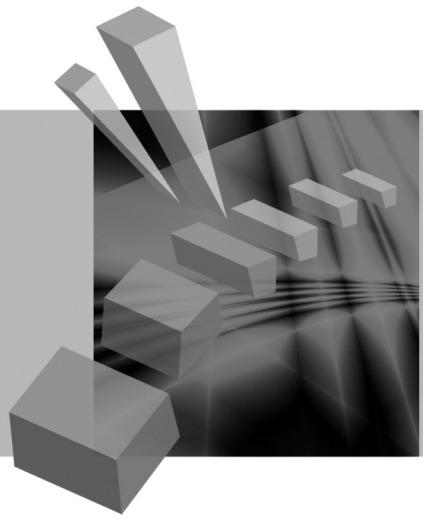
**PRACE NAUKOWE**
Uniwersytetu Ekonomicznego we Wrocławiu
**RESEARCH PAPERS**
of Wrocław University of Economics

**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 Pociecha,
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

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl



**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

## Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 242
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS 2012

Taksonomia 19. Klasyfikacja i analiza danych – teoria i zastosowania ISSN 1899-3192

**Kamila Migdał-Najman**
Uniwersytet Gdański

# PROPOZYCJA HYBRYDOWEJ METODY GRUPOWANIA OPARTEJ NA SIECIACH SAMOUCZĄCYCH

**Streszczenie:** W artykule autorka dokonuje prezentacji hybrydowej metody grupowania opartej na sieciach neuronowych samouczących typu SOM i GNG. Autorka weryfikuje potencjał proponowanej hybrydowej metody grupowania typu: sieć SOM + metoda *k*-średnich. Proponowane podejście weryfikuje na przykładzie badania preferencji i zachowań komunikacyjnych mieszkańców Gdyni w 2010 r.

**Słowa kluczowe:** sieć samoorganizująca się Kohonena (SOM), sieć typu gaz neuronowy (GNG), wskaźniki jakości grupowania, badanie preferencji, segmentacja.

## 1. Wstęp

Sztuczna sieć neuronowa typu SOM (*Self Organizing Map*), nazywana również siecią samouczącą, siecią lub mapą Kohonena, samoorganizującym się odwzorowaniem lub mapą cech (por. [Berthold, Hand 1999; Kohonen 1995;1997; 2001; 2006; Fort, Pagès 1996; Yin 2002; Fausett 1994; Deboeck, Kohonen 1998]), zaproponowana została w 1982 r. przez fińskiego profesora Teuvo Kohonena. Sieć SOM jest jedną z bardziej popularnych i efektywnych aplikacji *data mining*, która znajduje zastosowanie w zagadnieniach, takich jak: klasyfikacja, grupowanie, redukcja wymiarowości, wyszukiwanie anomalii i odchyleń od wartości typowych, wizualizacja wielowymiarowych zbiorów danych i badanie dynamiki zjawisk (por. [Papadimitriou i in. 2002; Delgado 2000; Fessant, Midenet 2002; Deventer, Moolman, Aldrich 1996; Migdał-Najman, Najman 2004]. W wyjściowej przestrzeni cech umieszczone zostają neurony, które lokalnie aproksymują analizowane obiekty. Neurony te uporządkowane są w pewną strukturę nazywaną siecią, w której są ze sobą w określonych związkach. Każdy wektor wejściowy połączony jest równolegle z wszystkimi neuronami na sieci przez wektor wag (współrzędnych). Wagi początkowo są liczbami losowymi z przedziału od zera do jednego. Dopasowanie tych wag jest istotą mechanizmu uczenia sieci. Uczenie sieci można nazwać adaptacyjnym procesem doboru wag sieci. Neurony, które są sąsiadami w przestrzeni, wykazują skłonność

do rozpoznawania podobnych (bliskich) do siebie obiektów wejściowych. Neurony sąsiadujące ze sobą na mapie mają podobne wektory wag. Każdy neuron, a dokładnie jego wagi, staje się pewnym wzorcem grupy bliskich sobie sygnałów (obiektów) wejściowych.

Jednym ze sposobów wizualizacji wyników sieci SOM jest macierz U (*unified distance matrix, U-matrix*), nazywana macierzą ujednoliconych odległości. Odległość między sąsiadującymi neuronami prezentowana jest różną kolorystyką. Ciemna kolorystyka[1] między neuronami odpowiada dużym odległościom, a zatem luce między wektorami wag w przestrzeni wejściowej. Jasna kolorystyka między neuronami oznacza, że wektory wag są blisko innych w przestrzeni wejściowej (por. [Fessant, Midenet 2002]). Technika ta umożliwia poszukiwanie skupień w danych wejściowych bez posiadania *a priori* żadnych informacji o tych klasach, ujawniając na mapie „pasma gór" i „wąwozów". Pierwsze są często strefą nieregularnie ukształtowaną z wysoką tendencją do tworzenia skupień, podczas gdy drugie rozdzielają zbiór danych na obszary, które mają odmienne właściwości. Wizualna eksploracja macierzy ujednoliconych odległości w celu poszukiwania skupień, którą zaproponował Kohonen, jest jednak subiektywna. Zależy od doboru kolorów i umiejętności „czytania" mapy SOM. Jest ona również często niepraktyczna, szczególnie kiedy prowadzonych jest wiele badań i budowanych jest wiele sieci SOM. Aby zobiektywizować i zautomatyzować proces wyróżniania skupień w literaturze, proponuje się zastosowanie metody dwustopniowej. Na pierwszym stopniu buduje się sieć SOM, a na drugim uzyskane neurony klasyfikuje się metodą *k*-średnich. Dla uzyskanych skupień neuronów możliwa jest identyfikacja jednostek, które one odwzorowują. Podejście to ma jednak wady wynikające z zastosowanej na drugim stopniu metody k-średnich. Jest ona wrażliwa na początkowe centra skupień, wyróżnia jedynie skupienia sferyczne, może prowadzić do otrzymania skupień o zerowej liczbie jednostek. Wymaga także ustalenia *a priori* liczby skupień – aby proces ten zautomatyzować, trzeba testować różne konfiguracje skupień, a następnie uzyskaną strukturę trzeba sprawdzić jednym z kilkudziesięciu wskaźników jakości grupowania. Dla uniknięcia tych niedogodności proponuje się, aby na drugim stopniu użyć innej metody grupowania, która nie ma wad metody *k*-średnich. Celem artykułu jest propozycja zautomatyzowanej metody wyróżniania skupień na neuronach sieci SOM na bazie sieci neuronowej o zmiennej strukturze typu GNG (*Growing Neural Gas*) B. Fritzkego (por. [Fritzke 1994; 1995; Kohonen 1995; 1999; 2001; 2006; Vesanto 1997; Deboeck, Kohonen 1998; Migdał-Najman 2009]). Uzyskana klasyfikacja dla zaproponowanej hybrydy: sieć SOM + sieć GNG, porównana zostanie z podejściem: sieć SOM + metoda *k*-średnich. Propozycja zautomatyzowanej metody zaprezentowana zostanie na zbiorze danych przedstawiających preferencje i zachowania komunikacyjne mieszkańców Gdyni.

[1] Jest to sprawa umowna, kolory w macierzy U można oznaczyć całkowicie odmiennie.

## 2. System transportu publicznego w Gdyni

Od 1998 r. Zarząd Komunikacji Miejskiej (ZKM) w Gdyni przy współpracy z Uniwersytetem Gdańskim co dwa lata prowadzi badanie preferencji i zachowań komunikacyjnych mieszkańców Gdyni. Ocenia różne aspekty komunikacji miejskiej oraz poglądy użytkowników dotyczące określonych rozwiązań polityki transportowej realizowanej przez władze samorządowe. Badania są prowadzone na podstawie reprezentatywnej próby mieszkańców Gdyni, biorąc pod uwagę liczbę ludności w poszczególnych dzielnicach miasta, proporcje kobiet i mężczyzn, a także ich wykształcenie. Liczebność próby w kolejnych edycjach badania różni się nieznacznie i oscyluje na poziomie ok. 2000 osób. Struktura kwestionariusza jest stabilna i liczy do 45 pytań.

System transportu publicznego w Gdyni tworzą dwa podsystemy: sieci linii drogowej transportu zbiorowego (trolejbusy i autobusy) i linia szybkiej kolei miejskiej (SKM). Podsystemy te nie są zintegrowane. Drogowy transport zbiorowy jest organizowany przez Zarząd Komunikacji Miejskiej (ZKM) w Gdyni – zakład budżetowy gminy Gdynia. SKM jest organizowana i obsługiwana przez PKP Szybką Kolej Miejską w Trójmieście sp. z o.o. Spółka ta funkcjonuje w ramach holdingu PKP SA. SKM na obszarze Gdyni obejmuje 16 km zelektryfikowanej dwutorowej linii z 9 przystankami: Orłowo, Redłowo, Wzgórze św. Maksymiliana, Główna, Stocznia, Grabówek, Leszczynki, Chylonia i Cisowa. Na rysunku 1 przedstawiono przebieg trakcji SKM na terenie miasta Gdynia.


**Rys. 1.** Przebieg trakcji SKM w Gdyni

Źródło: opracowanie własne.

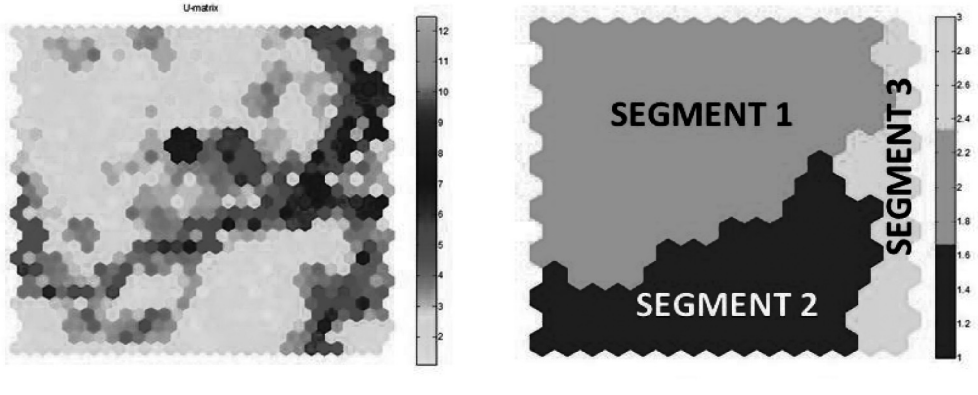
ZKM w Gdyni oferuje przewozy na obszarach: Gdyni, Sopotu, Rumii, Kosakowa, Żukowa, Wejherowa. Realizuje obsługę przewozową na 90 liniach, w tym:

12 trolejbusowych, 78 autobusowych, w tym: 64 zwykłych, 6 nocnych, 3 bezpłatnych, 3 specjalnych i 2 wodnych.

## 3. Eksperyment badawczy i wyniki analizy porównawczej

W celu wyróżnienia jednorodnych grup użytkowników komunikacji miejskiej w Gdyni ze względu na ich opinie dotyczące podstawowych aspektów podróżowania autobusem, trolejbusem i SKM wykorzystano samouczącą się sieć neuronową typu SOM. Wybór metody wynikał z jej własności, a także struktury zbioru danych. Zbiór składał się z 1975 jednostek, którymi są badane osoby i 18 cech wyrażonych na skali porządkowej (skala 5-stopniowa) opisujących ich opinie w zakresie: cen biletów, szybkości, wygody, czystości, punktualności i bezpieczeństwa podróżowania (np. 1 – bardzo niebezpieczny, 2 – niebezpieczny, 3 – ani niebezpieczny, ani bezpieczny, 4 – bezpieczny, 5 – bardzo bezpieczny). W zbiorze znajdowała się niewielka liczba braków danych.

**Pierwsza hybryda.** Przy budowie sieci SOM testowano różne jej topologie. Budowano sieci o heksagonalnej strukturze połączeń o rozmiarach od 10×10 do 15×15. Uwzględniano cztery funkcje sąsiedztwa: gaussowską, uciętą gaussowską, prostokątną i wykładniczą. W każdym wariancie uczono sieć od 1000 do 5000 iteracji. Dla każdej symulacji wyznaczano trzy miary jakości sieci: błąd kwantyzacji, topograficzny i dystorsji. Optymalną w sensie minimalizacji miar błędów siecią jest sieć SOM o następujących parametrach: typ połączeń neuronów: heksagonalny, rozmiar sieci: 15×15 neuronów, funkcja sąsiedztwa: ucięta gaussowska, błąd kwantyzacji: 1,54, błąd topograficzny: 0,26, błąd dystorsji: 3,44, liczba iteracji uczących: 5000. Macierz U i uzyskaną klasyfikację metodą *k*-średnich na neuronach sieci SOM przedstawiono na rys. 2.




a) b)

**Rys. 2.** a) macierz U i b) uzyskana klasyfikacja metodą *k*-średnich na neuronach sieci SOM

Źródło: opracowanie własne.

W wyniku zastosowania metody *k*-średnich na neuronach sieci SOM wyodrębniono trzy segmenty podróżujących komunikacją zbiorową. Liczbę badanych w wyróżnionych segmentach przedstawiono w tab. 1.

**Tabela 1.** Liczba badanych w wyróżnionych segmentach dla hybrydy: sieć SOM + metoda *k*-średnich

| Segment | Segment 1 – S1 | Segment 2 – S2 | Segment 3 – S3 | Razem |
|---|---|---|---|---|
| Liczba badanych w segmentach | 834 | 606 | 535 | 1975 |

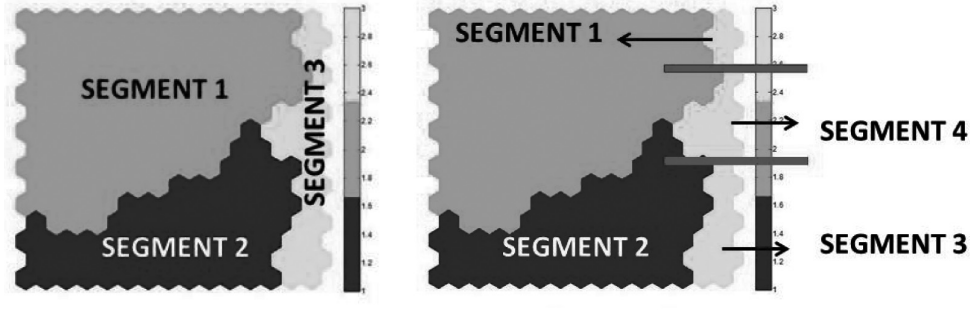
Źródło: opracowanie własne.

**Druga hybryda.** Drugą klasyfikację podróżujących komunikacją zbiorową w Gdyni uzyskano dla hybrydy: sieć SOM + sieć GNG. Parametry optymalnej w sensie minimalizacji miar błędów sieci SOM przedstawiono powyżej (por. hybryda pierwsza). Na etapie drugim na neuronach sieci SOM przeprowadzono grupowanie za pomocą sieci GNG. Przy budowie sieci GNG testowano: parametr lambda: 90 (liczba iteracji, po których jest wstawiany nowy neuron), maksymalny wiek połączenia: 89 (ile iteracji pozostaje neuron, który się nie uczy), krok uczenia neuronu zwycięzcy: 0,05, krok uczenia neuronów połączonych ze zwycięzcą: 0,006. Sieć GNG składała się z 35 neuronów. Ze względu na brak możliwości wizualizacji uzyskanej klasyfikacji na sieci GNG wynik tego podejścia przedstawiono na uzyskanej w kla-

**Tabela 2.** Liczba badanych w wyróżnionych segmentach dla hybrydy: sieć SOM + sieć GNG

| Segment | Segment 1 – S1 | Segment 2 – S2 | Segment 3 – S3 | Segment 4 – S4 | Razem |
|---|---|---|---|---|---|
| Liczba badanych w segmentach | 878 | 606 | 136 | 535 | 1975 |

Źródło: opracowanie własne.




a) b)

**Rys. 3.** Klasyfikacja a) metodą *k*-średnich na neuronach sieci SOM i b) siecią GNG na neuronach sieci SOM

Źródło: opracowanie własne.

syfikacji metodą *k*-średnich na neuronach sieci SOM. W wyniku zastosowania sieci GNG na neuronach sieci SOM wyróżniono 4 segmenty podróżujących komunikacją zbiorową. Liczbę badanych w wyróżnionych segmentach przedstawiono w tab. 2. Uzyskane klasyfikacje dla dwóch proponowanych podejść przedstawiono na rys. 3.

**Analiza porównawcza.** Porównując dwie uzyskane klasyfikacje, można zauważyć, że w wyniku zastosowania hybrydy: sieć SOM + sieć GNG wyróżniono o jeden segment więcej niż w hybrydzie: sieć SOM + metoda *k*-średnich. Co ciekawe, sieć GNG wyróżniła dwa skupienia (segment 1 i segment 2) identyczne jak metoda *k*-średnich. Natomiast segment 3, który wyróżniono w metodzie *k*-średnich, został przez sieć GNG podzielony na trzy części. 44 osoby tego segmentu sieć GNG zaklasyfikowała do skupienia 1, a pozostałe zostały rozdzielone na dwie grupy i przydzielone do segmentu 3 i segmentu 4. W wyniku zastosowania sieci GNG na neuronach sieci SOM udało się zauważyć ważne różnice między segmentami, które zostały „zatarte" w wyróżnionym segmencie 3 w metodzie *k*-średnich. Klasyfikację krzyżową dla dwóch proponowanych hybryd przedstawiono w tab. 3.

W tabeli 4 przedstawiono wskaźniki oceny podobieństwa dwóch porównywanych klasyfikacji: wskaźnik Jaccarda, Randa, korygowany Randa i Fowlkesa Mallowsa (por. [Jaccard 1908; Rand 1971; Fowlkes, Mallows 1983]).

Uzyskane wyniki wskazują na istotne różnice w uzyskanych podziałach. Ich ocena będzie wynikała z możliwości interpretacji uzyskanych różnic.

**Tabela 3.** Klasyfikacja krzyżowa dla dwóch hybryd

| Hybrydy | | **SOM + GNG** | | | | Razem |
|---|---|---|---|---|---|---|
| | | S1 | S2 | S3 | S4 | |
| **SOM + metoda *k*-średnich** | S1 | **834** | 0 | 0 | 0 | 834 |
| | S2 | 0 | **606** | 0 | 0 | 606 |
| | S3 | **44** | 0 | **136** | **355** | 535 |
| Razem | | 878 | 606 | 136 | 355 | 1975 |

Źródło: opracowanie własne.

**Tabela 4.** Wskaźniki oceny podobieństwa wyników grupowania

| Porównywane hybrydy | Jaccard | Rand | korygowany Rand | Fowlkes Mallows |
|---|---|---|---|---|
| sieć SOM + *k*-średnich/sieć SOM + sieć GNG | 0,85 | 0,95 | 0,88 | 0,92 |

Źródło: opracowanie własne.

## 4. Segmentacja podróżujących komunikacją zbiorową

W badaniu preferencji i zachowań komunikacyjnych mieszkańców Gdyni zastosowano hybrydę: sieć SOM + sieć GNG. W badaniu wyróżniono cztery segmenty

umożliwiające analizę preferencji i zachowań komunikacyjnych mieszkańców Gdyni. W tabeli 5 przedstawiono profile wyróżnionych segmentów.

**Tabela 5.** Charakterystyka wyróżnionych segmentów

| Profil | Segment 1 | Segment 2 | Segment 3 | Segment 4 |
|---|---|---|---|---|
| Preferowany środek transportu | autobus, trolejbus, SKM | autobus<br>trolejbus | autobus | nie mam zdania |
| Dzielnica | Chylonia, Cisowa, Grabówek, Działki Leśne, Karwiny, Leszczynki, Wzgórze św. Maksymiliana, Śródmieście, Redłowo, Orłowo, Witomino | Babie Doły, Cisowa, Dąbrowa, Karwiny, Leszczynki, Śródmieście | Oksywie, Obłuże, Pogórze | Mały Kack, Wielki Kack, Chwarzno--Wiczlino |
| Wiek | 16-20<br>21-30<br>31-40 | 51-60<br>61-70<br>71-75 | 31-40<br>51-60<br>61-70 | 31-40<br>41-50<br>51-60 |
| Płeć | 54% kobiet | 59% kobiet | 57% kobiet | 37% kobiet |
| Prawo jazdy | 53% | 46% | 48% | 93% |
| Status zawodowy | pracuje, pracuje i uczy się/ studiuje, uczy się / studiuje | pracuje, jest na emeryturze | pracuje, jest na emeryturze | pracuje |
| Realizacja podróży | zawsze komunikacją zbiorową, przeważnie komunikacją zbiorową | zawsze komunikacją zbiorową | zawsze komunikacją zbiorową, w równym stopniu komunikacją zbiorową i samochodem osobowym | zawsze samochodem osobowym |
| Cechy realizowane najlepiej | dostępność, częstotliwość, punktualność | dostępność, punktualność, częstotliwość | punktualność, dostępność, bezpośredniość | nie mogę ocenić najlepszej |
| Standard wygody | miejsca stojące w nieuciążliwych warunkach, przeważnie miejsca siedzące | przeważnie miejsca siedzące, miejsca stojące w nieuciążliwych warunkach | przeważnie miejsca siedzące, miejsca stojące w nieuciążliwych warunkach | nie mam zdania, przeważnie miejsca siedzące |
| Ocena gdyńskiej komunikacji miejskiej | dobra | dobra | dobra | dobra, nie jestem w stanie ocenić (42% respondentów) |

Źródło: opracowanie własne.

Środki komunikacji zbiorowej, takie jak: autobus, trolejbus i SKM, zostały ocenione przez mieszkańców Gdyni „dobrze". Większość badanych oceniła autobusy i trolejbusy jako neutralnie drogie, szybkie, wygodne, czyste, punktualne i bezpieczne. To, co różniło SKM od dwóch pozostałych środków transportu, to neutralna ocena czystości SKM i bezpieczeństwa. Oceny trzech analizowanych środków transportu przedstawiono w tab. 6.

**Tabela 6.** Ocena środków transportu komunikacji zbiorowej wśród mieszkańców Gdyni

| Środek transportu | Autobus | Trolejbus | SKM |
|---|---|---|---|
| Cena | neutralna | neutralna | neutralna |
| Szybkość | szybki | szybki | szybki |
| Wygoda | wygodny | wygodny | wygodny |
| Czystość | czysty | czysty | neutralnie czysty |
| Punktualność | punktualny | punktualny | punktualny |
| Bezpieczeństwo | bezpieczny | bezpieczny | neutralnie bezpieczny |

Źródło: opracowanie własne.

## 5. Wnioski

Liczne badania poświęcone sztucznej sieci neuronowej SOM wskazują na jej duży potencjał w zagadnieniach związanych z grupowaniem. Propozycja zautomatyzowanej metody grupowania neuronów sieci SOM oparta na sieci GNG wydaje się posiadać większy potencjał w rozpoznawaniu szczegółów struktury grupowej niż metoda *k*-średnich. Sieć GNG jest zaawansowanym narzędziem analizy skupień i w tym zakresie w większości przypadków jest skuteczniejsza od metody *k*-średnich. Zastosowanie sieci GNG zamiast metody *k*-średnich na neuronach sieci SOM pozwoliło uniknąć niedogodności metody *k*-średnich. Uczenie się sieci GNG jest szybsze od metody *k*-średnich i nie wymaga poświęcania dodatkowego czasu związanego z analizowaniem różnych konfiguracji skupień i stosowaniem wskaźników ustalania liczby skupień. Sieć GNG sama modyfikuje swoją strukturę, automatycznie ustala optymalną liczbę neuronów i automatycznie ustala liczbę skupień w zbiorze danych. Trudnością w zastosowaniu sieci samouczących się jest brak standardowego oprogramowania.

Zastosowanie proponowanego podejścia wyróżniania skupień opartego na sieciach samoorganizujących: sieci SOM i GNG pozwoliło zaobserwować interesujące prawidłowości w analizie preferencji i zachowań komunikacyjnych mieszkańców Gdyni. Szczegóły zaobserwowane dzięki sieci GNG, a pominięte przez metodę *k*-średnich są istotne. Wydaje się, że hybryda sieć SOM + sieć GNG jest wartościowym narzędziem analizy skupień, wartym dalszych badań.

## Literatura


Berthold M., Hand D.J., *Intelligent Data Analysis* Springer-Verlag, Berlin Heidelberg 1999.

Deboeck G., Kohonen T., *Visual Explorations in Finance with Self-Organizing Maps*, Springer-Verlag, London 1998.

Delgado A., *Control of nonlinear systems using a self-organizing neural network*, „Neural Computing&Applications" 2000, no 9.

Deventer J.S.J., Moolman D.W., Aldrich C., *Visualisation of plant disturbances using Self-Organizing Maps*, „Computers Chemical Engineering" 1996, no 20 .

Fessant F., Midenet S., *Self-Organizing map for data imputation and correction in surveys*, „Neural Computing&Applications" 2002, no 10.

Fausett L., *Fundamentals of Neural Networks, Architectures, Algorithms, and Applications*, Florida Institute of Technology, Prentice Hall International, Inc., 1994.

Fowlkes E.B., Mallows C.L., *A method for comparing two hierarchical clusterings*, „Journal of the American Statistical Association" 1983, no 78, 383.

Fort J.C., Pagès G., *About the Kohonen algorithm: strong or weak self-organization*?, „Neural Networks"1996, no 9, 5.

Fritzke B., *Growing cell structures – a self-organizing network for unsupervised and supervised learning*, „Neural Networks" 1994, no 7, 9 .

Fritzke B., *A Growing Neural Gas Network Learns Topologies*, Advances in Neural Information Processing Systems, 7th edn., MIT Press, Redmond, Washington 1995.

Jaccard P., *Nouvelles recherches sur la distribution florale*, [w:] Bulletin de la Société Vaudoise des Sciences Naturelles, 44, 1908.

Kohonen T., *Self-Organizing Maps*, Springer Series in Information Sciences, Springer-Verlag, Berlin Heidelberg 1995; 1999; 2001.

Kohonen T., *Self-organizing neural projections*, „Neural Networks" 2006, no 19, 6.

Migdał-Najman K., Najman K., *Diagnozowanie kondycji finansowej spółek notowanych na GPW w Warszawie w oparciu o sieć SOM*, Zeszyty Naukowe nr 389, Rynek Kapitałowy. Skuteczne inwestowanie, część I, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin 2004.

Migdał-Najman K., *Zastosowanie nienadzorowanych sieci neuronowych typu Growing Neural Gas w analizie skupień*, [w:] Taksonomia 16, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 47, Wydawnictwo UE, Wrocław 2009.

Papadimitriou S., Mavroudi S., Vladutu L., Pavlides G., Bezerianos A., *The supervised network Self-Organizing Map for classification of large data sets*, „Applied Intelligence" 2002, no 16 .

Rand W.M., *Objective criteria for the evaluation of clustering methods*, „Journal of the American Statistical Association" 1971, no 66, 336.

Vesanto J., *Data Mining Techniques Based on the Self-Organizing Map*, Thesis for the degree of Master of Science in Engineering, Helsinki University of Technology, Department of Engineering Physics and Mathematics, Espoo, Finland 1997.

Yin H., *Data visualization and manifold mapping using the ViSOM*, „Neural Network" 2002, no 15.

## A PROPOSAL OF THE HYBRID CLUSTERING METHOD BASED ON SELF-LEARNING NETWORKS

**Summary:** In the article a hybrid clustering method based on a self-learning neural networks, SOM and GNG, is presented. The author verified the potential of the proposed hybrid clustering method such as: network SOM + k-means method. The proposed approach is verified on the example of behavioral research and communication preferences of the inhabitants of Gdynia in 2010.

**Keywords:** Self Organizing Map (SOM), Growing Neural Gas (GNG), validity indexes, preferences research, segmentation.