

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 384

Taksonomia 24

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

Redaktorzy naukowi

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

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 stronie internetowej Wydawnictwa
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 2015

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

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
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
Krzysztof Jajuga, Józef Pociecha, Marek Walesiak: 25 lat SKAD.....	15
Beata Basiura, Anna Czapkiewicz: Symulacyjne badanie wykorzystania entropii do badania jakości klasyfikacji.....	25
Andrzej Bąk: Zagadnienie wyboru optymalnej procedury porządkowania liniowego w pakiecie <code>pllord</code>	33
Justyna Brzezińska: Analiza klas ukrytych w badaniach sondażowych.....	42
Grażyna Dehnel: Rejestr podatkowy oraz rejestr ZUS jako źródło informacji dodatkowej dla statystyki gospodarczej – możliwości i ograniczenia ..	51
Sabina Denkowska: Wybrane metody oceny jakości dopasowania w <i>Propensity Score Matching</i>	60
Marta Dziechciarz-Duda, Klaudia Przybysz: Zastosowanie teorii zbiorów rozmytych do identyfikacji pozafiskalnych czynników ubóstwa.....	75
Iwona Foryś: Potencjał rynku mieszkaniowego w Polsce w latach dekonjunktury gospodarczej.....	84
Eugeniusz Gatnar: Statystyczna analiza konwergencji krajów Europy Środkowej i Wschodniej po 10 latach członkostwa w Unii Europejskiej.....	93
Ewa Genge: Zaufanie do instytucji publicznych i finansowych w polskim społeczeństwie – analiza empiryczna z wykorzystaniem ukrytych modeli Markowa.....	100
Alicja Grześkowiak: Wielowymiarowa analiza uwarunkowań zaangażowania Polaków w kształcenie ustawiczne o charakterze pozaformalnym.....	108
Monika Hamerska: Wykorzystanie metod porządkowania liniowego do tworzenia rankingu jednostek naukowych.....	117
Bartłomiej Jefmański: Zastosowanie modeli IRT w konstrukcji rozmytego systemu wag dla zmiennych w zagadnieniu porządkowania liniowego – na przykładzie metody TOPSIS.....	126
Tomasz Józefowski, Marcin Szymkowiak: Wykorzystanie uogólnionej miary odległości do porządkowania liniowego powiatów województwa podkarpackiego w świetle funkcjonowania specjalnej strefy ekonomicznej Euro-Park Mielec.....	135
Krzysztof Kompa: Zastosowanie testów parametrycznych i nieparametrycznych do oceny sytuacji na światowym rynku kapitałowym przed kryzysem i po jego wystąpieniu.....	144
Mariusz Kubus: Rekurencyjna eliminacja cech w metodach dyskryminacji....	154

Marta Kuc: Wpływ sposobu definiowania macierzy wag przestrzennych na wynik porządkowania liniowego państw Unii Europejskiej pod względem poziomu życia ludności	163
Paweł Lula: Kontekstowy pomiar podobieństwa semantycznego	171
Iwona Markowicz: Model regresji Feldsteina-Horioki – wyniki badań dla Polski	182
Kamila Migdał-Najman: Ocena wpływu wartości stałej Minkowskiego na możliwość identyfikacji struktury grupowej danych o wysokim wymiarze	191
Małgorzata Misztal: O zastosowaniu kanonicznej analizy korespondencji w badaniach ekonomicznych.....	200
Krzysztof Najman: Zastosowanie przetwarzania równoległego w analizie skupień	209
Edward Nowak: Klasyfikacja danych a rachunkowość. Rozważania o relacjach	218
Marcin Pelka: Adaptacja metody <i>bagging</i> z zastosowaniem klasyfikacji pojęciowej danych symbolicznych.....	227
Józef Pocięcha, Mateusz Baryła, Barbara Pawelek: Porównanie skuteczności klasyfikacyjnej wybranych metod prognozowania bankructwa przedsiębiorstw przy losowym i nielosowym doborze prób	236
Agnieszka Przedborska, Małgorzata Misztal: Wybrane metody statystyki wielowymiarowej w ocenie jakości życia słuchaczy uniwersytetu trzeciego wieku	246
Wojciech Roszka: Konstrukcja syntetycznych zbiorów danych na potrzeby estymacji dla małych domen	254
Aneta Rybicka: Połączenie danych o preferencjach ujawnionych i wyrażonych	262
Elżbieta Sobczak: Poziom specjalizacji w sektorach intensywności technologicznej a efekty zmian liczby pracujących w województwach Polski	271
Andrzej Sokołowski, Grzegorz Harańczyk: Modyfikacja wykresu radarowego	280
Marcin Szymkowiak, Marek Witkowski: Wykorzystanie mediany do klasyfikacji banków spółdzielczych według stanu ich kondycji finansowej ..	287
Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski: Wpływ wyboru metody klasyfikacji na identyfikację zależności przestrzennych – zastosowanie testu <i>join-count</i>	296
Dorota Witkowska: Wykorzystanie drzew klasyfikacyjnych do analizy zróżnicowania płac w Niemczech	305
Artur Zaborski: Analiza niesymetrycznych danych preferencji z wykorzystaniem modelu punktu dominującego i modelu grawitacji.....	315

Summaries

Krzysztof Jajuga, Józef Pocięcha, Marek Walesiak: XXV years of SKAD	24
Beata Basiura, Anna Czapkiewicz: Simulation study of the use of entropy to validation of clustering.....	32
Andrzej Bąk: Problem of choosing the optimal linear ordering procedure in the p_llord package.....	41
Justyna Brzezińska-Grabowska: Latent class analysis in survey research...	50
Grażyna Dehnel: Tax register and social security register as a source of additional information for business statistics – possibilities and limitations.....	59
Sabina Denkowska: Selected methods of assessing the quality of matching in Propensity Score Matching	74
Marta Dziechciarz-Duda, Klaudia Przybysz: Applying the fuzzy set theory to identify the non-monetary factors of poverty.....	83
Iwona Foryś: The potential of the housing market in Poland in the years of economic recessions.....	92
Eugeniusz Gatnar: Statistical analysis of the convergence of CEE countries after 10 years of their membership in the European Union.....	99
Ewa Genge: Trust to the public and financial institutions in the Polish society – an application of latent Markov models.....	107
Alicja Grześkowiak: Multivariate analysis of the determinants of Poles' involvement in non-formal lifelong learning	116
Monika Hamerska: The use of the methods of linear ordering for the creating of scientific units ranking.....	125
Bartłomiej Jefmański: The application of IRT models in the construction of a fuzzy system of weights for variables in the issue of linear ordering – on the basis of TOPSIS method	134
Tomasz Józefowski, Marcin Szymkowiak: GDM as a method of finding a linear ordering of districts of Podkarpackie Voivodeship in the light of the operation of the Euro-Park Mielec special economic zone	143
Krzysztof Kompa: Application of parametric and nonparametric tests to the evaluation of the situation on the world financial market in the pre- and post-crisis period.....	153
Mariusz Kubus: Recursive feature elimination in discrimination methods ...	162
Marta Kuc: The impact of the spatial weights matrix on the final shape of the European Union countries ranking due to the standard of living.....	170
Paweł Lula: The impact of context on semantic similarity.....	181
Iwona Markowicz: Feldstein-Horioka regression model – the results for Poland.....	190

Kamila Migdal-Najman: The assessment of impact value of Minkowski's constant for the possibility of group structure identification in high dimensional data.....	199
Małgorzata Misztal: On the use of canonical correspondence analysis in economic research.....	208
Krzysztof Najman: The application of the parallel computing in cluster analysis.....	217
Edward Nowak: Data classification and accounting. A study of correlations	226
Marcin Pelka: The adaptation of bagging with the application of conceptual clustering of symbolic data.....	235
Józef Pociecha, Mateusz Baryła, Barbara Pawelek: Comparison of classification accuracy of selected bankruptcy prediction methods in the case of random and non-random sampling technique.....	244
Agnieszka Przedborska, Małgorzata Misztal: Selected multivariate statistical analysis methods in the evaluation of the quality of life of the members of the University of the Third Age.....	253
Wojciech Roszka: Construction of synthetic data sets for small area estimation.....	261
Aneta Rybicka: Combining revealed and stated preference data.....	270
Elżbieta Sobczak: Specialization in sectors of technical advancement vs. effects of workforce number changes in Poland's voivodships.....	279
Andrzej Sokółowski, Grzegorz Harańczyk: Modification of radar plot.....	286
Marcin Szymkowiak, Marek Witkowski: Classification of cooperative banks according to their financial situation using the median.....	295
Justyna Wilk, Michał B. Pietrzak, Roger S. Bivand, Tomasz Kossowski: The influence of classification method selection on the identification of spatial dependence – an application of join-count test.....	304
Dorota Witkowska: Application of classification trees to analyze wages disparities in Germany.....	314
Artur Zaborski: Asymmetric preference data analysis by using the dominance point model and the gravity model.....	323

Andrzej Sokółowski

Uniwersytet Ekonomiczny w Krakowie
e-mail: andrzej.sokolowski@uek.krakow.pl

Grzegorz Harańczyk

StatSoft Polska
e-mail: gharancyk@gmail.com

MODYFIKACJA WYKRESU RADAROWEGO¹

Streszczenie: Wykres radarowy to jeden z najpopularniejszych wykresów pozwalających ilustrować dane wielowymiarowe, stosowany w wielu dziedzinach wiedzy. Na osiach regularnie wychodzących z jednego punktu, przy kącie pomiędzy sąsiednimi osiami równym $360^\circ/\text{liczba cech}$, zaznacza się wartości poszczególnych cech uprzednio znormalizowane do przedziału $[0,1]$. Punkty na sąsiednich osiach łączy się i w ten sposób powstaje nieregularny wielokąt. Proponowana modyfikacja przewiduje:

- poszukiwanie optymalnego porządku osi,
- odejście od równych kątów między osiami.

W pracy przedstawiono różne algorytmy poszukiwania optymalnego układu osi zmodyfikowanego wykresu radarowego. Jako przykład zastosowania pokazano wykorzystanie tego wykresu do oceny odporności państw na kryzys ekonomiczny w zależności od identyfikatorów inteligentnego rozwoju.

Słowa kluczowe: wykres radarowy, dane wielowymiarowe, metody graficzne.

DOI: 10.15611/pn.2015.384.30

1. Propozycja modyfikacji

1.1. Wprowadzenie

Wykres radarowy został po raz pierwszy użyty przez Georga von Mayra w pracy *Die Gesetzmäßigkeit im Gesellschaftsleben*, opublikowanej w Oldenbourg w 1877. Od tego czasu wykorzystywany jest jako narzędzie ilustracyjne w wielu zagadnieniach statystyki opisowej. Najczęstszym zastosowaniem jest wykorzystanie wykresu radarowego do ilustrowania danych wielowymiarowych. Z punktu centralnego

¹ Praca finansowana ze środków NCN (DEC-2013/09/B/HS4/00509).

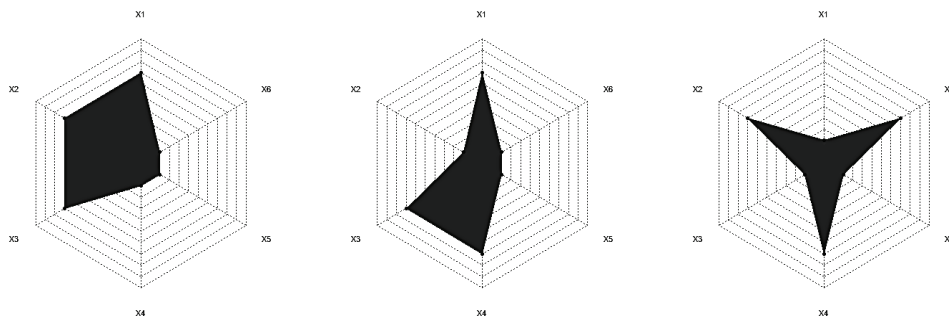
wychodzi tyle osi, ile jest cech statystycznych (m). Kąty pomiędzy osiami są równe i mają po $360/m$. Cechy statystyczne zostają znormalizowane do wspólnego przedziału niemianowanego – zazwyczaj $[0;1]$. W tym momencie można (choć nie jest to konieczne) zamienić destymulanty na stymulanty. Najczęściej wykorzystuje się tu wzory

$$\text{dla stymulant} \quad \rightarrow \quad x'_{ij} = \frac{x_{ij} - \min_i \{x_{ij}\}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}},$$

$$\text{dla destymulant} \quad \rightarrow \quad x'_{ij} = \frac{\max_i \{x_{ij}\} - x_{ij}}{\max_i \{x_{ij}\} - \min_i \{x_{ij}\}}.$$

Na osiach zaznacza się punkty odpowiadające wartościom znormalizowanym. Po połączeniu tych punktów odcinkami otrzymujemy wykres radarowy.

Na rysunku 1 przedstawiono trzy wykresy ilustrujące ten sam obiekt, dla tych samych danych znormalizowanych, a różniące się kolejnością cech przyporządkowanych kolejnym osiom.



Rys. 1. Przykładowe wykresy radarowe dla tych samych danych

Źródło: obliczenia własne.

Takich różnych wykresów w ilustrowanym przykładzie można sporządzić aż 60, gdyż liczba możliwych uporządkowań osi wynosi $(m - 1)!/2$.

1.2. Propozycja modyfikacji wykresu radarowego

Modyfikacje wykresu radarowego mogą iść w dwóch kierunkach:

- Optymalizacja rozmieszczenia cech na osiach. Osie cech skorelowanych dodatnio powinny na wykresie być położone „blisko” siebie, natomiast osie cech skorelowanych ujemnie – „daleko”.

– Optymalizacja kątów. Odejdźcie od równych kątów między osiami na rzecz kątów proporcjonalnych do siły korelacji między cechami.

Strategia wspomnianych powyżej optymalizacji obejmuje cztery kroki:

1. Przygotowanie tzw. bazy, czyli macierzy danych X o n wierszach (obiekty) i m cechach (kolumny).

2. Wyznaczanie macierzy korelacji cech R na podstawie macierzy X .

3. Wyznaczanie macierzy odległości D , gdzie elementy tej macierzy są równe $d_{ij} = \arccos(r_{ij})$ na podstawie macierzy R .

4. Poszukiwanie takiego wielowymiarowego układu współrzędnych biegunowych, że wyznaczona nań macierz odległości D^* jest najbardziej zbliżona do macierzy D .

Pierwszą narzucającą się możliwością realizacji punktu (4) jest przeszukanie wszystkich możliwych przyporządkowań zmiennych osiom w układzie symetrycznym, a następnie zmiana kątów. To podejście nie zapewnia jednak znalezienia minimum globalnego (najlepszej zgodności macierzy D oraz D^*), a ponadto liczba możliwych uporządkowań rośnie lawinowo wraz ze wzrostem liczby cech.

Poszukiwanie rozkładu punktów na płaszczyźnie tak, aby macierz D^* była jak najbardziej podobna do macierzy wyznaczonej z przestrzeni oryginalnej, to domena skalowania wielowymiarowego. Jednak tam odległości wyznaczone są w przestrzeniach euklidesowych, a nie w biegunowym układzie współrzędnych.

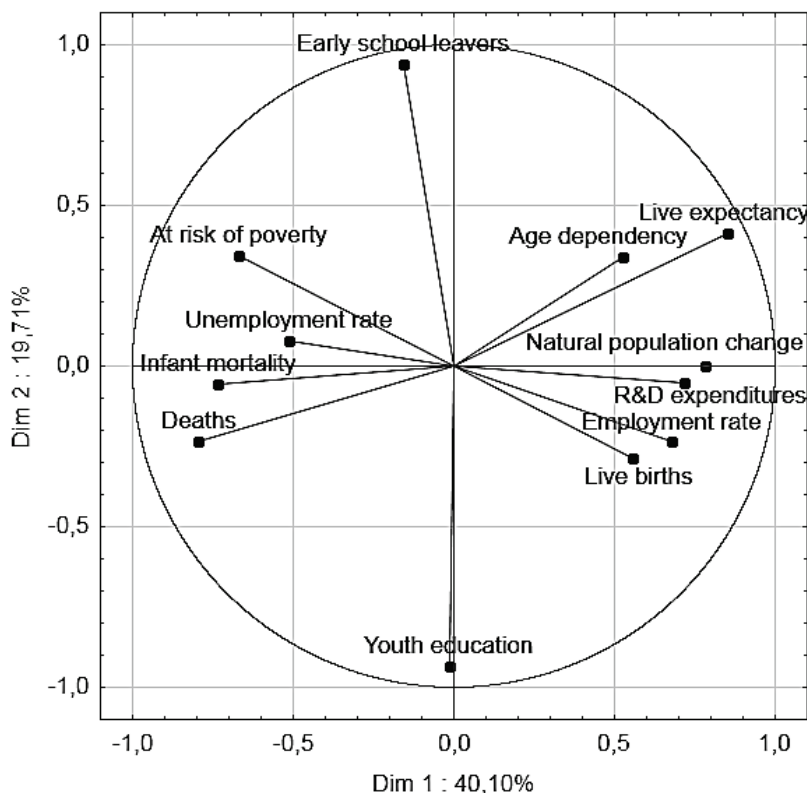
Kolejna zbliżona metoda, jaką jest biplot, umożliwia jednocześnie pokazanie na rysunku zarówno cech, jak i obiektów. Może być on sporządzony z wykorzystaniem różnych metod, jednak cel jest inny niż w naszym przypadku. W biplocie współrzędne punktów reprezentujących obiekty są określane poprzez rzutowanie punktów na osie reprezentujące cechy. W naszym zagadnieniu współrzędne obiektów wyznaczone są bezpośrednio na osiach.

Wydaje się, że metodą najbardziej odpowiadającą sugerowanej metodzie optymalizacji jest analiza głównych składowych. Jej idea to transformacja układu współrzędnych w przestrzeni wielowymiarowej na taki układ w przestrzeni o niższej liczbie wymiarów, aby ten nowy układ współrzędnych (w klasycznym ujęciu są to współrzędne ortogonalne) pozwalał wyjaśnić jak największy procent wariancji oryginalnego układu współrzędnych. Ponieważ wykres radarowy sporządzany jest na płaszczyźnie, proponujemy wykorzystanie tylko dwóch pierwszych głównych składowych. Tzw. ładunki są odpowiednikami współczynników korelacji zmiennych oryginalnych ze składowymi głównymi. Wartości ładunków wyznaczają współrzędne punktów, a kolejne osie wykresu radarowego to półproste przechodzące przez te punkty, a wychodzące z początku układu współrzędnych. Taki sposób konstrukcji wykresu radarowego czyni zadość postulatowi optymalizacji porządku i kątów między osiami. Zmienne skorelowane ujemnie są reprezentowane przez osie skierowane w przeciwnych kierunkach, zaś mocne korelacje dodatnie skutkują małymi kątami ostrymi między osiami. Zauważmy ponadto, że pomnożenie wartości ładunków czynnikowych przez -1 nie zmienia globalnego układu osi i kątów.

2. Przykład empiryczny

Do zilustrowania zmodyfikowanego wykresu radarowego wykorzystano dane wzięte z bazy EUROSTATU, dotyczące 27 krajów Unii Europejskiej w roku 2010. Wzięto pod uwagę następujące cechy statystyczne:

- współczynnik urodzeń żywych na 1000 mieszkańców (*Live birth*),
- współczynnik zgonów na 1000 mieszkańców (*Deaths*),
- współczynnik zgonów niemowląt na 1000 urodzeń żywych (*Infant mortality*),
- przyrost naturalny (*Natural population change*),
- przeciętne dalsze trwanie życia (*Life expectancy*),
- liczba osób wieku nieprodukcyjnym (0-14 oraz od 65) przypadająca na 100 osób w wieku produkcyjnym (15-64) (*Age dependency*),
- zatrudnieni jako procent populacji w wieku 20-64 lata (*Employment rate*),



Rys. 2. Układ współrzędnych biegunowych do sporządzenia wykresów radarowych dla rozważanego przykładu

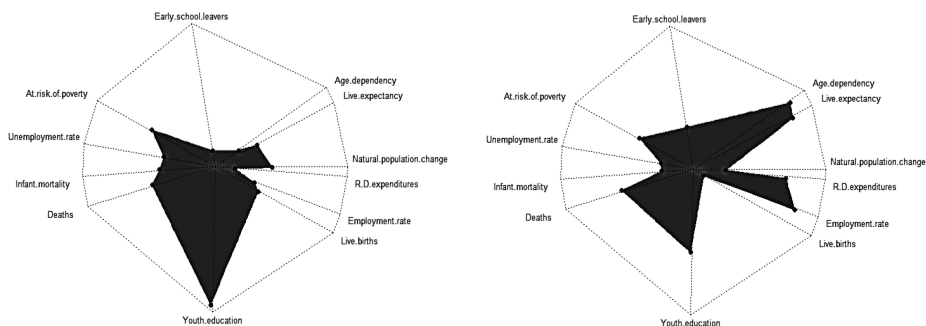
Źródło: obliczenia własne.

- stopa bezrobocia (*Unemployment rate*),
- procent osób o wykształceniu przynajmniej zawodowym w grupie wieku 20-24 lata (*Youth education*),
- procent osób w grupie wieku 18-24 lata, które zakończyły edukację na poziomie równoważnym z naszym gimnazjum (*Early school leavers*),
- procent osób zagrożonych ubóstwem (*At risk of poverty*),
- wydatki na badania i rozwój w procencie PKB (*R&D expenditures*).

Na ilustracjach osie zostały podpisane skrótowymi nazwami w języku angielskim odpowiadającymi informacyjnie pełnym definicjom EUROSTATU.

Oryginalna przestrzeń była 12-wymiarowa, a dane ze wszystkich 27 krajów wykorzystano do wyznaczenia macierzy **R**. Na jej podstawie przeprowadzono analizę głównych składowych. Wyodrębniono dwie pierwsze składowe główne, a wartości ładunków wyznaczają układ osi przedstawiony na rys. 2. Jakość optymalizacji jest mierzona procentem wyjaśnianej wariancji.

Zauważmy, że osie zmiennych skorelowanych dodatnio (np. zagrożenie ubóstwem, stopa bezrobocia, śmiertelność niemowląt, współczynnik zgonów) tworzą swojego rodzaju wiązki osi, między którymi kąty są małe. Kąt zaś między osiami reprezentującymi zmienne naturalnie mocno skorelowane ujemnie (procent młodych ludzi kończących edukację na niższym szczeblu średnim oraz procent młodych o wykształceniu przynajmniej zawodowym) jest równy niemal 180° .



Rys. 3. Zmodyfikowane wykresy radarowe dla Niemiec i Polski

Źródło: obliczenia własne.

Na rysunku 3 przedstawiono dwa wykresy radarowe dla dwóch krajów UE na układzie współrzędnych zdefiniowanym powyżej. Wykres po lewej stronie dotyczy Niemiec, zaś po prawej stronie – Polski.

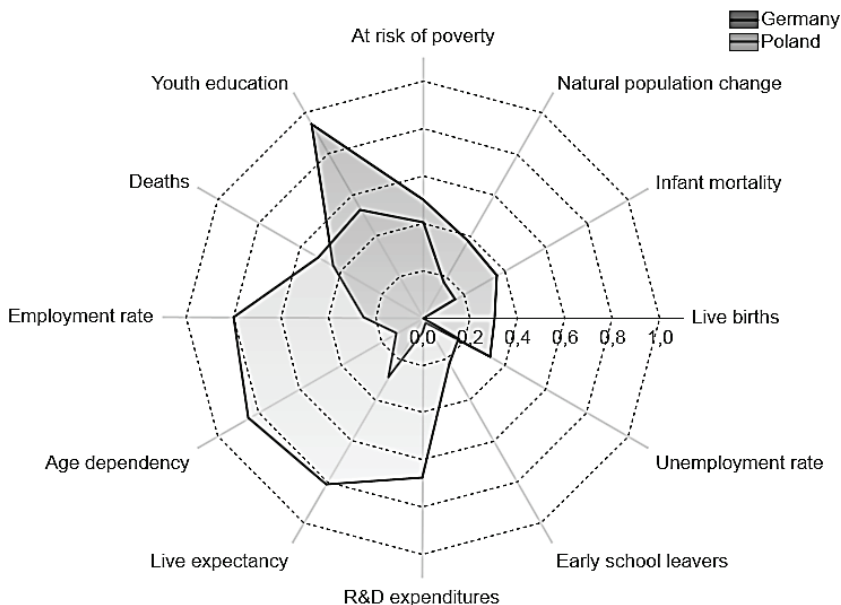
3. Optymalizacja klasycznego wykresu radarowego dla celów porównawczych

Przez pojęcie klasycznego wykresu radarowego rozumiemy wykres, w którym kąty między osiami są równe. Optymalizacji podlega wówczas tylko przyporządkowanie zmiennych osiom. Jeżeli porównujemy wykresy radarowe odpowiadające dwóm obiektom, to różnica między nimi będzie najlepiej widoczna, jeżeli część wspólna dwóch wykresów będzie najmniejsza. Tutaj niestety nie potrafimy zaproponować żadnego bardziej efektywnego algorytmu, poza przeszukiwaniem wszystkich możliwych uporządkowań zmiennych (osi) na okręgu. Dla przykładu omawianego powyżej wszystkich różnych uporządkowań jest prawie 20 milionów. Jeżeli chcemy zoptymalizować układ osi dla porównania dwóch wykresów radarowych A oraz B, to powinniśmy maksymalizować wielkość S.

$$S = \frac{Pole A + Pole B - 2 \cdot Pole (A \cap B)}{Pole (A \cap B)}$$

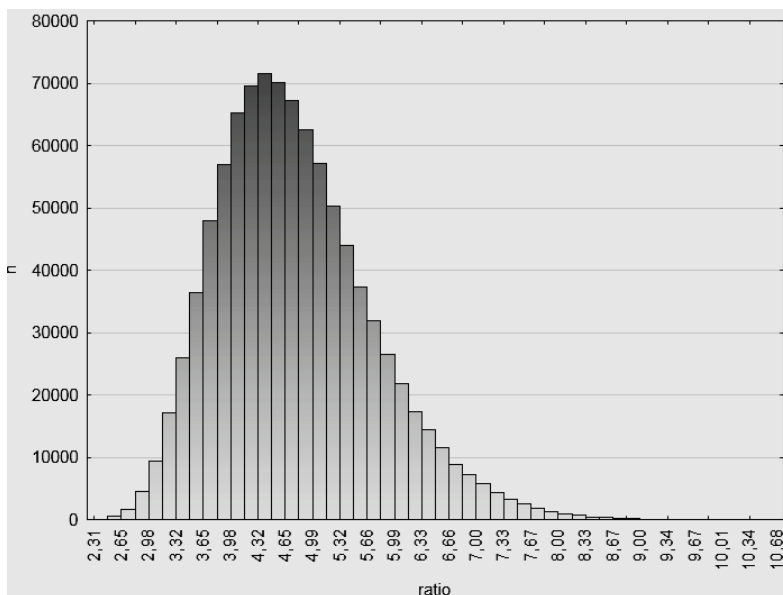
S to stosunek części niepokrywających się do części wspólnej. Zoptymalizowany układ osi dla porównania Niemiec i Polski przedstawia rys. 4.

Przeszukiwanie wszystkich możliwych układów osi dla dużych m jest czasochłonne, dlatego trwają prace nad opracowaniem efektywnego algorytmu znajdowania optymalnego układu osi. Symulacje pokazują, że rozkłady wielkości S mają



Rys. 4. Zoptymalizowany klasyczny wykres radarowy dla porównania Niemiec i Polski

Źródło: obliczenia własne.



Rys. 5. Rozkład wartości kryterium S dla porównania Niemiec i Polski

Źródło: obliczenia własne.

duży rozrzut i są prawostronnie asymetryczne. Na rysunku 5 zaprezentowano rozkład wielkości S dla 1 miliona różnych układów osi.

Gładki histogram sugeruje bardzo dobrą aproksymację rozkładu kryterium S. Jeżeli przy dużej liczbie cech ocenimy tylko część losowo wybranych układów osi, to i tak jest mało prawdopodobne, że rzeczywiste maksimum S będzie się różniło od tego *quasi*-optymalnego w sposób znaczny.

MODIFICATION OF RADAR PLOT

Summary: Radar plot is one of the most popular graphs used to illustrate the multidimensional data. It is used in various scientific disciplines. Axes emerging from the centre of the graph for coordinates with angles between consecutive ones equal $360^\circ/\text{number of variables}$. Variables are normalized to unit interval and for given objects its values define points on axes. While joining them we get the radar plot. The following modifications are proposed in the paper:

- Finding the best ordering of variables (their assignment to axis).
- Making angles between axes proportional to distances calculated through the correlation analysis.

Some optimization algorithms are discussed in the paper. An empirical example used here concerns EU countries data on smart development.

Keywords: radar plot, multivariate data.

Marcin Szymkowiak, Marek Witkowski

Uniwersytet Ekonomiczny w Poznaniu

e-mails: m.szymkowiak@ue.poznan.pl; marek.witkowski@ue.poznan.pl

WYKORZYSTANIE MEDIANY DO KLASYFIKACJI BANKÓW SPÓŁDZIELCZYCH WEDŁUG STANU ICH KONDYCJI FINANSOWEJ

Streszczenie: W artykule przeprowadzono klasyfikację banków spółdzielczych należących do zrzeszenia regionalnego SGB pod względem stanu ich kondycji finansowej w kilkuletnim horyzoncie czasowym. Dzięki temu możliwe było ustalenie, które z banków najlepiej wykorzystały szanse i poprawiły swoją konkurencyjność na rynku. Do przeprowadzenia przedmiotowej klasyfikacji wykorzystano metodę opartą na medianie zaproponowaną przez D. Strahl. Wydaje się to zasadne, gdyż banki spółdzielcze charakteryzują się mocno zróżnicowaną skalą i strukturą swojej działalności.

Słowa kluczowe: kondycja finansowa banków spółdzielczych, klasyfikacja banków spółdzielczych, metody klasyfikacji.

DOI: 10.15611/pn.2015.384.31

1. Wstęp

Segment banków spółdzielczych jest jedną z podstawowych składowych sekcji bankowego w Polsce. Jego udział w tym sektorze, jakkolwiek ciągle jeszcze jest niewielki, to jednak systematycznie rośnie. Rozwój technologii informatycznych w bankach spółdzielczych spowodował, że mogły one podnieść jakość swoich operacji bankowych, usprawnić procedury rozliczeniowe oraz stworzyć nowoczesne produkty bankowe i kanały dystrybucji. Wskutek tego banki spółdzielcze jeszcze bardziej, niż to było do niedawna, różnią się między sobą wielkością oraz skalą i strukturą swojej działalności, a tym samym i kondycją finansową.

Celem artykułu jest przeprowadzenie klasyfikacji banków spółdzielczych należących do zrzeszenia SGB-bank pod względem ich kondycji finansowej w czteroletnim horyzoncie czasowym. Dzięki temu możliwe będzie ustalenie, które z badanych banków najlepiej wykorzystują swoje szanse i poprawiają swoją konkuren-

cyjność na rynku, a których siła finansowa była słaba bądź stawała się coraz słabsza w analizowanym przedziale czasowym. Będziemy mogli zidentyfikować też mocne i słabe zmienne diagnostyczne, od których zależy przede wszystkim pozycja finansowa badanych banków. Zrealizowanie założonych wyżej celów wymaga zastosowania odpowiedniej metody postępowania badawczego. Metoda ta powinna być wybrana zgodnie z kardynalną zasadą poznania naukowego mówiącą, że przedmiot badania i jego właściwości determinują metodę badania. Sądzimy w kontekście powyższego, że zastosowanie do przeprowadzenia przedmiotowej klasyfikacji metody opartej na medianie będzie w pełni zasadne. Mediana kumuluje bowiem w sobie wrażliwość na dyspersję cech diagnostycznych w każdym obiekcie badanym oraz uwzględnia pozycyjną wartość tych cech w poszczególnych obiektach [Strahl (red.) 2006, s. 187]. Praca stanowi kontynuację badań autorów dotyczących weryfikacji przydatności metod WAS do badania kondycji finansowej banków spółdzielczych. Niektóre wyniki tych badań były prezentowane również na konferencjach SKAD.

2. Metoda badania

Przeprowadzenie klasyfikacji banków spółdzielczych należących do przedmiotowego zrzeszenia ze względu na stan ich kondycji finansowej wymaga, w pierwszej kolejności, wyspecyfikowania zmiennych diagnostycznych tę kondycję opisujących, jako że kondycja finansowa jest, ze statystycznego punktu widzenia, zjawiskiem złożonym [Jajuga 1993, s.15]. W związku z tym wybrano najpierw potencjalny zestaw cech diagnostycznych uwzględniając podstawowe obszary działalności bankowej, a mianowicie obszar operacyjny i finansowy. Obszar operacyjny charakteryzuje rentowność i wykorzystanie zasobów banku, obszar finansowy zaś skoncentrowany jest na płynności banku i jego bezpieczeństwie finansowym [Grabczan 1996, s.136]. Przy jego specyfikacji wzięto również pod uwagę ograniczenia leżące po stronie dostępności materiału źródłowego.

Dla określenia optymalnego, w danych warunkach, wektora zmiennych diagnostycznych wykorzystano metodę ortogonalizacji zmiennych [Malina, Zeliaś 1997, s. 245]. Metodę tę zastosowano oddzielnie dla każdego roku. W skład optymalnego wektora zmiennych diagnostycznych weszły te zmienne, które powtarzały się najczęściej w badanych latach.

Wybrany zgodnie z powyższą procedurą finalny wektor zmiennych diagnostycznych posłużył do przeprowadzenia klasyfikacji banków według stanu ich kondycji finansowej. Zastosowano metodę zaproponowaną przez D. Strahl [Strahl 2008, s. 9-12]. Proces klasyfikacji według tej metody przebiega następująco:

1) zakładamy, że każdy obiekt P_i ($i = 1, 2, \dots, n$) opisany jest wektorem m zmiennych diagnostycznych X_j ($j = 1, 2, \dots, m$) w t jednostkach czasu ($t=1, 2, \dots, T$),

2) wśród wybranych zmiennych diagnostycznych muszą występować tylko stymulanty. Jeśli występują destymulanty i nominanty, muszą być przekształcone na stymulanty. Żeby tego dokonać można zastosować jedną z metod normalizacji. W pracy zastosowano metodę unitaryzacji zerowanej, która pozwala sprowadzić zmienne o różnym charakterze do porównywalności [Kukuła 2000, s. 226]. Przekształcenia tego dokonuje się oddzielnie dla każdej jednostki czasowej t . Wtedy każdy obiekt P_i opisany jest w jednostce czasowej t macierzą w postaci:

$$P_i(t) = \{z_{ij}(t)\},$$

3) dla każdej zmiennej Z_j w jednostce czasowej t oblicza się medianę $Me(j)$,

4) dokonujemy klasyfikacji obiektów P_i w każdym okresie t oddzielnie według następujących zasad:

- do klasy G_1 wchodzi te obiekty ze zbioru P , w których wartości wszystkich m zmiennych diagnostycznych Z_j są wyższe od $Me(j)$ bądź jej równe,
- do grupy G_2 wchodzi te obiekty ze zbioru P , w których wartości $(m - 1)$ cech diagnostycznych Z_j są wyższe od mediany $Me(j)$ bądź jej równe,
- do grupy G_m wchodzi te obiekty ze zbioru P , w których wartość tylko jednej zmiennej diagnostycznej Z_j jest większa od mediany $Me(j)$ bądź jest jej równa,
- do grupy $G_{(m+1)}$ wchodzi te obiekty ze zbioru P , w których wartość żadnej cechy diagnostycznej Z_j nie jest większa od mediany $Me(j)$ bądź jej równa.

Postępując w podany wyżej sposób, otrzymuje się, jak łatwo zauważyć, podział obiektów badanych na $(m+1)$ rozłącznych grup.

3. Wyniki badania empirycznego

Omówioną w punkcie 2 metodę klasyfikacji zastosowano do przeprowadzenia dynamicznej klasyfikacji banków spółdzielczych należących do zrzeszenia SGB w latach 20RA-20RD¹. Liczba tych banków nie była stała. W każdym badanym roku było ich jednakże ponad 150. Podstawę klasyfikacji stanowiła ich kondycja finansowa. Źródłami danych były uproszczone bilanse i rachunki zysków i strat pochodzące ze wszystkich banków. W pierwszym etapie prowadzonego postępowania badawczego dokonano specyfikacji optymalnego wektora cech diagnostycznych opisujących kondycję przedmiotowych banków.

Optymalny wektor cech diagnostycznych wybrany według procedury podanej w punkcie 2 zawiera z kolei tab. 1. W wektorze tym brak jest współczynnika wypłacalności, który uznawany jest za podstawowy miernik bezpieczeństwa i wiarygodności banku [Świdorski 1999, s. 81-82]. Było to spowodowane wyłącznie brakiem możliwości jego obliczenia dla wszystkich badanych banków.

¹ Na podanie rzeczywistego przedziału czasowego badania nie wyraził zgody bank zrzeszający SGB, który udostępnił autorom dane.

Tabela 1. Optymalny zestaw cech diagnostycznych opisujących kondycję finansową badanych banków spółdzielczych w latach 20RA-20RD

Lp.	Zmienna diagnostyczna	Formuła obliczeniowa
1	wskaźnik kapitału własnego – X_1	kapitał własny/aktywa
2	produktywność wynagrodzeń – X_2	aktywa/wynagrodzenia
3	wskaźnik jakości portfela kredytowego – X_3	kredyty zagrożone/kredyty ogółem
4	wskaźnik dochodu odsetkowego – X_4	dochód odsetkowy/aktywa
5	rentowność kapitału własnego – X_5	zysk netto/kapitał własny
6	jednostkowy koszt kapitału trwałego – X_6	koszty pozostałe/aktywa trwałe

Źródło: zestawienie własne.

Zmienna diagnostyczna X_3 , czyli wskaźnik jakości portfela kredytowego, uwzględnia ryzyko kredytowe banku. Wzrost tego wskaźnika oznacza konieczność zwiększonych odpisów na rezerwy celowe, a tym samym spadek zysku banku [Kopiński 2008, s. 147]. Z kolei zmienna X_1 , to jest wskaźnik kapitału własnego, określa udział stałych źródeł finansowania banku w finansowaniu ogółem i stanowi jego zabezpieczenie przed niewypłacalnością [Kopiński 2008, s.148]. Zmienne diagnostyczne X_4 i X_5 mierzą rentowność działalności bankowej, przy czym X_4 pokazuje rentowność działalności podstawowej banku. Wzrost tych wskaźników jest więc zawsze pożądany. Dwie ostatnie zmienne diagnostyczne, tj. X_2 i X_6 , określają sprawność działania banku. Pierwsza z nich, czyli X_2 pokazuje, jaka jest efektywność pracy personelu banku, druga zaś (X_6) informuje o tym, jaki jest koszt jednostkowy zaangażowanego kapitału fizycznego (trwałego).

Jak z powyższego opisu wynika, wybrane zmienne reprezentują różne obszary działalności bankowej, a mianowicie: rentowność, sprawność działania, bezpieczeństwo banku oraz jakość aktywów i pasywów. Zmienne te spełniały również podstawowe postulaty, jakich wymaga się od cech diagnostycznych, czyli: uniwersalność, mierzalność, dostępność, ekonomiczność i interpretowalność [Zeliaś (red.) 2000, s. 37-38].

W tabeli 2 zaprezentowane zostały wartości mediany obliczone dla poszczególnych cech diagnostycznych w latach 20RA-20RD po uprzednim sprowadzeniu ich do postaci stymulant w drodze przekształcenia unitaryzacyjnego.

Tabela 2. Mediana zmiennych diagnostycznych opisujących kondycję finansową badanych banków spółdzielczych w latach 20RA-20RD

Zmienna	20RA	20RB	20RC	20RD
X_1	0,3428	0,2613	0,2573	0,2603
X_2	0,1683	0,1631	0,2206	0,2552
X_3	0,8973	0,8592	0,8737	0,8902
X_4	0,5786	0,4629	0,4161	0,4385
X_5	0,5225	0,3755	0,3776	0,2495
X_6	0,6899	0,7352	0,7447	0,6828

Źródło: obliczenia własne.

Tabela 3. Klasyfikacja banków spółdzielczych wchodzących w skład zrzeszenia SGB według stanu ich kondycji finansowej w latach 20RA–20RD (grupy G1 i G2)

Lata			
20RA	20RB	20RC	20RD
8539 Grabów	8558 Wierzbiniek	8537 Aleksandrów Kuj.	9251 Działoszyn
8674 Jutrosin	8937 Białośliwie	9271 Wartkowice	8328 Kościerzyna
8328 Kościerzyna	9326 Człuchów	8937 Białośliwie	8174 Tuchola
8175 Nowe n/W	8346 Gniew	9326 Człuchów	9537 Aleksandrów Kuj.
9271 Wartkowice	8674 Jutrosin	9527 Dzierżoniów	9326 Człuchów
8937 Białośliwie	8166 Kcynia	8346 Gniew	8346 Gniew
9326 Człuchów	8367 Międzyrzecz	9377 Gryfino	9075 Kostrzyn
9251 Działoszyn	8173 Nowe n/W	8328 Kościerzyna	8410 Krotoszyn
9491 Grębocin	9423 Ożarów	8410 Krotoszyn	8682 Poniec
8529 Kleczew	8682 Poniec	8943 Lipka	8320 Susz
8943 Lipka	8337 Pszczółki	8557 Malanów	9084 Śrem
9617 Lubcza	8543 Strzałkowo	8682 Poniec	8558 Wierzbiniek
8384 Lwówek Śl.	8168 Świecie n/W	9549 Radziejów	
8324 Sierakowice	9271 Wartkowice	8168 Świecie n/W	
8308 Stegna	8544 Zagórów	8174 Tuchola	
8174 Tuchola		8558 Wierzbiniek	
8558 Wierzbiniek			

Źródło: opracowanie własne.

Tabela 4. Banki spółdzielcze o słabej i bardzo słabej kondycji finansowej (grupy G6 i G7)

Lata			
20RA	20RB	20RC	20RD
8363 Gorzów Wlkp.	8566 Darłowo	8562 Białogard	8566 Darłowo
9048 Mosina	9375 Goleniów	8213 Ciechanów	9375 Goleniów
8159 Strzelno	8144 Koronowo	8566 Darłowo	8678 Gostyń
8164 Szubin	9068 Łubowo	9375 Goleniów	9068 Łubowo
8581 Świdwin	9048 Mosina	8678 Gostyń	9081 Rakoniewice
9511 Toruń	8371 Rzepin	9048 Mosina	8164 Szubin
9393 Wolin	8431 Skalmierzyce	8179 Nakło n/Notecią	8345 Tczew
8876 Wołczyn	8542 Słupca	8369 Ośno Lubuskie	9315 Ustka
8181 Żnin	8345 Tczew	8170 Pruszczy Pom.	8162 Więcbork
8163 Więcbork	9511 Toruń	9085 Środa Wlkp.	9681 Września
	9550 Włocławek	8345 Tczew	8876 Wołczyn
	9393 Wolin	9511 Toruń	
	8876 Wołczyn	9393 Wolin	
	8562 Białogard	9681 Września	
	8935 Jastrowie	8164 Szubin	
	8164 Szubin	8162 Więcbork	
		8876 Wołczyn	

Źródło: opracowanie własne.

Obserwacja zawartych w tab. 2 charakterystyk liczbowych wskazuje, że w badanym okresie zmniejszyła się przeciętna rentowność banków, poprawiła się natomiast ich sprawność działania oraz bezpieczeństwo. Charakterystyki liczbowe zawarte w tab. 2 dały możliwość dokonania klasyfikacji banków przedmiotowego zrzeszenia według stanu ich kondycji finansowej za pomocą omówionej w punkcie 2 metody. Wyniki tej klasyfikacji znajdują się w tab. 3-4.

Ze względu jednakże na dużą liczebność zbiorowości banków w tab. 3-4 przedstawiona została przynależność przedmiotowych banków do grup skrajnych, a więc do grup G1 i G2 oraz G6 i G7. Pierwsze dwie grupy reprezentują banki o najlepszej i bardzo dobrej kondycji finansowej, dwie pozostałe zaś o słabej i bardzo słabej kondycji finansowej.

Otrzymane rezultaty dają asumpt do sformułowania następujących wniosków i spostrzeżeń:

1) w subpopulacji banków o najlepszej i najgorszej kondycji finansowej znalazły się banki położone w różnych województwach Polski Północnej i Zachodniej, w miejscowościach o różnej wielkości. Sugeruje to, iż o kondycji finansowej badanych banków nie decydowało ich usytuowanie przestrzenne,

2) poprawa kondycji finansowej banków, jak również jej pogorszenie w badanym horyzoncie czasowym nie było zdeterminowane wyłącznie przez zmiany sumy bilansowej banków w tym okresie. Okazało się bowiem, że zarówno w grupie banków „najlepszych”, jak i „najgorszych” aktywa wzrosły w analizowanych latach o ponad 100%, nieznacznie bardziej przy tym w bankach z pierwszej grupy. Średni poziom aktywów był wszakże większy w grupie banków „najgorszych” pod względem ich kondycji finansowej,

3) pozycja finansowa banków była w dużym stopniu uzależniona od tego, jaka była w nich dynamika akcji kredytowej i depozytowej oraz relacja między nimi. Otóż, generalnie rzecz biorąc, w grupie banków „najlepszych” znalazły się banki, które nacisk położyły na równomierny przyrost akcji kredytowej i depozytowej. Dynamika akcji kredytowej w bankach o najlepszym stanie kondycji finansowej wynosiła bowiem 153,1%, zaś w podzbiorowości banków „najgorszych” pod tym względem 132,7%. Dynamika ekspozycji depozytowych wynosiła natomiast odpowiednio 153,8% i 104,6%,

4) stan kondycji finansowej był też lepszy w tych bankach spółdzielczych, które bardziej rozwinęły u siebie technologie informatyczne, dzięki czemu mogły wprowadzić nowe produkty bankowe i podnieść jakość oferowanych usług,

5) stopień wykorzystania depozytów do prowadzenia akcji kredytowej w bankach należących do grup G1 i G2 był zdecydowanie większy niż w bankach należących do grup G6 i G7. W tych dwóch pierwszych grupach współczynnik kreatywności wynosił około 86%, podczas gdy w grupach G6 i G7 był on na poziomie 69,7% w roku 20RA i 79,3% w roku 20RD,

6) w bankach należących do podzbiorowości banków o najlepszym stanie kondycji finansowej bardziej korzystnie kształtował się wskaźnik jakości portfela kredy-

towego. Otóż kredyty zagrożone stanowiły w nich średnio 2% kredytów ogółem w roku 20RA i 0,48% kredytów ogółem w roku 20RD. W bankach natomiast należących do grup G6 i G7 wskaźnik ten wynosił odpowiednio 9,3% i 3,5%,

7) przez wszystkie lata w grupie banków o najlepszej kondycji finansowej znajdowały się banki w Człuchowie i Wierzbinku, począwszy zaś od roku 20RB banki spółdzielcze w Gniewie, Pońcu, a od roku 20RC dodatkowo banki w Tucholi, Kościerzynie i Krotoszynie,

8) w podzbiorowości banków o najgorszej sytuacji finansowej we wszystkich latach znajdowały się natomiast banki w Szubinie i Wołczynie, począwszy zaś od roku 20RB dodatkowo banki w Darłowie, Goleniowie i Tczewie,

9) w stosunku do lat wcześniejszych poprawiły swoją sytuację finansową w roku 20RD banki w Wolinie, Toruniu i Mosinie. Przez pierwsze trzy lata znajdowały się one bowiem w subpopulacji banków o najgorszej sytuacji finansowej. Pogorszyły zaś swoją pozycję banki spółdzielcze w Wartkowicach i Białośliwiu.

W tabeli 5 przedstawiona została z kolei struktura badanej zbiorowości banków według stanu ich kondycji finansowej w latach 20RA-20RD.

Tabela 5. Struktura banków spółdzielczych według ich kondycji finansowej w analizowanym horyzoncie czasowym (w %)

Grupa	20RA	20RB	20RC	20RD
G1 (najlepsza kondycja)	3,21	0,65	1,32	1,97
G2 (kondycja bardzo dobra)	7,69	9,15	9,21	5,92
G3 (kondycja dobra)	17,30	24,18	25,00	18,42
G4 (kondycja przeciętna)	36,54	35,95	30,26	45,39
G5 (kondycja poniżej przeciętnej)	28,85	19,61	23,03	21,05
G6 (kondycja słaba)	5,77	8,50	9,21	6,58
G7 (kondycja bardzo słaba)	0,64	1,96	1,97	0,67
Ogółem	100,00	100,00	100,00	100,00

Źródło: opracowanie własne.

Zawarte w tej tabeli wskaźniki struktury wskazują, że najliczniejszą grupą była grupa zawierająca banki o przeciętnej kondycji finansowej. W roku 20RD banków tych było aż 45,39%, to jest o ca 15 p.p. więcej niż w roku poprzednim i o 9 p.p. więcej niż w roku 20RA. Spadła natomiast frakcja banków o najlepszym i bardzo dobrym stanie kondycji finansowej. O ile bowiem w roku 20RA było ich 10,9%, o tyle w roku 20RD tylko 7,9%. Nieznacznie wzrósł z kolei odsetek banków o słabej i bardzo słabej kondycji finansowej.

Generalnie jednak zmiany strukturalne w badanej zbiorowości banków w roku 20RD w porównaniu z rokiem 20RA nie były zbyt duże. Obliczony bowiem wskaźnik zmian strukturalnych wynosił 0,805 [Nowak (red.) 2001, s. 21].

Oznacza to, że badane struktury były w porównywanych okresach podobne do siebie w ponad 80%. Większe różnice miały miejsce w roku 20RD w porównaniu do roku 20RC i 20RB. W pierwszym przypadku podobieństwo struktur wyniosło 72,7% , a w drugim 78,3%.

4. Wnioski

Rezultaty przeprowadzonego badania empirycznego dają asumpt do sformułowania następujących wniosków i spostrzeżeń:

- zastosowanie mediany jako kryterium w procedurze klasyfikacyjnej było uzasadnione faktem występowania mocno zróżnicowanej i o dużym, przeciętnie biorąc, natężeniu dyspersji cech diagnostycznych w analizowanej zbiorowości banków. Dyspersja ta mierzona współczynnikiem zmienności była rzędu 40-50%. Wynikało to z faktu, że banki te charakteryzowały się mocno zróżnicowaną wielkością, a co za tym idzie – skalą i strukturą swojej działalności,
- za zaletę wykorzystanej w badaniu metody klasyfikacji należy uznać fakt, że liczba klas jest jednoznacznie ustalona,
- zastosowana metoda klasyfikacji dała możliwość dokonania zobiektywizowanego podziału przedmiotowych banków na grupy typologiczne, zobiektywizowanego w tym sensie, że uwzględniającego równocześnie wielowymiarowy charakter kondycji finansowej oraz jej zmiany w czasie,
- stan kondycji finansowej badanych banków nie zależał w prostej linii od ich wielkości mierzonej sumą bilansową, jako że w grupie banków o najlepszej sytuacji finansowej nie znalazły się wcale banki o największej sumie bilansowej, zaś w subpopulacji banków o najgorszej kondycji finansowej tylko banki o najmniejszej sumie bilansowej,
- o stanie kondycji finansowej poszczególnych banków nie decydowało ich usytuowanie przestrzenne,
- o przynależności banków do określonych grup w dużym stopniu decydowała dynamika akcji kredytowej i depozytowej oraz relacja między nimi, a także jakość portfela kredytowego.

Literatura

- Grabczan W., 1996, *Rachunkowość menadżerska w zarządzaniu bankiem*, Fundacja Rozwoju Rachunkowości w Polsce, Warszawa.
- Jajuga K., 1993, *Statystyczna analiza wielowymiarowa*, Wydawnictwo Naukowe PWN, Warszawa.
- Kopiński A., 2008, *Analiza finansowa banku*, PWE, Warszawa.
- Kukuła K., 2000, *Metoda unitaryzacji zerowanej*, Wydawnictwo Naukowe PWN, Warszawa.
- Malina A., Zeliaś A., 1997, *O budowie taksonomicznej miary jakości życia*, [w:] *Klasyfikacja i analiza danych. Teoria i zastosowania*, Taksonomia 4, red. Jajuga K., Walesiak M., Wydawnictwo AE, Wrocław.

- Nowak E. (red.), 2001, *Metody statystyczne w analizie działalności przedsiębiorstwa*, PWE, Warszawa.
- Strahl D. (red.), 2006, *Metody oceny rozwoju regionalnego*, Wydawnictwo AE, Wrocław.
- Strahl D., 2008, *Klasyfikacja pozycyjna w analizach ekonomicznych*, [w:] *Zastosowania metod ilościowych*, *Ekonometria* 22, red. Dziechciarz J., Wydawnictwo UE, Wrocław.
- Świdorski J., 1999, *Pomiar poziomu dywersyfikacji portfela i kwantyfikacji wpływu na współczynnik bezpieczeństwa banku*, BFG, Warszawa.
- Zeliaś A., 2000, *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*, Wydawnictwo AE, Kraków.

CLASSIFICATION OF COOPERATIVE BANKS ACCORDING TO THEIR FINANCIAL SITUATION USING THE MEDIAN

Summary: The main goal of the article is to classify cooperative banks belonging to the regional association SGB taking into account their financial situation in the period of several years. Thanks to this approach it was possible to establish which banks had taken the opportunity the best and improved their competitiveness on the market. To achieve the main aim the method proposed by D. Strahl and based on the median was used. It seems reasonable because the cooperative banks are characterized by a diverse scale and the structure of their activity.

Keywords: financial situation of cooperative banks, classification of cooperative banks, classification methods.