BADANIE OBIEKTÓW WIELOCECHOWYCH ZA POMOCĄ METOD TAKSONOMII NUMERYCZNEJ Z WYKORZYSTANIEM OBIEKTU WZORCOWEGO

TESTING OF OBJECTS USING METHODS MULTIVARIATE TAXONOMY NUMERICAL USING THE OBJECT MODEL

Krzysztof FICOŃ

Grzegorz KRASNODĘBSKI

Akademia Marynarki Wojennej

Wydział Dowodzenia i Operacji Morskich

Streszczenie. *W pracy przedstawiono praktyczne podejście do badania złożonych obiektów wielocechowych za pomocą metod taksonomii numerycznej, których użyteczność zilustrowano prostym przykładem liczbowym. Mechanizm działania metody taksonomicznej przedstawiono na przykładzie tzw. obiektu wzorcowego, symbolizującego pewien system pożądany względem, którego zostały zrelatywizowane pozostałe badane obiekty. Wielka uniwersalność metod taksonomicznych wywodzących się de facto z nauk przyrodniczych implikuje rozliczne kierunki ich zastosowań, także w naukach technicznych czy społecznych.*

Summary. *The paper presents a practical approach to the study of complex objects by means of multivariate methods of numerical taxonomy, the usefulness of which illustrates a simple numerical example. The mechanism of action of the taxonomic method is exemplified by the so-called. object model, symbolizing a system of desired terms, which have been qualified by other test objects. Great versatility of taxonomic methods-derived de facto natural sciences implies numerous directions of their applications, also in technical sciences or social.*

Słowa kluczowe: *cecha, destymulanta, norma, obiekt, standaryzacja, stymulanta, taksonomia, topologia, zbiór.*

Keywords: *feature destimulant, standard, object, standardization, stimulant, taxonomy, topology collection.*

Wprowadzenie

Burzliwy rozwój nauki powoduje przesuwanie granic ludzkiego poznania na coraz dalsze horyzonty, a istotą podejścia naukowego jest holistyczne, całościowe rozpatrywanie różnych zjawisk i procesów w strukturach wielkich systemów. Proporcjonalnie do stanu zaawansowania cywilizacyjnego dysponujemy coraz bardziej złożonymi systemami, układami i procesami, które wymagają precyzyjnych badań, zaczynających się zazwyczaj od etapu dekompozycji na elementy i czynniki prostsze. Istotnym warunkiem podejścia naukowego jest także konieczność porządkowania i systemowej klasyfikacji zarówno wiedzy teoretycznej, jak też rzeczywistych faktów i zdarzeń. Praktycznie we wszystkich dziedzinach naszego życia klasyfikacja i porządkowanie są podstawą racjonalnych działań i optymalnych decyzji. Problematyką tą zajmuje się współczesna taksonomia i liczne jej metody i narzędzia, którymi aktualnie dysponuje (Grabiński, Wydymus, Zeliaś. 1983).

Pojęcie taksonomii pochodzi od dwóch greckich słów: *taksis* oznaczający układ, porządek oraz *nomos* – zwyczaj, prawo, zasada. Początkowo terminem tym określano dział biologii zajmujący się podziałem świata flory i fauny na gatunki, rodzaje, rodziny, rzędy, klasy, gromady i typy. Obecnie pojęcia tego używa się do definiowania zasad porządkowania i klasyfikacji różnych mniej lub bardziej złożonych systemów, obiektów i zjawisk – technicznych, społecznych, gospodarczych (Jajuga, Walesiak, 2011).

Taksonomię definiuje się jako naukę o zasadach porządkowania i klasyfikacji złożonych obiektów, systemów wielocechowych. Początkowo znalazła zastosowanie głównie w badaniach antropologicznych i geograficznych, a ostatnio jest szeroko wykorzystywana w wielu dyscyplinach ekonomicznych (Pluta, 1986), a także w naukach technicznych i społecznych. Pierwotnie na gruncie nauk przyrodniczych taksonomia była definiowana jako "(...) dział biologii zajmujący się zasadami klasyfikacji organizmów zwierzęcych i roślinnych oraz ulepszaniem aktualnych ich klasyfikacji" (Leksykon, 1984, s. 984). Metodami taksonomicznymi interesują się obecnie zaawansowane technologie informatyczne, a przedmiotem szczególnych badań są perspektywy wykorzystania jej w systemach sztucznej inteligencji, bazujących w dużej części na różnych kryteriach klasyfikacyjnych i systemach ocenowych. Metody taksonomiczne umożliwiają sprawne przechodzenie od miar ilościowych do ocen jakościowych, co zgodnie z zasadami dialektyki generuje nieraz nowe, jakościowe kategorie wartości.

Szczególne znaczenie, zwłaszcza w naukach społeczno-ekonomicznych odgrywa dział taksonomii numerycznej, definiowany niekiedy jako wielowymiarowa analiza porównawcza, która zajmuje się klasyfikacją i porządkowaniem wielowymiarowych (wielocechowych) obiektów statystycznych, opisanych za pomocą pewnych zbiorów zmiennych (cech) (Hellwig, 1968). Taksonomia numeryczna dostarcza wielu praktycznych metod i narzędzi do badania złożonych systemów zarówno społeczno-ekonomicznych, jak też organizacyjno-funkcjonalnych, a także polityczno-militarnych. Pozwala na wyprowadzenie syntetycznej oceny dowolnego obiektu wielocechowego i wzajemne relatywizowanie tych obiektów w pewnej przestrzeni topologicznej. Końcowym efektem badań i analiz taksonomicznych są zazwyczaj systemowe klasyfikacje i różne konwencje porządkujące określone zbiorowości i populacje według założonych kryteriów ocenowych (Ficoń, 2006). Przedmiotem klasyfikacji niekoniecznie muszą być pojedyncze obiekty, znacznie większe zapotrzebowanie istnieje na klasyfikacje grupowe dzielące badaną zbiorowość na pewne grupy heterogeniczne, zawierające klasy obiektów o zbliżonych właściwościach i określonych kryteriach podobieństwa.

Z uwagi na ogromną złożoność zjawisk, procesów i systemów, opisywanych przez coraz bardziej rozbudowane zbiory rozmaitych właściwości i cech istnieje pilna potrzeba badania ich za pomocą spójnych, naukowych metod, opartych głównie na aparacie współczesnej statystyki, wspomaganej narzędziowo przez technologię komputerową (Młodak, 2006). Wykorzystanie naukowych metod badawczych i sprawnych narzędzi informatycznych zdecydowanie zwiększyło efektywność procedur taksonomicznych i poprawiło ich dostępność dla wielu badaczy i specjalistów zajmujących się szczegółowymi problemami w ramach własnych zainteresowań naukowo-badawczych. Dlatego przybliżenie ogólnych założeń metodologicznych niełatwego podejścia taksonomicznego oraz jego zalet, ale też pewnych wyzwań stało się celem niniejszej pracy.

1. Podstawowe pojęcia taksonomii

Znaczna część nietrywialnych pojęć, obiektów, systemów, procesów zarówno teoretycznych jak też rzeczywistych, z uwagi na swoją złożoność jest opisywana, a często nawet definiowana za pomocą pewnego zbioru różnorodnych czynników, atrybutów czy cech charakteryzujących bardziej precyzyjnie ich złożoność i naturę. Konieczność operowania dodatkowymi atrybutami i właściwościowymi (cechami) celem precyzyjnego opisania danego obiektu (systemu, procesu) wynika z wielkiej ich złożoności i dużej skali skomplikowania wzajemnych współzależności. Dlatego w takich sytuacjach mówi się wprost o wielowymiarowości badanego układu (obiektu, systemu) (Jajuga, Walesiak, 2011). Badanie złożonych systemów, zwłaszcza techniczno-konstrukcyjnych, społeczno-ekonomicznych czy organizacyjno-funkcjonalnych z reguły nie można ograniczyć do badania tylko jednego aspektu i zamknąć w jednym wymiarze, czyli wyczerpująco scharakteryzować za pomocą jednej miary czy nawet najbardziej uniwersalnej cechy (właściwości). Szczególnym przypadkiem takiego skomplikowanego układu (obiektu) są tzw. wielkie systemy obejmujące wzajemnie powiązane struktury społeczne, gospodarcze, a także techniczne, wyodrębnione z otaczającej rzeczywistości według dodatkowych kryteriów i miar klasyfikacyjnych (Ficoń, 1996).

Klasyczna definicja systemu (S) formułowa jest wprost jako wyodrębniony w pewien sposób zbiór elementów (E) połączonych między sobą różnymi relacjami (związkami) (R) funkcjonalnymi, organizacyjnymi czy systemowymi (Duda, 1986):

$$S = \{E = \{E_i;\ i = \overline{1, I}\},\ R \subseteq E \times E\} \tag{1}$$

gdzie:

S – system, także obiekt, układ, proces, zjawisko itp.,

E – zbiór elementów tworzących dany system S,

R – zbiór relacji jedno, dwu i wieloczłonowych łączących elementy E,

I – liczność zbioru elementów $|E|$.

W tym przypadku zarówno zbiór wyodrębnionych elementów np. strukturalnych, organizacyjnych, jak też zbiór łączących je relacji np. funkcjonalnych należą do szerszej klasy pojęć zaliczanych do kategorii cech (właściwości). Zbiór cech C opisujących dany obiekt czy system powinien być zbiorem skończonym i przeliczalnym o pewnej liczności, np. I:

$$C = \{C_1, C_2, \dots, C_i, \dots C_I\} = \{C_i;\ i = \overline{1, I}\} \in S \tag{2}$$

gdzie:

C – zbiór cech opisujących badany system S,

C_i – i-ta cecha należąca do zbioru cech $C_i \in C$.

Każdy złożony obiekt w szczególności wielki system S opisujemy za pomocą zbioru właściwych mu cech, jednoznacznie charakteryzujących ten obiekt. W najprostszym ujęciu system lub obiekt S możemy zdefiniować jako pewną funkcję zbioru cech C:

$$S = f(C) = f(C_i;\quad i = \overline{1, I}) \tag{3}$$

Obiekt (system) prosty można formalnie opisać za pomocą jednej cechy, co jest przypadkiem niezwykle rzadkim, a typowo teoretycznym przypadkiem może być obiekt, który definiowany jest za pomocą zerowego (pustego) zbioru cech.

Na ogół pojęcia obiektu, systemu (S) czy cechy (C) należą do pojęć pierwotnych, intuicyjnie definiowanych za pomocą potocznego języka w kategorii pojęć obiegowych. W badaniach naukowych obiektem może być zarówno wielki system społeczno-gospodarczy, skomplikowany układ techniczno-konstrukcyjny, jak też pojedyncze zjawisko, a także indywidualna osoba. W tym sensie obiekt ma charakter atrybutu jakościowego, natomiast pojęcie cechy raczej utożsamiane jest z pewnymi miarami ilościowymi (Pluta, 1986). Cechami są przykładowo takie kategorie jak: długość, szerokość, wysokość, waga, ciężar, powierzchnia, a także wartość rynkowa, cena sprzedaży, wielkość produkcji, wyposażenie itp. Formalnie cechy opisujące dany obiekt mogą być interpretowane jako odwzorowanie zbioru obiektów S w zbiór liczb rzeczywistych R^+.

$$C\colon S \to R^+ \tag{4}$$

Przy powyższym założeniu każdemu elementowi zbioru obiektów $S_k \in S$ przyporządkowana jest pewien zbiór cech $C_i \in C$, co w ogólności prowadzi do zbudowania następującej macierzy cech $\mathbb{C}$, zwanej powszechnie pod nazwą macierzy cech diagnostycznych, albowiem na ich podstawie będą klasyfikowane (porządkowane) obiekty, będące ich nosicielami:

$$\mathbb{C} = \begin{bmatrix} c_{11} & c_{12} \dots & c_{1i} \dots c_{1I} \\ c_{21} & c_{22} \dots & c_{2i} \dots c_{2I} \\ c_{k1} & c_{k2} \dots & c_{k} \dots c_{kI} \\ c_{K1} & c_{K2} \dots & c_{Ki} \dots c_{KI} \end{bmatrix} \tag{5}$$

Wiersze macierzy cech $C_{k(.)} \in \mathbb{C}$ tzn. wektory $\left[c_{k1,}c_{k2}, \dots c_{ki}, \dots c_{kI}\right]$ będziemy nazywać obrazami obiektów, zaś kolumny $C_{(.)i} \in \mathbb{C}$ tzn. wektory $\left[c_{1i,}c_{2i}, \dots c_{ki}, \dots c_{Ki}\right]$ obrazami cech (Grabiński, Wydymus , Zeliaś, 1983).

Obrazy k-tego obiektu $C_{k(.)} \in \mathbb{C}$ stanowią odwzorowanie Ψ_k, w którym każdemu obiektowi $S_k \in \$$ jest przyporządkowany wektor cech:

$$\Psi_k: S_k \longrightarrow C_{k(.)} = \left[c_{k1,}c_{k2}, \dots c_{ki}, \dots c_{kI}\right] \in \mathbb{C} \subseteq R^{+} \tag{6.1}$$

Wobec tego obraz k-tego obiektu $C_{k(.)} \in \mathbb{C}$ można rozumieć jako synonim wartości – realizacji cech opisujących dany obiekt, który na mocy przekształcenia (6.1) nadaje temu obiektowi określoną wartość liczbową należącą do kategorii liczb rzeczywistych $R_k \in R^{+}$.

Obrazy i-tej cechy $C_{k(.)} \in \mathbb{C}$ stanowią odwzorowanie Ψ^{i}, w którym każdej grupie cech $C_i \in \mathbb{C}$ jest przyporządkowany wektor obiektów:

$$\Psi^{i}: C_i \longrightarrow S_{(.)i} = \left[c_{1i,}c_{2i}, \dots c_{ki}, \dots c_{Ki}\right] \in \$ \subseteq R^{+} \tag{6.2}$$

Wobec tego obraz i-tej cechy$S_{(.)i} \in \$$ można rozumieć jako synonim wartości – realizacji obiektów opisanych za pomocą określonej cechy $C_{k(.)} \in \mathbb{C}$, który na mocy przekształcenia (6.2) nadaje temu obiektowi określoną wartość liczbową należącą do kategorii liczb rzeczywistych $R^{i} \in R^{+}$.

2. Cechy stymulanty, destymulanty i nominanty

Tradycyjnie w teorii wielowymiarowej analizy porównawczej zbiór cech $\mathbb{C}$ dzieli się na trzy zasadnicze kategorie – stymulanty, destymulanty i nominanty (Hellwig, 1968):

$$\mathbb{C} = C^{S} \cup C^{D} \cup C^{N} \text{ przy czym: } C^{S} \cap C^{D} \cap C^{N} = \emptyset \tag{7}$$

Pojęciem stymulanty C^{S}określa się taką cechę, której liczbowy wzrost wartości oceniamy jest pozytywnie, zaś spadek – negatywnie w świetle badanego systemu (procesu). Stymulanta C^{S} wpływa pozytywnie na obraz danego obiektu i powoduje relatywny wzrost jego wartości w przestrzeni pomiarowej.

$$C_{i(.)} \in C^S \Leftrightarrow c_{i,k} \geq c_{i,k+1} \Longrightarrow s_{k,i} \succcurlyeq s_{k+1,i} \tag{8}$$

W powyższym wyrażeniu relacja $s_{k,i} \succcurlyeq s_{k+1,i}$ oznacza, że wyższa wartość i-tej cechy stymulanty $c_{i,k} \geq c_{i,k+1}$ powoduje, że k-ty obiekt dominuje nad obiektem k+1. W efekcie obiekt k-ty uzyskuje lepszą pozycję klasyfikacyjną w stosunku do obiektu k+1 w danej przestrzeni pomiarowej, ze względu na badaną cechę $C_{i(.)} \in C^S$.

Destymulanta oznacza cechę, której bezwzględny wzrost wartości oceniany jest negatywnie, zaś spadek jej wartości pozytywnie z punktu widzenia danego obiektu. Destymulanta wpływa negatywnie na obraz danego obiektu i powoduje zmniejszenie jego wartości w przestrzeni pomiarowej.

$$C_{i(.)} \in C^D \Leftrightarrow c_{i,k} \geq c_{i,k+1} \Longrightarrow s_{k,i} \preccurlyeq s_{k+1,i} \tag{9}$$

W powyższym wyrażeniu relacja $s_{k,i} \preccurlyeq s_{k+1,i}$ oznacza, że wyższa wartość i-tej cechy destymulanty $c_{i,k} \geq c_{i,k+1}$ powoduje, że k-ty obiekt zostaje zdominowany przez obiekt k+1. W efekcie obiekt k-ty uzyskuje gorszą pozycję klasyfikacyjną w stosunku do obiektu k+1 w danej przestrzeni pomiarowej ze względu na badaną cechę $C_{i(.)} \in C^D$.

Ze względu na to, że optymalna wartość stymulanty oznacza jej wartość maksymalną, zaś w przypadku destymulanty rolę tą pełni wartość minimalna, takie typy cech nazywa się niekiedy odpowiednio maksymantą i minimantą (Nowak, 1990).

Wśród cech opisujących wielowymiarowe obiekty złożone mogą występować takie, których zarówno wzrost, jak też i spadek powyżej lub poniżej określonej normy (poziomu) należy uznać za zjawisko niekorzystne. Cechy te powinny przyjmować wartości z pewnego dość ściśle określonego przedziału zmienności. Cecha jest nominantą $C_{i(.)} \in C^N$ jeśli spełniana jest następująca zależność:

$$C_{i(.)} \in C^N \Leftrightarrow c_i^{min} \ll c_{ik} \leq c_i^{max}; \quad k \in \overline{1,K} \tag{10}$$

Podział cech na stymulanty, destymulanty i nominanty ma charakter względny i odnosi się jedynie do liczbowych wartości przyjmowanych w pewnej przestrzeni obrazów obiektów $[c_{k1,}c_{k2}, \dots c_{ki}, \dots c_{kI}]\ k \in \overline{1,K}$. Faktycznie specyfika badanych obiektów determinuje zarówno zbiór cech jak też, ich relatywne wartości względne. Zdarza się niekiedy, że cecha nominanta $C_{i(.)} \in C^N$przyjmuje we wszystkich badanych obiektach wartość większą lub mniejszą do wyznaczonego przedziału zmienności (normy). W praktyce wszystkie kategorie cech można stosunkowo prosto przekształcić do postaci stymulanty (Ficoń, 1995).

Dla każdej destymulanty $C_{i(.)} \in C^D$ istnieje przekształcenie zmieniające ją w stymulantę $C_{i(.)} \in C^S$, zgodnie z poniższą formułą:

$$C^S = \frac{1}{C^D} \quad \text{lub} \quad C^S = maxC^D - C^D \tag{11}$$

Podobnie każdą nominantę można przekształcić w destymulantę za pomocą następującego wyrażenia:

$$C^D = |\, C^S - C^N| \tag{12}$$

Według Nowaka (1990, s.14) „(…) klasyfikacją obiektów nazywamy niepustą rodzinę podzbiorów $< S_1, S_2, \dots, S_l, \dots , S_L > \in S$ spełniającą warunki rozłączności (13.1) i zupełności (13.2)":

$$S_i \cap S_j = \emptyset; \;\; i \neq j \tag{13.1}$$

$$S_1 \cup S_2 \dots \cup S_l, \dots \cup S_L = S \tag{13.2}$$

Klasyfikacja oznacza zatem wyodrębnienie w danym zbiorze takich podzbiorów, które nie mają elementów wspólnych oraz takich, których suma wyczerpuje zbiór klasyfikowany. Szczególnym przypadkiem klasyfikacji jest uporządkowanie, które oznacza przypisanie klasyfikowanym obiektom pewnej miary wartościującej w określonej przestrzeni topologicznej (Duda, 1986). Uporządkowanie wzajemnie relatywizuje poszczególne obiekty, co pozwala na ich liniową prezentację w przyjętej skali porównawczej (rankingowej). Ogólną formułę porządkowania można zapisać jako:

$$S_i \preccurlyeq S_j \cup S_i \succcurlyeq S_j; \;\; i \neq j \tag{14}$$

Podstawą porządkowania wielowymiarowych obiektów jest syntetyczna miara taksonomiczna, określająca ich położenie w pewnej przestrzeni topologicznej charakteryzującej się określoną siatką wymiarów, np. odległości. Klasyfikacja i porządkowanie obiektów wielowymiarowych wymaga określenia ilościowego sposobu pomiaru podobieństwa w sensie przyjętego kryterium, którym najczęściej są odległości topologiczne lub metrykalne (Jajuga, Walesiak, 2011). Najczęściej do realizacji tego celu przyjmuje się metrykę odległości, którą formalnie można zdefiniować w sposób następujący. Metryką pary zbiorów $< S_i, S_j >$ należących do S nazywa się nieujemną funkcję rzeczywistą $H(S_i, S_j)$ spełniającą następujące warunki: tożsamości (15.1), symetrii (15.2), nierówności trójkąta (15.3):

$$H(S_i, S_j) = 0; \quad S_i = S_j \tag{15.1}$$

$$H(S_i, S_j) = H(S_j, S_i); \quad S_i, S_j \in S \tag{15.2}$$

$$H(S_i, S_j) \leq H(S_i, S_k) + H(S_j, S_k); \quad S_i, S_j, S_k \in S \tag{15.3}$$

Przestrzeń z ustaloną metryką nosi nazwę przestrzeni metrykalnej, zaś wartość $H(S_i, S_j)$ nazywa się odległością między badanymi obiektami $< S_i, S_j >$. W tak zdefiniowanej przestrzeni metrykalnej możemy wyznaczać odległości od przyjętych punktów odniesienia, np. wzorca Δ, co prowadzi do wyznaczenia jednowymiarowego wektora odległości dla każdego badanego obiektu – względem wzorca (16.1) lub względem poszczególnych obiektów (16.2):

$$\overrightarrow{h_i} = H(S_i, \Delta) \in [\Delta]; \quad i = \overline{1, I} \tag{16.1}$$

$$\overrightarrow{h_i} = H(S_i, S_j) \in [H]; \quad i \neq j \tag{16.2}$$

Wyznaczając wszystkie odległości poszczególnych obiektów $S_i, S_j \in S$ od obiektu wzorcowego otrzymamy kwadratową macierz $[\Delta]$, natomiast wzajemnie od siebie otrzymamy kwadratową, symetryczną macierz odległości metrykalnych $[H]$. Tak wyznaczone macierze odległości $[\Delta]$ oraz $[H]$ stanowią punkt wyjścia dla badania różnych, wielowymiarowych koncepcji klasyfikacji obiektów w danej przestrzeni metrykalnej.

3. Przykładowa procedura badań taksonomicznych

W celu praktycznego zilustrowania przeprowadzonych powyżej rozważań teoretycznych i sprawdzenia ich zasadności rozpatrzymy prosty przykład liczbowy. Sposób jego rozwiązania zostanie oparty na klasycznej procedurze badania taksonomicznego obejmującej następujące etapy (Nowak, 1990, s.15):

1. Wstępna analiza badanego problemu.
2. Dobór cech diagnostycznych i skal ich pomiaru.
3. Przygotowanie danych statystycznych.
4. Ocena podobieństwa klasyfikowanych jednostek.
5. Wybór metody klasyfikacyjnej.
6. Klasyfikacja obiektów za pomocą wybranej metody.
7. Weryfikacja wyników badań.
8. Merytoryczna interpretacja wyników klasyfikacji.

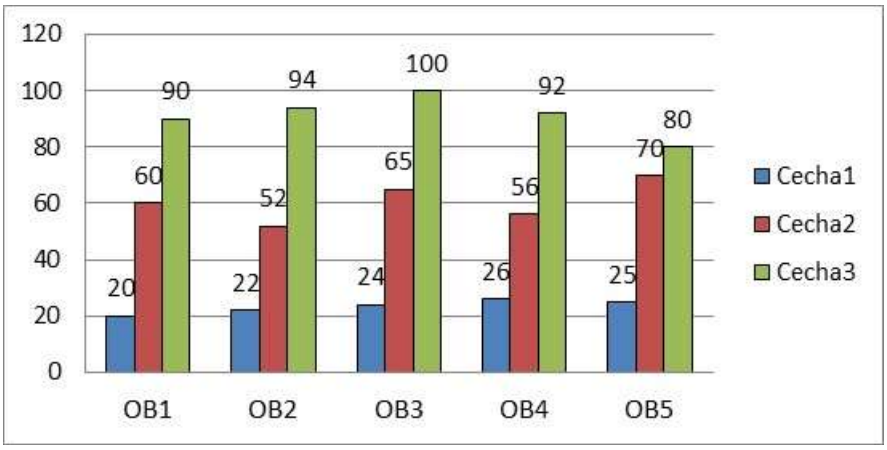
Niech przedmiotem analizy będzie zbiór złożony z 5 obiektów, z których każdy jest przykładowo opisany za pomocą wyselekcjonowanych 3 cech diagnostycznych, spełniających pewne kryteria istotności ze względu na poziom analizowanego zjawiska złożonego (Rys.1.).

Zakładamy, że cechy $C_{1(.)}$ i $C_{2(.)}$ należą do kategorii destymulant $C_{1(.)}, C_{2(.)} \in C^D$, natomiast cecha $C_{2(.)}$ reprezentuje kategorię stymulant $C_{3(.)} \in C^S$:

Tabela 1. Przykładowa realizacja macierzy cech diagnostycznych $C_{ij} \in \mathbb{C}$ dla zbioru obiektów $S_1, S_2, S_3, S_4, S_5 \in S$

	Obiekty	$C_{1(.)} \in C^D$	$C_{2(.)} \in C^D$	$C_{3(.)} \in C^S$
1.	Obiekt 1	20	60	90
2.	Obiekt 2	22	52	94
3.	Obiekt 3	24	65	100
4.	Obiekt 4	26	56	92
5.	Obiekt 5	25	70	80
Średnia arytmetyczna - $\widetilde{C_j}$		23,4	60,6	91,2
Odchylenie standardowe - $\mathfrak{S}_j$		2,15	6,37	6,52

Źródło: Opracowanie własne.



Rys. 1. Macierz przykładowych cech diagnostycznych [C] badanych obiektów

Źródło: Opracowanie własne.

Formalnie powyższą macierz, stanowiącą obraz obiektów $S_1, S_2, S_3, S_4, S_5 \in S$ możemy zapisać w następującej postaci:

$$\mathbb{C} = [C_{ij}]_{3\times 5} \tag{17}$$

Wobec tego każdy obiekt $S_i \in S;\ \ i = \overline{1, I}$ jest scharakteryzowany za pomocą zbioru trzech cech $C_{ij} \in \mathbb{C};\ \ i = \overline{1{,}5},\ j = \overline{1{,}3}$ przy czym, pierwsze dwie cechy $C_{i1}, C_{i2} \in \mathbb{C}$ mają charakter destymulant, co oznacza, że niepożądana jest ich wysoka ich wartość liczbowa. Natomiast cecha stymulanta $C_{i3} \in \mathbb{C}$ powinna osiągnąć możliwie najwyższy poziom, w sensie wartości liczbowej.

Bezpośrednie porównywanie bezwzględnych wartości liczbowych jest niemożliwe ze względu na odmienny charakter stymulant / destymulant oraz z uwagi na różne jednostki miar i rozmaite miana jakie praktycznie mogą być przypisane poszczególnym cechom. Dlatego stosuje się różne sposoby normalizacji (standaryzacji) pierwotnych wartości cech, co pozwala spełnić wymagany od nich warunek addytywności (Ficoń, 1995). Najczęściej stosowaną metodą normalizacji jest ich standaryzacja według następującego wzoru:

$$X_{ij} = \frac{C_{ij} - \widetilde{C_J}}{\mathfrak{G}_j} \tag{18}$$

gdzie: X_{ij} – wystandaryzowane wartości j-tej cechy w i-tym obiekcie:

$$\widetilde{C_J} = \frac{\sum_{j=1}^{J} C_{ij}}{J} \tag{19}$$

$$\mathfrak{G}_j = \sqrt{\frac{\sum_{j=1}^{J} (C_{ij} - \widetilde{C_J})^2}{J}} \tag{20}$$

Po obliczeniu wymaganych wartości średnich (19) oraz odchylenia standardowego (20) otrzymujemy następującą macierz wystandaryzowanych cech diagnostycznych $\mathbb{X}$:

$$[X_{ij}]_{5\times3} = \begin{bmatrix} -1{,}58 & 0{,}09 & -0{,}18 \\ -0{,}65 & -1{,}35 & 0{,}43 \\ 0{,}27 & 0{,}69 & 1{,}35 \\ 1{,}21 & -0{,}72 & 0{,}12 \\ 0{,}74 & -1{,}47 & -1{,}71 \end{bmatrix} = \mathbb{X} \tag{21}$$

Macierz wystandaryzowanych cech diagnostycznych $[X_{ij}]_{5\times3}$ jest podstawą taksonomicznego klasyfikowania i porządkowania badanych obiektów ze względu na poziom interesującego nas zjawiska. Poszczególne obiekty $S_1, S_2, S_3, S_4, S_5 \in S$ możemy interpretować jako punkty w trójwymiarowej przestrzeni topologicznej, których współrzędne są określone przez wystandaryzowane wartości cech diagnostycznych.

4. Konstrukcja obiektu wzorcowego

Stosownie do wymagań za pomocą macierzy $[X_{ij}]_{5\times3}$ możemy określić, jak daleko położone są od siebie poszczególne obiekty lub w jakiej odległości znajdują się od przyjętego obiektu wzorcowego (Hellwig, 1968, Młodak, 2006). W tym przypadku ideałem jest hipotetyczny punkt (obiekt) o współrzędnych równych maksymalnym wartościom stymulant i minimalnym wartościom destymulant. Współrzędne obiektu wzorcowego można określić także za pomocą miar poza statystycznych, wykorzystując do tego celu np., specjalne oceny komisji ekspertów, które posługują się kryteriami heurystycznymi adekwatnymi do obiektów, będących przedmiotem badań i analiz klasyfikacyjnych. Zgodnie z powyższym postulatem współrzędne obiektu wzorcowego $\mathfrak{W} \in S$ wyznaczymy znajdując minimalne wartości w kolumnie 1 i 2 oraz maksymalne wartości w kolumnie 3:

$$\min X_{i1} = \min\{-1{,}58;\ -0{,}65;\ 0{,}27;\ 1{,}21;\ 0{,}74\} = -1{,}58$$

$$\min X_{i2} = \min\{-0{,}09;\ -1{,}35;\ 0{,}69;\ -0{,}72;\ -1{,}47\} = -1{,}47$$

$$\max X_{i3} = \max\{-0{,}18;\ 0{,}43;\ 1{,}35;\ 0{,}12;\ -1{,}71\} = 1{,}35$$

Współrzędne obiektu wzorcowego $\mathfrak{W}$ są więc następujące:

$$[\mathfrak{W}]_{k_1,k_2,k_3} = [-1{,}58;\ -1{,}47,\ \ 1{,}35]$$

Zgodnie z wyrażeniem (21) widzimy, że żaden z 5 klasyfikowanych obiektów nie jest obiektem idealnym, gdyż współrzędne każdego z nich odbiegają w różnym stopniu od współrzędnych obiektu wzorcowego $[-1{,}58;\ -1{,}47,\ 1{,}35]$. Należy zatem obliczyć w jakiej odległości od wzorca znajdują się badane obiekty rzeczywiste $S_1, S_2, S_3, S_4, S_5 \in S$. Do tego celu wykorzystuje się w teorii taksonomii różne metryki odległościowe, oparte najczęściej na ogólnych zasadach geometrii analitycznej. Najbardziej popularne w literaturze (Pluta, 1986, Młodak, 2006, Nowak, 1990) formuły służące do metrykowania odległości topologicznych związane są z takimi nazwiskami jak: metryka Hamninga (miejska), metryka Jeffreysona i Matusita, metryka Braya i Curtisa, metryka Clarka. Równie popularne to odległość „Canbera" czy odległość kątowa (Grabinski, Wydymus, Zielniaś, 1983).

W dalszych rozważaniach skorzystamy z klasycznej metryki euklidesowej wyznaczanej na podstawie formuły:

$$Y_{ik} = \sqrt{\sum_{j=1}^{3}(X_{ik} - X_{ij})^2}; \qquad i = \overline{1{,}5} \tag{22}$$

gdzie: Y_{ik} – odległość i-tego obiektu od obiektu wzorcowego$[\mathfrak{W}]_{k_1,k_2,k_3}$.

Przeprowadzając odpowiednie obliczenia dla wszystkich badanych obiektów według formuły (22) otrzymamy następujące miary odległościowe od obiektu wzorcowego:

$$Y_{1,k} = 2{,}05$$
$$Y_{2,k} = 1{,}31$$
$$Y_{3,k} = 2{,}84$$

$$Y_{4,k} = 3{,}14$$

$$Y_{5,k} = 3{,}82$$

Najmniejszą odległość od obiektu wzorcowego $[\mathfrak{W}]_{k_1,k_2,k_3}$ uzyskał obiekt $S_2 \in S$ ponieważ:

$$\min\{2{,}05; 1{,}31; 2{,}84; 3{,}14; 3{,}82\} = 1{,}31 \in S_2$$

W największej odległości od obiektu wzorcowego$[\mathfrak{W}]_{k_1,k_2,k_3}$ pozostaje obiekt $S_5 \in S$, gdyż:

$$\max\{2{,}05; 1{,}31; 2{,}84; 3{,}14; 3{,}82\} = 3{,}82 \in S_5$$

Przeprowadzona powyżej klasyfikacja obiektów$S_1, S_2, S_3, S_4, S_5 \in S$ ma charakter klasyfikacji porządkującej i pozwala na liniowe uporządkowanie wszystkich badanych obiektów w pewnej, rosnącej lub malejącej kolejności względem taksonomicznych miar odległościowych, w tym przypadku odniesionych do tzw. obiektu wzorcowego $[\mathfrak{W}]_{k_1,k_2,k_3}$.

5. Analiza wyników badań

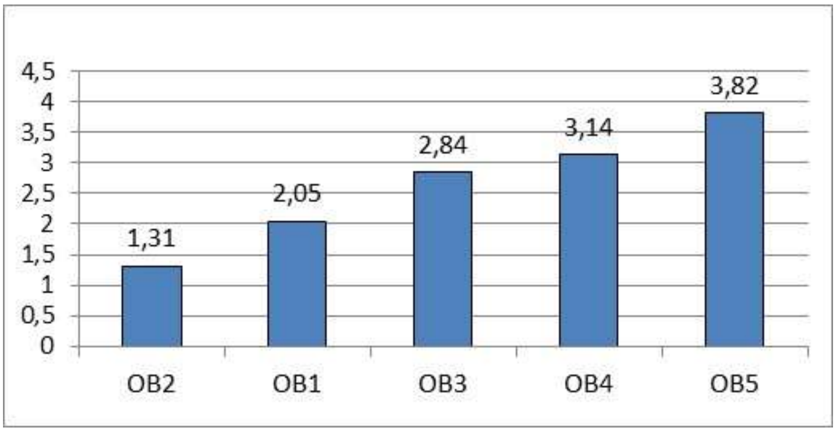
Zgodnie z miarami odległościowymi $Y_{i,k}$ rosnące uporządkowanie obiektów $S_1, S_2, S_3, S_4, S_5 \in S$ przedstawia następujący wektor $\vec{U}$ (Rys.2.):

$$\vec{U} =< S_2, S_1, S_3, S_4, S_5 >$$

Malejący porządek badanych obiektów przedstawia wektor $\overleftarrow{U}$

$$\overleftarrow{U} =< S_5, S_4, S_3, S_1, S_2 >$$

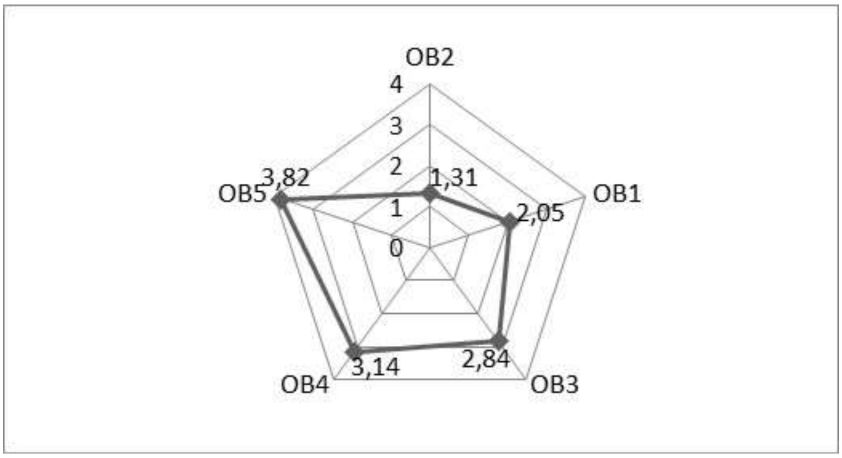
Taksonomiczne uporządkowanie badanych obiektów typu $\vec{U}$ lub $\overleftarrow{U}$ może być szeroko wykorzystywane w praktyce do wartościowania pewnych zbiorowości obiektów rzeczywistych ze względu na przyjęte kryterium ocenowe (Ficoń, Krasnodębski, Szubrycht, 2008).



Rys.2. Wektor odległości badanych obiektów od wzorca

Źródło: Opracowanie własne.

Topologiczne rozproszenie badanej zbiorowości obiektów w 5-wymiarowej przestrzeni odległości od początku układu współrzędnych (0,0,0,0,0) przedstawia tzw. wykres radarowy zobrazowany na rys.3.



Rys. 3. Topologiczne oddalenie badanych obiektów

Źródło: Opracowanie własne.

Dotychczas zajmowaliśmy się wyznaczaniem taksonomicznych miar odległościowych od pewnego obiektu wzorcowego $[\mathfrak{W}]_{k_1,k_2,k_3}$ którego współrzędne metrykalne $< k_1; k_2; k_3 >$ zostały wyznaczone na podstawie ekstremalnych wartości współrzędnych badanych obiektów. Wybór koncepcji wzorca – realny czy hipotetyczny, pozytywny czy negatywny, stały czy zmienny zależy od przedmiotu badań oraz dostępnych informacji merytorycznych i musi być poprzedzony wnikliwą analizą potrzeb jakie stoją przed danym programem badawczym.

Istnieje wiele innych sposobów budowania przestrzeni metrykalnej, niekoniecznie bazujących na obiektach wzorcowych typu $[\mathfrak{W}]_{k_1,k_2,k_3}$ (Jajuga, Walesiak. 2011). Równie popularną przestrzenią metrykalną jest odnoszenie poszczególnych obiektów, np. do początku pewnego układu współrzędnych, w naszym przypadku będzie to punkt $[\mathfrak{O}]_{k_1=0,k_2=0,\,k_3=0}$. Jeszcze innym sposobem jest wzajemne relatywizowanie poszczególnych obiektów względem siebie, co prowadzi do zbudowania kwadratowej macierzy odległości bezpośrednich (Ficoń, 2001). Podstawą konstruowania macierzy $\mathbb{D}$ jest wystandaryzowana macierz cech diagnostycznych (21) oraz ciąg analogicznych formuł, jak w przypadku macierzy odległości od obiektu wzorcowego. W efekcie końcowym uzyskujemy następującą macierz kwadratową:

$$[d_{ij}]_{I\times J} = \begin{bmatrix} 0 & d_{12} & \dots & d_{1K} \\ d_{21} & 0 & \dots & d_{2K} \\ \dots & \dots & \dots & \dots \\ d_{K1} & d_{K2} & \dots & 0 \end{bmatrix} = \mathbb{D}$$

gdzie: $[d_{ij}]_{I\times J}$ – odległość i-tego obiektu od obiektu j-tego,

przy czym: $d_{i=j,j=i} = 0 \cap d_{i,j} = d_{j,i}$

Macierz odległości bezpośrednich $\mathbb{D} = [d_{ij}]_{I\times J}$ jest stosowana w sytuacji, w której niecelowe jest wprowadzanie obiektu wzorcowego, a wzajemne relacje miedzy badanymi obiektami mają jakieś głębsze uzasadnienie badawcze (Ficoń, 1995). Przykładowo w badaniach pionierskich i innowacyjnych trudno jest powoływać się na wzorce, kiedy brak jest jakiekolwiek doświadczenia w tej dziedzinie.

Zaprezentowany powyżej przykład liczbowy dotyczył prostej zbiorowości obejmującej 5 obiektów opisanych za pomocą 3 cech diagnostycznych i miał na celu poglądowe zilustrowanie koncepcji badań taksonomicznych. W ogólności przedmiotem badań są duże K-elementowe populacje opisane za pomocą rozbudowanych N-elementowych charakterystyk w formie złożonych zbiorów cech diagnostycznych. Powstaje wówczas ogromny problem merytorycznego doboru, selekcji i weryfikacji zbioru cech diagnostycznych pod kątem zamierzonych celów i programów badawczych. Zagadnienie doboru cech w badaniach taksonomicznych jest problemem najwyższej rangi i należy do skomplikowanych procedur wstępnych, wspieranych równie rozbudowanym aparatem statystycznym (Nowak, 1990). Równie złożonym problemem jest wybór adekwatnej do rozpatrywanych zagadnień metody badań taksonomicznych, zwłaszcza że ich repertuar ciągle burzliwie się rozwija, głównie w nurcie wielowymiarowej analizy porównawczej. Doskonałym stymulatorem tego rozwoju są osiągnięcia statystyki matematycznej i powszechna dostępność komputerowych programów usprawniających żmudne obliczenia numeryczne (Ficoń, Krasnodębski, 1999).

W ogólnym przypadku praktyczne zadanie klasyfikacji i porządkowania dostatecznie złożonej zbiorowości raczej nie polega na liniowym porządkowaniu poszczególnych elementów (obiektów) tej zbiorowości według pewnego wskaźnika taksonomicznego, a znacznie częściej na wyodrębnieniu pewnych, w miarę jednorodnych podzbiorów (grup) klasyfikacyjnych. Wymaga się, aby obiekty należące do tego samego podzbioru były możliwie najbardziej do siebie podobne – w sensie przyjętej miary podobieństwa, a obiekty znajdujące się w różnych podzbiorach istotnie różniły się od siebie. Ogromne zapotrzebowanie na tego typu metody i klasyfikacje zgłaszają różne sektory gospodarki narodowej, rozmaite programy ekologiczne, a także cała sfera badań socjologicznych i społecznych, a także nauki związane z obronnością i praktyka sztabowa (Krasnodębski, 2001).

Uwagi i wnioski końcowe

1. Podstawowym pojęciem taksonomii jest pojęcie złożonego obiektu wielocechowego, będącego z reguły przedmiotem klasyfikacji i topologicznego porządkowania, rozpatrywanego na tle pewnej zbiorowości statystycznej. Zasadniczym aparatem pojęciowym taksonomii numerycznej jest współczesna topologia i statystyka matematyczna oraz oferowane przez nią liczne procedury badawcze.
2. Badane procesy, obiekty i systemy charakteryzuje się za pomocą rozbudowanych zbiorów cech, opisujących właściwości tych obiektów, które mogą być dowolnie konfigurowane w określone struktury hierarchiczne lub organizacyjno-funkcjonalne.
3. W badaniach obiektów rzeczywistych bezpośrednie wykorzystanie „surowych" cech do badań taksonomicznych, ze względu na ich fizyczne konteksty i różne układy mianowania jest najczęściej niemożliwe i dlatego poddawane są one normalizacji i standaryzacji.
4. Istnieje wiele sposobów definiowania przestrzeni topologicznej, a najbardziej popularny wykorzystuje bazowy obiekt wzorcowy, który może być konstruowany teoretycznie lub zadawany aplikacyjnie, jako najbardziej pożądany obiekt odniesienia.
5. Bardzo efektywne i jednocześnie uniwersalne metody i narzędzia taksonomiczne powinny być szerzej stosowane w badaniach naukowych, a także w praktyce społeczno-gospodarczej. Wysoce perspektywicznym kierunkiem wykorzystania metod taksonomicznych jest wspomaganie procesów prognozowania, programowania i planowania społeczno-gospodarczego.
6. Z uwagi na skomplikowane i żmudne operacje statystyczno-numeryczne efektywne wykorzystanie założeń teoretycznych i proponowanych metod taksonomicznych jest uzależnione od stopnia komputeryzacji bardzo licznych czynności statystyczno-arytmetycznych.
7. Zaprezentowany prosty przykład liczbowy posłużył do praktycznej ilustracji założeń teoretycznych i mechanizmu działania wybranej metody i równocześnie potwierdził ich poprawność i zasadność.
8. Metody taksonomiczne znajdują szerokie zastosowania w wielu dziedzinach nauki i praktyki społecznej czy gospodarczej, jako efektywne narzędzia badania złożonych, wielowymiarowych procesów i obiektów, podlegających wzajemnej relatywizacji na tle pewnej zbiorowości.

9. Na tym polu szczególne osiągnięcia odnotowali polscy uczeni i wykreowane przez nich szkoły naukowe, takie jak np. słynna taksonomia wrocławska czy szkoła krakowska. Dużymi osiągnięciami, głównie w taksonomii numerycznej legitymują się przykładowo takie nazwiska jak: Z. Hellwig, T. Grabiński, S. Wydymus czy A. Zeliaś.

Literatura

[1] Duda R. (1986), *Wprowadzenie do topologii*, PWN Warszawa.

[2] Ficoń K. (2006), *Badania operacyjne stosowane. Modele i aplikacje*, BEL Studio Warszawa.

[3] Ficoń K., Krasnodębski G. (1999), *Komputerowy system analizy i oceny potencjału bojowego sił morskich SI "Expert"*. AMW-DMW Gdynia.

[4] Ficoń K., Krasnodębski G., Szubrycht T. (2008), *Wykorzystanie metod taksonomicznych do określenia potencjału morskiego państwa*. [w:] Gry i symulacje jako przedmiot i metoda badań w naukach społecznych, Wyd. Collegium Civitas, Warszawa.

[5] Ficoń K. (1995), *Propozycja jednostkowego wskaźnika jakości w badaniach taksonomicznych*. ZN AMW nr 2 (125), 1995.

[6] Ficoń K. (1996), *Syntetyczny wskaźnik kwantyfikacji potencjału bojowego okrętu*. ZN AMW nr 1/96, 1996.

[7] Ficoń K. (2001), *Taksonomiczne modele złożonych systemów broni*. ZN AMW Wewn. 1013/2001.

[8] Ficoń K. (1995), *Zmodyfikowana metoda normowania cech diagnostycznych w badaniach taksonomicznych*. ZN AMW nr 2/95.

[9] Grabiński T., Wydymus S., Zeliaś A. (1983), *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWE, Warszawa.

[10] Hellwig Z. (1968), *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju*, Przegląd Statystyczny XV z. 4 Warszawa.

[11] Jajuga K., Walesiak M. (red.) (2011), *Taksonomia 18. Klasyfikacja i analiza danych - teoria i zastosowania*, Wyd. UE Wrocław.

[12] Krasnodębski G. (2001), *Modelowanie potencjałów bojowych sił morskich za*

pomocą metod eksperckich, AMW Gdynia.

[13] *Leksykon Naukowo-Techniczny*, (1984), WN-T, Warszawa 1984.

[14] Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, Difin Warszawa.

[15] Nowak E. (1990), *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE Warszawa.

[16] Pluta W. (1986), *Wielowymiarowa analiza porównawcza w modelowaniu ekonometrycznym*, PWN Warszawa.