

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 328

Taksonomia 23

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

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

w Dolnośląskiej Bibliotece Cyfrowej www.dbc.wroc.pl,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

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

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2014

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	11
Małgorzata Rószkiewicz , Wykorzystanie metaanalizy w budowaniu modelu pomiarowego w przypadku braku niezmienniczości zasad pomiaru na przykładzie pomiaru zadowolenia z życia.....	13
Elżbieta Sobczak , Harmonijność inteligentnego rozwoju regionów Unii Europejskiej	21
Ewa Roszkowska, Renata Karwowska , Analiza porównawcza województw Polski ze względu na poziom zrównoważonego rozwoju w roku 2010.....	30
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel , Analiza porównawcza wybranych filtrów w analizie synchronizacji cyklu koniunkturalnego.....	41
Marcin Salamaga , Próba konstrukcji tablic „wymierania scenicznego” spektakli operowych na przykładzie Metropolitan Opera.....	51
Iwona Foryś , Wykorzystanie analizy dyskryminacyjnej do typowania rynków podobnych w procesie wyceny nieruchomości niemieszkalnych	59
Jerzy Korzeniewski , Selekcja zmiennych w klasyfikacji – propozycja algorytmu	69
Sabina Denkowska , Testowanie wielokrotne przy weryfikacji wieloczynnikowych modeli proporcjonalnego hazardu Coxa.....	76
Ewa Chodakowska , Teoria równań strukturalnych w klasyfikacji zmiennych jawnych i ukrytych według charakteru ich wzajemnych oddziaływań	85
Iwona Konarzewska , Model PCA dla rynku akcji – studium przypadku	94
Katarzyna Wójcik, Janusz Tuchowski , Dobór optymalnego zestawu słów istotnych w opiniach konsumentów na potrzeby ich automatycznej analizy	106
Aleksandra Łuczak , Zastosowanie metody AHP-LP do oceny ważności determinant rozwoju społeczno-gospodarczego w jednostkach administracyjnych	116
Aleksandra Witkowska, Marek Witkowski , Klasyfikacja pozycyjna banków spółdzielczych według stanu ich kondycji finansowej w ujęciu dynamicznym	126
Adam Depta , Zastosowanie analizy korespondencji do oceny jakości życia ludności na podstawie kwestionariusza SF-36v2	135
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej , Indukcja reguł dla danych niekompletnych i niezbalansowanych: modele klasyfikatorów i próba ich zastosowania do predykcji ryzyka operacyjnego w torakochirurgii	146

Małgorzata Misztal , Wybrane metody oceny jakości klasyfikatorów – przegląd i przykłady zastosowań.....	156
Anna M. Olszewska , Wykorzystanie wybranych metod taksonomicznych do oceny potencjału innowacyjnego województw	167
Iwona Bąk , Porównanie jakości grupowań powiatów województwa zachodniopomorskiego pod względem atrakcyjności turystycznej.....	177
Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn , Segmentacja gospodarstw domowych według wydatków na turystykę zorganizowaną.....	186
Agnieszka Wałęga , Podejście syntetyczne w analizie spójności ekonomicznej gospodarstw domowych.....	196
Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek , Zastosowanie analizy korespondencji do badania wpływu elektrowni wiatrowych na jakość życia ludności	205
Joanna Banaś, Krzysztof Małecki , Klasyfikacja punktów pomiarów ankietowych kierowców na granicy Szczecina z wykorzystaniem zmiennych symbolicznych.....	214
Aneta Becker , Wykorzystanie informacji granularnej w analizie wymagań rynku pracy.....	222
Katarzyna Cheba, Joanna Holub-Iwan , Wykorzystanie analizy korespondencji w segmentacji rynku usług medycznych.....	230
Adam Depta, Iwona Staniec , Identyfikacja czynników decydujących o jakości życia studentów łódzkich uczelni.....	238
Katarzyna Dębowska, Jarosław Kilon , Reguły asocjacyjne w analizie wyników badań metodą Delphi.....	247
Anna Domagała , O wykorzystaniu analizy głównych składowych w metodzie <i>Data Envelopment Analysis</i>	254
Alicja Grześkowiak , Analiza wykluczenia cyfrowego w Polsce w ujęciu indywidualnym i regionalnym.....	264
Anna M. Olszewska, Anna Gryko-Nikitin , Pomiar postrzegania jakości kształcenia uczelni wyższej na danych porządkowych z wykorzystaniem środowiska R.....	273
Karolina Paradysz , Hierarchiczna metoda grupowania powiatów jako podejście benchmarkowe w ocenie bezrobocia według BAEL-u w wybranych typach małych obszarów	282
Radosław Pietrzyk , Porównanie metod pomiaru efektywności zarządzania portfelami funduszy inwestycyjnych.....	290
Agnieszka Przedborska, Małgorzata Misztal , Wybrane metody statystyki wielowymiarowej w ocenie skuteczności terapeutycznej głębokiej stymulacji elektromagnetycznej u pacjentów z chorobą zwyrodnieniową stawów.....	299

Wojciech Roszka, Marcin Szymkowiak , Podejście kalibracyjne w statystycznej integracji danych	308
Iwona Skrodzka , Zastosowanie wybranych metod klasyfikacji do analizy kapitału ludzkiego krajów Unii Europejskiej	316
Agnieszka Stanimir , Wielowymiarowa analiza czynników sprzyjających włączeniu społecznemu	326
Dorota Strózik, Tomasz Strózik , Przestrzenne zróżnicowanie poziomu życia w województwie wielkopolskim.....	334
Izabela Szamrej-Baran , Identyfikacja przyczyn ubóstwa energetycznego w Polsce przy wykorzystaniu modelowania miękkiego.....	343
Janusz Tuchowski, Katarzyna Wójcik , Klasyfikacja obiektów w systemie Krajowych Ram Kwalifikacji opisanych za pomocą ontologii	353
Aleksandra Matuszewska-Janica , Grupowanie krajów Unii Europejskiej ze względu na poziom feminizacji sektorów gospodarczych	361
Monika Rozkrut, Dominik Rozkrut , Identyfikacja strategii innowacyjnych przedsiębiorstw usługowych w Polsce	369

Summaries

Małgorzata Rószkiewicz , The use of meta-analysis in building the measurement model in case of the absence of measurement invariance on the example of measuring of life satisfaction.....	20
Elżbieta Sobczak , Harmonious smart growth of European Union regions.....	29
Ewa Roszkowska, Renata Karwowska , The comparative analysis of Polish voivodeships with respect to sustainable development in 2010.....	40
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel , Comparative analysis of chosen filters in business cycles analysis	50
Marcin Salamaga , The attempt of construction of the life tables for opera works on the example of the Metropolitan Opera	58
Iwona Foryś , Using discriminant analysis to select similar markets in non-residential property valuation process.....	68
Jerzy Korzeniewski , Variable selection in classification – algorithm proposal	75
Sabina Denkowska , Multiple testing in the verification process of multifactorial Cox proportional hazards models	84
Ewa Chodakowska , The theory of structural equations modelling in the classification of observed variables and latent constructs according to the character of their relationship.....	93
Iwona Konarzewska , Modelling stock market by PCA factor model – case study	105

Katarzyna Wójcik, Janusz Tuchowski , Selection of the optimal set of relevant words in consumers opinions in the context of the opinion mining ..	115
Aleksandra Łuczak , Application of AHP-LP to the evaluation of importance of determinants of socio-economic development in the administrative units	125
Aleksandra Witkowska, Marek Witkowski , A dynamic approach to the ranking of cooperative banks by their financial condition	134
Adam Depta , Application of correspondence analysis for the measurement of quality of life – questionnaire SF-36v2 based research	145
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej , Classification rules extraction for missing and imbalance data: models of classifiers and initial results in the rules-based thoracic surgery risk prediction.....	155
Małgorzata Misztal , Selected methods for assessing the performance of classifiers – an overview and examples of applications.....	166
Anna M. Olszewska , The application of selected quantitative methods to the evaluation of voivodeship innovation level potential.....	176
Iwona Bąk , The comparison of the quality of groupings of poviats of West Pomeranian Voivodeship in terms of tourism attractiveness	185
Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn , Household segmentation with respect to the expenditure on organized tourism.....	195
Agnieszka Wałęga , Synthetic approach in the analysis of economic coherence of households	204
Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek , Using the correspondence analysis to examine the impact of wind turbines on the quality of life.....	213
Joanna Banaś, Krzysztof Małecki , Classification of measurement survey points of drivers on the boundary of Szczecin using symbolic variables...	221
Aneta Becker , The use granular information in the analysis of the requirements of the labor market.....	229
Katarzyna Cheba, Joanna Hołub-Iwan , The application of the correspondence analysis of patients segmentation on the medical service market	237
Adam Depta, Iwona Staniec , Identification of the factors that determine the quality of students life at universities in Lodz.....	246
Katarzyna Dębkowska, Jarosław Kilon , Association rules in the analysis of research results the Delphi method	253
Anna Domagała , About using Principal Component Analysis in Data Envelopment Analysis	263
Alicja Grześkowiak , Analysis of the digital divide in Poland at the individual and regional level	272

Anna M. Olszewska, Anna Gryko-Nikitin, Assessment of perception of quality of teaching at an institution of higher learning based on the ordinal data with the utilization of R environment.....	281
Karolina Paradysz, The hierarchical method of grouping poviats as a benchmark approach in the assessment of unemployment by BAEL in selected types of small areas	289
Radosław Pietrzyk, Comparison of methods of measuring the performance of investment funds portfolios.....	298
Agnieszka Przedborska, Małgorzata Misztal, Selected multivariate statistical analysis methods in the evaluation of efficacy of deep electromagnetic stimulation in patients with degenerative joint disease	307
Wojciech Roszka, Marcin Szymkowiak, A calibration approach in statistical data integration	315
Iwona Skrodzka, Application of some methods of classification to the analysis of human capital in the European Union.....	325
Agnieszka Stanimir, Multivariate analysis of social inclusion factors.....	333
Dorota Strózik, Tomasz Strózik, Spatial differentiation of the standard of living in Great Poland Voivodeship	342
Izabela Szamrej-Baran, Identification of fuel poverty causes in Poland using soft modelling	352
Janusz Tuchowski, Katarzyna Wójcik, Classification of objects in the National Classification Framework described by the ontology.....	360
Aleksandra Matuszewska-Janica, Clustering of European Union states taking into consideration the levels of feminization of economic sectors..	368
Monika Rozkrut, Dominik Rozkrut, Identification of service sector innovation strategies in Poland.....	379

Aneta Becker

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

WYKORZYSTANIE INFORMACJI GRANULARNEJ W ANALIZIE WYMAGAŃ RYNKU PRACY

Streszczenie: W artykule zaprezentowano przebieg i wyniki procedury badawczej potrzeb kompetencyjnych polskiego rynku pracy. Głównym zadaniem była analiza zapotrzebowania na kompetencje (umiejętności i kwalifikacje) zgłaszane przez polskich przedsiębiorców. W badaniu poszukiwano zależności pomiędzy rodzajem specjalności, doprecyzowanej profilem kompetencyjnym, a wymogiem znajomości języka angielskiego. Materiał empiryczny pochodził z badań realizowanych w 2012 r. w ramach projektu „Bilans Kapitału Ludzkiego” (III edycja). Obliczenia zrealizowano z wykorzystaniem analizy skupień i teorii zbiorów przybliżonych za pomocą Rough Set Exploration System 2.1.

Słowa kluczowe: granule informacyjne, analiza skupień, zbiory przybliżone, rynek pracy.

1. Wstęp

Pojęcie granuli informacyjnej, nazywanej także infogranulą lub informacją ziarnistą (*information granule*) zostało wprowadzone przez L.A. Zadeha [Zadeh 1979, 1997] w drugiej połowie XX wieku. Termin ten dotyczy pewnych podzbiorów (klas) uniwersum, do których należą obiekty charakteryzujące się: nierozróżnialnością, podobieństwem lub zbliżonym zachowaniem. Ważnym pojęciem są obliczenia granularne (*granular computing*), definiowane jako ogólna teoria obliczeń, która ma na celu skuteczne stosowanie tzw. ziaren, takich jak klasy, klastry, podzbiory, grupy i przedziały (interwały), w celu konstruowania bardziej efektywnych modeli obliczeniowych na potrzeby złożonych zastosowań ze znaczącą ilością danych, informacji i wiedzy [Borowik, Borowik, Laird 2009, s. 154]. Obliczenia granularne mogą być postrzegane jako pewien sposób strukturalnego myślenia lub procedura strukturalnego rozwiązywania problemów. W zależności od przyjętej ziarnistości (agregacji) w badanym zbiorze danych zauważa się prawidłowości (powiązania, zależności) niewidoczne przy innym podziale, co umożliwia szersze przeanalizowanie problemu. Na jakość rezultatów uzyskanych w wyniku zastosowania wspomnianych obliczeń wpływa przede wszystkim konstrukcja granul oraz metody realizacji obliczeń granularnych.

Celem artykułu jest badanie potrzeb kompetencyjnych polskiego rynku pracy. Głównym zadaniem jest analiza zapotrzebowania na kompetencje (umiejętności i kwalifikacje) zgłaszane przez polskich przedsiębiorców. W pracy przeprowadzono badania z wykorzystaniem systemu RSES 2.1 (Rough Set Exploration System 2.1), który jest narzędziem komputerowym, skonstruowanym przez zespół kierowany przez Z. Skowrona. Umożliwia on analizę danych w postaci tablicowej z zastosowaniem teorii zbiorów przybliżonych.

W badaniach przedstawionych w artykule wykorzystano zbiory przybliżone często stosowane w teorii obliczeń granularnych. Skoncentrowano się na poszukiwaniu zależności pomiędzy rodzajem specjalności, doprecyzowanej profilem kompetencyjnym, a wymogiem znajomości języka angielskiego. Próbowano także odpowiedzieć na trzy pytania związane z analizowanym problemem.

- Czy występują wymagania pracodawców wobec specjalistów dotyczące znajomości języka angielskiego?
- Od których specjalistów (szczególnie) wymagana jest znajomość języka angielskiego?
- Czy prawidłowość ta jest jednakowa w regionach Polski (granul informacyjnych) charakteryzujących się różnym poziomem rozwoju gospodarczego?

2. Wykorzystanie informacji granularnej w teorii zbiorów przybliżonych

Obliczenia granularne są nakierowane na wiedzę, przeciwieństwo tradycyjnych obliczeń numerycznych zorientowanych na dane. Powoduje to, że stosuje się je w aplikacjach związanych z odkrywaniem wiedzy i eksploracją danych (*Knowledge Discovery i Data Mining*) [Borowik, Borowik, Laird 2009, s. 155]. Pierwsze wzmianki dotyczące pojęcia granulacji informacji przedstawił w 1979 r. L.A. Zadeh. Natomiast w 1982 r. Z. Pawlak zaproponował wykorzystanie tej informacji w postaci teorii zbiorów przybliżonych (*rough sets*) [Yao 2005].

Teoria zbiorów przybliżonych pomaga kontrolować duże zbiory danych oraz dostarcza narzędzi matematycznych potrzebnych do formalnego opisu wiedzy, szczególnie wiedzy niepełnej i niedokładnej. Z uwagi na to, że granularność dostępnej informacji może powodować niespójność opisu obiektów, teoria ta została zaproponowana jako narzędzie analizy informacji granularnej [Nowak 2013].

Zdaniem Z. Pawlaka z praktycznego punktu widzenia teoria zbiorów przybliżonych jest nową metodą analizy danych. Umożliwia ona między innymi: szukanie zależności między danymi, redukcję danych, określenie wagi danych, generowanie reguł decyzyjnych z danych [Pawlak 1982]. W metodzie tej regułowa reprezentacja wiedzy (w formie JEŻELI ... TO) jest rezultatem wykonania procesu indukcji reguł decyzji na podstawie zbioru danych reprezentujących modele uczenia występujące w opisie obiektów. Opis ten ma postać tablicy decyzyjnej, w której wiersze reprezentują kolejne przykłady, a kolumny wybrane cechy (atrybuty opisujące).

Wśród metod bezpośredniej indukcji reguł decyzji można wyróżnić procedury oparte na sekwencyjnym pokrywaniu (*sequential covering*) zbioru przypadków uczących. W konsekwencji uzyskuje się model uczenia, mający postać zbioru reguł, który pokrywa wszystkie przypadki ze zbioru uczącego. Przykładem algorytmu bezpośredniej indukcji reguł jest zastosowany w badaniach LEM2 [Grzymala-Busse 1992]. To popularna procedura tworzenia minimalnego zbioru reguł, dogodnych do celów klasyfikacyjnych. Tworzy ona symboliczny opis dyskryminujący przybliżenia każdej klasy decyzyjnej za pomocą oryginalnej propozycji zbliżonej do zasady generowania kolejnych pokryć. Algorytm LEM2 na wejściu otrzymuje aproksymację rozpatrywanej klasy decyzyjnej. Jeżeli jest to dolna aproksymacja, generowane są reguły pewne, jeżeli górna – reguły możliwe. W badaniu można także rozpatrywać obszar brzegowy, wówczas generowane są reguły przybliżone.

Innym algorytmem indukcji reguł wykorzystanym w badaniach praktycznych jest procedura tworząca wyczerpujący ich zbiór. Jest ona nastawiona na poszukiwanie zbioru wszystkich reguł w przyjętej składni, które można wygenerować z danej tablicy. W metodzie tej koncentrujemy się na redukcji tablicy decyzyjnej, wykorzystując do tego celu najczęściej postępowanie wnioskowania boolowskiego, oparte na funkcjach rozróżnialności [Skowron 1993].

3. Oferty pracy dla zawodów specjalistycznych

Materiał empiryczny wykorzystany w przeprowadzonych analizach pochodził z badań realizowanych w 2012 r. w ramach projektu „Bilans Kapitału Ludzkiego” – BKL (III edycja) [<http://bkl.parp.gov.pl/dane>]. Objęto nim oferty pracy (z wykluczeniem staży i praktyk dla studentów i uczniów) zgromadzone w powiatowych urzędach pracy (PUP) wszystkich województw Polski oraz na Careerjet.pl, ogólnokrajowym portalu internetowym pośrednictwa pracy (wyszukiwarka ofert pracy). W III edycji BKL wylosowano 160 urzędów (10 placówek na województwo). Próba była dobrana w sposób wyczerpujący, a oferty aktualne w wybranym dniu – 26 marca 2012 r. (w przypadku PUP wzięto pod uwagę ofertę aktualną w dniu przeprowadzania badania, a w przypadku Careerjet.pl jako pierwsze kodowane były oferty zarejestrowane w tym dniu).

W artykule skoncentrowano się na informacjach dotyczących ofert kierowanych do następujących specjalistów:

- nauk fizycznych, matematycznych i technicznych (S1),
- do spraw zdrowia (S2),
- nauczania i wychowania (S3),
- do spraw ekonomicznych i zarządzania (S4),
- do spraw technologii informacyjno-komunikacyjnych (S5),
- z dziedziny prawa, dziedzin społecznych i kultury (S6).

Liczebność próby, wykorzystana w badaniach prezentowanych w pracy, wynosiła 1709 ofert, które pochodziły z poszczególnych województw Polski.

Analiza danych z projektu BLK wskazała, że w latach 2010-2012 największe zapotrzebowanie odnotowano na robotników wykwalifikowanych. Nieznaczny wzrost dotyczył także operatorów i monterów oraz robotników niewykwalifikowanych. Natomiast popyt na specjalistów w 2010 r. wynosił 26%. W następnym roku spadł do 15%, a w 2012 r. ponownie wzrósł do 23%.

Formułowane przez pracodawców zapotrzebowanie na fachowców wśród wszystkich poszukiwanych pracowników było największe w firmach i instytucjach zatrudniających powyżej 50 osób. Przede wszystkim poszukiwano specjalistów do spraw ekonomicznych i zarządzania, inżynierów (ekspertów w zakresie nauk fizycznych, chemicznych i technicznych) oraz specjalistów do spraw zdrowia. Zestawienie popytu na stanowiska specjalistyczne zgłaszane przez różnej wielkości podmioty gospodarcze przedstawia tabela 1.

Tabela 1. Deklarowany przez pracodawców popyt na specjalistów

Stanowiska specjalistyczne		Zapotrzebowanie na specjalistów (w %) próby połączone z lat 2010-2012		
		liczba zatrudnionych		
		1-9	10-49	pow. 50
Specjaliści	nauk fizycznych, matematycznych i technicznych	5	4	12
	do spraw zdrowia	5	3	9
	nauczania i wychowania	2	5	6
	do spraw ekonomii i zarządzania	5	6	15
	do spraw technologii informacyjno-komunikacyjnych	3	2	4
	do spraw prawa, dziedzin społecznych i kultury	2	1	2
N		7335	441	188

Źródło: opracowanie własne na podstawie [Strzebońska, Kocór 2012, s. 57].

Wymagania pracodawców stawiane profesjonalistom dotyczyły najczęściej odpowiedniego wykształcenia, doświadczenia i znajomości języka obcego, szczególnie języka: angielskiego, niemieckiego i francuskiego. Mniejszą wagę przywiązywano do płci, chociaż mężczyźni częściej byli poszukiwani do pracy jako inżynierowie i specjaliści IT. Natomiast kobiety preferowano na stanowisku specjalisty do spraw ekonomii i zarządzania oraz w zawodzie nauczyciela. Język obcy był wymagany przede wszystkim od ekspertów do spraw: technologii IT oraz ekonomii i zarządzania.

W latach 2010-2012 zaobserwowano, że najbardziej dla pracodawców liczyły się kompetencje: interpersonalne (porozumiewanie się, współpraca w grupie i komunikatywność), samoorganizacyjne (właściwa organizacja swojej pracy, przedsiębiorczość, inicjatywa) oraz zawodowe. Dodatkowo w przypadku specjalistów IT oraz inżynierów wymagano kompetencji: komputerowych oraz kognitywnych (umiejętności analizowania informacji i formułowania wniosków, szybkiego myślenia i uczenia się). Natomiast w przypadku lekarzy i pielęgniarek stawiano przede wszystkim na kompetencje

zawodowe. Od nauczycieli i specjalistów wychowania oczekiwano umiejętności związanych ze znajomością przepisów prawa, dotyczących zamówień publicznych oraz zarządzania przedsiębiorstwem. Tymczasem w zawodach społecznych należało posiadać kompetencje biurowe. Najrzadziej liczone, że specjalista będzie dysponował kompetencjami matematycznymi, fizycznymi i kierowniczymi.

4. Badanie potrzeb kompetencyjnych polskiego rynku pracy

Prezentowane w artykule badanie dotyczące poszukiwania zależności pomiędzy rodzajem specjalności, doprecyzowanej profilem kompetencyjnym, a wymogiem znajomości języka angielskiego polegało na zastosowaniu dwóch algorytmów indukcji reguł decyzyjnych generujących dla każdej granuli wyczerpujący oraz minimalny (LEM2) ich zbiór. Dane ujęto w tablicy decyzyjnej, zawierającej 11 atrybutów warunkowych, które przyjmowały wartości {tak/nie} dotyczących kompetencji: indywidualnych (k1), artystycznych (k2), fizycznych (k3), interpersonalnych (k4), kierowniczych (k5), dyspozycyjnych (k6), biurowych (k7), technicznych (k8), komputerowych (k9), matematycznych (k10), zawodowych (k11) oraz dwunasty atrybut związany ze specjalnością – Spec{S1, S2, ..., S6}. Atrybutem decyzyjnym była znajomość języka angielskiego {tak/nie}.

Badanie podzielono na dwa etapy. W pierwszym poszukiwano prawidłowości w całej zbiorowości ofert kierowanych do specjalistów (granuli głównej – 1709 ofert). W drugim zbiorowość tę podzielono na 4 subgranule, które uzyskano w wyniku klasyfikacji województw Polski pod względem stopnia zamożności. Materiał empiryczny do grupowania uwzględniał wybrane kategorie makroekonomiczne i pochodził z GUS [*Produkt krajowy...*]. Podziału dokonano za pomocą metody *k*-średnich [Grabiński 1992, s. 124-127], z wykorzystaniem programu Statistica 10. Uzyskane subgranule (klasy) przedstawiono na rysunku 1.

Każdy z dwóch etapów przebiegał według tej samej procedury. Analizę rozpoczynano od zbadania spójności zbioru danych. Po stwierdzeniu występowania danych niespójnych, za pomocą metody ilościowej usuwano oferty będące jej przyczyną. W wyniku tych czynności w etapie pierwszym uzyskano końcowy zestaw, który zawierał 1498 ofert spójnych. Natomiast suma przypadków po wyeliminowaniu niespójności odrębnie w każdej subgranuli (etap 2) wynosiła 1613 ofert (por. tabela 2). Można zauważyć, że pewne przypadki sprzeczne i wyeliminowane na poziomie całej zbiorowości (etap 1) zostały uwzględnione w drugim etapie badań. Oferty te odizolowano od siebie w poszczególnych subgranulach, co ogólnie wpłynęło na zmniejszenie liczby ofert niespójnych.

W drugim kroku procedury zbadano występowanie reduktów w każdym zbiorze. W efekcie tylko w etapie drugim uzyskano zredukowane zbiory atrybutów, zapewniające dotychczasową klasyfikację ofert. W przypadku granuli pierwszej wyeliminowano kompetencje artystyczne (k2), natomiast w przypadku trzeciej i czwartej – kompetencje matematyczne (k10).



Rys. 1. Klasyfikacja województw pod względem stopnia zamożności

Źródło: opracowanie własne na podstawie danych GUS.

Tabela 2. Podsumowanie wyników badań

Zakres badań	Liczba ofert	Liczba atrybutów w redukcje	Usunięte atrybuty	Liczebność wyczerpującego zbioru reguł	Liczebność minimalnego zbioru reguł (oraz dla wsp. pokrycia $\geq 0,5$)	Podział liczby ofert – język angielski
						tak/nie
Granula główna	1498	12	-	325	180 (12)	947/551
Subgranula 1	281	11	{K2}	120	68 (10)	199/82
Subgranula 2	498	12	-	184	92 (11)	314/184
Subgranula 3	543	11	{K10}	152	81 (8)	326/217
Subgranula 4	291	11	{K10}	93	40 (4)	163/128

Źródło: opracowanie własne.

Trzeci krok procedury badawczej polegał na generowaniu wyczerpującego i minimalnego zbiorów reguł decyzyjnych dla każdej granuli z możliwie zredukowaną liczbą atrybutów. Tabela 2 zawiera podsumowanie liczby uzyskanych reguł na tle liczebności granuli i reduktów. Należy wyjaśnić, że kolumna szósta zawiera liczebność minimalnego zbioru reguł, przy czym w nawiasie ujęto liczbę tych, które posiadały najwyższe pokrycie. Obliczono wskaźnik wsparcia, który jest wyrażonym procentowo wskaźnikiem liczebności spójnego zbioru przypadków potwierdzających regułę, oraz wskaźnik wiarygodności (dokładności, bezwzględne)

wsparcia), wyrażający odsetek liczebności niepustego podzbioru przypadków reprezentujących dane pojęcie („tak” lub „nie”) dla atrybutu decyzyjnego. W wyczerpującym zbiorze reguł najwyższe wartości wskaźnika wsparcia kształtowały się następująco: w granuli głównej – 6,54%; w subgranulach 1, 2, 3, 4 kolejno 6,41%, 6,83%, 6,45% i 6,87%. W przypadku wskaźnika wiarygodności otrzymano: w granuli głównej – 17,79%; w subgranulach 1, 2, 3, 4 stosownie 9,05%, 18,48%, 16,13% i 15,63%. Natomiast w minimalnym zbiorze reguł wartości wskaźnika wsparcia były wyższe i wyniosły: w granuli głównej – 14,75%; w subgranulach 1, 2, 3, 4 odpowiednio 14,23%, 15,26%, 16,76% i 30,24%. Z kolei wskaźniki wiarygodności były na poziomie: w granuli głównej – 23,34%; w subgranulach 1, 2, 3, 4 kolejno 20,73%, 24,20%, 27,91% i 53,99%.

5. Wnioski

W badaniu obejmującym zarówno wszystkie oferty kierowane do specjalistów, pochodzące z terenu całej Polski (granula główna), jak i oferty uzyskane z regionów zgrupowanych w poszczególnych subgranulach, na uwagę zasługują najsilniejsze prawidłowości wynikające z minimalnego zbioru reguł. We wszystkich przypadkach silne reguły dotyczyły specjalistów do spraw: technologii informacyjno-komunikacyjnych oraz ekonomicznych i zarządzania.

W ofertach pochodzących z całej Polski (granuli głównej) i województw reprezentujących wszystkie subgranule, jeżeli propozycja pracy dotyczyła pracowników z branży IT, gdzie oprócz kompetencji komputerowych nie sprecyzowano innych wymagań, to oczekiwano od nich znajomości języka angielskiego. W przypadku propozycji, które dotyczyły regionów sklasyfikowanych w subgranulach 1, 2 i 3 od kandydatów specjalizujących się w teleinformatyce, dysponujących umiejętnościami komputerowymi i dodatkowo zawodowymi, wymagano znajomości języka angielskiego. Natomiast jeżeli oferty kierowano do specjalistów ekonomii i zarządzania, posiadających kompetencje: indywidualne, interpersonalne i kierownicze, i jeżeli pochodziły one z województw skupionych w subgranulach 1, 3 i 4, to nie była wymagana znajomość języka angielskiego. Natomiast w ofertach pochodzących z terenów skupionych w subgranuli 2, jeżeli stawiano na umiejętności: indywidualne, interpersonalne, kierownicze i komputerowe, to nie oczekiwano znajomości języka angielskiego. W przypadku granuli głównej, czyli propozycji obejmujących cały kraj, a których adresatami byli specjaliści do spraw ekonomii i zarządzania o kompetencjach indywidualnych, interpersonalnych i komputerowych, umiejętność posługiwania językiem angielskim nie była konieczna.

Grupa zawodowa specjalistów jest wewnątrznie zróżnicowana, dlatego wymagania formułowane wobec nich różnią się w zależności od stanowiska, na które poszukiwani są pracownicy. Duże znaczenie dla pracodawców ma poziom wykształcenia, doświadczenie oraz profil kompetencyjny kandydatów. Od specjali-

stów pracodawcy wymagają posiadania przede wszystkim kompetencji: samoorganizacyjnych, interpersonalnych i zawodowych. Znajomość języków obcych wśród kandydatów na stanowiska specjalistyczne ma znaczenie przede wszystkim na etapie rozmowy z pracodawcami, podczas której jest ona weryfikowana.

Literatura

- Bilans Kapitału Ludzkiego, bazy z wynikami III edycji badań BKL realizowanych w 2012 r., <http://bkl.parp.gov.pl/dane> (24.04.2013).
- Borowik B., Borowik B., Laird S. (2009), *Powiązania między rozmytymi pamięciami asocjacyjnymi i rozmytą morfologią matematyczną*, „Czasopismo Techniczne, Zeszyt nr 8, s. 139-157. <http://bkl.parp.gov.pl/dane> (9.09.2013).
- http://bkl.parp.gov.pl/system/files/Downloads/20121128143313/BKL_Raport_2013_int_m.pdf?1364281897 (9.09.2013).
- Nowak A. (2013), *Zbiory przybliżone w obszarze systemów ekspertowych*, www.zsi.tech.us.edu.pl/~anowak/index.php?s=file_download&id=22 (8.08.2013).
- Grabiński T. (1992), *Metody taksonometrii*, Wyd. Akademii Ekonomicznej w Krakowie, Kraków.
- Grzymala-Busse J.W. (1992), *LERS – a system for learning from examples based on rough sets*, [w:] R. Słowiński (red.), *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, Dordrecht, s. 3-18.
- Pawlak Z. (1982), *Rough sets*, „International Journal of Computer and Information Sciences”, 11, s. 341-356.
- Produkt krajowy brutto – Rachunki regionalne w 2010 r.* (2013), GUS, http://www.stat.gov.pl/gus/5840_3594_PLK_HTML.htm (8.08.2013).
- Skowron A. (1993), *Boolean reasoning for decision rules generation*, [w:] J. Komorowski, Z. Ras (red.), *Methodologies for Intelligent Systems*, LNAI 689, Springer-Verlag, Berlin, s. 295-305.
- Strzebońska A., Kocór M. (2012), *Specjaliści na rynku pracy. Wymagania stawiane przez pracodawców*, [w:] J. Górniak (red.), *Młodość czy doświadczenie? Kapitał ludzki w Polsce*. Raport podsumowujący III edycję badań BKL z 2012 r.
- Yao Y.Y. (2005), *Perspectives of Granular Computing*, Proceedings of 2005 IEEE International Conference on Granular Computing, Vol. 1, s. 85-90.
- Zadeh L.A. (1979), *Fuzzy sets and information granularity*, [w:] M. Gupta i in. (red.), *Advances in Fuzzy Set Theory and Applications*, North-Holland Publishing Co., Amsterdam, s. 3-18.
- Zadeh L.A. (1997), *Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic*, Fuzzy Sets and Systems 90, s. 111-127.

THE USE GRANULAR INFORMATION IN THE ANALYSIS OF THE REQUIREMENTS OF THE LABOR MARKET

Summary: The paper presents the progress and results of the test procedure needs competence polish labor market. The main task was to analyze the need for competence (skills and qualifications) reported by polish entrepreneurs. The study looked for the relationship between the type of specialty, fine-tuning the competence profile and there quirked knowledge of English. The empirical material came from a study carried out in 2012 under the project „Human Capital” (third edition). The calculation was carried out with the use of cluster analysis and rough set theory, using the system RSES2.1.

Keywords: information granules, cluster analysis, rough sets, labor market.