

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 327

Taksonomia 22

**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	9
Eugeniusz Gatnar , Balance of payments statistics and external competitiveness of Poland.....	15
Andrzej Sokolowski, Magdalena Czaja , Efektywność metody k -średnich w zależności od separowalności grup.....	23
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Wielosektorowa analiza ukrytych przejść w modelowaniu zagrożenia upadłością polskich przedsiębiorstw	30
Elżbieta Gołata , Zróżnicowanie procesu starzenia i struktur demograficznych w Poznaniu i aglomeracji poznańskiej na tle wybranych dużych miast Polski w latach 2002-2011.....	39
Aleksandra Łuczak, Feliks Wysocki , Ustalanie systemu wag dla cech w zagadnieniach porządkowania liniowego obiektów	49
Marek Walesiak , Wzmacnianie skali pomiaru dla danych porządkowych w statystycznej analizie wielowymiarowej	60
Paweł Lula , Identyfikacja słów i fraz kluczowych w tekstach polskojęzycznych za pomocą algorytmu <i>RAKE</i>	69
Mariusz Kubus , Propozycja modyfikacji metody złagodzonego LASSO.....	77
Andrzej Bąk, Tomasz Bartłomowicz , Wielomianowe modele logitowe wyborów dyskretnych i ich implementacja w pakiecie <i>DiscreteChoice</i> programu R.....	85
Justyna Brzezińska , Wykorzystanie modeli logarytmiczno-liniowych do analizy bezrobocia w Polsce w latach 2004-2012.....	95
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Zastosowanie pakietu <i>dcMNM</i> programu R w badaniach preferencji konsumentów wódki	104
Barbara Batóg, Jacek Batóg , Analiza stabilności klasyfikacji polskich województw według sektorowej wydajności pracy w latach 2002-2010	113
Małgorzata Markowska, Danuta Strahl , Klasyfikacja europejskiej przestrzeni regionalnej ze względu na filary inteligentnego rozwoju z wykorzystaniem referencyjnego systemu granicznego.....	121
Kamila Migdał-Najman, Krzysztof Najman , Formalna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	131
Kamila Migdał-Najman, Krzysztof Najman , Graficzna ocena jakości odwzorowania struktury grupowej na mapie Kohonena	139
Beata Basiura, Anna Czapkiewicz , Badanie jakości klasyfikacji szeregów czasowych	148
Michał Trzęsiok , Wybrane metody identyfikacji obserwacji oddalonych.....	157

Grażyna Dehnel, Tomasz Klimanek , Taksonomiczne aspekty estymacji pośredniej uwzględniającej autokorelację przestrzenną w statystyce gospodarczej.....	167
Michał Bernard Pietrzak, Justyna Wilk , Odległość ekonomiczna w modelowaniu zjawisk przestrzennych z wykorzystaniem modelu grawitacji.....	177
Maciej Beręsewicz , Próba zastosowania różnych miar odległości w uogólnionym estymatorze Petersena	186
Marcin Szymkowiak, Tomasz Józefowski , Konstrukcja i praktyczne wykorzystanie estymatorów typu SPREE na przykładzie dwuwymiarowych tabel kontyngencji	195
Marcin Pelka , Klasyfikacja pojęciowa danych symbolicznych w podejściu wielomodelowym	202
Małgorzata Machowska-Szewczyk , Ocena klas w rozmytej klasyfikacji obiektów symbolicznych.....	210
Justyna Wilk , Problem wyboru liczby klas w taksonomicznej analizie danych symbolicznych.....	220
Andrzej Dudek , Metody analizy skupień w klasyfikacji markerów map Google	229
Ewa Roszkowska , Ocena ofert negocjacyjnych w słabo ustrukturyzowanych problemach negocjacyjnych z wykorzystaniem rozmytej procedury SAW	237
Marcin Szymkowiak, Marek Witkowski , Zastosowanie analizy korespondencji do badania kondycji finansowej banków spółdzielczych.....	248
Bartłomiej Jefmański , Budowa rozmytych indeksów satysfakcji klientów z zastosowaniem programu R.....	257
Karolina Bartos , Odkrywanie wzorców zachowań konsumentów za pomocą analizy koszykowej danych transakcyjnych	266
Joanna Trzęsiok , Taksonomiczna analiza krajów pod względem dzietności kobiet oraz innych czynników demograficznych	275
Beata Bal-Domańska , Próba identyfikacji większych skupisk regionalnych oraz ich konwergencja.....	285
Beata Bieszk-Stolorz, Iwona Markowicz , Wpływ zasiłku na proces poszukiwania pracy	294
Marta Dziechciarz-Duda, Klaudia Przybysz , Wykształcenie a potrzeby rynku pracy. Klasyfikacja absolwentów wyższych uczelni.....	303
Tomasz Klimanek , Problem pomiaru procesu dezagrarnizacji wsi polskiej w świetle wielowymiarowych metod statystycznych.....	313
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Wybrane metody analizy danych wzdluznych.....	321
Artur Zaborski , Zastosowanie miar odległości dla danych porządkowych do agregacji preferencji indywidualnych	330
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Zastosowanie analizy korespondencji do badania sytuacji mieszkańców strefy podmiejskiej Warszawy na rynku pracy.....	338

Katarzyna Wawrzyniak , Klasyfikacja województw według stopnia realizacji priorytetów Strategii Rozwoju Kraju 2007-2015 z wykorzystaniem wartości centrum wierszowego	346
---	-----

Summaries

Eugeniusz Gatnar , Statystyka bilansu płatniczego a konkurencyjność gospodarki Polski	22
Andrzej Sokółowski, Magdalena Czaja , Cluster separability and the effectiveness of k -means method	29
Barbara Pawelek, Józef Pocięcha, Adam Sagan , Multisectoral analysis of latent transitions in bankruptcy prediction models.....	38
Elżbieta Golata , Differences in the process of aging and demographic structures in Poznań and the agglomeration compared to selected Polish cities in the years 2002-2011	48
Aleksandra Łuczak, Feliks Wysocki , Determination of weights for features in problems of linear ordering of objects	59
Marek Walesiak , Reinforcing measurement scale for ordinal data in multivariate statistical analysis	68
Paweł Lula , Automatic identification of keywords and keyphrases in documents written in Polish.....	76
Mariusz Kubus , The proposition of modification of the relaxed LASSO method.....	84
Andrzej Bąk, Tomasz Bartłomowicz , Microeconomic multinomial logit models and their implementation in the <code>DiscreteChoice</code> R package .	94
Justyna Brzezińska , The analysis of unemployment data in Poland in 2004-2012 with application of log-linear models	103
Andrzej Bąk, Marcin Pelka, Aneta Rybicka , Application of the MMLM package of R software for vodka consumers preference analysis.....	112
Barbara Batóg, Jacek Batóg , Analysis of the stability of classification of Polish voivodeships in 2002-2010 according to the sectoral labour productivity	120
Małgorzata Markowska, Danuta Strahl , Classification of the European regional space in terms of smart growth pillars using the reference limit system.....	130
Kamila Migdał Najman, Krzysztof Najman , Formal quality assessment of group structure mapping on the Kohonen's map	138
Kamila Migdał Najman, Krzysztof Najman , Graphical quality assessment of group structure mapping on the Kohonen's map	147
Beata Basiura, Anna Czapkiewicz , Validation of time series clustering	156
Michał Trzęsiok , Selected methods for outlier detection.....	166

Grażyna Dehnel, Tomasz Klimanek , Taxonomic aspects of indirect estimation accounting for spatial correlation in enterprise statistics	176
Michał Bernard Pietrzak, Justyna Wilk , Economic distance in modeling spatial phenomena with the application of gravity model.....	185
Maciej Beręsewicz , An attempt to use different distance measures in the Generalized Petersen estimator	194
Marcin Szymkowiak, Tomasz Józefowski , Construction and practical using of SPREE estimators for two-dimensional contingency tables.....	201
Marcin Pelka , The ensemble conceptual clustering for symbolic data.....	209
Małgorzata Machowska-Szewczyk , Evaluation of clusters obtained by fuzzy classification methods for symbolic objects.....	219
Justyna Wilk , Problem of determining the number of clusters in taxonomic analysis of symbolic data	228
Andrzej Dudek , Clustering techniques for Google maps markers.....	236
Ewa Roszkowska , The evaluation of negotiation offers in ill structure negotiation problems with the application of fuzzy SAW procedure	247
Marcin Szymkowiak, Marek Witkowski , The use of correspondence analysis in analysing the financial situation of cooperative banks.....	256
Bartłomiej Jefmański , The construction of fuzzy customer satisfaction indexes using R program.....	265
Karolina Bartos , Discovering patterns of consumer behaviour by market basket analysis of the transactional data.....	274
Joanna Trzęsiok , Cluster analysis of countries with respect to fertility rate and other demographic factors	284
Beata Bal-Domańska , An attempt to identify major regional clusters and their convergence	293
Beata Bieszk-Stolorz, Iwona Markowicz , The influence of benefit on the job finding process	302
Marta Dziechciarz-Duda, Klaudia Przybysz , Education and labor market needs. Classification of university graduates	312
Tomasz Klimanek , The problem of measuring deagrarianisation process in rural areas in Poland using multivariate statistical methods.....	320
Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska , Selected methods for an analysis of longitudinal data.....	329
Artur Zaborski , The application of distance measures for ordinal data for aggregation individual preferences	337
Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek , Application of correspondence analysis to examine the situation of the inhabitants of Warsaw suburban area in the labour market	345
Katarzyna Wawrzyniak , Classification of voivodeships according to the level of the realization of priorities of <i>the National Development Strategy 2007-2015</i> with using the values of centroid of the rows	355

Andrzej Dudek

Uniwersytet Ekonomiczny we Wrocławiu

METODY ANALIZY SKUPIEŃ W KLASYFIKACJI MARKERÓW MAP GOOGLE

Streszczenie: Klasyfikacja markerów na mapach Google jest ciekawym przykładem zastosowań algorytmów analizy skupień, w którym ostateczna liczba otrzymanych skupień jest wynikiem nie tylko założeń wstępnych i zastosowanego algorytmu, ale również skali, w której jest wyświetlana mapa. Ostatecznym celem klasyfikacji nie jest wyłącznie otrzymanie względnie homogenicznych skupień, ale również zapobieganie zjawisku „zlewania się” markerów na mapie. W artykule zaproponowano modyfikację klasycznego algorytmu k -średnich, uwzględniającą odległość między punktami opisanymi szerokością i długością geograficzną, opisano zagadnienie powiązania skali, w jakiej mapa jest wyświetlana, z otrzymaną strukturą klas oraz zaproponowano algorytm przydziału punktów do zadanej trasy na podstawie odległości od niej.

Słowa kluczowe: analiza skupień, mapy Google.

1. Wstęp

Technologia map Google i pokrewnych, takich jak *Open street maps*, rozwija się od kilkunastu lat bardzo gwałtownie, podobnie szybko wzrasta liczba aplikacji komputerowych wykorzystujących mapy komputerowe do prezentacji danych. W praktyce coraz częściej się zdarza, że w informatycznych systemach zarządzania przedsiębiorstwami [por. np. Dudek (red.) 2011, s. 13], prócz tradycyjnych sposobów ewidencjonowania kontrahentów, możliwe jest ich wizualizowanie na mapach, a także inne operacje, takie jak planowanie tras dla przedstawicieli handlowych. Funkcjonalność ta jest bardzo przydatną cechą nowoczesnych systemów informatycznych, jednak obciążoną pewną wadą, którą sama firma Google nazywa „*too many clusers*”. Chodzi o sytuację, gdy liczba markerów (odpowiadających na mapie adresom opisanym szerokością i długością geograficzną) jest zbyt duża, a prezentowane dane są prawie całkowicie nieczytelne (por. rys. 1).

W artykule zaproponowano modyfikację klasycznego algorytmu k -średnich, uwzględniającą odległość między punktami opisanymi szerokością i długością geograficzną, omówiono zagadnienie powiązania skali, w jakiej mapa jest wyświet-



Rys. 1. Zjawisko „zlewania się” markerów na mapie

Źródło: opracowanie własne z wykorzystaniem map Google.

łana, z otrzymaną strukturą klas oraz zaproponowano algorytm przydziału punktów do zadanej trasy na podstawie odległości od niej. Całość zakończona jest krótkim podsumowaniem i przedstawieniem problemów otwartych.

2. Proste metody klasyfikacji dla markerów map Google

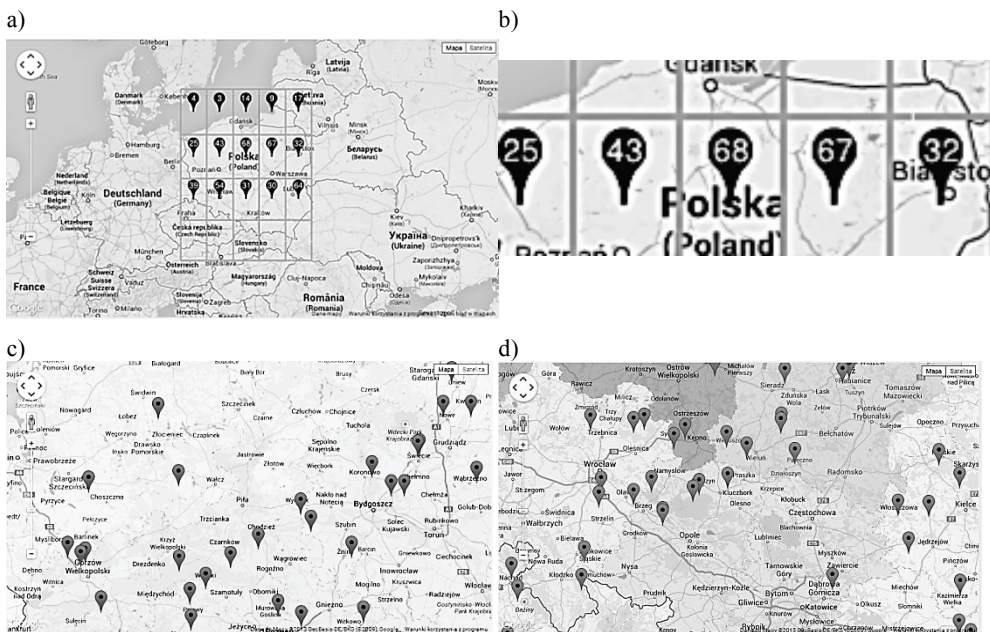
Problem nieczytelności markerów na mapie przy zbyt dużym ich zagęszczeniu został zauważony przez firmę Google. W dokumencie [Too Many Markers..., s. 1-3] proponowane jest zastosowanie w takim przypadku trzech procedur, które nazwano w nim „procedurami klasyfikacyjnymi”. Choć nie są to klasyczne algorytmy analizy skupień, należałoby określić je jako „proste” lub „nawne” metody. Są to:

- podział wg siatki geograficznej,
- podział wg jednostek administracyjnych,
- podział wg kodów pocztowych lub prefiksów telefonicznych.

Rysunek 2 przedstawia przykłady zastosowania tych metod do danych z rysunku 1. (rys. 2a, b) oraz do rzeczywistych kontrahentów firmy z Dolnego Śląska (rys. 2c, d).

3. Algorytm klasyfikacja markerów na mapie bez uwzględniania skali

Metody opisane w poprzednim punkcie są częścią standardowego programistycznego API map Google. W jego dokumentacji autorzy wspominają również o klasycznych metodach analizy skupień, nadając im wspólną nazwę *distance based clustering*, jednak nie podając przykładów ich implementacji.



a, b – podział według siatki geograficznej; c – podział według jednostek administracyjnych – poszczególne odcienie szarości odpowiadają województwom; d – podział według prefiksów telefonicznych – poszczególne odcienie szarości odpowiadają miejscowościom z tej samej grupy prefiksów telefonicznych.

Rys. 2. Proste metody klasyfikacji dla map Google

Źródło: opracowanie własne z wykorzystaniem map Google.

Zaproponowany algorytm jest modyfikacją klasycznej metody k -średnich, w której do mierzenia odległości między punktami zamiast odległości euklidesowej wykorzystywana jest odległość między punktami opisanymi przez szerokość i długość geograficzną zgodnie z wzorem (1).

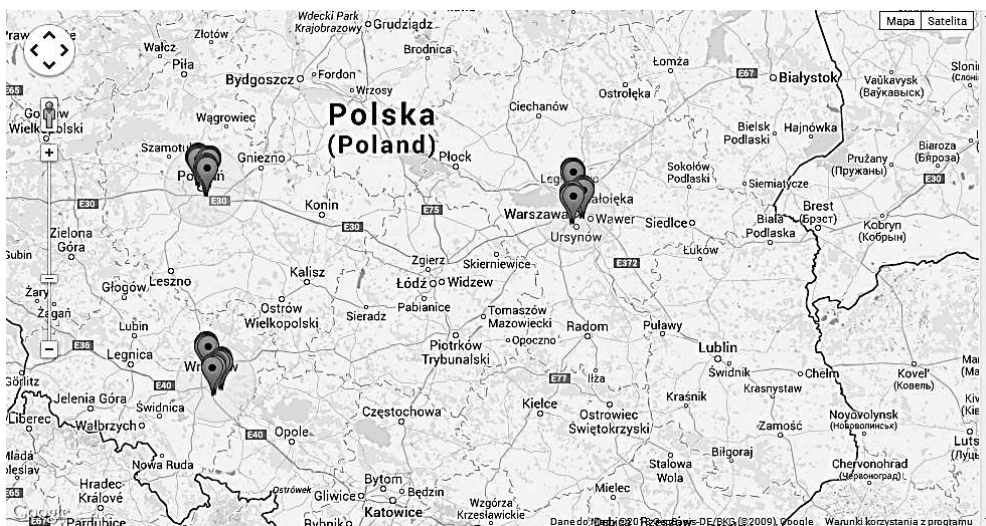
Niech $r_{earth} = 6371$ oznacza promień Ziemi, natomiast $p_1 = (\phi_1, \nu_1)$, $p_2 = (\phi_2, \nu_2)$ – dwa punkty na mapie opisane długością i szerokością geograficzną. Odległość między tymi punktami można zdefiniować jako:

$$GD(p_1, p_2) = 2r_{earth} \operatorname{atgh} \left(\frac{\sqrt{\sin(\Delta\phi)^2 \sin(\Delta\nu)^2 \sin(\phi_1) \cos(\phi_1)}}{\sqrt{1 - \sin(\Delta\phi)^2 \sin(\Delta\nu)^2 \sin(\phi_1) \cos(\phi_1)}} \right), \quad (1)$$

gdzie: $\Delta\phi = \phi_2 - \phi_1$, $\Delta\nu = \nu_2 - \nu_1$.

Proponowany algorytm można sformułować w czterech etapach [na podst. Waleśiak, Gatnar (red.) 2009, s. 331]:

- a) punktem wyjścia jest ustalony losowo wstępny podział zbioru obiektów na s klas; dla każdej klasy wstępnego podziału oblicza się środki ciężkości oraz odległości GD każdego obiektu od środków ciężkości tych klas.
- b) zmienia się przyporządkowanie obiektów do klas o najbliższym, według odległości GD , środku ciężkości;
- c) oblicza się nowe środki ciężkości dla każdej klasy;
- d) powtarza się kroki (b) i (c) do chwili, aż nastąpi przesunięcie obiektów między klasami.



Rys. 3. Skupienia otrzymane w wyniku zastosowania proponowanej metody

Źródło: opracowanie własne z wykorzystaniem map Google i funkcji własnych języka JavaScript.

Efekt działania algorytmu w przypadku przykładowych rzeczywistych danych przedstawia rysunek 3.

4. Dynamiczna klasyfikacja markerów w zależności od skali mapy

Specyfiką klasyfikacji markerów na mapach jest to, że wraz ze zmianą skali i obszaru wyświetlania mapy może zmieniać się też struktura skupień. Przykładowo markery odpowiadające kontrahentom firmy na mapie Polski mogą grupować się wokół dużych miast i ośrodków przemysłowych, tworząc skupienia odpowiadające tym miastom. Natomiast po powiększeniu mapy i ograniczeniu do jednego miasta markery odpowiadające firmom mogą koncentrować się wokół dzielnic, a jedno duże skupienie jest naturalnie dzielone na kilka mniejszych.

Można wyróżnić co najmniej dwie strategie poprawnej klasyfikacji markerów w zależności od skali mapy. Pierwsza z nich to każdorazowe przeprowadzanie

procedury analizy skupień po zmianie skali mapy lub przesunięciu środka mapy z wykorzystaniem konstrukcji programistycznej:

```
google.maps.event.addListener(map, 'zoom_changed', function() {
doClustering(map.getZoom());
});
```

Efekt takiego podejścia przedstawiony jest na rysunku 4. Rysunek 4a przedstawia nieprawidłową strukturę skupień, powtórzoną z rysunku 3, natomiast rysunek 4b zawiera prawidłową strukturę skupień otrzymaną po ponownym przeprowadzeniu analizy skupień, ograniczonej do markerów z danego wycinka mapy. Jedynym problemem związany z tym podejściem, to konieczność dokonywania obliczeń po każdym przeskalowaniu i przesunięciu mapy, co znacznie zwiększa czas ładowania się map i pracy z nimi.



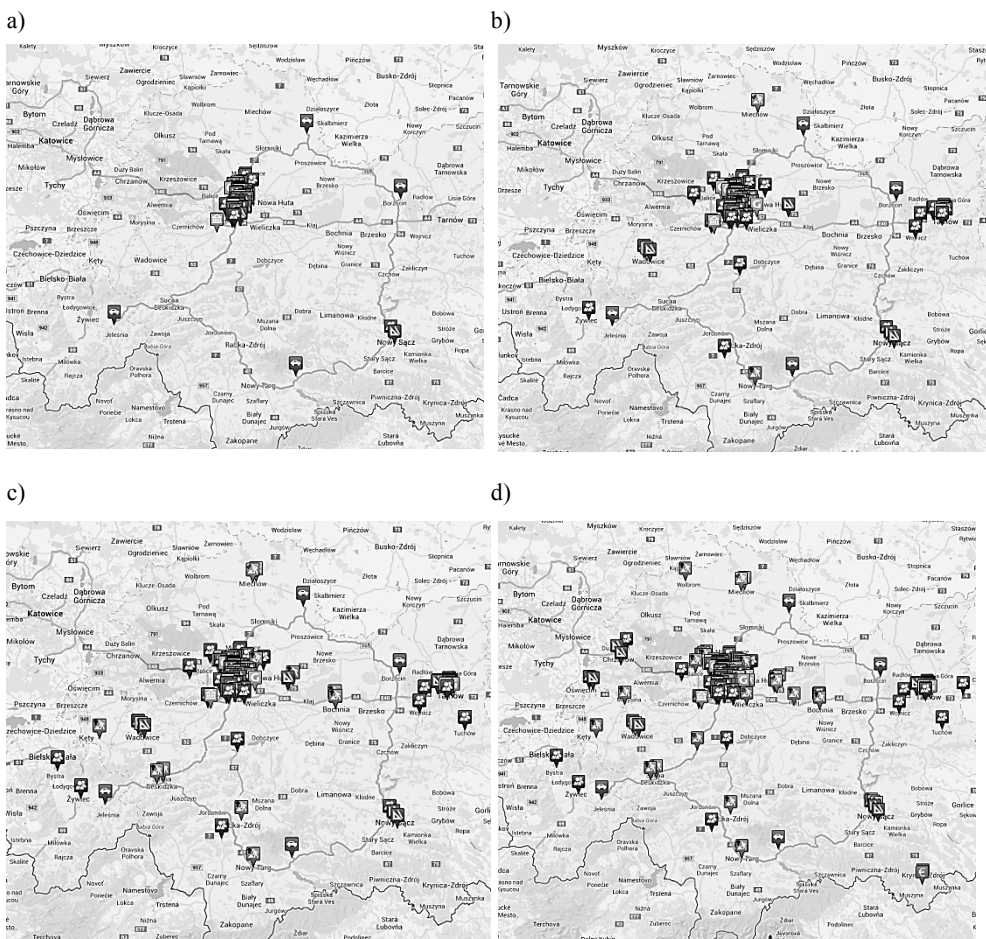
Rys. 4. Powiązanie struktury skupień ze skali mapy: a – dane z rysunku 3 w powiększonej skali – nieprawidłowa struktura – jedno skupienie; b – dane z rysunku 3 w powiększonej skali po przeprowadzeniu ponownej procedury klasyfikacyjnej – prawidłowa struktura

Źródło: opracowanie własne z wykorzystaniem map Google i funkcji własnych języka JavaScript.

Drugie podejście realizujące dynamiczną klasyfikację w zależności od skali polega na jednorazowym przeprowadzeniu klasyfikacji, której efektem nie byłby jeden podział na klasy, ale hierarchia dająca różne wyniki w zależności od poziomu odcięcia. Niestety, nie mogą do tego celu zostać wykorzystane klasyczne metody hierarchiczne aglomeracyjne (hierarchia jest konstruowana w inny sposób), natomiast pierwsze próby z metodami hierarchicznymi deglomeracyjnymi nie dały zadowalających rezultatów. Opracowanie takiego algorytmu wydaje się więc problemem otwartym.

5. Klasyfikacja markerów w zależności od odległości od zadanej trasy

Innym ciekawym zagadnieniem, mogącym mieć zastosowanie praktyczne, jest przydział markerów do zadanych tras według kryterium odległości. Zagadnienie to może mieć praktyczne zastosowanie na przykład przy planowaniu tras dla przedstawicieli handlowych firmy.



a) promień = 2 km; b) promień = 18 km; c) promień = 30 km; d) promień = 40 km.

Rys. 5. Przydział markerów według zadanej odległości od trasy

Źródło: opracowanie własne z wykorzystaniem map Google i funkcji własnych języka Java

Niech $p_0 = (\phi_0, \nu_0)$ oznacza punkt, do którego liczona jest odległość, $O = \{p_1, p_2\}$ odcinek trasy reprezentowany przez początek i koniec, natomiast $\{p_1, p_2, p_3, \dots, p_n\}$ – całą trasę reprezentowaną przez n punktów ścieżki (*path-points*), czyli takich punktów, które jednoznacznie identyfikują klasę, a pomiędzy kolejnymi punktami nie ma miejsc, w których możliwa byłaby zmiana trasy (skrzyżowań, rozgałęzień, rond itp.). Odległość punktu (markera) od trasy liczona jest zgodnie z równaniem (2).

$$D(p_0, \{p_1, p_2, p_3, \dots, p_n\}) = \min \{D(p_0, \{p_1, p_2\}), D(p_0, \{p_2, p_3\}), \dots, D(p_0, \{p_{n-1}, p_n\})\}, \quad (2)$$

gdzie: $D(p_0, O) = \min \{GD(p_0, p_1), GD(p_0, p_2), GD(p_0, p'_0)\}$ – odległość punktu od odcinka trasy;

$$p'_0 = \left(\phi_0 - \frac{\phi_2 - \phi_1}{v_2 - v_1} k, v_0 - k \right) - \text{rzut } p_0 \text{ na prostą wyznaczoną przez } \{p_1, p_2\};$$

$$k = \left(\frac{\phi_2 - \phi_1}{v_2 - v_1} \phi_0 + v_0 - \frac{\phi_2 - \phi_1}{v_2 - v_1} v_0 - \frac{\phi_2 - \phi_1}{v_2 - v_1} \phi_1 \right) / \left(\frac{\phi_2 - \phi_1}{v_2 - v_1} \frac{\phi_2 - \phi_1}{v_2 - v_1} + 1 \right).$$

Sama reguła klasyfikacyjna to porównanie obliczonej odległości z zadany promieniem i przydział tych punktów, dla których odległość jest mniejsza od promienia.

Rysunek 5 przedstawia rezultaty zaproponowanej metody przydzielania punktów do trasy przy zwiększającym się promieniu.

6. Podsumowanie i problemy otwarte

W artykule zaproponowano modyfikację klasycznego algorytmu k -średnich, realizującą podział markerów (punktów zdefiniowanych przez szerokość i długość geograficzną) na skupienia zgodnie z postulatem maksymalnej czytelności na mapie Google. Zaproponowany algorytm daje dobre rezultaty dla rzeczywistych danych, choć z uwagi na specyfikę danych brakuje kryterium liczbowego (takiego jak indeks Randa [Hubert, Arabie 1985, s. 193]), oceniającego zgodność otrzymanych wyników z rzeczywistą strukturą klas. Wydaje się, że jednak do celów wizualizacyjnych kryterium oceny wzrokowej jest wystarczające. Należy również podkreślić, że sama modyfikacja nie jest znaczącą zmianą w stosunku do klasycznej metody k -średnich, natomiast wypełnia pewną lukę, zwłaszcza że jej implementacja w języku JavaScript jest w pełni zintegrowana z API map Google.

Problemem otwartym jest powiązanie algorytmu klasyfikacyjnego ze skalą, w jakiej aktualnie jest wyświetlana mapa. W podejściu zaproponowanym w artykule każdorazowo po zmianie skali mapy procedura klasyfikacyjna jest przeprowadzana ponownie, wydaje się jednak, że docelowy algorytm powinien mieć „wbudowane” tworzenie hierarchicznej struktury z możliwością odcięcia na poziomie odpowiadającym aktualnej skali mapy.

Prezentowane w pracy algorytmy zaimplementowane są w językach JavaScript i Java. Czytelnicy zainteresowani ich kodami źródłowymi proszeni są o kontakt mailowy z autorem. W przyszłości planowane jest stworzenie biblioteki programistycznej i udostępnienie jej na licencji GNU/GPL.

Literatura

- Dudek A. (red.) (2011), *Systemy Informatyczne Zarządzania. Microsoft Business Solutions Navision*, Wydawnictwo UE we Wrocławiu, Wrocław.
- Everitt B.S., Landau S., Leese M. (2001), *Cluster analysis*, Edward Arnold, London.
- Gordon A.D. (1999), *Classification*, Chapman & Hall/CRC, London.
- Hubert L.J., Arabie P. (1985), *Comparing partitions*, „Journal of Classification”, no. 2, s. 193-218.
- Too Many Markers! Google Maps API*, <https://developers.google.com/maps/articles/toomanymarkers#gridbasedclustering> (27.04.2014).
- Walesiak M., Gatnar E. (red.) (2009), *Statystyczna analiza danych z wykorzystaniem programu R*, Wyd. Naukowe PWN, Warszawa.

CLUSTERING TECHNIQUES FOR GOOGLE MAPS MARKERS

Summary: Classification of geographical points on Google maps is an interesting example of the use of cluster analysis algorithm in which the final number of clusters is obtained not only by presuppositions and the algorithm used, but also by the scale, on which the map is actually displayed. The ultimate goal of classification is not only to obtain relatively homogeneous clusters, but also to prevent the phenomenon of "blurring" partitions on the map. In the paper a modification of the classical *k-means* algorithm, taking into account the distance between the two points described by latitude and longitude, is proposed. Next the issue of relationship between the scale at which the map is displayed with the final structure of the classes is discussed. Finally the algorithm of allocation of points to a predetermined route is proposed.

Keywords: cluster analysis, Google maps.