

Elżbieta Sobczak

Akademia Ekonomiczna we Wrocławiu

PROPOZYCJA DYNAMICZNO-PRZESTRZENNEJ ANALIZY PORÓWNAWCZEJ STRUKTUR EKONOMICZNYCH

1. Wprowadzenie

Współcześnie podkreśla się dużą rolę struktur ekonomicznych dla funkcjonowania i dynamiki gospodarki regionalnej. Struktura ekonomiczna jest jednym z czynników kształtujących proces rozwoju społeczno-gospodarczego każdego regionu. Struktura dynamizująca powoduje przyspieszenie rozwoju gospodarczego, a struktura stagnacyjna przyczynia się do zmniejszenia efektywności wykorzystania innych czynników rozwoju.

Wyodrębnienie podobnych struktur ekonomicznych i łączenie ich w homogeniczne grupy umożliwia uporządkowanie otaczającej człowieka rzeczywistości i poznawanie istoty badanych zjawisk gospodarczych. Jest to jeden z podstawowych problemów wielowymiarowej analizy strukturalnej. Do jego realizacji stosuje się metody klasyfikacji. Dyscypliną naukową zajmującą się zasadami i metodami klasyfikacji jest taksonomia. Spośród wielu znanych metod taksonomicznych można wyodrębnić metody ilościowe i jakościowe. Rodzaj stosowanych metod zależy od typu wykorzystywanych informacji wyjściowych. Przedmiotem dalszych rozważań będą ilościowe metody taksonomiczne składające się na taksonomię numeryczną.

Klasyfikacja jest jednym z podstawowych zagadnień wchodzących w zakres zainteresowań zarówno taksonomii numerycznej jak i wielowymiarowej analizy strukturalnej. Pojęcie „klasyfikacja” może być pojmowane w sposób wieloznaczny. Formulowane są semiotyczne, metodologiczne i teoriomnogościowe definicje tego terminu (por. [6, 26]). Różne znaczenia pojęcia „klasyfikacja” przedstawiają w swoich pracach T. Wójcik [41] i T. Borys [4]. W taksonomii numerycznej wykorzystuje się definicje klasyfikacji opartą na teorii mnogości. Klasyfikacja rozumiana jest tutaj jako [13] „(...) niepusta rodzina podzbiorów

określona na zbiorze obiektów i spełniająca warunki rozłączności i zupełności". Na podstawie różnych definicji klasyfikacji, które można odnaleźć w literaturze, sformułowano określenie klasyfikacji struktur ekonomicznych [6, 26, 13, 29, 25, 37].

Klasyfikacja struktur ekonomicznych oznacza zupełny i rozłączny podział danego, niejednorodnego z punktu widzenia określonych cech strukturalnych, zbioru struktur ekonomicznych na pewną liczbę niepustych podzbiorów, wewnętrznie bardziej jednorodnych, dokonywany na podstawie relacji podobieństwa struktur ekonomicznych. Klasyfikacja struktur ekonomicznych oznacza zarówno czynność podziału zbioru struktur opisanych na obiektach jak i efekt tej czynności.

Klasyfikacja powinna spełniać trzy podstawowe warunki [6, 13, 25, 37, 28]:

1) Warunek zupełności, zwany inaczej warunkiem adekwatności polegający na tym, że suma podzbiorów struktur ekonomicznych uzyskanych w wyniku klasyfikacji musi być równa wyjściowemu zbiorowi struktur.

2) Warunek rozłączności oznaczający, że każda z badanych struktur ekonomicznych może należeć tylko do jednego podzbioru.

3) Warunek niepustości oznaczający, że w każdym podzbiorze znajduje się co najmniej jedna struktura ekonomiczna.

Podzbiory uzyskane w wyniku klasyfikacji nazywa się klasami lub grupami. Liczba klas może być ustalona za pomocą określonych kryteriów lub może pozostać nieznana, jak to najczęściej ma miejsce.

Jedynym wewnętrznym kryterium podziału wyjściowego zbioru struktur ekonomicznych jest matematycznie zdefiniowane podobieństwo między nimi. Dlatego klasyfikacja powinna być dokonywana w taki sposób, aby struktury ekonomiczne należące do tej samej klasy były podobne, a struktury ekonomiczne należące do różnych klas były niepodobne. Należy jednak pamiętać, że każda klasyfikacja jest umowna i nie zmienia tego zastosowanie metod statystycznych formalnie obiektywizujących jej procedurę.

Problematyka metod klasyfikacji została poruszona m.in. w następujących publikacjach autorów polskich i zagranicznych: J. Czekanowski [10], K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus, S. Zubrzycki [11], W. Bukietyński, Z. Hellwig, U. Królik, A. Smoluk [5], U. Siedlecka [32], W. Pluta [27], Z. Hellwig [15], D. Strahl [35], S. Chomątowski, A. Sokołowski [7], S. A. Ajwazjan, Z. T. Bieżajewa, O. W. Starowierow [1], C. H. Anderberg [2], P. H. A. Sneath, R. R. Socal [33], J. van Ryzin [8], J. A. Hartigan [14], G. N. Lance, W. T. Williams [19, 20], M. Rand [31], J. H. Ward [39], D. Wishard [40], R. C. Prim [30], B. J. L. Berry [3]. Szerszy przegląd literatury na ten temat zawierają m.in. monografie: K. Jajugi [18], E. Nowaka [25], M. Walesiaka [36].

Klasyfikacja struktur ekonomicznych nie stanowi odrębnego zagadnienia klasyfikacji, dlatego do ustalenia jednorodnych grup struktur ekonomicznych mogą być wykorzystywane ogólnie znane metody klasyfikacji.

Uporządkowanie i pogrupowanie znanych metod taksonomicznych może być dokonywane na podstawie wielu różnorodnych kryteriów. Szczegółowa klasyfikacja metod taksonomicznych przedstawiona jest m.in. w następujących pracach: J. Pocięcha, B. Podolec, A. Sokołowski, K. Zajac [28], T. Grabiński, S. Wydymus, A. Zeliaś [13], T. Marek [21].

W artykule ograniczono się do zwięzłego omówienia metod taksonomicznych wyodrębnionych według dwóch następujących kryteriów:

- 1) hierarchia grupowania,
- 2) rozpoczęcie procesu grupowania.

Bardzo rozpowszechnionym kryterium podziału metod taksonomicznych jest hierarchia grupowania. Z tego względu można wyodrębnić hierarchiczne i niehierarchiczne metody klasyfikacji. W efekcie zastosowania metody hierarchicznej uzyskuje się ciąg podziałów, przy czym dla każdego podziału zachowany jest warunek zupełności, rozłączności i niepustości. Po zastosowaniu metody niehierarchicznej otrzymuje się jeden podział wynikowy. Przy stosowaniu metod hierarchicznych należy z uzyskanego ciągu podziałów dokonać wyboru tzw. podziału optymalnego, czyli ustalić optymalną liczbę klas, na którą należy podzielić badane struktury ekonomiczne. W metodach niehierarchicznych zasadniczą trudnością jest określenie tzw. odległości krytycznej, którą określa odległość graniczna dla struktur ekonomicznych uznanych za podobne.

Rozpoczęcie procesu grupowania jest kryterium umożliwiającym wyodrębnienie spośród metod hierarchicznych metod aglomeracyjnych i deglomeracyjnych. Punktem wyjścia w metodach aglomeracyjnych, zwanych niekiedy indukcyjnymi, wstępującymi lub metodami od dołu, jest moment, w którym każda z badanych struktur ekonomicznych stanowi odrębną grupę. Struktury najbardziej podobne według określonego kryterium łączone są w klasy. Klasyfikacja kończy się wówczas, gdy wszystkie obiekty znajdują się w jednej klasie.

Punktem wyjścia procesu klasyfikacji dokonywanego z wykorzystaniem metod deglomeracyjnych, zwanych również dedukcyjnymi, zstępującymi lub metodami od góry, jest moment, w którym wszystkie obiekty należą do jednej klasy. W wyniku klasyfikacji następuje stopniowe rozbijanie tej klasy na kilka mniej licznych i bardziej jednorodnych. Efektem ostatniego etapu klasyfikacji, dokonanego według metody deglomeracyjnej, jest podział na jednoelementowe klasy.

Najbardziej rozpowszechnione w badaniach empirycznych są hierarchiczne metody aglomeracyjne, dlatego zostaną krótko scharakteryzowane. Hierarchiczne metody aglomeracyjne działają według schematu zwanego centralną procedurą aglomeracyjną, której algorytm jest następujący:

- 1) Strukturę ekonomiczną każdego obiektu w ($r = 1, \dots, n$) traktuje się jako odrębną klasę.
- 2) Łączy się dwie klasy najmniej oddalone od siebie.
- 3) Wyznacza się odległość nowo utworzonej klasy od wszystkich pozostałych, odpowiednio do stosowanej metody.

4) Powtarza się kroki 2-3 ($n-1$) razy, czyli do momentu, gdy wszystkie struktury utworzą jedna grupę.

Po każdorazowej realizacji kroków 2-3 procedury aglomeracyjnej liczba klas zmniejsza się o jeden.

Hierarchiczne metody aglomeracyjne różnią się między sobą sposobem ustalania odległości nowo utworzonej klasy od wszystkich pozostałych klas. Zwięźle opisano te spośród znanych hierarchicznych metod aglomeracyjnych, w których odległość między dwiema klasami daje się wyrazić za pomocą odległości między strukturami ekonomicznymi obiektów należących do tych klas.

1) Metoda najbliższego sąsiada (metoda prostych połączeń) – odległość między klasami ustalona jest jako minimalna odległość między strukturami ekonomicznymi obiektów należących do tych klas.

2) Metoda najdalszego sąsiada (metoda zupełnych połączeń) – odległość między klasami rozumiana jest jako maksymalna odległość między strukturami obiektów należących do tych klas.

3) Metoda medianowa – odległość między dwoma klasami to mediana odległości między strukturami ekonomicznymi obiektów należących do tych klas.

4) Metoda średniej grupowej (metoda średniej międzyklasowej) – odległość między klasami oblicza się jako średnią arytmetyczną z odległości między strukturami ekonomicznymi obiektów należących do tych klas.

5) Metoda środka ciężkości (metoda centroidalna) – odległość między klasami jest równa odległości między środkami ciężkości tych klas. Przez środek ciężkości klasy rozumie się wektor średnich arytmetycznych z odpowiadających sobie unormowanych elementów struktur prostych obiektów należących do tej klasy, lub macierz średnich arytmetycznych z odpowiadających sobie unormowanych elementów struktur złożonych obiektów należących do tej klasy.

G. N. Lance i W. T. Williams [19, 20] zaproponowali uogólnioną formułę obliczania odległości dla wymienionych wyżej metod aglomeracyjnych, którą cytuje się powszechnie w literaturze (por. [13, 25, 28, 18]).

Do grupy hierarchicznych metod deglomeracyjnych należą m.in. metoda taksonomii wrocławskiej [11], metoda R. C. Prima [30] i metody L. Huberta [16].

2. Metodologia przestrzennej i przestrzenno-czasowej klasyfikacji struktur ekonomicznych

Przyjmuje się, że klasyfikacja przestrzenna obejmuje podział na klasy zbioru struktur ekonomicznych, opisanych za pomocą jednej cząstkowej cechy strukturalnej dotyczącej jednego okresu badania. Klasyfikacja przestrzenno-czasowa to podział zbioru struktur ekonomicznych na klasy w przestrzeni t cząstkowych cech strukturalnych składających się na złożoną cechę strukturalną (t-liczba okresów badania).

Do przeprowadzenia przestrzennej i przestrzenno-czasowej klasyfikacji struktur ekonomicznych proponuje się wykorzystanie następującego schematu postępowania:

1) Normalizacja cech strukturalnych umożliwiającą prezentację skali i kształtu struktur ekonomicznych.

2) Ustalenie miary zróżnicowania struktur prostych i złożonych.

Jako miarę odległości struktur prostych proponuje się zastosowanie jednej z miar odzwierciedlających zróżnicowanie zarówno kształtu jak i skali struktur. Do kwantyfikacji zróżnicowania struktur złożonych proponuje się średnią arytmetyczną z odległości między strukturami prostymi, składającymi się na dane struktury złożone.

3) Wybór metody klasyfikacji.

Klasyfikację przestrzenną i przestrzenno-czasową badanych obiektów ze względu na kształt i skalę struktur ekonomicznych można przeprowadzić stosując równolegle następujące hierarchiczne metody aglomeracyjne: najbliższego sąsiada, najdalszego sąsiada, medianową, średniej grupowej i środka ciężkości.

Zdecydowano się na wybór metod hierarchicznych, ponieważ w wyniku ich zastosowania otrzymuje się ciąg klasyfikacji, od podziału w którym każdy obiekt stanowi odrębną klasę, aż do podziału w którym wszystkie obiekty znajdują się w jednej klasie. Umożliwia to kontrolę procesu klasyfikacji i wybór optymalnego jej etapu.

Przyjęto, że podział na klasy uznaje się za przydatny, jeżeli spełnia n]w warunki:

- liczba klas jednoelementowych nie jest zbyt duża,
- nie występują klasy bardzo liczne.

Zastosowanie hierarchicznej metody klasyfikacji obiektów stwarza konieczność ustalenia tzw. reguły stop do wyboru z uzyskanego ciągu podziałów, podziału optymalnego. W hierarchicznych metodach aglomeracyjnych stosuje się różne sposoby ustalania podziału optymalnego. W literaturze można znaleźć m.in. propozycje R. Mojeny [23], L. Berry'ego [6], J.J. Fortiera i H. Solomona [12], G.N. Żitkowa [42] czy B.G. Mirkina [22]. Więcej informacji na temat ustalania tzw. reguły stop można znaleźć m.in. w pracach M. Walesiaka [38], E. Nowaka [25], K. Jajugi [17].

Metody R. Mojeny i L. Berry'ego znajdują zastosowanie w tych spośród hierarchicznych metod aglomeracyjnych, w których wartości poziomu połączenia klas rozłożone są monotonicznie rosnąco. Przez poziom połączenia klas rozumie się odległość między klasami najbardziej podobnymi, łączonymi w danej iteracji. Sposoby te nie mogą być wykorzystane do klasyfikacji struktur metodą medianową, gdyż rozkład poziomu połączenia klas uzyskany tą metodą nie jest monotoniczny. Zaletą pozostałych propozycji jest możliwość ich zastosowania do klasyfikacji uzyskanych wszystkimi metodami hierarchicznymi. Ich podstawową wadą jest konieczność arbitralnego ustalenia tzw. krytycznej odległości między

strukturami obiektów. Ustalenie poziomu krytycznej odległości w istotny sposób wpływa na wybór optymalnego podziału z ciągu klasyfikacji.

Następnie dokonuje się wyboru najlepszego podziału uzyskanego każdą z metod i porównuje się te podziały za pomocą wskaźników jakości klasyfikacji. Przesądza to o ostatecznym wyborze metody dającej optymalne podziały.

4) Rejestracja zmian w czasie w klasyfikacjach przestrzennych struktur ekonomicznych.

Porównania wyników klasyfikacji struktur uzyskanych jedną ze stosowanych metod, na podstawie informacji statystycznych pochodzących w dwóch różnych okresach czasu, można dokonać posługując się różnymi sposobami. Efektem zastosowania metody hierarchicznej jest ciąg wyników klasyfikacji, z którego po zastosowaniu określonej reguły wybiera się jeden optymalny podział. W takim przypadku do porównania klasyfikacji struktur pochodzących z dwóch różnych okresów można wykorzystać następujące podejścia:

1) Porównanie dwóch ciągów klasyfikacji uzyskanych na podstawie informacji statystycznych pochodzących z l -tego i l -tego okresu badania.

2) Porównanie dwóch optymalnych podziałów wybranych z ciągów klasyfikacji uzyskanych dla l -tego i l -tego okresu badania.

Propozycje miar umożliwiających porównanie ciągów klasyfikacji można znaleźć m.in. w pracach [2, 9, 34, 14].

Zgodność dwóch optymalnych podziałów wybranych z ciągów klasyfikacji można ocenić pod względem składu porównywanych klas. Propozycje miar oceniających zgodność wyników klasyfikacji, w takim właśnie rozumieniu, przedstawili m.in. W. M. Rand [31], R. R. Anderberg [2], E. Nowak [24].

5) Analiza współzależności występujących między podziałami otrzymanymi ze względu na skalę i kształt analizowanych struktur.

Do pomiaru stopnia natężenia omawianych współzależności można wykorzystać sposoby omówione w poprzednim punkcie.

Analizę wymienionych współzależności można przeprowadzić zarówno na podstawie klasyfikacji przestrzenno-czasowych jak i klasyfikacji przestrzennych uzyskanych w odpowiadających sobie latach.

Literatura

- [1] S.A. Ajwazjan, Z.T. Bieżajewa, O.W. Starowierow: *Klassifikacyja mnogomiernych nabljudenij*. „Statistika” Moskwa 1974.
- [2] M.R. Anderberg: *Cluster analysis for applications*. New York, San Francisco, London: Academic Press 1973.
- [3] B.J.L. Berry: *A method for deriving multifactor uniform regions*. „Przegląd Geograficzny” 1962 nr 2.

- [4] T. Borys: *Kategoria jakości w statystycznej analizie porównawczej*. Wrocław: AE 1984. Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 284.
- [5] W. Bukietyński, Z. Hellwig, U. Królik, A. Smoluk: *Uwagi o dyskryminacji zbiorów skończonych*. Wrocław: WSE 1969. Prace Naukowe Wyższej Szkoły Ekonomicznej we Wrocławiu nr 21.
- [6] Z. Chojnicki, T. Czyż: *Metody taksonomii numerycznej w regionalizacji geograficznej*. Warszawa: PWN 1973.
- [7] S. Chomątowski, A. Sokołowski: *Taksonomia struktur*. „Przegląd Statystyczny” 1978 nr 2.
- [8] *Classification and Clustering*. Red. J. Van Ryzin. New York: Academic Press 1977.
- [9] R.M. Cormack: *A review of classification (with discussion)*. „Journal of the Royal Statistical Society” 1971, vol. 134, part 3, ser. A.
- [10] J. Czekanowski: *Zarys metod statystycznych w zastosowaniu do antropologii*. „Prace Towarzystwa Naukowego Warszawskiego” 1913 nr 5.
- [11] K. Florek, J. Łukasiewicz, J. Perkal, H. Steinhaus, S. Zubrzycki: *Taksonomia wroclawska*. „Przegląd Antropologiczny” 1951 nr 17.
- [12] J.J. Fortier, H. Solomon: *Clustering Procedures*, w: *Multivariate Analysis*. Red. P.R. Krishnaiah. New York: Academic Press 1966 nr 62.
- [13] T. Grabiński, S. Wydymus, A. Zeliaś: *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*. Warszawa: PWN 1989.
- [14] J.A. Hartigan: *Clustering Algorithms*. Wiley, New York 1975.
- [15] Z. Hellwig Z.: *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*. „Przegląd Statystyczny” 1968 nr 4.
- [16] L. Hubert: *Monotone invariant clustering procedures*. „Psychometrica” 1973 nr 38.
- [17] K. Jajuga: *O sposobach określania ilości klas w zagadnieniu klasyfikacji i klasyfikacji rozmytej*. Wrocław: AE 1984. Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 262.
- [18] K. Jajuga: *Statystyczna teoria rozpoznawania obrazów*. Warszawa: PWN 1990.
- [19] G.N. Lance, W.T. Williams: *A general theory of classificatory sorting strategies. 1. Hierarchical systems*. „Computer Journal” 1967 nr 9.
- [20] G.N. Lance, W.T. Williams: *A general theory of classificatory sorting strategies. 2. Clustering systems*. „Computer Journal” 1968 nr 10.
- [21] T. Marek: *Analiza skupień w badaniach empirycznych. Metody SAHN*. Warszawa: PWN 1989.
- [22] B.G. Mirkin: *Analiz kaczestwiennych признаков*. „Statistika” Moskwa 1976.
- [23] R. Mojena: *Hierarchical grouping methods and stopping rules: an evaluation*. „The Computer Journal” 1977 nr 20.
- [24] E. Nowak: *Wskaźniki podobieństwa wyników podziałów*. „Przegląd Statystyczny” 1985 nr 1.
- [25] E. Nowak: *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*. Warszawa: PWE 1990.
- [26] W. Ostasiewicz: *Dyskryminacja, klasyfikacja, rozpoznanie*. Wrocław: AE 1980. Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 165.
- [27] W. Pluta: *O pewnej metodzie klasyfikacji przedsiębiorstw*. „Przegląd Statystyczny” 1972 nr 1.
- [28] J. Pocięcha, B. Podolec, A. Sokołowski, K. Zając: *Metody taksonomiczne w badaniach społeczno-ekonomicznych*. Warszawa: PWN 1988.
- [29] B. Podolec, K. Zając: *Ekonometryczne metody ustalania rejonów konsumpcji*. Warszawa: PWE 1978.
- [30] R.C. Prim: *Shortest connection networks and some generalizations*. „The Bell System Technical Journal” 1957 nr 6.
- [31] W.M. Rand: *Objective criteria for the evaluation of clustering methods*. „Journal of the American Statistical Association” 1971 nr 66.
- [32] U. Siedlecka: *Zastosowanie metody taksonomii stochastycznej do dyskryminacji zbiorów skończonych*. „Przegląd Statystyczny” 1976 nr 3.

- [33] P.H.A. Sneath, R.R. Sokal: *Numerical taxonomy. The principles and practice of numerical classification*. San Francisco: W.H.Freeman and Co. 1973.
- [34] R.R. Sokal, F.J. Rohlf: *The comparison of dendrograms by objective methods*. „Taxon” 1962 nr 11.
- [35] D. Strahl: *Dyskryminacja zbioru z kryterium stabilnego podobieństwa*. „Przegląd Statystyczny” 1981 nr 1.
- [36] M. Walesiak: *Metody klasyfikacji w badaniach strukturalnych*. Rozprawa doktorska. Wrocław: AE 1985 (maszynopis).
- [37] M. Walesiak: *Sposoby rejestracji zmian w czasie w wynikach klasyfikacji*. Wrocław: AE 1985. Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 311.
- [38] M. Walesiak: *Sposoby wyznaczania optymalnej liczby klas w zagadnieniu klasyfikacji hierarchicznej*. Wrocław: AE 1988. Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 449.
- [39] J.H. Ward: *Hierarchical grouping to optimize an objective function*. „Journal of the American Statistical Association” 1963 nr 58.
- [40] D. Wishard: *Algorithm for hierarchical classifications*. „Biometrics” 1969 nr 25.
- [41] T. Wójcik: *Zarys teorii klasyfikacji*. Warszawa: PWN 1965.
- [42] G.N. Żitkow: *Niekatoryje metody awtomatycznej klassifikacyi*, w: *Strukturyjne metody opoznawaniya i awtomatyczekoje cztienije*. „Statistika” Moskwa 1970.