

## KLASYFIKACJE ROZMYTE W MODELOWANIU ZJAWISK EKONOMICZNYCH

AGATA WAWRZYNIAK  
Uniwersytet Szczeciński

### Streszczenie

*W artykule przedstawiono problematykę klasyfikacji danych źródłowych w modelowaniu zjawisk ekonomicznych, zarówno w ujęciu klasycznym, jak i nieklasycznym (klasyfikacja rozmyta). Zaprezentowane w artykule wyniki badań dotyczą klasyfikacji gmin województwa zachodniopomorskiego ze względu na poziom korzystania z outsourcingu informatycznego w jednostkach samorządu terytorialnego. Celem artykułu jest wykazanie, że przyjęta metoda klasyfikacji rozmytej pozwala na pełniejsze odwzorowanie badanego zjawiska.*

**Słowa kluczowe:** klasyfikacja rozmyta, algorytm rozmytych c-średnich, algorytm FCM, usługi informatyczne, outsourcing informatyczny, jednostka samorządu terytorialnego.

### 1. Wprowadzenie

Kryzys gospodarczy jaki obecnie obserwujemy, zmiany społeczne, rynkowe oraz technologiczne, coraz częściej o zasięgu globalnym, zmuszają organizacje do ciągłego uaktualniania strategii i wypracowywania nowych sposobów funkcjonowania. Aby sprostać konkurencji, zredukować koszty, zwiększyć wydajność i przeorientować krytyczne obszary działalności muszą one poszukiwać nowoczesnych rozwiązań. Z drugiej strony uświadamiają sobie, że biorąc pod uwagę ograniczenia ich wewnętrznych zasobów nie są w stanie przeprowadzić we wszystkich realizowanych funkcjach. W tej sytuacji korzystne jest dla nich zwrócenie się do wyspecjalizowanych firm zewnętrznych i przekazanie im do realizacji pewnych funkcji lub procesów dotąd wykonywanych we własnym zakresie. Ten sposób działania określa się mianem profesjonalnych usług zewnętrznych lub krótko – outsourcingu. Outsourcing polega na „oddaniu na zewnątrz” zadań niezwiązanych bezpośrednio z podstawową działalnością organizacji, jaką może być na przykład jednostka samorządu terytorialnego. Dla każdego podmiotu najważniejsze jest realizowanie misji i skoncentrowanie się na strategii. Outsourcingiem można objąć funkcje pomocnicze, wspierające strategię. W przypadku urzędu gminnego takim obszarem działalności, który można oddać w outsourcing, jest między innymi obsługa informatyczna. Wstępna analiza problemu wykazała, że w przypadku większości usług informatycznych świadczonych przez firmy zewnętrzne wybór ich zakresu opiera się na nie w pełni sprecyzowanych przesłankach. Często jedynym kryterium wyboru oferty jest tylko cena usługi. Fakt ten skłonił autorkę artykułu do zajęcia się tematem outsourcingu informatycznego w odniesieniu do jednostek samorządu terytorialnego. Zaobserwowano tu istnienie luki badawczej – jak dotąd wszelkie opracowania dotyczące outsourcingu IT dotyczyły jedynie przedsiębiorstw. Podstawowymi przesłankami podjęcia badań były zmiany w podejściu do prowadzenia działalności przez organizacje samorządowe (coraz częściej decydenci dochodzą do wniosku, że nie wszystko trzeba realizować samodzielnie, można korzystać z usług specjalistów – jest to efektywniejsze), skala, powszechność i powtarzalność korzystania z usług outsourcingowych oraz brak

sprecyzowanych kryteriów oceny czy dla konkretnej organizacji korzystanie z outsourcingu jest opłacalne.

W związku z tym autorka przeprowadziła analizę rynku w zakresie wyboru profesjonalnych usług informatycznych (outsourcingu IT) w jednostkach samorządowych województwa zachodniopomorskiego. Aby zrealizować cel podjętych badań przyjęto trój etapową procedurę badawczą. W pierwszym etapie dokonano analizy strukturalnej stanu usług informatycznych w jednostkach samorządowych. W badaniu tym zastosowano metodę wywiadu kwestionariuszowego i przeprowadzono statystyczną analizę zebranych danych. Celem drugiego etapu procedury badawczej było stwierdzenie istnienia prawidłowości i podobieństw ze względu na korzystanie z outsourcingu na poziomie gmin. Dokonano tego poprzez klasyfikację gmin ze względu na stan ich informatyzacji oraz rodzaj stosowanego outsourcingu informatycznego. Ostatni, trzeci etap przyjętej procedury miał na celu znalezienie ukrytych, niejawnych reguł, które legły u podstaw wyboru outsourcingu w gminach. Aby móc dokonać identyfikacji tych reguł zastosowano teorię zbiorów przybliżonych.

Celem niniejszego artykułu jest przybliżenie problematyki klasyfikacji danych źródłowych w modelowaniu zjawisk ekonomicznych, zarówno w ujęciu klasycznym, jak i nieklasycznym (klasyfikacja rozmyta). Zaprezentowane w artykule wyniki badań dotyczą klasyfikacji gmin województwa zachodniopomorskiego ze względu na poziom korzystania z outsourcingu informatycznego w jednostkach samorządu terytorialnego (drugi etap procedury badawczej). Na podstawie otrzymanych wyników stwierdzono, że przyjęta metoda klasyfikacji rozmytej pozwala na pełne odwzorowanie badanego zjawiska, a otrzymane, w wyniku grupowania, podziały badanej zbiorowości posłużyły w dalszej części badania do stworzenia wzorca obsługi informatycznej organizacji samorządowych.

## 2. Pojęcie klasyfikacji oraz charakterystyka metod

**Klasyfikacja** to podział zupełny danego zbioru na pewną liczbę rozłącznych podzbiorów. Przedmiotem klasyfikacji są zbiory obserwacji – obiektów opisanych zwykle wieloma cechami zarówno mierzalnymi (ilościowymi), jak i niemierzalnymi (jakościowymi) [12]. Głównym celem klasyfikacji jest badanie podobieństwa lub odrębności obiektów i ich zbiorów, a więc podział zbioru na klasy zawierające obiekty podobne ze względu na obserwacje na zmiennych (tzw. **klasy względnie jednorodne**). Ponadto obiekty znajdujące się w różnych klasach powinny być jak najmniej podobne. Postuluje się, aby wyodrębnione klasy spełniały dwa kryteria: **wewnętrznej spójności i zewnętrznej izolacji** [2].

W literaturze przedmiotu i języku potocznym zamiast określenia *klasyfikacja* używa się zamiennie takich terminów, jak grupowanie, podział, delimitacja, taksonomia, taksonomia numeryczna, taksonometria, analiza skupień, klasteryzacja czy identyfikacja. Ta różnorodność terminologiczna wynika z tego, że metody klasyfikacji są tworzone i stosowane przez przedstawicieli różnych dyscyplin badawczych. Nie wnikając szczegółowo w zakres pojęciowy wymienionych terminów, autorka artykułu używa terminu klasyfikacja (w odniesieniu do ogólnego pojęcia) i klasteryzacja (w odniesieniu do metody klasyfikacji zastosowanej w badaniach).

W modelowaniu zjawisk ekonomicznych klasyfikacja danych jest procesem, który przypisuje obiekty posiadające wspólne cechy do określonych klas. Jest ona jedną z podstawowych operacji wykonywanych na danych źródłowych zgromadzonych w bazach danych. Bazy te stanowią często ogromne zbiory danych stąd też wynika duże znaczenie odpowiednich, wydajnych metod przeprowadzania klasyfikacji tych danych [11]. Prawidłowo przeprowadzona klasyfikacja pomaga

odkryć charakterystyki danych w sposób zrozumiały dla użytkownika, uogólniać dane lub też organizować dane zgodnie z zakładanymi strukturami zorientowanymi na wiedzę.

Podstawowe przesłanki, z których wynika konieczność przeprowadzania klasyfikacji, to przede wszystkim [3]:

1. zredukowanie dużej ilości nagromadzonych informacji do kilku podstawowych kategorii, które mogą być traktowane jako przedmiot dalszej analizy (przesłanka poznawcza),
2. określenie jednorodnych grup obiektów, ze względu na charakteryzujące je właściwości, co ułatwia wyodrębnienie czynników systematycznych oraz ustalenie zasadniczych związków przyczynowo-skutkowych (przesłanka metodologiczna),
3. zmniejszenie nakładów czasu i kosztów badania przez ograniczenie rozważań do najbardziej typowych faktów, zjawisk, obiektów przy stosunkowo niewielkim zniekształceniu wyników badań i stratach informacji (przesłanka ekonomiczna).

Nawet pobieżny przegląd literatury przedmiotu pokazuje, że istnieją setki metod klasyfikacji i nie jest sprawą prostą dokonanie ich systematyzacji. Szczegółowe rozważania na ten temat można znaleźć w pracy P.H. Sneatha i R.R. Sokala [16]. Nie istnieje jednak jeden rozłączny podział tych metod, uwzględniający choćby większość propozycji (godne uwagi monografie i artykuły traktujące o metodach klasyfikacji wymienione są w pracy [1]).

Rozpatrując problem klasyfikacji z perspektywy sposobu jej przeprowadzania, można wyróżnić dwa podstawowe jej rodzaje – klasyfikację wykluczającą się i klasyfikację nie wykluczającą się (por. rysunek 1). **Klasyfikacja wykluczająca się** (ang. *exclusive classification*) jest takim podziałem zbioru obiektów, w którym każdy obiekt należy do jednego i tylko jednego podzbioru (klasy). **Klasyfikacja niewykluczająca się** (ang. *non-exclusive* lub *overlapping classification*) może przydzielać obiekt do kilku klas. Wśród klasyfikacji wykluczających się wyróżnia się dwa typy – klasyfikację z nadzorem i bez nadzoru. **Klasyfikacja bez nadzoru** (ang. *unsupervised* lub *intrinsic classification*), nazywana także klasyfikacją bezwzorcową, analizą skupień, klasteryzacją (ang. *clustering*), wykorzystuje jedynie macierz odległości do przeprowadzenia klasyfikacji. Z kolei **klasyfikacja z nadzorem** (klasyfikacja wzorcowa, ang. *supervised* lub *extrinsic classification*) wykorzystuje zarówno macierz odległości, jak i etykiety kategorii (do wyznaczenia pierwotnego podziału). Problemem, który pojawia się potem jest wyznaczenie przestrzeni dyskryminacji, która rozdziela obiekty według kategorii. Innymi słowy, klasyfikacja z nadzorem wykorzystuje kategorie jako „nauczyciela” (stąd inna jej nazwa – klasyfikacja z nauczycielem), natomiast klasyfikacja bez nadzoru ma do dyspozycji jedynie macierz odległości.

Należy w tym miejscu odnotować, że pewna część autorów nie uznaje klasteryzacji za rodzaj klasyfikacji, lecz traktuje te dwa pojęcia odrębnie (por. [9], [11]). Jednak autorka artykułu, opierając się na dostępnych źródłach, przyjmuje postulat traktowania klasteryzacji za typ klasyfikacji (używając tych pojęć zamiennie) i opiera się na systematyce klasyfikacji podanej powyżej.

Klasyfikację bez nadzoru można podzielić na hierarchiczną (ang. *hierarchical classification*) i niehierarchiczną (ang. *nonhierarchical classification*). **Hierarchiczne metody klasyfikacji** prowadzą do wyodrębnienia pełnej hierarchii skupień (klas) z monotonicznie wzrastającym współczynnikiem ich podobieństwa. We wszystkich metodach hierarchicznych możliwe jest graficzne przedstawienie wyników podziału w postaci tzw. dendrogramu (drzewa połączeń), który ilustruje kolejne połączenia skupień coraz wyższego rzędu. Uzyskana hierarchia pozwala na określenie wzajemnego położenia skupień i obiektów w nich zawartych. Metody hierarchiczne należą do najpopularniejszych metod klasyfikacji i najlepiej opracowanych pod względem metodologicznym (por. [1, 12]). Istotną wadą tych metod jest brak oczywistego kryterium stopu dla ustalenia liczby

skupień względnie jednorodnych oraz w pewnych przypadkach tendencję do tworzenia skupień w postaci łańcucha, a więc do powstawania skupień dość odległych od siebie [12].



Rys. 1. Systematyka klasyfikacji

Źródło: Opracowanie własne na podstawie [2, 5, 9, 11, 12].

Oprócz metod hierarchicznych rozwinęły się również **niehierarchiczne metody klasyfikacji**. Metody te dają niehierarchicznie ułożone skupienia. Skupienia wyodrębnione na drodze analizy niehierarchicznej mogą być rozłączne lub mieć wspólne obiekty. Podstawową wadą klasyfikacji niehierarchicznej jest to, że informacja o strukturze danych uzyskiwana na podstawie skupień niehierarchicznych jest uboższa w porównaniu z informacją uzyskaną dzięki metodom hierarchicznym. Z drugiej strony klasyfikacja niehierarchiczna ma przewagę nad hierarchiczną w zbiorach zawierających duże ilości danych, dla których tworzenie dendrogramu byłoby procesem bardzo złożonym obliczeniowo. Wśród wielu stosowanych metod niehierarchicznych do najpopularniejszych można zaliczyć algorytmy: *k*-średnich (ang. *k-means*) i rozmytych *k*-średnich (lub rozmytych *c*-średnich (ang. *fuzzy c-means*, FCM). Algorytm rozmytych *k*-średnich zostanie szczegółowo omówiony w dalszej części niniejszego artykułu, ze względu na to, że metoda ta została użyta przez autorkę w opisywanych badaniach. Jednak aby pojęcia stosowane do opisu tego algorytmu były wystarczająco jasne, należy przedtem przybliżyć zagadnienia związane z zastosowaniem teorii zbiorów rozmytych w klasyfikacji.

### 3. Teoria zbiorów rozmytych w zagadnieniach klasyfikacji

Z dotychczasowych rozważań dotyczących pojęcia klasyfikacji wynika, że klasyfikacja klasyczna jest oparta na założeniu istnienia dychotomii: dany element należy do określonego zbioru lub nie należy. Takie podejście jest równoznaczne z założeniem istnienia dwuwartościowej **funkcji przynależności**, której argumentami są elementy zbioru. Funkcja ta przyjmuje wartość równą jedności, jeśli element należy do zbioru, wartość równą zero – jeśli nie należy [4].

W praktyce często występują jednak trudności związane z jednoznacznym przyporządkowaniem obiektu do określonej klasy. Dzieje się tak, gdy dany obiekt jest bardzo podobny do obiektów znajdujących się w różnych klasach. W takiej sytuacji granice między wydzielonymi klasami są nieostre, rozmyte i postulat różnicowania między obiektami różnych klas nie jest spełniony [7, 10]. Problemy te spowodowały zainteresowanie badaczy możliwością zastosowania w klasyfikacji

elementów teorii zbiorów rozmytych, której twórcą jest L.A. Zadeh. Podstawowym pojęciem tej teorii jest **zbiór rozmyty**, o następującej definicji: zbiór rozmyty  $A$  w pewnej przestrzeni (obszarze) rozważań  $X=\{x\}$ , co zapisujemy  $A\subset X$ , jest to zbiór uporządkowanych par:  $A=\{(\mu_A(x),x)\}$ ,  $\forall x\in X$ , gdzie:  $\mu_A:X\rightarrow[0,1]$  jest funkcją przynależności, która każdemu elementowi przestrzeni  $X$  przyporządkowuje stopień przynależności do danego zbioru rozmytego: od nieprzynależności ( $\mu_A(x)=0$ ), przez przynależność częściową ( $0<\mu_A(x)<1$ ), do całkowitej przynależności ( $\mu_A(x)=1$ ) [6, 8, 13]. Na zbiorach rozmytych, jako uogólnieniu zbiorów zwykłych, można zdefiniować szereg operacji matematycznych, które są uogólnieniem podobnych operacji obowiązujących dla zbiorów nierozmytych. Podobnie jak dla zbiorów w zwykłym sensie, tak i dla zbiorów rozmytych słuszne są: prawa przemienności sumy i iloczynu, prawa rozłączności sumy i iloczynu, prawa rozdzielnosci sumy względem iloczynu i iloczynu względem sumy oraz prawa de Morgana [6].

W ramach pojęć teorii zbiorów rozmytych zagadnienie klasyfikacji może zostać uogólnione na tzw. **klasyfikację rozmytą**. Można je przedstawić następująco: Niech zbiór  $\Omega$  zawiera  $N$  obiektów  $O_1, O_2, \dots, O_N$ . Każdy z tych obiektów jest opisywany przez wartości cech diagnostycznych  $X_1, X_2, \dots, X_K$ , tak że dane są obserwacje  $x_{ik}$  ( $i=1, 2, \dots, N, k=1, 2, \dots, K$ ). Zadanie klasyfikacji sprowadza się do określenia na elementach zbioru  $\Omega$  rodziny zbiorów rozmytych (klas, grup)  $A_1, A_2, \dots, A_p$  (gdzie liczba  $P$  jest dana i  $1 < P < N$ ) tak, aby były spełnione warunki [6, 10]:

1.  $0 \leq \mu_{ip} \leq 1, (i=1, 2, \dots, N, p=1, 2, \dots, P)$ ,
2.  $\sum_{p=1}^P \mu_{ip} = 1, (i=1, 2, \dots, N)$ ,
3. obiekty, dla których wartości funkcji przynależności do tego samego zbioru rozmytego są duże – są bardzo podobne do siebie, natomiast obiekty odznaczające się dużymi stopniami przynależności do różnych klas – są mało do siebie podobne.

Wartość  $\mu_{ip}$  oznacza stopień przynależności obiektu  $O_i$  do zbioru rozmytego  $A_p$ :  $\mu_{ip} = \mu_{A_p}(O_i)$ , co można interpretować jako podobieństwo obiektu  $O_i$  do klasy rozmytej  $A_p$ . W klasyfikacji rozmytej obiekt należy do różnych klas z różnymi stopniami przynależności. Klasy rozmyte w wyniku operacji sumowania algebraicznego dają zbiór  $\Omega$ .

Jak już wspomniano wcześniej, w klasyfikacji w zwykłym sensie należenie do klasy jest zmienną zerowydkową (1 – obiekt należy do danej klasy, 0 – obiekt nie należy do danej klasy). Natomiast w klasyfikacji rozmytej przynależność do klasy jest zmienną ciągłą, która może przyjmować wartości z przedziału  $\langle 0, 1 \rangle$ . Funkcja przynależności przybiera wartości bliższe skrajnym 0 i 1, gdy istnieją „ostre” granice między klasami.

Pewną niedogodnością metod klasyfikacji rozmytej jest założenie, że liczba klas jest z góry znana. W badaniach praktycznych takiej sytuacji nie spotyka się zbyt często. Zazwyczaj w metodach klasyfikacji rozmytej liczba klas jest ustalana przez badacza, który dysponuje wiedzą merytoryczną o badanym zjawisku i potrafi zwykle określić pewien niezbyt szeroki przedział możliwych wartości  $K$  (czyli liczby klas) [14].

#### 4. Klasyfikacja rozmyta gmin woj. zachodniopomorskiego ze względu na poziom korzystania z outsourcingu

Jak wspomniano we wprowadzeniu do niniejszego artykułu, głównym celem badania była analiza rynku w zakresie wyboru profesjonalnych usług informatycznych (outsourcingu IT) w jednostkach samorządowych województwa zachodniopomorskiego (badaniem objęte były

wszystkie urzędy gminne województwa). W związku z tym opracowany kwestionariusz wywiadu koncentrował się na stopniu informatyzacji urzędu, liczbie zatrudnionych informatyków, szczególnych danych dotyczących usług informatycznych, z których korzysta urząd oraz określeniu czynników decydujących o wyborze usługodawcy i ocenie jakości świadczonych przez niego usług. Kwestionariusz liczył 22 pytania i był podzielony na dwie części. Pierwsza część kwestionariusza zawierała pytania dotyczące ogólnej charakterystyki urzędu. Pytania te odnosiły się do stopnia informatyzacji urzędu oraz liczby zatrudnionych specjalistów z zakresu informatyki. Celem ich było także sprawdzenie czy urząd korzysta z usług podmiotów zewnętrznych w zakresie obsługi informatycznej (tj. czy występuje w urzędzie outsourcing informatyczny). Pytania zawarte w drugiej części kwestionariusza miały na celu umożliwienie analizy urzędów gminnych pod względem stosowania outsourcingu informatycznego w ich obsłudze. Po wyselekcjonowaniu z badanej zbiorowości grupy gmin, które korzystają z usług informatycznych świadczonych przez podmiot lub podmioty zewnętrzne, przystąpiono do badania cech charakteryzujących te usługi, wpływu stosowania obsługi zewnętrznej na działanie urzędów oraz oceny jakości usług przez respondentów reprezentujących jednostki samorządowe. W odniesieniu do tej części kwestionariusza liczebność badanej zbiorowości wyniosła 108, w pozostałych 6 jednostkach respondenci wykluczyli występowanie obsługi informatycznej ze strony podmiotów zewnętrznych.

Dane pozyskane w ramach badania ankietowego miały zróżnicowany charakter, były to zarówno dane liczbowe, jak i słowne. Zostały one w pierwszym etapie procedury badawczej uporządkowane w postaci elektronicznej tablicy i posłużyły do przeprowadzenia analizy strukturalnej stanu usług informatycznych w jednostkach samorządowych województwa zachodniopomorskiego (zastosowano tutaj standardowe metody statystyczne).

Zgodnie z przyjętą procedurą badawczą klasyfikacja rozmyta została przeprowadzona w etapie drugim. Celem tego etapu procedury było stwierdzenie istnienia prawidłowości i podobieństw ze względu na korzystanie z outsourcingu informatycznego na poziomie jednostek samorządowych województwa zachodniopomorskiego. Jako metodę klasyfikacji rozmytej w badaniu zastosowano metodę c-środków (FCM).

Algorytm rozmytych c-średnich (inaczej nazywany również metodą c-środków, ang. *Fuzzy C-Means*, FCM) pozwala na przypisanie tych samych obiektów do różnych klas (grup) z odpowiednimi stopniami przynależności. Algorytm FCM jest najczęściej stosowanym algorytmem klasyfikacji (klasteryzacji) rozmytej [13, 15]. Idea tej metody została opracowana przez J.C. Dunna w 1973 r., a następnie uogólniona w 1981 r. przez J. Bezdeka, którego uznaje się powszechnie za jej twórcę. Dzięki tej metodzie możliwe jest wykrycie grup o prototypach będących punktami w przestrzeni danych. Wszystkie grupy mają taki sam kształt zależny od przyjętej z góry normy, gdyż algorytm nie ma możliwości dostosowania macierzy  $A$  do istniejących danych [15]. Algorytm ten wyprowadza się poprzez minimalizację kryterium:

$$J(X; U, V) = \sum_{i=1}^c \sum_{k=1}^M (\mu_{ik})^m \|x_k - v_i\|_A^2,$$

przy czym  $U = [\mu_{ik}] \in Z_2$  jest macierzą podziału zbioru  $X$ , natomiast  $V = [v_1, v_2, \dots, v_c]$  jest wektorem środków, które mają być określone w wyniku działania algorytmu,  $v_i \in R^n$ ,  $i = 1, 2, \dots, c$ . Występujący we wzorze człon:

$$\|x_k - v_i\|_A^2 = (x_k - v_i)^T A (x_k - v_i)$$

pozwala obliczyć odległość między wektorem  $x_k$  i środkiem grupy  $v_i$ , a  $m \in (1, \infty)$  jest współczynnikiem mówiącym o stopniu rozmycia powstałych grup danych. Gdy  $m \rightarrow 1$ , podział staje się coraz bardziej ostry. Gdy  $m \rightarrow \infty$ , podział staje się coraz bardziej rozmyty (wówczas  $\mu_{ik} = 1/c$ ). W praktyce przyjmuje się wartość  $m=2$ . W celu realizacji algorytmu, mając dany zbiór danych  $X$ , należy wybrać liczbę klas  $c$ , stopień rozmycia  $m$ , parametr  $\varepsilon$  w kryterium zatrzymania algorytmu oraz zainicjować losowo macierz  $U^{(0)} \in Z_2$  i wektor prototypów klas  $V^{(0)}$ . Warunkiem zatrzymania algorytmu jest najczęściej odpowiednio mała zmiana wartości elementów macierzy  $U$ , czyli  $\|U^{(t+1)} - U^{(t)}\| < \varepsilon$ , gdzie  $\varepsilon$  jest ustaloną stałą. Alternatywnie można sprawdzać zmianę położenia środków grup, to znaczy  $\|V^{(t+1)} - V^{(t)}\| < \varepsilon$ . Kształt grup zależy od przyjętej miary odległości.

Uwagi dotyczące metody rozmytych  $c$ -średnich:

- uzyskane rozwiązanie jest optymalne (suboptymalne) w sensie lokalnym i uzależnione od parametrów inicjalizujących. Należy powtarzać klasteryzację startując z różnych zestawów parametrów;
- jeżeli dobrać się właściwą liczbę  $c$  klastrów, a dane posiadają dobrą strukturę klasteryzacyjną, to metoda  $c$ -średnich generuje zazwyczaj podobne klaster (tzw. stabilność rozwiązań);
- poszukiwanie właściwej liczby  $c$  klastrów może odbywać się metodą prób i błędów. Można także odrzucać klaster mniej istotne na podstawie oceny ich istotności, stosując takie miary jak: liczba kardynalna  $Card_i$ , stopień inkluzji (włączenia)  $f$  próbek przez klaster  $i$ , zasadność  $Val_i$  klastrera (por. [13]).

W ramach przeprowadzonych przez autorkę badań została dokonana, z zastosowaniem algorytmu rozmytych  $c$ -średnich (FCM), klasyfikacja rozmyta jednostek samorządowych w województwie zachodniopomorskim ze względu na poziom korzystania z zewnętrznych usług informatycznych (outsourcing IT). Jak wspomniano wyżej jego zadaniem jest podział zbioru danych na określoną liczbę grup (jest to zadany przez użytkownika parametr). Metoda FCM została użyta w badaniu dwukrotnie, po pierwsze do przeprowadzenia klasyfikacji ogólnego stanu informatyzacji urzędów gminnych, po drugie – do klasyfikacji rodzajowej outsourcingu informatycznego. Do dokonania klasyfikacji w oparciu o ten algorytm użyto biblioteki narzędziowej „Fuzzy Logic Toolbox” działającej w środowisku Matlab 6.0 firmy The MathWorks.

Na podstawie odpowiedzi udzielonych na pytania dotyczące ogólnej sytuacji w urzędzie w zakresie stopnia informatyzacji i zatrudnienia specjalistów z tego zakresu, 108 gmin zostało podzielonych, za pomocą algorytmu FCM, na sześć klas o liczebności od 7 do 29 gmin (taka liczba klas jest wynikiem szeregu eksperymentów przeprowadzonych przez autorkę, mających na celu uzyskanie tzw. stabilności rozwiązań). W sytuacji, gdy dana gmina należała do dwóch zbiorów rozmytych, odwzorowujących dwie różne klasy, zostawała ona przyporządkowana do klasy, dla której stopień przynależności był wyższy. Dodatkowo pytania zostały skumulowane w jeden atrybut warunkowy ( $Q = \{q\}$ ), przyjmujący 6 różnych wartości ( $q = 1, 2, 3, 4, 5, 6$ ), czyli tyle ile klas. Gminy w tych klasach różnią się liczbą pracowników, komputerów, liczbą zatrudnionych informatyków oraz faktem istnienia lub nie wyodrębnionych komórek organizacyjnych odpowiedzialnych za informatyzację danego urzędu.

W **klasie 1**, liczącej 29 gmin, znalazły się urzędy zatrudniające największą liczbę pracowników (od 44 do 300 pracowników) i wyposażone w największą liczbę komputerów (od 38 do 300 komputerów). Urzędy należące do tej klasy zatrudniają przeważnie 1 lub 2 informatyków i w większości nie mają wyodrębnionej komórki organizacyjnej odpowiedzialnej za informatyzację urzędu.

W **klasie 2** (24 gminy) znalazły się urzędy zatrudniające najmniej pracowników (w większości są to urzędy zatrudniające od 12 do 19 pracowników) i wyposażone w najmniejszą liczbę komputerów (od 7 do 22 komputerów). Urzędy należące do klasy 2 nie zatrudniają żadnych etatowych informatyków i w większości nie mają wyodrębnionej komórki organizacyjnej odpowiedzialnej za informatyzację urzędu.

W **klasie 3** (7 gmin) znalazły się urzędy zatrudniające niewielu pracowników (od 20 do 24 osób) i wyposażone w niedużą liczbę komputerów (od 23 do 37 komputerów), należy jednak zauważyć, że w tej klasie znajdują się urzędy, w których jest tyle samo lub więcej komputerów niż zatrudnionych urzędników. Urzędy należące do klasy 3 zatrudniają etatowych informatyków, ale nie mają wyodrębnionych komórek organizacyjnych odpowiedzialnych za informatyzację urzędu.

W **klasie 4** (19 gmin) znalazły się urzędy średniej wielkości zatrudniające od 25 do 43 osób. Należy zauważyć, że w tej klasie w większości urzędów jest mniej komputerów niż pracujących w nich urzędników. Urzędy należące do klasy 4 zatrudniają specjalistów z zakresu informatyki (1 etatowy informatyk), ale nie mają wyodrębnionych komórek organizacyjnych odpowiedzialnych za informatyzację urzędu.

Urzędy należące do **klasy 5**, liczącej 16 gmin, są niejednorodne pod względem liczby pracowników i komputerów. Większość urzędów należących do klasy 5 nie zatrudnia etatowych specjalistów z zakresu informatyki i nie posiadają one wyodrębnionych komórek organizacyjnych odpowiedzialnych za informatyzację urzędu.

W **klasie 6** (13 gmin) przeważają urzędy zatrudniające niewielu pracowników (w większości są to urzędy zatrudniające od 12 do 24 osób) i wyposażone w odpowiednią do tego liczbę komputerów. Wszystkie urzędy należące do klasy 6 zatrudniają etatowych informatyków i jest to klasa o najwyższym odsetku odpowiedzi twierdzących na pytanie o istnienie wyodrębnionej komórki organizacyjnej odpowiedzialnej za informatyzację urzędu.

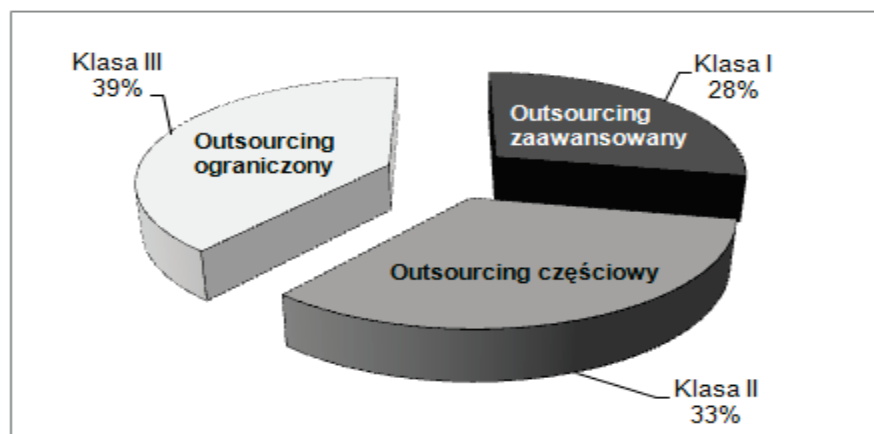
Algorytm grupujący FCM został zastosowany również w celu dokonania klasyfikacji rodzajowej outsourcingu informatycznego w badanych jednostkach samorządowych. Zmienna objaśniana, czyli typ outsourcingu, jest to zmienna skumulowana ( $D=\{d\}$ ). Zawiera ona odpowiedzi na pytania kwestionariusza dotyczące outsourcingu IT w obsłudze jednostek samorządowych. Zmienna ta została podzielona na 3 klasy ( $d=1,2,3$ ), z których każda określa konkretny rodzaj outsourcingu (por. rysunek 2).

Do **klasy I** należą te urzędy, w których występuje tzw. **outsourcing zaawansowany**. Jest to taki rodzaj outsourcingu, w którym większość elementów systemu informatycznego urzędu jest obsługiwana przez podmioty zewnętrzne (biorące odpowiedzialność za ich sprawne działanie) lub przez podmioty zewnętrzne we współpracy z pracownikami urzędu. Klasa I obejmuje 30 gmin.

W **klasie II** znalazły się te urzędy, w których występuje tzw. **outsourcing częściowy**. Jest to taki rodzaj outsourcingu, w którym pewne elementy systemu informatycznego urzędu są częściowo obsługiwane przez podmioty zewnętrzne (witryna WWW urzędu, oprogramowanie użytkowe, dostęp do Internetu) a inne częściowo przez same urzędy (sieć lokalna, komputery, serwery, systemy składowania i zabezpieczania danych). Klasa II obejmuje 36 gmin.

W **klasie III** zawarte są urzędy, w których występuje tzw. **outsourcing ograniczony**. Jest to taki rodzaj outsourcingu, w którym pewne elementy systemu informatycznego urzędu są najczęściej obsługiwane przez sam urząd (np. witryna WWW urzędu, sieć lokalna, komputery, systemy składowania i zabezpieczania danych) a tylko niektóre przez podmioty zewnętrzne (np. oprogramowanie użytkowe, dostęp do Internetu). Klasa III obejmuje 42 gminy.





Rys. 2. Podział gmin na klasy ze względu na typ występującego w nich outsourcingu

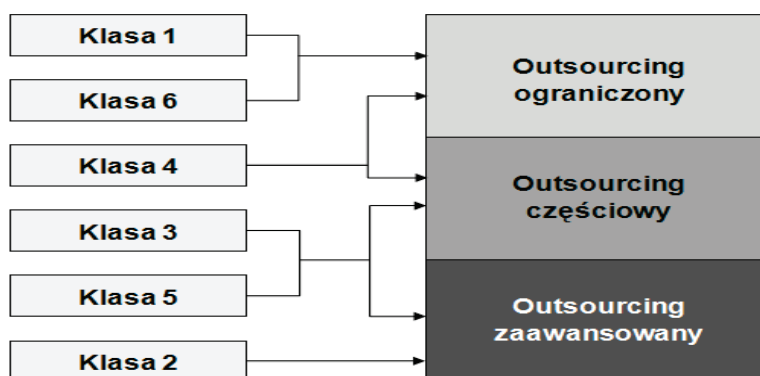
Źródło: Opracowanie własne.

Przeprowadzone podziały gmin na klasy ze względu na ich ogólną sytuację (klasyfikacja pierwsza) i ze względu na rodzaj stosowanego rodzaju usług zewnętrznych (klasyfikacja druga) posłużyły w dalszej części badania do wykrycia, przy pomocy teorii zbiorów przybliżonych, reguł łączących te dwa podziały. Kolejny etap przyjętej procedury badawczej, czyli poszukiwanie ukrytych, niejawnych reguł decyzyjnych, które stanowiły podstawę wyboru outsourcingu informatycznego w badanych jednostkach samorządu terytorialnego, został zrealizowany dzięki zastosowaniu metody zbiorów przybliżonych. Autorka zaprezentowała w zarysie analizę danych zbiorami przybliżonymi w pracy [17].

## 5. Podsumowanie

Na podstawie analizy dokonanej klasyfikacji gmin oraz otrzymanych, przy pomocy zbiorów przybliżonych, reguł sformułowano wzorzec obsługi informatycznej dla jednostek samorządowych (rysunek 3). Wyodrębnionych wcześniej sześć klas gmin w wyniku zastosowania teorii zbiorów przybliżonych zostało przyporządkowane trzem rodzajom outsourcingu w taki sposób, że określonym gminom został przypisany jeden lub co najwyżej dwa wzorce obsługi w postaci outsourcingu ograniczonego, częściowego i zaawansowanego.

Według tego wzorca gmina, która została zakwalifikowana do klasy 1 lub 6 powinna rozważyć zastosowanie outsourcingu ograniczonego. Gmina, która znalazła się w klasie 4 ma do wyboru albo outsourcing ograniczony albo częściowy. W przypadku gmin przyporządkowanych do klas 3 lub 5 należy rozważyć zastosowanie takiego zakresu usług, który jest właściwy dla outsourcingu częściowego lub zaawansowanego. Z kolei dla gmin w klasie 2 wybór wzorca obsługi informatycznej sprowadza się do skorzystania z najszerszego zakresu usług informatycznych, tj. outsourcingu zaawansowanego. Przedstawiony wzorzec został skonstruowany w celu ułatwienia podjęcia decyzji dotyczącej wyboru zakresu obsługi informatycznej przez osoby odpowiedzialne za informatyzację urzędów gminnych, które poszukują dla danego urzędu najlepszego rozwiązania.



Rys. 3. Wzorzec obsługi informatycznej organizacji samorządowych

Źródło: Opracowanie własne.

#### Bibliografia

- [1] Gatnar E. i in., *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Akademia Ekonomiczna we Wrocławiu, Wrocław 2004.
- [2] Gordon A.D., *Classification*, Chapman & Hall/CRC, Londyn 1999.
- [3] Grabiński T. i in., *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWN, Warszawa 1988.
- [4] Gutenbaum J., *Modelowanie matematyczne systemów*, EXIT, Warszawa 2003.
- [5] Jain A.K., Dubes R.C., *Algorithms for clustering data*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey 1988.
- [6] Jajuga K., *O pewnym uogólnieniu zagadnienia klasyfikacji*. W: Statystyka-ekonometria. Prace naukowe Akademii Ekonomicznej we Wrocławiu nr 285, Wrocław 1984, 5–18.
- [7] Jajuga K., *Zbiory rozmyte w zagadnieniu klasyfikacji*, Przegląd Statystyczny, Zeszyt nr 3/4, Warszawa 1984, 237–249.
- [8] Kacprzyk J., *Zbiory rozmyte w analizie systemowej*, PWN, Warszawa 1986.
- [9] Larose D.T., *Discovering Knowledge in Data. An Introduction to Data Mining*, John Wiley & Sons, Inc., New Jersey 2005.
- [10] Nowak E., *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa 1990.
- [11] Nycz M., *Klasyfikacja danych w procesie inteligentnego pozyskiwania wiedzy z baz danych*. W: Pozyskiwanie wiedzy i zarządzanie wiedzą, pod red. M. Nycz i M. Owoca, Akademia Ekonomiczna we Wrocławiu, Wrocław 2003, 351–361.
- [12] Ostasiewicz i in., *Statystyczne metody analizy danych*, Akademia Ekonomiczna we Wrocławiu, Wrocław 1998.
- [13] Piegat A., *Modelowanie i sterowanie rozmyte*, EXIT, Warszawa 1999.
- [14] Roubens M., *Pattern Classification Problems and Fuzzy Sets*, Fuzzy Sets and Systems 1/1978, pp. 239–253.
- [15] Rutkowski L., *Metody i techniki sztucznej inteligencji*, PWN, Warszawa 2005.

- [16] Sneath P.H., Sokal R.R., *Numerical Taxonomy. The Principles and Practice of Numerical Classification*, W. H. Freeman and Co., San Francisco 1973.
- [17] Wawrzyniak A., *Teoria zbiorów przybliżonych w określaniu wzorca obsługi informatycznej dla jednostek samorządowych*. W: *Metody ilościowe w badaniach ekonomicznych nr VIII*, SGGW, Warszawa 2007, pp. 399–408.

## FUZZY CLASSIFICATIONS IN MODELLING OF ECONOMIC PHENOMENA

### Summary

*The paper presents the problem of primary data classification in an economic phenomena modeling both in standard and non-standard approach (fuzzy classification). Presented research results refer to classification of self-government entities of zachodniopomorskie voivodship in consideration of the kind of outsourcing IT used by them. The main objective of the paper is the demonstration that adopted method of fuzzy classification enables better projecting of examined phenomenon.*

**Keywords:** fuzzy classification, fuzzy c-means, FCM algorithm, IT services, outsourcing IT, self-government entity.

Agata Wawrzyniak  
Instytut Informatyki w Zarządzaniu  
Uniwersytet Szczeciński  
ul. Mickiewicza 64, 71-101 Szczecin  
e-mail: agata.wawrzyniak@wneiz.pl