

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 427

Taksonomia 27

**Klasyfikacja i analiza danych –
teoria i zastosowania**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracnaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2016

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp	9
Beata Bal-Domańska: Propozycja procedury oceny zrównoważonego rozwoju w układzie <i>presja – stan – reakcja</i> w ujęciu przestrzennym / Proposal of the assessment of poviats sustainable development in the pressure – state – response system in spatial terms.....	11
Tomasz Bartłomowicz: Pomiar preferencji konsumentów z wykorzystaniem metody <i>Analytic Hierarchy Process</i> / Analytic Hierarchy Process as a method of measurement of consumers’ preferences.....	20
Maciej Beręsewicz, Marcin Szymkowiak: Analiza skupień wybranych lokalnych rynków nieruchomości w Polsce z wykorzystaniem internetowych źródeł danych / Cluster analysis of selected local real estate markets in Poland based on Internet data sources.....	30
Beata Bieszk-Stolorz: Wybrane modele przeciętnego efektu oddziaływania w analizie procesu wychodzenia z bezrobocia / Chosen average treatment effect models in the analysis of unemployment exit process.....	40
Justyna Brzezińska: Modele IRT i modele Rascha w badaniach testowych / IRT and Rasch models in test measurement.....	49
Mariola Chrzanowska, Nina Drejerska: Geograficznie ważona regresja jako narzędzie analizy poziomu rozwoju społeczno-gospodarczego na przykładzie regionów Unii Europejskiej / Geographically weighted regression as a tool of analysis of socio-economic development level of regions in the European Union.....	58
Sabina Denkowska: Zastosowanie analizy wrażliwości do oceny wpływu nieobserwowanej zmiennej w <i>Propensity Score Matching</i> / The application of sensitivity analysis in assessing the impact of an unobserved confounder in Propensity Score Matching.....	66
Adam Depta: Zastosowanie analizy czynnikowej do wyodrębnienia aspektów zdrowia wpływających na jakość życia osób jękających się / The application of factor analysis to the identification of the health aspects affecting the quality of life of stuttering people.....	76
Mariusz Doszyń, Sebastian Gnat: Taksonomiczno-ekonometryczna procedura wyceny nieruchomości dla różnych miar porządkowania / Taxonomic and econometric method of real estate valuation for various classification measures.....	84

Marta Dziechciarz-Duda, Anna Król: Segmentacja konsumentów smartfonów na podstawie preferencji wyrażonych / Segmentation of smartphones' consumers on the basis of stated preferences	94
Ewa Genge: Zmienne towarzyszące w ukrytym modelu Markowa – analiza oszczędności polskich gospodarstw domowych / Latent Markov model with covariates – Polish households' saving behaviour	103
Joanna Górna, Karolina Górna: Modelowanie wzrostu gospodarczego z wykorzystaniem narzędzi ekonometrii przestrzennej / Economic growth modelling with the application of spatial econometrics tools	112
Alicja Grześkowiak: Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności / Multivariate analysis of professional competencies with respect to the age groups of the population	122
Agnieszka Kozera, Feliks Wysocki: Problem ustalania współrzędnych obiektów modelowych w metodach porządkowania liniowego obiektów / The problem of determining the coordinates of model objects in object linear ordering methods	131
Mariusz Kubus: Lokalna ocena mocy dyskryminacyjnej zmiennych / Local evaluation of a discrimination power of the variables.....	143
Paweł Lula, Katarzyna Wójcik, Janusz Tuchowski: Analiza wydzźwięku polskojęzycznych opinii konsumenckich ukierunkowanych na cechy produktu / Feature-based sentiment analysis of opinions in Polish.....	153
Aleksandra Łuczak, Agnieszka Kozera, Feliks Wysocki: Ocena sytuacji finansowej jednostek samorządu terytorialnego z wykorzystaniem rozmytych metod klasyfikacji i programu R / Assessment of financial condition of local government units with the use of fuzzy classification methods and program R	165
Dorota Rozmus: Badanie stabilności taksonomicznej czynnikowej metody odległości probabilistycznej / Stability of the factor probability distance clustering method	176
Adam Sagan, Aneta Rybicka, Justyna Brzezińska: <i>Conjoint analysis</i> oparta na modelach IRT w zagadnieniu optymalizacji produktów bankowych / An IRT-approach for conjoint analysis for banking products preferences.....	184
Michał Stachura: O szacowaniu centrum populacji określonego obszaru na przykładzie Polski / On estimating centre of population of a given territory. Poland's case	195
Michał Stachura, Barbara Wodecka: Wybrane aspekty i zastosowania modeli zdarzeń ekstremalnych / Selected facets and application of models of extremal events	205
Iwona Staniec, Jan Żółtowski: Wykorzystanie analizy log-liniowej do wyboru czynników determinujących współpracę w przedsiębiorczości	

technologicznej / Use of log-linear analysis for the selection determinants of cooperation in technological entrepreneurship.....	215
Marcin Szymkowiak, Wojciech Roszka: Potencjał gospodarczy gmin aglomeracji poznańskiej w ujęciu taksonomicznym / The economic potential of municipalities of the Poznań agglomeration in the light of taxonomy analysis.....	224
Lucyna Wojcieszka: Zastosowanie modeli klas ukrytych w badaniu opinii respondentów na temat roli państwa w gospodarce / Implementation of latent class models in the respondents' survey on the role of the country in economy.....	234

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

Krzysztof Jajuga, Marek Walesiak

Aleksandra Łuczak, Agnieszka Kozera, Feliks Wysocki

Uniwersytet Przyrodniczy w Poznaniu
e-mail: {luczak; akozera; wysocki}@up.poznan.pl

**OCENA SYTUACJI FINANSOWEJ
JEDNOSTEK SAMORZĄDU TERYTORIALNEGO
Z WYKORZYSTANIEM ROZMYTYCH METOD
KLASYFIKACJI I PROGRAMU R**

**ASSESSMENT OF FINANCIAL CONDITION
OF LOCAL GOVERNMENT UNITS WITH THE USE
OF FUZZY CLASSIFICATION
METHODS AND PROGRAM R**

DOI: 10.15611/pn.2016.427.17

Streszczenie: Celem pracy jest przedstawienie możliwości wykorzystania rozmytych metod klasyfikacji do identyfikacji typów sytuacji finansowej samorządów gminnych w układzie powiatów województwa wielkopolskiego w 2013 r. W pracy wykorzystano dwie metody ustalania rozmytych klasyfikacji oparte na algorytmach iteracyjno-optymalizacyjnych: C-średnich (prowadzącą do klasyfikacji sferycznej) oraz C-medoidów, która jest rozszerzeniem idei algorytmu rozmytych C-średnich i wykazuje większą odporność na obserwacje odstające, co ma istotne znaczenie w analizie danych finansowych. Podstawę empiryczną przeprowadzonych badań stanowiły dane pochodzące z Głównego Urzędu Statystycznego oraz Ministerstwa Finansów. Badania przeprowadzono z wykorzystaniem programu statystycznego R.

Słowa kluczowe: klasyfikacja rozmyta, sytuacja finansowa, gminy.

Summary: The aim of the study was to present the possibilities of using fuzzy classification methods to identify the types of development financial condition of local governments in the counties of Wielkopolska province in 2013. In the paper two methods of the fuzzy classification based on iterative-optimization algorithms were used: the C-means method (leading to the spherical classification) and the C-medoids method which is an extension of the idea of an algorithm of fuzzy C-means method and has a reduced sensitivity to outliers, which is important in the analysis of financial data. The study was conducted using statistical program R.

Keywords: fuzzy classification, financial situation, communes.

1. Wstęp

W rozpoznawaniu typów rozwojowych zjawisk finansowych klasyczne metody klasyfikacji często nie dają zadowalających efektów, ze względu na trudności wyodrębnienia w zbiorowości jednostek samorządu terytorialnego (JST) wyraźnej struktury klas dotyczącej poziomu samodzielności finansowej. Ma to związek z wieloznacznością i wieloaspektowością wskaźników opisujących sytuację finansową tych podmiotów. Na ogół nie można wszystkich jednostek przyporządkować jednoznacznie tylko do jednej z klas. Często wykazują one podobieństwo (ze względu na wartości analizowanych wskaźników finansowych) do innych JST znajdujących się w różnych klasach. W rozwiązaniu tego problemu mogą być pomocne rozmyte metody klasyfikacji. Ich zastosowanie pozwala na przyporządkowanie JST do poszczególnych klas z różnymi stopniami przynależności [Jajuga 1993; Wysocki 2010].

Celem pracy jest przedstawienie możliwości wykorzystania rozmytych metod klasyfikacji do identyfikacji typów sytuacji finansowej samorządów gminnych w układzie powiatów województwa wielkopolskiego w 2013 r. W pracy wykorzystano dwie metody ustalania rozmytych klasyfikacji oparte na algorytmach iteracyjno-optymalizacyjnych:

- rozmytą metodę *C*-średnich¹ [Bezdek 1973] – prowadzącą do klasyfikacji sferycznych,
- rozmytą metodę *C*-medoidów [Krishnapuram i in. 2001], która jest rozszerzeniem idei algorytmu rozmytych *C*-średnich i wykazuje większą odporność na występowanie obserwacji odstających, co ma również istotne znaczenie w analizie danych finansowych, ponieważ niektóre wskaźniki finansowe dla jednostek samorządu terytorialnego charakteryzują się występowaniem wartości odstających.

Podstawę empiryczną przeprowadzonych badań stanowiły dane pochodzące z Głównego Urzędu Statystycznego [GUS 2013] oraz Ministerstwa Finansów [2014]. Analizy przeprowadzono z wykorzystaniem programu statystycznego R [Giordani, Ferraro 2015].

2. Metody badawcze

Klasyczne metody klasyfikacji stosowane w analizie sytuacji finansowej JST charakteryzują się tym, że dana jednostka samorządu terytorialnego (np. gmina, powiat) zostaje jednoznacznie zaklasyfikowana do jednej z wyodrębnionych klas typologicznych. W rzeczywistości często jednak granice między poszczególnymi

¹ Pierwszy algorytm z grupy rozmytych metod *C*-średnich (FCM – *fuzzy C-means*) zaproponował J.C. Dunn w 1973 r. w pracy *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters*.

klasami są „nieostre”, a przyporządkowanie danej jednostki tylko do jednej z klas może skutkować utratą znacznej części informacji w trakcie procedury klasyfikacji. W takiej sytuacji można zastosować metody klasyfikacji wykorzystujące teorię zbiorów rozmytych [Zadeh 1965]. Metody te pozwalają na uzyskanie znacznie większej liczby informacji o sklasyfikowanych obiektach niż metody klasyczne [Wysocki 2010].

Ideę rozmytej klasyfikacji można przedstawić następująco. Niech zbiór Ω zawiera N jednostek samorządu terytorialnego ($1, \dots, N$), a każda z tych jednostek jest opisana przez wartości cech prostych x_1, \dots, x_k tak, że dane są obserwacje x_{ik} ($i = 1, \dots, N, k = 1, \dots, K$). Zadanie klasyfikacji sprowadza się do określenia na elementach zbioru Ω rodziny klas rozmytych (zbiorów) A_1, \dots, A_C (gdzie liczba C jest dana i $1 < C < N$) tak, aby były spełnione warunki [Jajuga 1984; de Oliveira, Pedrycz (red.) 2007; Wawrzyniak 2010]:

$$u_{ic} = \mu_{A_c}(x) \in \langle 0, 1 \rangle, \quad \sum_{c=1}^C u_{ic} = 1, \quad (1)$$

gdzie: $\mu_{A_c} : \Omega \rightarrow \langle 0, 1 \rangle$ – funkcja przynależności; u_{ic} – stopień przynależności i -tej jednostki do c -tej klasy rozmytej.

W klasyfikacji rozmytej każda jednostka należy do każdej z klas z różnymi stopniami przynależności. Jednostki, dla których wartości stopni przynależności do tej samej klasy rozmytej są względnie duże, są podobne do siebie, natomiast jednostki odznaczające się dużymi stopniami przynależności do różnych klas są do siebie podobne w niewielkim stopniu.

W metodach klasyfikacji rozmytej pewien problem stanowi wybór optymalnej liczby klas. Zazwyczaj liczba klas jest ustalana przez badacza na podstawie wiedzy pozastatystycznej bądź też przyjmowana w oparciu o przesłanki statystyczne. Przyjmuje się wówczas liczbę klas ustaloną z użyciem wybranej metody klasyfikacji rozłącznej (np. Warda) dla tej samej macierzy danych lub przeprowadza klasyfikację rozmytą przy różnej liczbie klas, a następnie wybiera tę, dla której zastosowany wskaźnik jakości klasyfikacji osiąga poziom optymalny [Wysocki 2010].

W pracy wykorzystano dwie metody ustalania rozmytych klasyfikacji oparte na algorytmach iteracyjno-optymalizacyjnych, tj. rozmytą metodę C -średnich (*fuzzy c-means* – FCM) oraz rozmytą metodę C -medoidów (*fuzzy c-medoids* – FCMdd). Problem sferycznej klasyfikacji rozmytej został przedstawiony jako następujące zadanie programowania matematycznego [Bezdek 1973]:
zminimalizować

$$J_{FCM} = \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m \| \mathbf{x}_i - \mathbf{v}_c \|^2 \quad (2)$$

przy warunkach:

$$\sum_{c=1}^C u_{ic} = 1 \quad (i=1, \dots, N), \quad \sum_{i=1}^N u_{ic} > 0 \quad (c=1, \dots, C), \quad u_{ic} \geq 0, \\ (i=1, \dots, N; c=1, \dots, C) \quad (3)$$

gdzie: u_{ic} – stopień przynależności i -tej jednostki (\mathbf{x}_i) do c -tej klasy rozmytej; \mathbf{v}_c – środek ciężkości (prototyp) c -tej klasy rozmytej w K -wymiarowej przestrzeni cech; $d_{ic} = \|\mathbf{x}_i - \mathbf{v}_c\|$ – odległość euklidesowa pomiędzy i -tą jednostką (\mathbf{x}_i) a c -tym środkiem ciężkości (\mathbf{v}_c) w K -wymiarowej przestrzeni cech; m – parametr regulujący stopień rozmytości klasyfikacji.

Rozmyta metoda C -medoidów jest modyfikacją rozmytej metody C -średnich [Krishnapuram, Joshi, Nasraoui 2001]. Algorytm C -średnich jest czuły na występowanie wartości nietypowych – odstających, gdyż mogą wpływać one znacząco na wartości środków ciężkości klas, powodując ich zniekształcenia. Istotna różnica między metodą C -medoidów i C -średnich sprowadza się do wyboru środka klasy, który będzie reprezentował tę klasę. W przypadku metody C -medoidów jest to jedna z jednostek wielocechowych badanej zbiorowości (jednostka najbardziej reprezentatywna), która jest jedną z jednostek najbliższych położonych w stosunku do środka klasy, podczas gdy w metodzie C -średnich środek klasy, zwany środkiem ciężkości, oblicza się jako średnie ważone wartości cech dla jednostek należących do tej klasy, gdzie wagami są stopnie przynależności jednostek do klasy (zob. [Mei, Chen 2010]). Metoda C -medoidów jest w większym stopniu odporna na występowanie wartości odstających [Maulik, Bandyopadhyay, Mukhopadhyay 2011].

Problem klasyfikacji rozmytej z uwzględnieniem C -medoidów może być przedstawiony jako następujące zadanie programowania matematycznego (zob. [Krishnapuram, Joshi, Nasraoui 2001; Mei, Chen 2010]):
zminimalizować

$$J_{\text{FCMdd}} = \sum_{i=1}^N \sum_{c=1}^C u_{ic}^m \text{Dis}(\mathbf{x}_i, \delta_c) \quad (4)$$

przy warunkach (3),

gdzie $\text{Dis}(\mathbf{x}_i, \delta_c)$ oznacza niepodobieństwo pomiędzy i -tą jednostką (\mathbf{x}_i) a medoidem c -tej klasy (oznaczonym jako δ_c w K -wymiarowej przestrzeni cech).

Proponowana procedura rozpoznawania typów sytuacji finansowej samorządów gminnych w układzie powiatów województwa wielkopolskiego z wykorzystaniem rozmytych metod klasyfikacji obejmuje sześć etapów [Wysocki 2010]. Pierwszy etap to wybór cech opisujących badane jednostki samorządowe w oparciu o kryteria merytoryczne i statystyczne. Kolejny etap (II) obejmuje normalizację wartości cech prostych z wykorzystaniem procedury klasycznej standaryzacji. Trzeci etap to przeprowadzenie wstępnej klasyfikacji badanych jednostek (powiatów, $N = 31$) za pomocą metody Warda (odległość euklidesowa) i ich ocena na

podstawie kryterium ekstremalności wskaźników jakości klasyfikacji (Calińskiego-Harabasz, Krzanowskiego-Lai, wskaźnik sylwetkowy) [Walesiak, Dudek 2006]. W etapie IV dokonuje się klasyfikacji rozmytej badanych jednostek z wykorzystaniem wybranych metod rozmytych (C -średnich i C -medoidów). Następnym etapem (V) jest dokonanie porównania uzyskanych podziałów klasyfikacyjnych z wykorzystaniem wskaźnika podobieństwa struktur [Chomątowski, Sokołowski 1978], wskaźnika podobieństwa klasyfikacji rozłącznych [Nowak 1990] i znormalizowanego wskaźnika entropii [Wysocki 2010]. W kolejnym etapie (VI) wyróżnia się cechy charakterystyczne w klasach i na tej podstawie przeprowadza identyfikację oraz opis wyodrębnionych typów.

3. Wyniki badań

W pierwszym etapie dokonano wyboru 10 cech opisujących sytuację finansową samorządów gminnych w układzie powiatów, tj.: wskaźnik poziomu dochodów własnych (zł *per capita*) ($WDWM$), wskaźnik samodzielności finansowej (wydatkowej) pierwszego stopnia (%) ($WSFW I$), wskaźnik finansowej ingerencji państwa (%) ($WFIP$), wskaźnik autonomii podatkowej (%) (WAP), wskaźnik udziału wyniku finansowego (nadwyżki lub deficytu operacyjnego) w dochodach ogółem (%) ($WFWO$)², wskaźnik udziału wydatków inwestycyjnych w wydatkach ogółem (%) ($WIWO$), wskaźnik samofinansowania (%) (WS), wskaźnik poziomu nadwyżki operacyjnej (zł *per capita*) (NOM), wskaźnik udziału zobowiązań ogółem w dochodach ogółem (%) ($ZODO$), wskaźnik obciążenia dochodów ogółem (%) ($ODoOZ$) i dochodów własnych obsługą zadłużenia (%) ($ODwOZ$). Wartości wskaźników poddano procesowi standaryzacji klasycznej celem pozbawienia ich mian oraz ujednoczenia rzędów wielkości. Po przeprowadzeniu klasyfikacji powiatów metodą Warda (odległość euklidesowa) wyznaczono wartości wskaźników jakości klasyfikacji. Uzyskane wartości tych miar wskazały, że najlepszym byłby podział zbiorowości powiatów na 3 klasy. Świadczą o tym maksymalne wartości wskaźników dla trzyklasowego podziału (tab. 1).

Tabela 1. Wartości wskaźników jakości klasyfikacji w zależności od liczby klas (metoda Warda)

Wskaźniki jakości klasyfikacji	Liczba klas								
	2	3	4	5	6	7	8	9	10
Caliński-Harabasz	6,65	6,74	6,35	6,13	5,67	5,72	5,65	5,55	5,65
Wskaźnik sylwetkowy	0,15	0,16	0,14	0,13	0,09	0,11	0,12	0,12	0,12
Krzanowski-Lai	0,85	1,43	1,09	1,58	0,67	1,26	1,16	0,87	1,16

Źródło: obliczenia własne na podstawie danych pochodzących z Głównego Urzędu Statystycznego [GUS 2013] i Ministerstwa Finansów [2014].

² Obliczony za lata 2011–2013.

Wyniki klasyfikacji powiatów na 3 klasy uzyskane metodą Warda przyjęto jako klasyfikację początkową w metodach klasyfikacji rozmytej: *C*-średnich oraz *C*-medoidów (dla $m = 1, 3$). Ich zastosowanie pozwala na rozpoznanie, w jakim stopniu poszczególne powiaty przynależą do każdej z wyodrębnionych klas. W celu możliwości przeprowadzenia analizy porównawczej, otrzymane klasy typologiczne z wykorzystaniem obu zastosowanych metod klasyfikacji rozmytej uporządkowano w sposób następujący: klasa I reprezentuje powiaty o najkorzystniejszej, klasa II – o przeciętnej i klasa III – o najmniej korzystnej sytuacji finansowej samorządów gminnych. Uzyskane podziały powiatów oceniono z punktu widzenia ich zgodności między sobą. Podziały uzyskane z zastosowaniem rozmytej metody *C*-średnich okazały się w wysokim stopniu zgodne z podziałami uzyskanymi za pomocą rozmytej metody *C*-medoidów. Wskaźnik podobieństwa struktur wyniósł bowiem 0,86, natomiast wskaźnik podobieństwa dla klasyfikacji rozłącznej 0,84.

W wyniku przeprowadzonych obliczeń otrzymano dwie rozmyte klasyfikacje badanych powiatów o stosunkowo niskim stopniu rozmytości³ (tab. 2), jednakże wyższy stopień rozmytości charakteryzował klasyfikację powiatów uzyskaną metodą *C*-medoidów. Wskaźnik entropii dla tej klasyfikacji wyniósł 0,35, natomiast dla klasyfikacji uzyskanej metodą *C*-średnich stanowił 0,27. Z przeprowadzonej analizy stopni przynależności powiatów do klas rozmytych (u_{ic}) wynika, że w większości przypadków przynależą one wyraźnie do jednej klasy. W szczególności sytuacja ta dotyczy powiatu poznańskiego ($u_{21, 1} = 1,00$), który stanowi najlepiej rozwinięty obszar w województwie wielkopolskim i w rezultacie sytuacja finansowa samorządów gminnych w aglomeracji poznańskiej⁴ jest najkorzystniejsza (klasa I).

Przeprowadzone badania z wykorzystaniem rozmytych metod *C*-średnich i *C*-medoidów wykazały pewne różnice, między innymi w wynikach klasyfikacji, dotyczące wartości wskaźnika stopnia przynależności niektórych powiatów do klas II i III. Sytuacja ta dotyczy siedmiu powiatów, tj. czarnkowsko-trzcianeckiego, grodzkiego, kępińskiego, konińskiego, ostrzeszowskiego, ślupieckiego i złotowskiego. W przypadku zastosowania rozmytej metody *C*-średnich powiaty: czarnkowsko-trzcianecki, grodzki, kępiński, złotowski, a zwłaszcza koniński wyróżniał wysoki stopień przynależności do klasy III ($u_{i, 3} > 0,6$), reprezentującej najmniej korzystną sytuację finansową samorządów gminnych. Natomiast przy zastosowaniu rozmytej metody *C*-medoidów stopnie przynależności tych powiatów są wyższe w odniesieniu do klasy II, która charakteryzuje przeciętną sytuację finansową samorządów gminnych. Może to wynikać stąd, że rozmyta metoda *C*-średnich jest mniej odporna na występowanie w zbiorze wartości cech obserwa-

³ Stopień rozmytości klasyfikacji zmierzono z wykorzystaniem znormalizowanego wskaźnika entropii [Wysocki 2010, s. 104]. Wskaźnik ten przyjmuje wartości z przedziału [0, 1], a im wyższa jest jego wartość, tym większy jest stopień rozmytości otrzymanej klasyfikacji.

⁴ Aglomeracja poznańska położona jest w centrum regionu i skupia miasto Poznań i jego strefę podmiejską (najczęściejje utożsamianą z zasięgiem powiatu poznańskiego).

Tabela 2. Stopnie przynależności powiatów województwa wielkopolskiego do klas rozmytych sytuacji finansowej samorządów gminnych według zastosowanych metod

Powiaty	Metoda rozmyta / sytuacja finansowa					
	C-średnich ($m = 1,3$)			C-medoidów ($m = 1,3$)		
	I korzystna	II przeciętna	III niekorzystna	I korzystna	II przeciętna	III niekorzystna
Chodzieski	0,00	0,99	0,01	0,00	0,94	0,06
Czarnkowsko- -trzcianecki	0,00	0,22	0,78	0,00	0,62	0,38
Gnieźnieński	0,00	0,06	0,94	0,00	0,14	0,86
Gostyński	0,00	0,97	0,03	0,00	0,98	0,02
Grodziski	0,00	0,06	0,94	0,00	0,62	0,38
Jarociński	0,00	0,05	0,95	0,00	0,19	0,81
Kaliski	0,00	0,27	0,73	0,00	0,19	0,81
Kępiński	0,03	0,18	0,79	0,04	0,53	0,43
Kolski	0,00	0,04	0,96	0,00	0,02	0,98
Koniński	0,00	0,05	0,95	0,00	0,60	0,40
Kościański	0,00	0,94	0,06	0,00	0,90	0,10
Krotoszyński	0,00	0,96	0,04	0,00	1,00	0,00
Leszczyński	0,00	0,89	0,11	0,00	0,98	0,02
Międzychodzki	0,00	0,98	0,02	0,00	0,99	0,01
Nowotomyski	0,00	0,88	0,12	0,00	0,93	0,07
Obornicki	0,00	0,95	0,05	0,02	0,82	0,16
Ostrowski	0,00	0,99	0,01	0,00	0,97	0,03
Ostrzeszowski	0,00	0,69	0,31	0,00	0,49	0,51
Pilski	0,01	0,78	0,22	0,01	0,88	0,11
Pleszewski	0,00	0,06	0,94	0,00	0,16	0,84
Poznański	1,00	0,00	0,00	1,00	0,00	0,00
Rawicki	0,00	0,05	0,95	0,00	0,07	0,93
Słupecki	0,00	0,70	0,30	0,00	0,38	0,62
Szamotulski	0,01	0,89	0,10	0,03	0,82	0,15
Średzki	0,01	0,89	0,10	0,03	0,74	0,23
Śremski	0,01	0,84	0,15	0,02	0,77	0,21
Turecki	0,00	0,96	0,04	0,00	0,97	0,03
Wągrowiecki	0,00	0,01	0,99	0,00	0,08	0,92
Wolsztyński	0,00	0,00	1,00	0,00	0,00	1,00
Wrzesiński	0,00	0,67	0,33	0,00	0,53	0,47
Złotowski	0,00	0,33	0,67	0,00	0,66	0,34

Źródło: jak w tab. 1.

cji nietypowych – odstających, które często pojawiają się w analizie zjawisk finansowych, w tym w ocenie sytuacji finansowej JST. Przykładowo w 2013 w powiatach konińskim, kępińskim i grodzkim zaobserwowano bardzo wysoki poziom obciążenia dochodów ogółem samorządów gminnych obsługą zadłużenia (odpo-

wiednio wynoszący 9,6, 10 i 10,4%, przy przeciętnej dla ogółu samorządów gminnych w województwie wynoszącej 6,0%), które stanowiły wartości odstające. W powiecie kępińskim zaobserwowano także wysoki poziom zobowiązań ogółem samorządów gminnych, wynoszący 404 zł na mieszkańca, czyli o prawie 50% wyższy w stosunku do przeciętnej dla województwa [Ministerstwo Finansów 2014]. W rezultacie, pomimo wyższego w stosunku do przeciętnej dla województwa wielkopolskiego poziomu samodzielności finansowej samorządów gminnych, kwantyfikowanego wielkościami wskaźników *WDWM* oraz *WSFW*, wysokiej atrakcyjności finansowej mierzonej wielkością wskaźnika *WIWO*, powiaty te przy zastosowaniu metody *C*-średnich zostały zakwalifikowane do klasy III, reprezentującej jednostki o niekorzystnej sytuacji finansowej samorządów gminnych. Z kolei powiaty ostrzeszowski i słupecki, przy zastosowaniu rozmytej metody *C*-średnich, wyróżniał wyższy stopień przynależności do klasy II ($u_{i,2} > 0,6$), charakteryzującej powiaty o przeciętnej sytuacji finansowej. Przy zastosowaniu rozmytej metody *C*-medoidów stopnie przynależności tych powiatów są natomiast wyższe w odniesieniu do klasy III, wyróżniającej powiaty o niekorzystnej sytuacji finansowej samorządów gminnych. W 2013 r. w powiatach ostrzeszowskim i słupeckim zaobserwowano bowiem bardzo wysoki poziom wskaźnika samofinansowania⁵ (odpowiednio wynoszący 152,2 i 158,7%, przy średniej dla ogółu JST w województwie wynoszącej 114,0%), które stanowiły wartości odstające [Ministerstwo Finansów 2014]. W rezultacie, pomimo niższego w stosunku do przeciętnej dla województwa wielkopolskiego poziomu samodzielności finansowej, mierzonej wielkościami wskaźników *WDWM* i *WAP*, niskiej aktywności inwestycyjnej tych jednostek, kwantyfikowanej wielkością wskaźnika *WIWO*, powiaty te przy zastosowaniu metody *C*-średnich zostały zakwalifikowane do klasy II reprezentującej jednostki o przeciętnej sytuacji finansowej samorządów gminnych. Z przeprowadzonych badań wynika więc, że przyjęcie do badań przynajmniej jednej cechy charakteryzującej się obserwacjami odstającymi wpłynąć może w sposób znaczący na wyniki klasyfikacji rozmytej i przeprowadzonej na jej podstawie identyfikacji typów rozwojowych. W rezultacie można stwierdzić, że identyfikacja typów sytuacji finansowej samorządów gminnych przeprowadzona na podstawie metody *C*-medoidów byłaby bardziej poprawna pod względem merytorycznym.

Przeprowadzona analiza wewnętrznej struktury gmin ze względu na sytuację finansową pokazała (tab. 3), że pierwszą klasę typologiczną, charakteryzującą najkorzystniejszą sytuację finansową, utworzył tylko powiat poznański.

⁵ Wartości tego wskaźnika wskazują na stopień, w jakim JST finansuje inwestycje środkami własnymi. Im wyższe wartości tego wskaźnika, tym mniejsze ryzyko utraty płynności finansowej w związku z nadmiernymi kosztami obsługi zadłużenia. Wysoka jego wartość może świadczyć o niskim poziomie realizowanych inwestycji w stosunku do własnych możliwości [Ministerstwo Finansów 2014].

Tabela 3. Wyniki klasyfikacji powiatów województwa wielkopolskiego według sytuacji finansowej samorządów gminnych z wykorzystaniem metod klasyfikacji rozmytej

Klasa	Sytuacja finansowa	Metoda klasyfikacji – rozmyta metoda			
		C-średnich		C-medoidów	
		N_c^*	%	N_c	%
I	korzystna	1	3,2	1	3,2
II	przeciętna	17	54,9	20	64,5
III	niekorzystna	13	41,9	10	32,3

* N_c – liczebność w c -tej klasie.

Źródło: obliczenia własne na podstawie danych z tab. 2.

Na podstawie klasyfikacji uzyskanej z wykorzystaniem rozmytej metody C-średnich 54,9% powiatów zakwalifikowano natomiast do klasy II reprezentującej przeciętny i aż 41,9% powiatów do klasy III wyróżniającej powiaty o niekorzystnej sytuacji finansowej samorządów gminnych. Z kolei według wyników badań z wykorzystaniem rozmytej metody C-medoidów, odpornej na występowanie obserwacji odstających, 64,5% powiatów zakwalifikowano do klasy II, natomiast 32,3% powiatów do klasy III. Powiaty czarnkowsko-trzcianecki, grodziski, kępiński, koniński oraz złotowski w tym przypadku zostały zakwalifikowane do klasy II reprezentującej przeciętny, a nie jak w przypadku zastosowania rozmytej metody C-średnich do klasy reprezentującej niekorzystną sytuację finansową.

Tabela 4. Rozmyte średnie klasowe charakteryzujące klasy typologiczne powiatów województwa wielkopolskiego według sytuacji finansowej samorządów gminnych (metoda rozmyta C-medoidów)

Wskaźniki	Klasa typologiczna / Sytuacja finansowa			Średnia
	I	II	III	
	korzystna	przeciętna	niekorzystna	
Poziomu dochodów własnych (zł <i>per capita</i>) (<i>WDWM</i>)	2279,4	1497,8	1334,7	1464,2
Samodzielności finansowej (wydatkowej) (%) (<i>WSFW I</i>)	68,1	49,9	45,3	48,8
Autonomii podatkowej (%) (<i>WAP</i>)	30,0	21,4	18,8	20,7
Poziomu nadwyżki operacyjnej (zł <i>per capita</i>) (<i>NOM</i>)	344,5	263,7	281,9	273,7
Obciążenia dochodów własnych obsługą zadłużenia (%) (<i>ODwOZ</i>)	11,2	16,0	21,3	17,9
Udziału wydatków inwestycyjnych w wydatkach ogółem (%) (<i>WIWO</i>)	21,0	17,4	15,9	16,9

Wartości przeciętne w wyodrębnionych klasach obliczono przy zastosowaniu średniej ważonej, gdzie wagami są wartości wskaźników stopni przynależności powiatów do klas rozmytych.

Źródło: jak w tab. 1.

Identyfikację typów sytuacji finansowej samorządów gminnych przeprowadzono na podstawie klasyfikacji uzyskanej z wykorzystaniem metody *C-medoidów* (tab. 4), która jest bardziej odporna na występowanie obserwacji odstających. W klasie I znalazł się jeden powiat – poznański, reprezentujący najkorzystniejszą sytuację finansową samorządów gminnych. Świadczy o tym znacząca jego przewaga nad innymi powiatami w zakresie poziomu samodzielności finansowej ($WDWM = 2279,4$ zł *per capita*, $WSFWI = 68,1\%$), poziomu autonomii podatkowej ($WAP = 30,0\%$) i możliwości rozwoju lokalnego (kwantyfikowanego najwyższym udziałem wydatków inwestycyjnych w wydatkach ogółem – $WIWO = 21,0\%$). Do klasy II, reprezentującej przeciętny poziom sytuacji finansowej samorządów gminnych, zakwalifikowano 20 powiatów, stanowiących blisko 64,5% zbiorowości powiatów w województwie wielkopolskim. Klasę tę utworzyły głównie powiaty z zachodnio-północnej części województwa. Wartości wskaźników finansowych określających poziom samodzielności finansowej, atrakcyjności, jak i poziomu zadłużenia samorządów gminnych zbliżone były do wartości przeciętnych dla województwa. Z kolei do klasy III, reprezentującej niekorzystną sytuację finansową, zakwalifikowano 10 powiatów, stanowiących 32,3% ogółu badanej zbiorowości, głównie ze wschodniej części województwa. Powiaty te charakteryzowała najsłabsza sytuacja finansowa samorządów gminnych – względnie niska samodzielność finansowa ($WDWM = 1334,7$ zł *per capita*, $WSFWI = 45,3\%$), najniższy udział wydatków inwestycyjnych w wydatkach ogółem ($WIWO = 15,9\%$), jak i najwyższy poziom obciążenia dochodów własnych obsługą zadłużenia ($ODwOZ = 21,3\%$) (zob. tab. 3 i 4).

4. Zakończenie

Na podstawie przeprowadzonych obliczeń i analiz można sformułować następujące stwierdzenia i wnioski.

Ze względu na wieloaspektowość wskaźników opisujących sytuację finansową jednostek samorządu terytorialnego (JST) klasyczne metody klasyfikacji mogą nie dawać dobrych rezultatów (zarówno pod względem statystycznym, jak i merytorycznym). Na ogół granice pomiędzy poszczególnymi klasami są „nieostre”, a przyporządkowanie danej JST tylko do jednej z klas może skutkować utratą części informacji w trakcie procedury klasyfikacji dotyczących rzeczywistej sytuacji finansowej. W rozwiązaniu tego problemu mogą być pomocne rozmyte metody klasyfikacji, takie jak rozmyta metoda *C-średnich* i *C-medoidów*, w których każdy obiekt należy do poszczególnych klas z różnymi stopniami przynależności.

Analiza uzyskanych wyników klasyfikacji powiatów województwa wielkopolskiego ujawniła wyraźnie zarysowujący się podział województwa wielkopolskiego na trzy klasy różniące się pod względem sytuacji finansowej samorządów gminnych w 2013 roku i jednocześnie wskazała na pewne podobieństwa oraz różnice pomiędzy klasyfikacjami uzyskanymi rozmytymi metodami: *C-średnich* i *C-me-*

doidów. Wyrażają się one w strukturze wewnętrznej odpowiadających sobie klas powiatów wyodrębnionych ze względu na sytuację finansową samorządów gminnych. Różnice dotyczą klas II i III, reprezentujących odpowiednio przeciętną i niekorzystną sytuację finansową. Przyczyny tego stanu rzeczy można doszukiwać się w fakcie, że rozmyta metoda *C*-średnich nie jest odporna na występowanie obserwacji odstających. Natomiast zastosowanie rozmytej metody *C*-medoidów, którą wyróżnia zmniejszona wrażliwość na obserwacje nietypowe, prowadzi do typologii bardziej poprawnych pod względem merytorycznym. Powiaty, takie jak czarnkowsko-trzcianecki, grodziski, kępiński, koniński oraz złotowski w wyniku zastosowania rozmytej metody *C*-medoidów zostały poprawnie zakwalifikowane do klasy II reprezentującej przeciętny, a nie jak w przypadku metody *C*-średnich do klasy 0 niekorzystnej sytuacji finansowej samorządów gminnych.

Literatura

- Bezdek J.C., 1973, *Fuzzy mathematics in pattern classification*, Ph.D. dissertation, Cornell University, Ithaca, NY.
- Chomątowski S., Sokołowski A., 1978, *Taksonomia struktur*, Przegląd Statystyczny, nr 2, s. 217–225.
- Dunn J.C., 1973, *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters*, Cybernetics and Systems, vol. 3, no. 3, s. 32–57.
- Giordani P., Ferraro M.B., 2015, Package 'fclust', <https://cran.r-project.org/web/packages/fclust/fclust.pdf?>
- GUS, 2013, *Bank Danych Lokalnych*, www.stat.gov.pl/bdl (1.06.2015).
- Jajuga K., 1984, *O pewnym uogólnieniu zagadnienia klasyfikacji*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 285, Statystyka – Ekonometria, s. 5–18.
- Jajuga K., 1993, *Statystyczna analiza wielowymiarowa*, Polskie Wydawnictwo Statystyczne, Warszawa.
- Krishnapuram R., Joshi A., Nasraoui O., 2001, *Low-complexity fuzzy relational clustering algorithms for web mining*, IEEE Transactions on Fuzzy Systems, vol. 9, no. 4, s. 595–607.
- Maulik U., Bandyopadhyay S., Mukhopadhyay A., 2011, *Multiobjective genetic algorithms for clustering*, Springer, Heidelberg.
- Mei J.-P., Chen L., 2010, *Fuzzy clustering with weighted medoids for relational data*, Pattern Recognition, vol. 43, no. 9, s. 1964–1974.
- Ministerstwo Finansów, 2014, *Wskaźniki do oceny sytuacji finansowej jednostek samorządu terytorialnego w latach 2011–2013*, www.finanse.mf.gov.pl (1.06.2015).
- Nowak E., 1990, *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa.
- Oliveira J.V. de, Pedrycz W. (red.), 2007, *Advances in Fuzzy Clustering and Its Applications*, John Wiley and Sons, New York.
- Walesiak M., Dudek, A., 2006, *Symulacyjna optymalizacja wyboru procedury klasyfikacyjnej dla danego typu danych – oprogramowanie komputerowe i wyniki badań*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 1126, s. 120–129.
- Wawrzyniak A., 2010, *Klasyfikacje rozmyte w modelowaniu zjawisk ekonomicznych*, Studies & Proceedings of Polish Association for Knowledge Management, tom 32, s. 81–91.
- Wysocki F., 2010, *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*, Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu, Poznań.
- Zadeh L.A., 1965, *Fuzzy sets*, Information and Control, vol. 8, no. 3, s. 338–353.