

Justyna Wilk

Uniwersytet Ekonomiczny we Wrocławiu

TAKSONOMICZNA ANALIZA RYNKU PRACY WOJEWÓDZTW POLSKI – PODEJŚCIE SYMBOLICZNE

Streszczenie: Celem artykułu jest klasyfikacja województw Polski ze względu na strukturę regionalnego rynku pracy w 2009 r. z wykorzystaniem podejścia symbolicznego. W artykule zaprezentowano sposób reprezentacji zjawisk z wykorzystaniem danych symbolicznych. Następnie przedstawiono metody statystyczne, jakie można zastosować w procesie analizy skupień na podstawie danych symbolicznych. Na podstawie zaproponowanej procedury wyodrębniono 4 klasy województw i zbadano sytuację regionalnych rynków pracy.

Słowa kluczowe: analiza danych symbolicznych, taksonomia, rynek pracy, badania regionalne.

1. Wstęp

Badania regionalnych rynków pracy prowadzi się najczęściej z wykorzystaniem wskaźników bazujących na danych udostępnianych przez statystykę publiczną. Problemem wielu analiz tego typu jest precyzja opisu badanych zjawisk. Dane regionalne są wartościami „uśrednionymi” i nie uwzględniają struktury terytorialnej zjawiska. Może to zniekształcać obraz rzeczywistej sytuacji regionu.

Propozycją rozwiązania tego problemu jest zastosowanie podejścia symbolicznego, w którym istnieje możliwość:

- opisanie obiektów (np. województw) za pomocą zmiennych symbolicznych o realizacjach w postaci przedziałów wartości, zbiorów kategorii lub struktur procentowych,
- reprezentacji obiektów niższego rzędu (np. powiatów) za pomocą obiektów wyższego rzędu (np. województw).

Taki sposób interpretacji zjawisk wymaga wprowadzenia specyficznych rozwiązań metodologicznych. W analizie taksonomicznej zastosowanie będą miały przede wszystkim metody zaproponowane na gruncie analizy danych symbolicznych, a także niektóre metody opracowane dla danych w ujęciu klasycznym.

Celem badania jest klasyfikacja województw Polski ze względu na strukturę regionalnego rynku pracy w 2009 r. na podstawie danych symbolicznych. W artykule zaprezentowane zostaną:

- sposób reprezentacji zjawisk z wykorzystaniem danych symbolicznych,
- metody, jakie można zastosować w analizie skupień na podstawie danych symbolicznych,
- propozycja zastosowania podejścia symbolicznego w badaniach regionalnych.

2. Sposoby agregacji danych

W ujęciu klasycznym obiektem jest elementarna jednostka badania (np. kraj, respondent, produkt). Obiekty opisywane są za pomocą zmiennych metrycznych lub niemetrycznych. Realizacją zmiennej może być pojedyncza kategoria lub wartość liczbową, a zbiór obserwacji zapisuje się w macierzy danych.

Cechą podejścia symbolicznego jest możliwość agregacji danych w ujęciu klasycznym do postaci danych symbolicznych. Realizacją zmiennej symbolicznej jest przedział wartości, struktura procentowa lub zbiór kategorii. Obiekty opisane zmiennymi symbolicznymi określa się jako obiekty symboliczne. Zbiór obserwacji zapisuje się w tablicy danych symbolicznych.

Wyróżnić można agregację zmiennych oraz agregację obiektów. Agregacja zmiennych jest uzasadniona w sytuacjach, gdy celem jest uzyskanie syntetycznego opisu zjawiska bądź redukcja liczby zmiennych. Agregacji można poddać zmienne zarówno metryczne, jak i niemetryczne [Wilk 2010, s. 89]. W tabeli 1 zawarto informacje na temat ludności województw w 2009 r. z wyróżnieniem ekonomicznych grup

Tabela 1. Przykład agregacji zmiennych

Województwo	Zmienne metryczne			Zmienna symboliczna
	Ludność w wieku przedprodukcyjnym	Ludność w wieku produkcyjnym	Ludność w wieku poprodukcyjnym	Struktura wiekowa ludności
1	2	3	4	5
Łódzkie	446 694	1 624 856	470 282	{przedprodukcyjny (17,6%), produkcyjny (63,9%), poprodukcyjny (18,5%)}
Mazowieckie	975 278	3 329 007	917 882	{przedprodukcyjny (18,7%), produkcyjny (63,7%), poprodukcyjny (17,6%)}
Małopolskie	659 907	2 099 386	538 977	{przedprodukcyjny (20,0%), produkcyjny (63,7%), poprodukcyjny (16,3%)}
Śląskie	806 093	3 034 503	800 129	{przedprodukcyjny (17,4%), produkcyjny (65,4%), poprodukcyjny (17,2%)}
Lubelskie	421 829	1 361 161	374 212	{przedprodukcyjny (19,6%), produkcyjny (63,1%), poprodukcyjny (17,3%)}

Źródło: opracowanie własne na podstawie danych BDL GUS.

wieku. Kolumny 2-4 reprezentują 3 zmienne w ujęciu klasycznym o charakterze metrycznym. Na ich podstawie wyznaczono udziały procentowe każdej grupy i uzyskano zmienną symboliczną charakteryzującą strukturę wiekową ludności województw.

Agregację obiektów stosuje się w celu redukcji opisu ze względu na bardzo liczny zbiór obiektów bądź uszczegółowienie informacji o obiektach. Umożliwia ona reprezentację obiektów niższego rzędu (np. powiatów) za pomocą obiektów wyższego rzędu (np. województw). Obiekty niższego rzędu opisywane są za pomocą zmiennych w ujęciu klasycznym, natomiast obiekty wyższego rzędu charakteryzowane są zmiennymi symbolicznymi (por. [Diday 2000]).

Agregację obiektów niższego rzędu, opisanych zmiennymi ilorazowymi, można przeprowadzić na dwa sposoby, uzyskując zmienne symboliczne o realizacjach w postaci przedziałów liczbowych lub struktur procentowych. W tabeli 2 zawarto informacje na temat stopy bezrobocia rejestrowanego w województwach w 2009 r. oraz wyznaczono wartości wybranych statystyk opisowych dla stopy bezrobocia w powiatach poszczególnych województw. W województwie mazowieckim wartość wskaźnika wynosiła 9,0%, co stanowiło najlepszy wynik na tle pozostałych województw. W takiej formie jest to jednak informacja niepełna, ponieważ wskaźnik przyjmował wartości między 2,8% (Warszawa) a 35,3% (powiat szydłowiecki), przy czym w $\frac{3}{4}$ powiatów osiągnął wartości wyższe niż 9,0%.

Tabela 2. Stopa bezrobocia w roku 2009

Województwo	Stopa bezrobocia w województwie	Wartości statystyk opisowych wyznaczonych dla stopy bezrobocia w powiatach* województwa			
		wartość minimalna	wartość maksymalna	mediana	klasyczny współczynnik zmienności (%)
Mazowieckie	9,0	2,8	35,3	14,2	44,2
Śląskie	9,4	3,3	18,5	10,3	35,4
Małopolskie	9,7	4,1	18,8	10,6	29,3
Łódzkie	11,9	7,0	18,0	12,2	23,1
Dolnośląskie	12,8	5,0	26,4	17,2	37,2
Podlaskie	12,8	7,2	19,9	13,0	29,8
Lubelskie	12,9	8,3	20,7	14,5	20,2

* Łącznie z miastami na prawach powiatu.

Źródło: opracowanie własne na podstawie danych BDL GUS.

Zastosowanie w tym przypadku zmiennych symbolicznych lepiej obrazowałoby sytuację województw, ponieważ pozwoliłoby uwzględnić strukturę terytorialną zjawiska. Zmienną symboliczną o realizacjach w postaci przedziałów liczbowych uzyskano na podstawie minimalnej i maksymalnej wartości stopy bezrobocia w powiatach danego województwa (kolumna 2 w tab. 3).

Zmienną symboliczną o realizacjach w postaci struktury procentowej można uzyskać w następujący sposób. W pierwszym kroku należy dokonać podziału wartości zmiennej (dla obiektów niższego rzędu, np. powiatów) na kilka przedziałów. Przedziały można ustalić w sposób arbitralny bądź z wykorzystaniem metod statystycznych. Na przykład jeśli zmienna symboliczna ma przyjmować dwie kategorie (np. wysoki, niski), to do wyznaczenia wartości środkowej można zastosować medianę. W drugim kroku dla każdego obiektu wyższego rzędu (np. województwa) należy wyznaczyć procent obiektów niższego rzędu (np. powiatów) zakwalifikowanych do danego przedziału. W analizowanym przykładzie ustalono trzy kategorie zmiennej symbolicznej opisującej stopę bezrobocia: poniżej 10% – niska, 10-20% – przeciętna, powyżej 20% – wysoka (kolumna 3 w tab. 3).

Tabela 3. Przykład agregacji obiektów

Województwo	Zmienna symboliczna	
	w postaci przedziałów liczbowych	w postaci struktury procentowej
1	2	3
Mazowieckie	[2,8; 35,3]	{niska (26,2%), przeciętna (59,5%), wysoka (14,3%)}
Śląskie	[3,3; 18,5]	{niska (44,4%), przeciętna (55,6%), wysoka (0,0%)}
Małopolskie	[4,1; 18,8]	{niska (36,4%), przeciętna (63,6%), wysoka (0,0%)}
Łódzkie	[7,0; 18,0]	{niska (25,0%), przeciętna (75,0%), wysoka (0,0%)}
Dolnośląskie	[5,0; 26,0]	{niska (20,7%), przeciętna (48,3%), wysoka (31,0%)}
Podlaskie	[7,2; 19,9]	{niska (23,5%), przeciętna (76,5%), wysoka (0,0%)}
Lubelskie	[8,3; 20,7]	{niska (12,5%), przeciętna (83,3%), wysoka (4,2%)}

Źródło: opracowanie własne na podstawie danych BDL GUS.

Rozpaprując wartości stopy bezrobocia w województwach zawarte w tab. 2, można stwierdzić, że w 2009 r. podobną sytuacją cechowały się województwa dolnośląskie (12,8%) i podlaskie (12,8%). Z kolei interpretując dane uzyskane w tab. 3, należy zauważyć, że w większości powiatów województwa dolnośląskiego odnotowano co najmniej przeciętny poziom bezrobocia, natomiast w większości powiatów województwa podlaskiego – co najwyżej przeciętny poziom bezrobocia.

3. Procedura badawcza

Przedmiotem badania jest taksonomiczna analiza województw Polski ze względu na strukturę regionalnego rynku pracy w 2009 r. W procedurze klasyfikacji obiektów opisanych zmiennymi symbolicznymi zastosowanie mają metody zarówno zaproponowane w ramach analizy danych symbolicznych, jak i niektóre opracowane dla danych w ujęciu klasycznym (tab. 4).

Tabela 4. Metody mające zastosowanie w procedurze klasyfikacji danych symbolicznych

Etap w procedurze klasyfikacji	Metody stosowane w procesie klasyfikacji obiektów symbolicznych	
	metody opracowane dla danych w ujęciu klasycznym	metody opracowane na gruncie analizy danych symbolicznych
1. Wybór obiektów i zmiennych	– adaptacja metody doboru zmiennych <i>HINoV</i> Carmone’a, Kary i Maxwell	– metoda doboru zmiennych Ichino
2. Wybór miary odległości	–	– miary Ichino-Yaguchi, de Carvalho itd.
3. Wybór metody klasyfikacji	– metody hierarchiczne, np. Warda, kompletnego połączenia – metody optymalizacyjne, np. <i>k</i> -medoidów	– metody hierarchiczne, np. Brito – metody optymalizacyjne, np. SCLUST, <i>k</i> -średnich Verde
4. Wybór liczby klas	– indeksy wyboru liczby klas, np. Bakera i Huberta, Huberta i Levine, Calińskiego i Harabasz	– indeksy wyboru liczby klas, np. <i>Q(P)</i> Verde, Lechevallier i Chavent, <i>CI</i> Gowdy i Didaya
5. Ocena wyników klasyfikacji	– indeks <i>silhouette</i> Rousseeuwa – adaptacja analizy replikacji z indeksem Randa	– metoda Bertranda i Bel-Mufti
6. Opis klas	–	– technika CLINT
7. Profilowanie klas	– adaptacja drzew klasyfikacyjnych – adaptacja analizy dyskryminacyjnej	–

Źródło: opracowanie na podstawie pracy: [Wilk 2011].

W klasyfikacji województw przyjęto następującą procedurę badawczą:

1. Wybór obiektów i zmiennych. W analizie uwzględniono 16 obiektów symbolicznych (województw), które powstały w wyniku agregacji obiektów niższego rzędu (powiatów). Dobór zmiennych miał charakter merytoryczny, był jednak ograniczony dostępnością danych na poziomie powiatów. Zbiór obiektów opisano 5 zmiennymi symbolicznymi o realizacjach w postaci udziałów procentowych (tab. 5). W pierwszej kolejności ustalono kategorie zmiennych. Zbiór realizacji zmiennych „chłonność rynku”, „dysproporcja”, „przedsiębiorczość” i „zarobki” podzielono na dwa przedziały z wykorzystaniem mediany. Z kolei zbiór realizacji zmiennej „profil gospodarczy” podzielono na trzy przedziały o zbliżonej liczebności powiatów. Następnie dla każdej zmiennej wyznaczono procent powiatów województwa reprezentujący daną kategorię. W ten sposób otrzymano tablicę danych symbolicznych (zob. załącznik).

2. Wybór liczby klas. Indeksy Bakera i Huberta, Huberta i Levine oraz *silhouette* (zob. [Walesiak 2009, s. 418]) wskazały podział zbioru obiektów na 4 klasy.

3. Klasyfikacja obiektów. Grupowanie obiektów przeprowadzono optymalizacyjną metodą taksonomii symbolicznej SCLUST (zob. [Verde 2004]).

4. Charakterystyka klas. W opisie klas zastosowano technikę CLINT (zob. [Brito 2004]). Dla każdej kategorii zmiennej symbolicznej wyznaczono minimalne, maksymalne i średnie udziały procentowe powiatów województw należących do danej klasy. Na przykład przyjmując, że do jednej klasy należą województwa podlaskie,

Tabela 5. Zbiór zmiennych symbolicznych

Lp.	Nazwa zmiennej symbolicznej	Zbiór realizacji zmiennej symbolicznej	
		kategorie zmiennej	% powiatów województwa reprezentujących kategorię zmiennej
1	Chłonność rynku (odsetek osób pracujących w ludności aktywnej zawodowo (%))	niska (poniżej 50%)	[14; 86]
		wysoka (powyżej 50%)	[14; 86]
2	Profil gospodarczy (odsetek osób pracujących w usługach* (%))	nieusługowy (poniżej 30%)	[0; 76]
		mieszany (30-50%)	[6; 86]
		usługowy (powyżej 50%)	[8; 47]
3	Dysproporcja (relacja liczby pracujących kobiet do liczby pracujących mężczyzn (osoba))	przewaga mężczyzn (poniżej 100 kobiet na 100 mężczyzn)	[29; 86]
		przewaga kobiet (powyżej 100 kobiet na 100 mężczyzn)	[14; 71]
4	Przedsiębiorczość (liczba osób fizycznych prowadzących działalność gospodarczą na 1000 osób w wieku produkcyjnym (osoba))	mała (poniżej 100 osób)	[5; 83]
		duża (powyżej 100 osób)	[17; 95]
5	Zarobki (przeciętne miesięczne wynagrodzenie brutto (zł))	niskie (poniżej 2700 zł)	[22; 86]
		wysokie (powyżej 2700 zł)	[14; 78]

* Dane za 2008 r. według klasyfikacji PKD 2004.

Źródło: opracowanie własne na podstawie BDL GUS.

Tabela 6. Przykład zastosowania techniki CLINT

Numer klasy	Nazwa zmiennej symbolicznej	Kategorie zmiennej symbolicznej	Wartości statystyk opisowych wyznaczone dla kategorii zmiennej symbolicznej		
			wartość minimalna	wartość maksymalna	średnia arytmetyczna
1	chłonność rynku	mała	82	86	84
		duża	14	18	16
	profil gospodarczy	nieusługowy	57	76	67
		mieszany	6	29	17
		usługowy	14	18	16

Źródło: opracowanie własne na podstawie danych zawartych w załączniku.

lubelskie i świętokrzyskie, dla zmiennych symbolicznych „chłonność rynku” i „profil gospodarczy” otrzymuje się opis klasy zawarty w tab. 6. Z danych zawartych w tej tabeli wynika, że województwa należące do klasy cechują się niskim zatrudnieniem (większość powiatów tych województw zakwalifikowana została do kategorii „mała chłonność rynku”) oraz bardzo niskim odsetkiem osób pracujących w usługach

(większość powiatów tych województw zakwalifikowana została do kategorii „nieusługowy profil gospodarczy”).

4. Wyniki badań

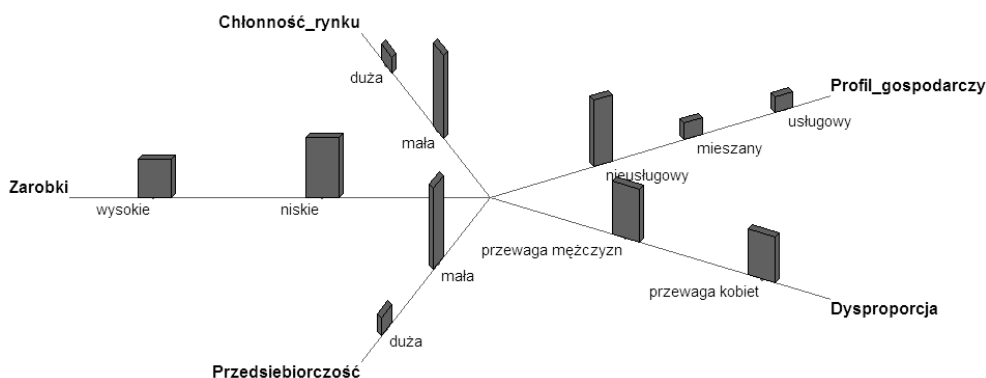
W wyniku zastosowanej procedury otrzymano 4 klasy województw. Interpretacje klas zaprezentowano w tab. 7. Symbol „x” oznacza brak możliwości jednoznacznej oceny sytuacji ze względu na to, że powiaty województw należących do klasy cechuje duże zróżnicowanie terytorialne względem kategorii zmiennej.

Tabela 7. Interpretacja klas

Numer klasy	Chłonność rynku	Profil gospodarczy	Dysproporcja	Przedsiębiorczość	Zarobki
1	bardzo niska	nieusługowy	x	bardzo niska	niskie
2	bardzo niska	mieszany	przewaga mężczyzn	niska	bardzo niskie
3	niska	x	przewaga kobiet	wysoka	x
4	średnia	usługowy	przewaga mężczyzn	wysoka	średnie

Źródło: opracowanie własne.

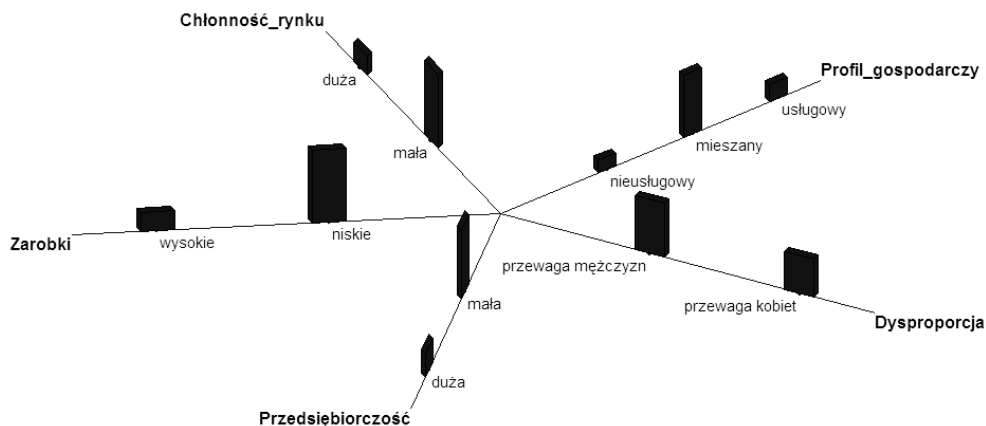
Skupienie 1 reprezentują województwa podlaskie, lubelskie i świętokrzyskie. Większość powiatów tych województw charakteryzuje się najniższym poziomem zatrudnienia oraz niskim poziomem zarobków (rys. 1). Największy odsetek osób pracuje w przemyśle lub rolnictwie. W województwach tych odnotowano najmniejszą liczbę osób fizycznych prowadzących działalność gospodarczą w przeliczeniu na 1000 osób w wieku produkcyjnym.



Rys. 1. Charakterystyka klasy 1

Źródło: opracowanie własne w programie SODAS.

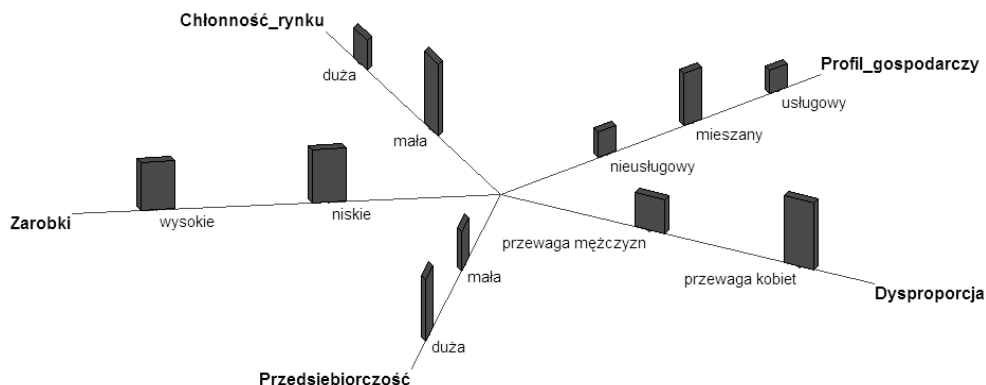
Bardzo niskim poziomem zatrudnienia, najniższym poziomem zarobków i niskim poziomem przedsiębiorczości charakteryzują się województwa klasy 2, tj. warmińsko-mazurskie, kujawsko-pomorskie i podkarpackie (rys. 2). Profil gospodarczy tych województw jest zróżnicowany (nie dominuje żaden sektor), a wśród osób pracujących przeważają mężczyźni.



Rys. 2. Charakterystyka klasy 2

Źródło: opracowanie własne w programie SODAS.

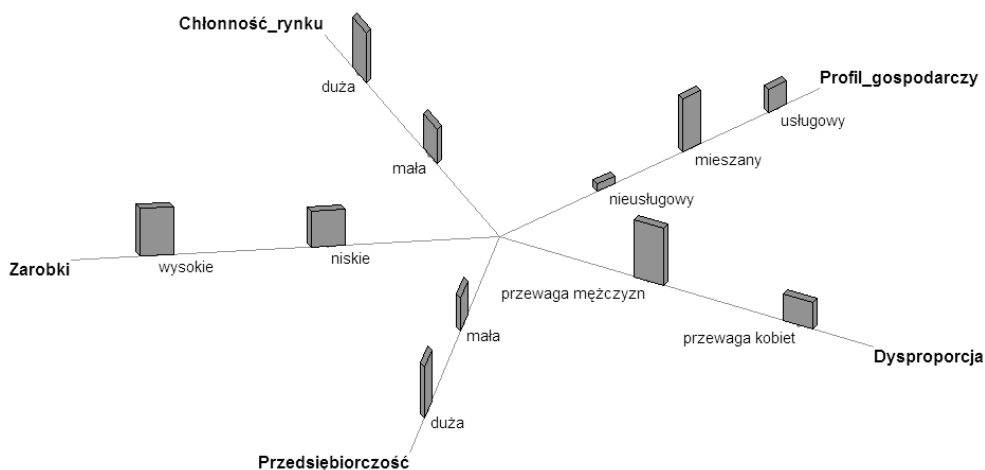
Województwa klasy 3, tj. zachodniopomorskie, mazowieckie, łódzkie, małopolskie, cechują się niskim poziomem zatrudnienia, ale relatywnie dużą liczbą osób fizycznych prowadzących działalność gospodarczą, a także przewagą kobiet na rynku pracy (rys. 3).



Rys. 3. Charakterystyka klasy 3

Źródło: opracowanie własne w programie SODAS.

Do klasy 4 należą województwa o profilu usługowym, tj. pomorskie, opolskie, wielkopolskie, śląskie, lubuskie i dolnośląskie (rys. 4). Charakteryzują się relatywnie korzystną sytuacją na rynku pracy ze względu na najwyższy poziom zatrudnienia oraz poziom zarobków, a także relatywnie dużą liczbą osób prowadzących działalność gospodarczą.



Rys. 4. Charakterystyka klasy 4

Źródło: opracowanie własne w programie SODAS.

Ocenę sytuacji na rynku pracy oraz poziomu spójności (mierzonego odsetkiem powiatów województw o zbliżonej sytuacji) przedstawiono w tab. 8. Najbardziej korzystna sytuacja na rynku pracy oraz relatywnie wysoka spójność terytorialna cechuje województwa klasy 4. W odwrotnej sytuacji znajdują się województwa klasy 1.

Tabela 8. Ocena klas województw

Sytuacja na rynku pracy	Spójność terytorialna	
	relatywnie mała	relatywnie duża
Relatywnie zła	klasa 1	–
Przeciętna	klasa 3	klasa 2
Relatywnie dobra	–	klasa 4

Źródło: opracowanie własne na podstawie tab. 6.

5. Podsumowanie

Zastosowanie podejścia symbolicznego umożliwiło taksonomiczną analizę sytuacji rynku pracy województw z uwzględnieniem ich zróżnicowania terytorialnego. Re-prezentacja zjawisk w formie danych symbolicznych dała pełniejszy obraz badanej

rzeczywistości ze względu na zastosowanie, w miejsce danych „uśrednionych”, danych zagregowanych. Jednocześnie złożona struktura danych symbolicznych wymagała zastosowania specyficznych rozwiązań metodologicznych w procedurze analizy skupień.

Literatura

- Brito P., *Clustering Interpretation. Interpreting Clusters by Using the Module CLINT*, [w:] User Manual for SODAS 2 Software, Software Report, Analysis System of Symbolic Official Data, M. Noirhomme-Fraiture (red.), Project no. IST-2000-25161, 2004.
- Diday E., *Symbolic Data Analysis and the Sodas Project: Purpose, History, Perspective*, [w:] *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, H.H. Bock, E. Diday (red.), Springer-Verlag, Berlin-Heidelberg 2000.
- Verde R., *Clustering Methods in Symbolic Data Analysis*, [w:] *Classification, Clustering and Data Mining Applications*, D. Banks, L. House, E. R. McMorris, P. Arabie, W. Gaul (red.), Springer-Verlag, Heidelberg 2004.
- Walesiak M., *Analiza skupień*, [w:] *Statystyczna analiza danych z wykorzystaniem programu R*, M. Walesiak, E. Gatnar (red.), PWN, Warszawa 2009.
- Wilk J., *Analiza skupień na podstawie danych symbolicznych*, [w:] *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, E. Gatnar, M. Walesiak (red.), C.H. Beck, Warszawa 2011.
- Wilk J., *Problemy segmentacji rynku z wykorzystaniem metod klasyfikacji i danych symbolicznych*, Uniwersytet Ekonomiczny we Wrocławiu, rozprawa doktorska (nieopublikowana), Jelenia Góra 2010.

Załącznik. Tablica danych symbolicznych

Województwo	Chłonność rynku	Profil gospodarczy	Dysproporcja	Przedsiębiorczość	Zarobki
Łódzkie	{mała (71%), duża (29%)}	{nieusługowy (42%), mieszany (45%), usługowy (13%)}	{przewaga mężczyzn (33%), przewaga kobiet (67%)}	{mała (54%), duża (46%)}	{niskie (71%), wysokie (29%)}
Mazowieckie	{mała (69%), duża (31%)}	{nieusługowy (43%), mieszany (33%), usługowy (24%)}	{przewaga mężczyzn (29%), przewaga kobiet (71%)}	{mała (55%), duża (45%)}	{niskie (33%), wysokie (67%)}
Małopolskie	{mała (73%), duża (27%)}	{nieusługowy (18%), mieszany (64%), usługowy (18%)}	{przewaga mężczyzn (45%), przewaga kobiet (55%)}	{mała (41%), duża (59%)}	{niskie (50%), wysokie (50%)}
Śląskie	{mała (14%), duża (86%)}	{nieusługowy (8%), mieszany (44%), usługowy (48%)}	{przewaga mężczyzn (81%), przewaga kobiet (19%)}	{mała (44%), duża (56%)}	{niskie (22%), wysokie (78%)}
Lubelskie	{mała (83%), duża (17%)}	{nieusługowy (66%), mieszany (17%), usługowy (17%)}	{przewaga mężczyzn (29%), przewaga kobiet (71%)}	{mała (83%), duża (17%)}	{niskie (75%), wysokie (25%)}
Podkarpackie	{mała (76%), duża (24%)}	{nieusługowy (16%), mieszany (60%), usługowy (24%)}	{przewaga mężczyzn (60%), przewaga kobiet (40%)}	{mała (76%), duża (24%)}	{niskie (68%), wysokie (32%)}
Podlaskie	{mała (82%), duża (18%)}	{nieusługowy (76%), mieszany (6%), usługowy (18%)}	{przewaga mężczyzn (47%), przewaga kobiet (53%)}	{mała (82%), duża (18%)}	{niskie (59%), wysokie (41%)}
Świętokrzyskie	{mała (86%), duża (14%)}	{nieusługowy (57%), mieszany (29%), usługowy (14%)}	{przewaga mężczyzn (86%), przewaga kobiet (14%)}	{mała (79%), duża (21%)}	{niskie (50%), wysokie (50%)}
Lubuskie	{mała (43%), duża (57%)}	{nieusługowy (0%), mieszany (71%), usługowy (29%)}	{przewaga mężczyzn (43%), przewaga kobiet (57%)}	{mała (43%), duża (57%)}	{niskie (50%), wysokie (50%)}
Wielkopolskie	{mała (43%), duża (57%)}	{nieusługowy (23%), mieszany (66%), usługowy (11%)}	{przewaga mężczyzn (80%), przewaga kobiet (20%)}	{mała (31%), duża (69%)}	{niskie (63%), wysokie (37%)}
Zachodniopomorskie	{mała (67%), duża (33%)}	{nieusługowy (0%), mieszany (62%), usługowy (38%)}	{przewaga mężczyzn (29%), przewaga kobiet (71%)}	{mała (5%), duża (95%)}	{niskie (57%), wysokie (43%)}
Dolnośląskie	{mała (52%), duża (48%)}	{nieusługowy (7%), mieszany (59%), usługowy (34%)}	{przewaga mężczyzn (59%), przewaga kobiet (41%)}	{mała (45%), duża (55%)}	{niskie (45%), wysokie (55%)}
Opolskie	{mała (50%), duża (50%)}	{nieusługowy (17%), mieszany (75%), usługowy (8%)}	{przewaga mężczyzn (83%), przewaga kobiet (17%)}	{mała (58%), duża (42%)}	{niskie (50%), wysokie (50%)}
Kujawsko-pomorskie	{mała (74%), duża (26%)}	{nieusługowy (26%), mieszany (57%), usługowy (17%)}	{przewaga mężczyzn (74%), przewaga kobiet (26%)}	{mała (70%), duża (30%)}	{niskie (83%), wysokie (17%)}
Pomorskie	{mała (45%), duża (55%)}	{nieusługowy (0%), mieszany (65%), usługowy (35%)}	{przewaga mężczyzn (65%), przewaga kobiet (35%)}	{mała (15%), duża (85%)}	{niskie (30%), wysokie (70%)}
Warmińsko-mazurskie	{mała (86%), duża (14%)}	{nieusługowy (0%), mieszany (86%), usługowy (14%)}	{przewaga mężczyzn (43%), przewaga kobiet (57%)}	{mała (76%), duża (24%)}	{niskie (86%), wysokie (14%)}

TAXONOMIC ANALYSIS OF LABOUR MARKET IN POLISH VOIVODESHIPS – SYMBOLIC APPROACH

Summary: The aim of this paper is to provide taxonomy analysis of labour market in Polish regions in 2009 with symbolic approach. The paper discusses a way of phenomena structure representation with symbolic data application. Next the statistical methods are presented that can be used in cluster analysis procedure based on a symbolic data set. Pursuant to suggested procedure four clusters of voivodeships were separated and regional labour markets were investigated.

Keywords: symbolic data analysis, taxonomy, labour market, regional investigations.