, East and West Sussex 6. Eastern Scotland
4/4	-	+	+	1. Cheshire 2. Shropshire and Staffordshire 3. Hampshire and Isle of Wight 4. Dorset and Somerset
5/5	+	-	-	1. Greater Manchester, 2. Merseyside 3. East Anglia 4. Kent 5. South Western Scotland
6/7	-	+	-	1. Tees Valley and Durham 2. Northumberland, Tyne and Wear 3. Lancashire 4. Derbyshire and Nottinghamshire 5. Leicestershire, Rutland and Northants 6. Lincolnshire 7. Herefordshire, Worcestershire and Warks
7/2	-	-	+	1. South Yorkshire 2. Devon
8/7	-	-	-	1. Cumbria 2. East Riding and North Lincolnshire 3. Cornwall and Isles of Scilly 4. West Wales and The Valleys 5. North Eastern Scotland 6. Highlands and Islands 7. Northern Ireland

Źródło: obliczenia własne na podstawie danych Eurostatu.

6. Zakończenie

Przedstawiona propozycja klasyfikacji obiektów badania ma charakter wartościujący oraz uwzględnia pozycje obiektów w zadanym okresie, co ma szczególne znaczenie w badaniach dynamicznych. Klasyfikacja wykorzystująca statystyki pozycyjne, w tym medianę i dominantę, pozwala na podział obiektów uwzględniający tendencje relacji, jakie zachodzą między obiektami badania w zadanym przedziale czasowym.

Literatura

- [1] Jajuga K., *Statystyczna analiza wielowymiarowa*, PWN, Warszawa 1993.
- [2] Luszniwicz A., Słaby T., *Statystyka stosowana*, PWE, Warszawa 1998.
- [3] Milasewic P., Ducharme G.R., *Uniqueness of the Spatial Median*, „The Annals of Statistics” 1987, vol. 15, no. 3.
- [4] Młodak A., *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa 2006.
- [5] *Poziom życia w Polsce i krajach UE*, red. A. Zeliaś, PWE, Warszawa 2004.
- [6] Strahl D., *Strukturalna miara rozwoju obiektów hierarchicznych*, [w:] *Ekonometria 16*, red. J. Dziechciarz, Prace Naukowe Akademii Ekonomicznej nr 1100, AE, Wrocław 2006.
- [7] Strahl D., *Klasyfikacja regionów z medianą*, [w:] *Ekonometria 10*, red. J. Dziechciarz, Prace Naukowe Akademii Ekonomicznej nr 950, AE, Wrocław 2002.
- [8] Strahl D., *Miejsce Polski w regionalnej przestrzeni UE*, [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, red. A. Zeliaś, AE, Kraków 2005.
- [9] Strahl D., Markowska M., *Klasyfikacja dynamiczno-przestrzenna obiektów hierarchicznych z wykorzystaniem statystyk pozycyjnych*. AE, Wrocław (złożone do druku).
- [10] *Taksonomia struktur w badaniach regionalnych*, red. D. Strahl, AE, Wrocław 1998.
- [11] *Taksonomiczna analiza przestrzennego różnicowania poziomu życia w Polsce w ujęciu dynamicznym*, red. A. Zeliaś, AE, Kraków 2000.
- [12] Walesiak M., *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, AE, Wrocław 2002.

POSITIONAL CLASSIFICATION IN DYNAMIC ANALYSES

Summary

The article presents the proposal of the studied objects classification in a dynamic perspective. The procedure of objects division is divided into two stages. In the first stage the classification of objects is performed by means of positional statistics. The basic division criterion becomes the evaluation of attributes relations which describe the studied objects up to the median value. The procedure is repeated for each observation moment of values referring to the studied objects' values. The second stage consists in the studied objects classification with reference to positions occupied by

objects in particular moments of the study in the first stage of the classification. The classification criterion in the second stage becomes the frequency dominant of the studied objects' classes, defined in the first stage of each moment of the study.

Danuta Strahl – prof. zw. dr hab., kierownik Katedry Gospodarki Regionalnej Uniwersytetu Ekonomicznego we Wrocławiu – Wydział w Jeleniej Górze.