

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

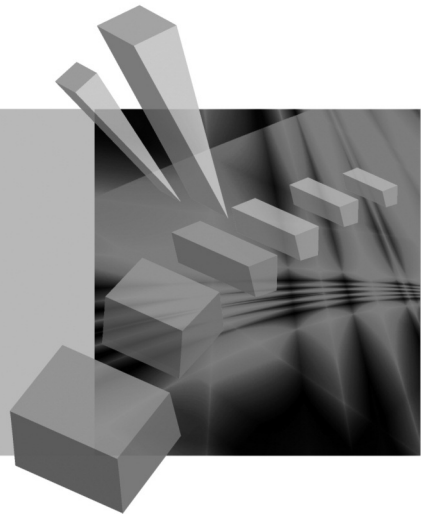
RESEARCH PAPERS

of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,
Mirosław Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w 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](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

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

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

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Wstęp	13
Stanisława Bartosiewicz , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej	17
Andrzej Sokolowski , Q uniwersalna miara odległości	22
Eugeniusz Gatnar , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)	31
Marek Walesiak , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
Krzysztof Jajuga, Marek Walesiak , XXV lat konferencji taksonomicznych – fakty i refleksje	47
Józef Pocięcha, Barbara Pawelek , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne	50
Paweł Lula , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych	58
Ewa Roszkowska , Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji.....	68
Andrzej Młodak , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne	76
Andrzej Bąk , Modele kategorii nieuporządkowanych w badaniach preferencji	86
Jacek Kowalewski , Zintegrowany model optymalizacji badań statystycznych.....	96
Jan Paradysz, Karolina Paradysz , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
Tomasz Szubert , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
Izabela Szamrej-Baran , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne	126
Sylwia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych	144
Hanna Dudek , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów	153

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka, Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
Ewa Chodakowska, Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
Bartosz Soliński, Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
Krzysztof Szwarz, Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
Elżbieta Gołata, Grażyna Dehnel, Rejestry administracyjne w analizie przedsiębiorczości.....	202
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień, Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
Katarzyna Dębkowska, Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
Alina Bojan, Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
Justyna Brzezińska, Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka, Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
Bartłomiej Jefmański, Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
Julita Stańczuk, Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
Jerzy Krawczuk, Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
Anna Czapkiewicz, Beata Basiura, Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
Radosław Pietrzyk, Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
Aleksandra Witkowska, Marek Witkowski, Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
Marcin Pelka, Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
Justyna Wilk, Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

Tomasz Bartłomowicz, Justyna Wilk , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
Kamila Migdał-Najman , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących	342
Dorota Rozmus , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	352
Krzysztof Najman , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG	361
Małgorzata Misztal , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna	370
Mariusz Kubus , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
Barbara Batóg, Jacek Batóg , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym	387
Katarzyna Wójcik, Janusz Tuchowski , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej	396
Iwona Staniec , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach	406
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami	416
Iwona Foryś , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
Ewa Genge , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
Jerzy Korzeniewski , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień	444
Andrzej Dudek , SMS – propozycja nowego algorytmu analizy skupień	451
Artur Mikulec , Metody oceny wyniku grupowania w analizie skupień.....	460
Małgorzata Machowska-Szewczyk , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych	469
Artur Zaborski , Analiza PROFIT i jej wykorzystanie w badaniu preferencji	479
Karolina Bartos , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena	488

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych .	496
Izabela Kurzawa , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś	505
Aleksandra Łuczak, Feliks Wysocki , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych	513
Agnieszka Sompolska-Rzechuła , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim	523
Joanna Banaś, Małgorzata Machowska-Szewczyk , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego	532
Iwona Bąk , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę	541
Aneta Becker , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
Katarzyna Dębowska , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej	562
Anna Domagała , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej	580
Hanna Gruchociak , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
Tomasz Klimanek, Marcin Szymkowiak , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy	601
Jarosław Lira , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce	610
Christian Lis , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku	619
Beata Bieszk-Stolorz, Iwona Markowicz , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
Lucyna Przezbórska-Skobiej, Jarosław Lira , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
Paweł Ulman , Model rozkładu wydatków a funkcje popytu.....	646
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Zastosowanie metod analizy statystycznej w badaniach mięczaków	655

Summaries

Stanisława Bartosiewicz , The effects of subjectivism in multivariate analysis revisited.....	21
Andrzej Sokółowski , Q universal distance measure	30
Eugeniusz Gatnar , Data quality in central banks' statistical systems (NBP example)	38
Marek Walesiak , Distance measures for ordinal data – strategies of proceedings.....	46
Krzysztof Jajuga, Marek Walesiak , XXV years of taxonomic conferences – some facts and remarks.....	49
Józef Pocięcha, Barbara Pawelek , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
Paweł Lula , Learning-based systems of information extraction from textual resources	67
Ewa Roszkowska , The application of the TOPSIS method to support the negotiation process	75
Andrzej Młodak , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
Andrzej Bąk , Models for unordered categories in preference analysis.....	95
Kowalewski Jacek , An integrated model of optimizing statistical surveys	105
Jan Paradysz, Karolina Paradysz , Areas of unemployment in Poland – benchmark problem	115
Tomasz Szubert , How to play to lose the least? Classification of systems in sports bets	125
Izabela Szamrej-Baran , Classification of EU member states in view of fuel poverty	134
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , An attempt to use the gravity model in the analysis of commuters.....	143
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households	152
Hanna Dudek , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study	172
Ewa Chodakowska , Selected methods of classification in schools' rating.....	181
Bartosz Soliński , Renewable energy sector in the European Union – classification in the light of change management strategy	191
Krzysztof Szwarc , Classification of Wielkopolska voivodeship due to the demographic situation	201

Elżbieta Gołata, Grażyna Dehnel , Administrative registers in business analysis.....	211
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
Katarzyna Dębowska , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
Alina Bojan , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
Justyna Brzezińska , Log-linear analysis in the study of mortality in EU.....	246
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Latent class analysis in student satisfaction surveys.....	254
Bartłomiej Jefmański , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
Julita Stańczuk , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
Jerzy Krawczuk , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
Anna Czapkiewicz, Beata Basiura , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
Radosław Pietrzyk , Timing and selectivity in mutual funds performance measurement.....	305
Aleksandra Witkowska, Marek Witkowski , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
Marcin Pelka , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
Justyna Wilk , Comparative study of symbolic data classification software.....	332
Tomasz Bartłomowicz, Justyna Wilk , Application of symbolic data analysis methods for domain database searching.....	341
Kamila Migdał-Najman , A proposal of hybrid clustering method based on self-learning networks.....	351
Dorota Rozmus , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
Krzysztof Najman , A dynamic grouping based on self-learning GNG networks.....	369
Małgorzata Misztal , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
Mariusz Kubus , The application of pre-conditioning of explanatory variable for feature selection.....	386
Barbara Batóg, Jacek Batóg , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395

Katarzyna Wójcik, Janusz Tuchowski , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
Iwona Staniec , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
Iwona Foryś , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
Ewa Genge , Trimming approach to the mixtures of normal distributions.....	443
Jerzy Korzeniewski , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
Andrzej Dudek , SMS – proposal of new clustering algorithm.....	459
Artur Mikulec , Evaluation methods for the grouping result in cluster analysis.....	468
Małgorzata Machowska-Szewczyk , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
Artur Zaborski , PROFIT analysis and its using in the research of preferences.....	487
Karolina Bartos , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
Izabela Kurzawa , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
Aleksandra Luczak, Feliks Wysocki , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
Agnieszka Sompolska-Rzechuła , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
Joanna Banaś, Małgorzata Machowska-Szewczyk , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
Iwona Bąk , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
Aneta Becker , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
Katarzyna Dębowska , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

Anna Domagała , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Statistical analysis in demand research of ICT services in mobile networks.....	589
Hanna Gruchociak , Construction of regression estimator for two-level data	600
Tomasz Klimanek, Marcin Szymkowiak , Application of spatial models in indirect estimation of some labor market characteristics	609
Jarosław Lira , Forecasting of hog livestock production profitability in Poland	618
Christian Lis , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports	627
Beata Bieszk-Stolorz, Iwona Markowicz , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers	636
Lucyna Przezbórska-Skobiej, Jarosław Lira , Agritourism space of Poland and its valuation.....	645
Paweł Ulman , Model of expenses distribution and demand functions.....	654
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Methods of statistical analysis in research of molluscs	663

Aleksandra Łuczak, Feliks Wysocki

Uniwersytet Przyrodniczy w Poznaniu

METODY PORZĄDKOWANIA LINIOWEGO OBIEKTÓW OPISANYCH ZA POMOCĄ CECH METRYCZNYCH I PORZĄDKOWYCH

Streszczenie: W artykule przedstawiono trzy podejścia do konstrukcji syntetycznego miernika rozwoju obiektów opisanych za pomocą cech mierzonych na skalach metrycznych i porządkowej. Podejście pierwsze polega na sprowadzeniu cech metrycznych i porządkowych do określeń lingwistycznych i liczb rozmytych (osłabienie skali cech metrycznych). Podejście drugie wiąże się z zastosowaniem uogólnionej miary odległości GDM zaproponowanej przez Walesiaka (zachowanie skal pomiarowych). Podejście trzecie polega na zamianie wartości cech porządkowych (lingwistycznych) na liczby rozmyte, co prowadzi do wyrażenia wartości tych cech w skali ilorazowej (wzmocnienie skali cech porządkowych). Proponowane podejścia zostały zastosowane do porządkowania liniowego powiatów województwa wielkopolskiego według stanu i jakości infrastruktury technicznej.

Słowa kluczowe: porządkowanie liniowe, cechy metryczne, cechy porządkowe, uogólniona miara odległości (GDM).

1. Wstęp

Celem artykułu jest prezentacja trzech podejść do konstrukcji syntetycznego miernika rozwoju obiektów opisanych za pomocą cech mierzonych na skalach metrycznych i porządkowej. Podejście pierwsze polega na sprowadzeniu cech metrycznych i porządkowych do określeń lingwistycznych i liczb rozmytych (osłabienie skali cech metrycznych) [Singh, Benyoucef 2011]. Podejście drugie wiąże się z zastosowaniem uogólnionej miary odległości GDM zaproponowanej przez Walesiaka [1993; 2006] (zachowanie skal pomiarowych). W podejściu trzecim wartości cech porządkowych (lingwistycznych) zamienia się na liczby rozmyte, co prowadzi do ich wyrażenia w skali ilorazowej (sztuczne wzmocnienie skali cech porządkowych) [Wysocki, Łuczak 2006].

W procedurze tworzenia cechy syntetycznej zastosowano miarę odległości między trójkątnymi liczbami rozmytymi zaproponowaną przez Chena [2000], a także uogólnioną miarę GDM Walesiaka [2006], a do agregacji cech metodę TOPSIS (*Technique for Order Preference by Similarity to an Ideal Solution*) – rozmytą (po-

dejsie pierwsze i trzecie) oraz klasyczną (podejście drugie) [Wysocki 2010]. Dzięki zastosowaniu do konstrukcji mierników syntetycznych teorii zbiorów rozmytych oraz uogólnionej miary odległości GDM można rozwiązywać problemy klasyfikacji jakościowej, które często pojawiają się w praktyce badań ekonomiczno-społecznych [Borys 1984].

Proponowane podejścia zostały zastosowane do porządkowania liniowego powiatów województwa wielkopolskiego według stanu i jakości infrastruktury technicznej.

2. Metodyka badań

W procesie tworzenia cechy syntetycznej można wyróżnić osiem etapów postępowania (rys. 1). Pierwszym z nich jest wybór cech opisujących wybrane obiekty, który jest podstawą ich uporządkowania liniowego (etap I). W zasadzie istnieją dwa podstawowe podejścia do wyboru cech: statystyczne i merytoryczne (zob. [Wysocki 2010]).

Podejścia			Etapy postępowania
I (osłabienie skal cech metrycznych)	II (zachowanie skal pomiarowych)	III (sztuczne wzmocnienie skali cech porządkowych)	
Dobór cech (metrycznych i porządkowych) oraz ich weryfikacja pod względem merytorycznym i/lub statystycznym			<i>I. Wybór cech</i>
Ustalenie kierunku preferencji cech prostych w stosunku do rozpatrywanego kryterium ogólnego			<i>II. Podział cech według kierunku preferencji</i>
Ustalenie wag dla cech na podstawie analizy statystycznej lub/i merytorycznej			<i>III. Przyjęcie systemu wag dla cech</i>
Sprowadzenie wartości cech metrycznych i porządkowych do określeń lingwistycznych i liczb rozmytych	Nie wymaga przekształcenia cech	Zamiana wartości cech porządkowych na liczby rozmyte	<i>IV. Przekształcenie wartości cech</i>
Nie wymaga normalizacji	Wymaga normalizacji cech metrycznych	Wymaga normalizacji wszystkich cech	<i>V. Normalizacja wartości cech</i>
Obliczenie odległości rozmytej metodą wierzchołkową (<i>vertex method</i>)	Obliczenie uogólnionej miary odległości GDM	Obliczenie odległości rozmytej metodą wierzchołkową (<i>vertex method</i>)	<i>VI. Obliczenie odległości od wzorca i antywzorca rozwoju</i>
Rozmyta TOPSIS	Klasyczna TOPSIS	Rozmyta TOPSIS	<i>VII. Agregacja cech – metody</i>
Wykorzystanie odchylenia standardowego i średniej arytmetycznej z wartości syntetycznego miernika rozwoju			<i>VIII. Uporządkowanie liniowe obiektów i wyodrębnienie ich klas typologicznych</i>

Rys. 1. Etapy konstrukcji cechy syntetycznej

Źródło: opracowanie własne.

Następnie należy ustalić kierunek preferencji cech prostych w stosunku do rozpatrywanego kryterium ogólnego (etap II). W konstrukcji cechy syntetycznej można również cechom prostym przyporządkować wagi (etap III). Mogą one być ustalone na podstawie analizy statystycznej lub merytorycznej. Pierwsze podejście wykorzystuje procedury statystyczne, bazując na informacjach o cechach tkwiących tylko w samej macierzy danych, a szczególnie wykorzystuje analizę zmienności cech oraz analizę korelacji między cechami albo tylko jedną z tych analiz [Wysocki 2010]. Drugie podejście opiera się na opiniach ekspertów.

W czwartym etapie następuje przekształcenie wartości cech. W podejściu I można zastosować metodę klasyfikacji wariancyjnej, m.in. w taki sposób, aby dla każdej cechy uzyskać szereg rozdzielczy z 5 przedziałami klasowymi o równej rozpiętości, a następnie przeprowadzić rangowanie uzyskanych przedziałów klasowych przez przyporządkowanie im poziomów cechy lingwistycznej (zob. [Wysocki, Łuczak 2006, tab. 1]). W przypadku destymulant rangowanie jest odwrócone. Następnie poziomy cechy lingwistycznej zostają zamienione na liczby rozmyte według zasad podanych w tab. 1.

Podejście II nie wymaga przekształcenia wartości cech. Natomiast w podejściu III wymagana jest zamiana wartości cech porządkowych na liczby rozmyte analogicznie jak w podejściu I.

Tabela 1. Cecha lingwistyczna – jej poziomy i odpowiadające im liczby rozmyte wykorzystywane do opisu cech porządkowych

Wartość cechy metrycznej	$\langle 0; 0,2 \rangle$	$\langle 0,2; 0,4 \rangle$	$\langle 0,4; 0,6 \rangle$	$\langle 0,6; 0,8 \rangle$	$\langle 0,8; 1,0 \rangle$
Poziomy cechy lingwistycznej	bardzo niski (BN)	niski (N)	średni (Ś)	wysoki (W)	bardzo wysoki (BW)
Trójkątne liczby rozmyte (a, b, c)	$(0, 0, 20)$	$(20, 30, 40)$	$(40, 50, 60)$	$(60, 70, 80)$	$(80, 100, 100)$

Źródło: [Chang, Yeh 2004].

Wybrane cechy zazwyczaj mają też różne miana, dlatego należy stosować odpowiednie procedury normalizacyjne (etap V) [Wysocki 2010]. W pracy zastosowane zostały przekształcenia ilorazowe. W podejściu I wartości cech zostają zamienione na liczby rozmyte według jednolitej zasady (tab. 1), a zatem nie wymagają normalizacji. Zastosowanie podejścia II wymaga wcześniejszej normalizacji cech metrycznych (zob. [Walesiak 2006]), a podejście III – normalizacji wszystkich cech (metrycznych i porządkowych), których wartości zostały uprzednio zamienione na liczby rozmyte [Wysocki 2010]. Normalizację cech przeprowadzono z wykorzystaniem klasycznych przekształceń ilorazowych (cechy metryczne) oraz ilorazowych wykonanych na liczbach rozmytych (cechy porządkowe) [Wysocki 2010].

W etapie VI oblicza się odległości każdego ocenianego obiektu ($i = 1, \dots, N$) od obiektu – wzorca $N + 1$ (+) i obiektu – antywzorca rozwoju $N + 2$ (+). Jeżeli w zbiorze cech prostych oprócz cech metrycznych występują cechy porządkowe, to do ob-

liczenia tych odległości można zastosować formułę odległości wierzchołkowej (*vertex method*) [Chen 2000], która jest zazwyczaj wykorzystywana w obliczaniu odległości między dwiema trójkątnymi liczbami rozmytymi (zob. [Wysocki 2010]) lub uogólnioną miarę odległości dla cech z różnych skal pomiaru (zob. [Walesiak 2006]). A zatem w podejściach I i III odległość każdego ocenianego obiektu od rozmytego obiektu – wzorca

$$\tilde{A}^+ = \left(\max_i(\tilde{x}_{i1}), \max_i(\tilde{x}_{i2}), \dots, \max_i(\tilde{x}_{iK}) \right) = (\tilde{x}_1^+, \tilde{x}_2^+, \dots, \tilde{x}_K^+)$$

i obiektu – antywzorca rozwoju

$$\tilde{A}^- = \left(\min_i(\tilde{x}_{i1}), \min_i(\tilde{x}_{i2}), \dots, \min_i(\tilde{x}_{iK}) \right) = (\tilde{x}_1^-, \tilde{x}_2^-, \dots, \tilde{x}_K^-):$$

$$d_i^+ = \sum_{k=1}^K d(\tilde{x}_{ik}; \tilde{x}_k^+), \quad d_i^- = \sum_{k=1}^K d(\tilde{x}_{ik}; \tilde{x}_k^-), \quad (i = 1, 2, \dots, N), \quad k = 1, \dots, K$$

oblicza się jako odległość między dwiema trójkątnymi liczbami rozmytymi: $\tilde{x}_{ik} = (a_{ik}, b_{ik}, c_{ik})$ i $\tilde{x}_k^+ = (a_k^+, b_k^+, c_k^+)$ oraz $\tilde{x}_{ik} = (a_{ik}, b_{ik}, c_{ik})$ i $\tilde{x}_k^- = (a_k^-, b_k^-, c_k^-)$, która jest definiowana jako:

$$d_i^{(\bullet)} = d(\tilde{x}_{ik}; \tilde{x}_k^{(\bullet)}) = \sqrt{\frac{1}{3} \left((a_{ik} - a_k^{(\bullet)})^2 + (b_{ik} - b_k^{(\bullet)})^2 + (c_{ik} - c_k^{(\bullet)})^2 \right)},$$

gdzie (\bullet) oznacza alternatywnie (+) lub (-).

W podejściu II stosuje się formułę [Walesiak 2006]:

$$d_i^{(\bullet)} = d_{ij} = \frac{w_1 d_{ij}^P + w_2 d_{ij}^I + w_3 d_{ij}^R}{w_1 + w_2 + w_3}, \quad i = 1, \dots, N, \quad j = N+1, N+2,$$

gdzie: d_{ij} – miara odległości GDM i -tego obiektu od obiektu – wzorca ($N+1$) oraz od obiektu – antywzorca ($N+2$), (\bullet) oznacza alternatywnie obiekt – wzorzec (+) lub obiekt antyworzec (-), P, I, R – podzbiór cech porządkowych, przedziałowych, ilorazowych, $d_{ij}^P, d_{ij}^I, d_{ij}^R$ – miara odległości dla cech porządkowych, przedziałowych, ilorazowych, w_1, w_2, w_3 – wagi przyporządkowane odległościom wyznaczonym na podstawie cech porządkowych, przedziałowych, ilorazowych, przy czym $w_1 + w_2 + w_3 = K$ (liczba cech); wagi te mogą również wyrażać merytoryczną ważność poszczególnych cech i ich podzbiorów.

Konstrukcja uogólnionej miary odległości określona jest wówczas wzorem [Walesiak 2006]:

$$d_{ij}^{(\bullet)} = \frac{1}{2} - \frac{\sum_{k=1}^K a_{ijk} b_{jik} + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq i, j}}^{N+2} a_{ilk} b_{jlk}}{2 \left[\left(\sum_{k=1}^K a_{ijk}^2 + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq i, j}}^{N+2} a_{ilk}^2 \right) \left(\sum_{k=1}^K b_{jik}^2 + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq i, j}}^{N+2} b_{jlk}^2 \right) \right]^{\frac{1}{2}}},$$

gdzie: (\bullet) oznacza alternatywnie P, I, R ; $i, l = 1, \dots, N, N+1, N+2, j = N+1, N+2$ – numeruje obiekty, $N+1$ jest numerem obiektu wzorcowego, a $N+2$ numerem obiektu antywzorcowego, $k = 1, \dots, K$ – numeruje cechy.

Odległości d_{ij}^P dla cech mierzonych na skali porządkowej wyznacza się, stosując podstawienie:

$$a_{iuk}(b_{jtk}) = \begin{cases} 1 & x_{ik} > x_{uk} \left(x_{jk} > x_{tk} \right), \\ 0 & x_{ik} = x_{uk} \left(x_{jk} = x_{tk} \right), \\ -1 & x_{ik} < x_{uk} \left(x_{jk} < x_{tk} \right), \end{cases} \text{ dla } u = j, l; t = i, l,$$

gdzie: $x_{ik}(x_{jk}, x_{lk}, x_{uk}, x_{tk})$ – i -ta (j -ta, l -ta, u -ta, t -ta) obserwacja na k -tej cesze.

Odległości d_{ij}^I oraz d_{ij}^R wyznacza się dla cech mierzonych na skali ilorazowej i/lub przedziałowej, podstawiając w powyższej formule:

$$a_{iuk} = x_{ik} - x_{uk} \text{ dla } u = j, l, \quad b_{jtk} = x_{jk} - x_{tk} \text{ dla } t = i, l,$$

gdzie $x_{ik}(x_{jk}, x_{lk})$ – i -ta (j -ta, l -ta) obserwacja na k -tej cesze.

W kolejnym etapie VII następuje agregacja cech. Oblicza się wartości syntetycznego miernika rozwoju:

$$S_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad 0 \leq S_i \leq 1, \quad (i = 1, 2, \dots, N).$$

Im mniejsza jest odległość danego obiektu od obiektu modelowego – wzorca rozwoju, a tym samym większa od drugiego bieguna – antywzorca rozwoju, tym wartość miernika syntetycznego jest bliższa 1.

Ostatnim VIII etapem jest uporządkowanie liniowe obiektów i wyodrębnienie klas typologicznych.

3. Wyniki badań

W badaniach wykorzystano dane pierwotne pochodzące z ankiety przeprowadzonej w starostwach powiatowych województwa wielkopolskiego nt. *Stanu i możliwości*

rozwojowych powiatów województwa wielkopolskiego oraz dane statystyczne z Głównego Urzędu Statystycznego (*Bank Danych Lokalnych* 2009). W pierwszym etapie oceny poziomu rozwoju infrastruktury technicznej dokonano wyboru 10 cech reprezentujących stan i jakość różnych elementów infrastruktury w przekroju powiatów województwa wielkopolskiego obejmujących *cechy metryczne*: 1) odsetek ludności mającej dostęp do sieci wodociągowej (%), 2) odsetek ludności mającej dostęp do sieci kanalizacyjnej (%), 3) odsetek ludności mającej dostęp do sieci gazowej (%), 4) odsetek ludności obsługiwanej przez oczyszczalnie ścieków (%), 5) drogi publiczne o twardej nawierzchni (powiatowe i gminne) w km na 100 km² i *cechy porządkowe* wyrażające jakość elementów infrastruktury w punktach: 6) dróg gminnych i powiatowych, 7) sieci kanalizacyjnej, 8) sieci wodociągowej, 9) sieci gazowej, 10) oczyszczalni ścieków.

W drugim etapie przyjęto, że wszystkie cechy mają charakter stymulant, charakteryzują najistotniejsze elementy infrastruktury technicznej, a ich współczynniki wagowe w procesie konstrukcji cechy syntetycznej będą jednakowe (etap III). Następnie cechy metryczne i porządkowe (podejście I) oraz tylko cechy porządkowe (podejście III) zostały przekształcone na liczby rozmyte (etap IV). W kolejnym etapie wartości cech metrycznych w podejściu II oraz cech metrycznych i porządkowych w podejściu III poddano normalizacji przez przekształcenia ilorazowe (etap V).

W etapie VI w podejściu I i III obliczono rozmyte odległości od wzorca i antywzorca rozwoju metodą wierzchołkową, a w podejściu II do obliczenia odległości zastosowano uogólnioną miarę odległości GDM¹. Stanowiły one podstawę do obliczenia wartości syntetycznych mierników poziomu rozwoju infrastruktury technicznej S_i ($i = 1, 2, \dots, N$) dla każdego powiatu.

Tabela 2 pokazuje wartości syntetycznego miernika rozwoju infrastruktury technicznej dla powiatów według trzech proponowanych podejść oraz uzyskane w tym procesie rangi powiatów. Jednym z powiatów, dla którego wystąpiły znaczne różnice w rangowaniu, jest powiat ostrzeszowski. W podejściu I powiat ten uzyskał rangę 7,5, a w podejściach II i III uplasował się dopiero na 13 pozycji. Dodatkowo, gdyby nie zostały uwzględnione w procesie porządkowania liniowego cechy porządkowe, to powiat ten uplasowałby się dopiero na 25 miejscu (podejście II*) lub 24 miejscu (podejście III*), czyli odpowiednio o 12 i 11 miejsc niżej niż w analogicznych podejściach uwzględniających w konstrukcji cechy syntetycznej zarówno cechy metryczne, jak i porządkowe.

Przykładem dużych rozbieżności w uporządkowaniu może być również powiat gnieźnieński. Gdyby ocenę oprzeć tylko na cechach metrycznych, to uplasowałby się na pierwszym miejscu pod względem poziomu rozwoju infrastruktury technicznej w województwie wielkopolskim. Biorąc pod uwagę również jej jakość, należy stwierdzić, że sytuacja powiatu jest znacznie gorsza, gdyż zajmuje on dopiero 9 (podejście II) lub 8 (podejście III) pozycję.

¹ W obliczeniach wykorzystano program R.

Tabela 2. Wartości cechy syntetycznej poziomu rozwoju infrastruktury technicznej według proponowanych podejść dla powiatów województwa wielkopolskiego

Lp.	Powiaty	Wartości syntetycznego miernika/podejścia					Rangi powiatów/podejścia				
		I	II	III	II*	III*	I	II	III	II*	III*
1	chodzieski	0,502	0,671	0,551	0,785	0,556	9	6	6	4	6
2	czarnkowsko-trzecieński	0,327	0,382	0,361	0,14	0,261	25	23	24	30	29
3	gnieźnieński	0,536	0,647	0,548	0,864	0,584	3,5	9	8	1	1
4	gostyński	0,603	0,775	0,630	0,758	0,560	1,5	2	2	7	4
5	grodziski	0,503	0,645	0,551	0,421	0,403	7,5	10	6	14	14
6	jarociński	0,536	0,672	0,555	0,771	0,535	3,5	5	4	5	7
7	kaliski	0,193	0,196	0,194	0,141	0,258	31	29	31	29	30
8	kępiński	0,369	0,343	0,390	0,325	0,370	21	24	23	20	19
9	kolski	0,406	0,408	0,401	0,332	0,368	17	21	22	19	20
10	koniński	0,373	0,447	0,424	0,312	0,360	20	18	18	21	21
11	kościański	0,463	0,669	0,539	0,637	0,474	12,5	7	9	12	12
12	krotoszyński	0,509	0,65	0,534	0,715	0,521	6	8	10	8	8
13	leszczyński	0,193	0,151	0,251	0,063	0,207	30	31	29	31	31
14	międzychodzki	0,463	0,538	0,467	0,348	0,373	12,5	15	16	18	18
15	nowotomyski	0,362	0,441	0,411	0,266	0,338	22	19	20	24	25
16	obornicki	0,261	0,204	0,285	0,374	0,385	27	28	28	17	17
17	ostrowski	0,435	0,521	0,473	0,708	0,514	16	16	14,5	10	9
18	ostrzeszowski	0,503	0,566	0,504	0,256	0,343	7,5	13	13	25	24
19	pilski	0,462	0,594	0,514	0,769	0,558	15	12	12	6	5
20	pleszewski	0,333	0,306	0,352	0,298	0,352	24	25	25	22	23
21	poznański	0,603	0,825	0,651	0,812	0,565	1,5	1	1	3	3
22	rawicki	0,496	0,680	0,551	0,667	0,489	10	4	6	11	11
23	stłupecki	0,209	0,152	0,247	0,238	0,323	29	30	30	26	26
24	szamotulski	0,334	0,383	0,408	0,287	0,353	23	22	21	23	22
25	średzki	0,463	0,626	0,520	0,714	0,503	12,5	11	11	9	10
26	śremski	0,529	0,752	0,588	0,827	0,567	5	3	3	2	2
27	turecki	0,394	0,416	0,420	0,226	0,321	19	20	19	27	27
28	wągrowiecki	0,396	0,450	0,448	0,455	0,413	18	17	17	13	13
29	wolsztyński	0,463	0,558	0,473	0,382	0,388	12,5	14	14,5	15	15,5
30	wrzesiński	0,260	0,248	0,334	0,377	0,388	28	27	26,5	16	15,5
31	złotowski	0,294	0,271	0,334	0,199	0,301	26	26	26,5	28	28
	max	0,603	0,825	0,651	0,864	0,584	×				
	min	0,193	0,151	0,194	0,063	0,207					
	Rozstęp	0,410	0,674	0,457	0,801	0,377					

* W konstrukcji cechy syntetycznej uwzględniono tylko cechy metryczne. Oznacza to, że podejście I* jest tożsame z III*.

Źródło: obliczenia własne na podstawie ankiety przeprowadzonej w starostwach powiatowych w województwie wielkopolskim nt. *Stanu i możliwości rozwojowych powiatów województwa wielkopolskiego* (2000) oraz dane statystyczne Głównego Urzędu Statystycznego z *Banku Danych Lokalnych* (2009), <http://www.stat.gov.pl/bdl>.

Tabela 3. Wartości współczynników korelacji τ -Kendalla pomiędzy wartościami cechy syntetycznej w proponowanych podejściach

		Podejście				
		I	II	III	II*	III*
Podejście	I	1,000	0,846	0,861	0,628	0,636
	II		1,000	0,938	0,630	0,637
	III			1,000	0,631	0,639
	II*				1,000	0,956
	III*					1,000

* W konstrukcji cechy syntetycznej uwzględniono tylko cechy metryczne.

Źródło: obliczenia własne na podstawie tab. 2.

Następnie zbadano zgodność wyników uporządkowań powiatów w prezentowanych podejściach (tab. 3). Za miarę zgodności uporządkowań przyjęto wartości współczynnika korelacji rang τ -Kendalla². Są one wysokie (0,85-0,95) dla uporządkowań powiatów według wartości mierników uwzględniających cechy metryczne i porządkowe, a znacznie niższe dla uporządkowań opartych tylko na cechach metrycznych (0,63).

4. Wnioski

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

1. Zaproponowane podejścia do porządkowania liniowego obiektów metodą TOPSIS mogą być zastosowane do wyznaczenia syntetycznego miernika rozwoju dla różnych typów danych – metrycznych oraz porządkowych. Wprowadzenie cech porządkowych zmieniło uporządkowanie liniowe większości powiatów (niektórych istotnie) ze względu na poziom rozwoju infrastruktury technicznej. Świadczą o tym wartości współczynnika korelacji rang τ -Kendalla obliczone dla rang powiatów uzyskanych na podstawie wartości mierników syntetycznych uwzględniających cechy metryczne i porządkowe oraz tylko metryczne.

2. Zastosowane trzy podejścia dostarczyły podobnych uporządkowań liniowych powiatów według rozwoju infrastruktury technicznej, o czym świadczą obliczone dla nich wysokie wartości współczynnika korelacji τ -Kendalla. Największy rozstęp pomiędzy wartością największą i najmniejszą cechy syntetycznej wystąpił w przypadku podejścia II (ponad 0,67), gdzie zastosowano uogólnioną miarę odległości GDM do mierzenia odległości od wzorca i antywzorca rozwoju oraz metodę TOPSIS agregacji cech. W podejściach I i III zakresy zmienności wartości cechy syntetycznej są zbliżone i wynosiły odpowiednio 0,41 i 0,46.

² Współczynnik korelacji τ -Kendalla przyjmuje wartości z przedziału $[-1,1]$. Im jego wartość jest bliższa liczbie 1, tym większe podobieństwo uporządkowań powiatów, zob. [Steczowski, Zeliaś 1997].

3. W podejściu I istnieje możliwość pojawienia się rang powiązanych ze względu na pewną utratę informacji przy przekształcaniu cech metrycznych na określenia lingwistyczne i liczby rozmyte.

4. Najwyższy poziom rozwoju infrastruktury technicznej wystąpił na obszarze powiatu poznańskiego znajdującego się w zasięgu oddziaływania aglomeracji miejskiej Poznania. Natomiast najniższy jest charakterystyczny dla obszarów obejmujących powiaty zdominowane przez gospodarkę rolną, tj.: słupecki, kaliski – ziemski, leszczyński – ziemski.

Literatura

Bank Danych Lokalnych, <http://www.stat.gov.pl/bdl>, 2009.

Borys T., *Kategoria jakości w statystycznej analizie porównawczej*, [w:] Monografie i Opracowania, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 23, Wydawnictwo AE, Wrocław 1984.

Chang Y.H., Yeh C.H., *A new airline safety index*, "Transportation Research", Part B, 38, 2004.

Chen C.T., *Extensions of the TOPSIS for group decision-making under fuzzy environment*, "Fuzzy Sets and Systems" 2000, no 114 (1).

Singh R.K., Benyoucef L., *A fuzzy TOPSIS based approach for e-sourcing*, "Engineering Applications of Artificial Intelligence" 2011, no 24.

Stan i możliwości rozwojowe powiatów województwa wielkopolskiego, Ankieta przeprowadzona w starostwach powiatowych w województwie wielkopolskim, 2000.

Steczkowski J., Zeliaś A., *Metody statystyczne w badaniach zjawisk jakościowych*, Wydawnictwo AE, Kraków 1997.

Walesiak M., *Statystyczna analiza wielowymiarowa w badaniach marketingowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 654, Seria: Monografie i Opracowania nr 101, Wydawnictwo AE, Wrocław 1993.

Walesiak M., *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, Wydawnictwo AE, Wrocław 2006.

Wysocki F., Łuczak A., *Rozmyta wielokryterialna metoda porządkowania liniowego obiektów*, [w:] Taksonomia 13, *Klasyfikacja i analiza danych. Teoria i zastosowania*, AE, Wrocław 2006.

Wysocki F., *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*, Wydawnictwo UP, Poznań 2010.

LINEAR ORDERING METHODS OF OBJECTS DESCRIBED BY A SET OF METRIC AND ORDINAL CHARACTERISTICS

Summary: The article presents three approaches to linear ordering of objects described by a set of metric and ordinal characteristics. In the first approach the metric and ordinal characteristics are transformed to linguistic qualifications and the fuzzy numbers (the weakness of scale of metric characteristics). The second approach is joined with the use of generalized distance measure (GDM) proposed by Walesiak (the constant of measuring scales). In the third approach the value of ordinal characteristics (linguistic) is exchanged on fuzzy numbers (strengthening of the scale of ordinal characteristics). The proposed approaches were applied to the arrangement of linear ordering of administrative districts of the Wielkopolska voivodeship according to the state and quality of technical infrastructure.

Keywords: linear ordering, metric characteristics, ordinal characteristics, generalized distance measure (GDM).