

Analiza nastrojów gospodarczych z wykorzystaniem sieci Kohonena

Andrzej Burda

Wyższa Szkoła Zarządzania i Administracji w Zamościu

Streszczenie: W pracy dokonano bezwzorcowej klasyfikacji respondentów badań nastrojów gospodarczych realizowanych przez Wyższą Szkołę Zarządzania i Administracji w Zamościu na terenie woj. lubelskiego. Badania prowadzono z wykorzystaniem metodologii sztucznych sieci neuronowych. Klasyfikacji dokonano przy pomocy sieci Kohonena. Wyodrębniono i opisano trzy charakterystyczne profile respondentów oraz dokonano analizy zmian ich liczebności w latach 2003–2007. Wyniki badań oraz indeks nastrojów gospodarczych wykorzystano do poszerzonej, opisowej analizy zachowań respondentów.

Wstęp

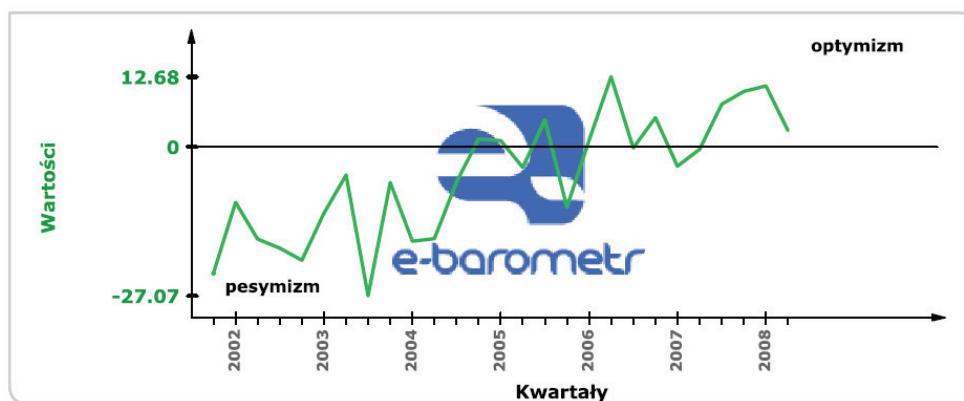
Jedną z możliwości uzyskania aktualnych informacji o procesach zachodzących w gospodarce są sondaże wśród przedsiębiorców i konsumentów. Mają one charakter badań jakościowych, służących analizie gospodarczej. Ich atrakcyjność wynika z faktu, że są zazwyczaj dostępne wcześniej niż skorelowane z nimi dane ilościowe pochodzące z innych źródeł.

Wyniki sondaży stanowią ważne źródło informacji dla wszystkich zainteresowanych rozwojem sytuacji gospodarczej: instytucji publicznych, zarządów przedsiębiorstw, naukowców, a przede wszystkim dla podejmujących decyzje z zakresu polityki gospodarczej na obszarze, którego sondaże dotyczą. Mogą być wykorzystywane w formie nieprzekształconych danych lub w sposób zagregowany, w postaci złożonych wskaźników określających nastroje gospodarcze i zaufanie konsumentów.

Jeśli wskaźniki nastrojów gospodarczych mają być przydatne, to muszą charakteryzować się szeregiem właściwości takich jak: spójność, aktualność, porównywalność, itp. Ich jakość musi być poddawana ciągłej kontroli poprzez weryfikację, na ile dobrze odzwierciedlają one reprezentowane agregaty makroekonomiczne. W zależności od wyników tej oceny podlegają odpowiednim zmianom i ulepszeniom [Komisja... 2006].

Cel i zakres badań

Badania koniunktury gospodarczej woj. lubelskiego, realizowane od 2001 roku przez Wyższą Szkołę Zarządzania i Administracji w Zamościu, doprowadziły do wystandaryzowania branżowo/rodzajowych indeksów nastrojów gospodarczych. Jeden z nich, odnoszący się do gospodarstw domowych dotyczących prognoz konsumentów, przedstawiono na Rys.1.



Rys. 1. Zmiana indeksu nastrojów gospodarczych dla gospodarstw domowych woj. lubelskiego w latach 2001–2008 (prognoza – horyzont 1 kwartał)

Źródło: [WSZIA 2008]

Celem podjętych badań jest wyodrębnienie spośród grona respondentów charakterystycznych profili konsumentów, przedstawienie ich w postaci statystyczno-opisowej oraz próba odpowiedzi na pytania, jakim zmianom podlegały określone profile w badanym przedziale czasu.

Badania przeprowadzono na danych kwartalnych z lat 2003–2007 z wykorzystaniem metodologii samoorganizującej się sieci odwzorowań Kohonena (SOM, ang. Self Organising Map).

Opis eksperymentu

Opis bazy informacyjnej

Dane sondażowe wykorzystane do budowy modelu gromadzone są przez Wyższą Szkołę Zarządzania i Administracji w ramach projektu pt.: Barometr koniunktury gospodarczej województwa lubelskiego. Do oceny prognostycznych nastrojów konsumentów (indeks pokazano na Rys. 1.), z ankiet wybrano odpowiedzi na 5 niżej zaprezentowanych pytań:

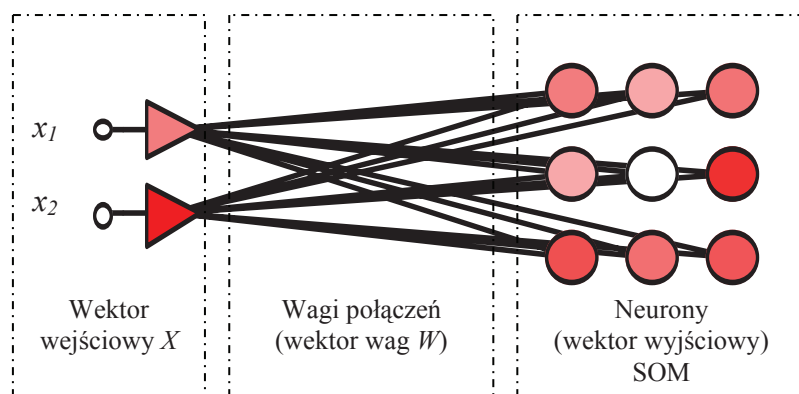
1. Jak przewidują Państwo sytuację finansową swojego gospodarstwa domowego w ciągu najbliższych 3 miesięcy: poprawi się (3 pkt.), nie zmieni się (2 pkt.), pogorszy się (1 pkt.)?
2. Czy wydatki na dobra konsumpcyjne w ciągu nadchodzących 3 miesięcy: zwiększą się (1), nie zmienią się (2), zmniejszą się (3)?
3. Czy przewidują Państwo, że stan oszczędności w Państwa gospodarstwie domowym w ciągu najbliższych 3 miesięcy: zwiększy się (3), nie ulegnie zmianie (2), zmniejszy się (1)?
4. Jak przewidują Państwo sytuację gospodarczą w województwie lubelskim w ciągu najbliższych 3 miesięcy: poprawi się (3), nie zmieni się (2), pogorszy się (1)?
5. Czy Państwa zdaniem, poziom bezrobocia w województwie lubelskim w ciągu najbliższych 3 miesięcy: zwiększy się (1), nie ulegnie zmianie (2), zmniejszy się (3)?

Do analizy danych metodami uczenia maszynowego wykorzystuje się dwa typy danych. Do budowy tablicy decyzyjnej na potrzeby realizacji przedstawionych celów badań wykorzystano typ I. Ten typ danych pozwala na: uzyskanie wglądu w ich strukturę, wyszukiwanie skupisk zawierających podobne do siebie obiekty, ustalanie hierarchii istotności cech opisujących badane obiekty oraz wykrywanie obiektów odległych [Varmuza 1998, s. 346–366].

Każdy przypadek w tablicy decyzji opisany jest przez wektor, którego elementy przyjmują wartości całkowite i przedstawiają oceny punktowe kolejnych odpowiedzi respondentów na wyżej przedstawione pytania. Tak skonstruowana tablica decyzyjna przyjęła postać macierzy o wymiarach 7115×5 .

Opis badań

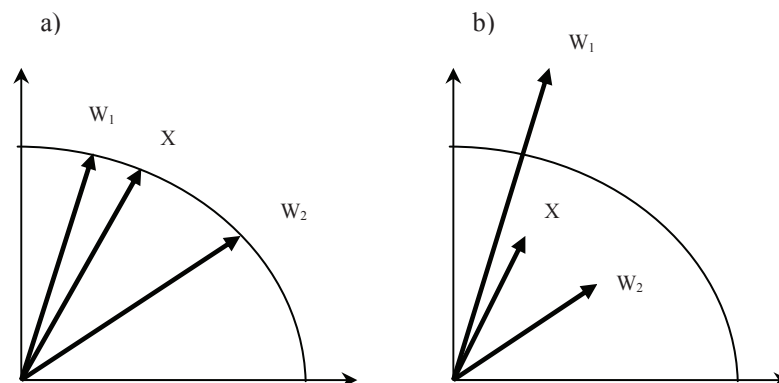
Badania przeprowadzono zgodnie z metodologią tworzenia sieci neuronowych Kohonena. Architektura sieci dla dwuwymiarowej przestrzeni wejściowej, przekształcaniej w 3×3 wymiarową mapę wyjściową, prezentuje Rys. 2.



Rys. 2. Budowa sieci Kohonena (SOM)

Sieć uczono w trybie nienadzorowanym z zastosowaniem algorytmu Kohonena, który należy do grupy algorytmów typu WTM (ang. *Winner Takes Most*). W pierwszym kroku, w zależności od charakteru

danych wejściowych, może być dokonywana ich normalizacja wraz z wagami neuronów sieci (wektory X i W z Rys. 2.). Jeżeli wektory wejściowe znacznie różnią się długością, może to utrudnić interpretację ich podobieństwa do wektorów wagowych reprezentujących neurony w sieci (Rys. 3b).



Rys. 3. Wektory wag dwóch neuronów (W_1 i W_2) oraz wektor wejściowy (X) w przestrzeni dwuwymiarowej: a) znormalizowane, b) bez normalizacji.
Źródło: [Rutkowski 2006, s. 206–215]

W następnym kroku wyznaczane są odpowiedzi wszystkich neuronów sieci na podany sygnał wejściowy i wybierany jest neuron zwycięski. Ten, którego odległość (iloczyn skalarny wektorów) jest najmniejsza od wektora wejściowego. Na Rys. 3a byłby to neuron reprezentowany przez wektor wagowy W_1 . Zwycięski neuron modyfikowany jest zgodnie z wzorem (1).

$$[1] \quad W_i(t+1) = W_i(t) + \eta G(i, x)[X - W_i],$$

gdzie:

W_i – oznacza wektor wag i -tego neuronu,

t – numer kroku iteracji,

η – współczynnik uczenia,

$G(i, x)$ – funkcja sąsiedztwa określona wzorem (2).

$$[2] \quad G(i, x) = \begin{cases} 1, & \text{dla } d(i, j) \leq R, \\ 0, & \text{dla pozostałych} \end{cases}$$

gdzie:

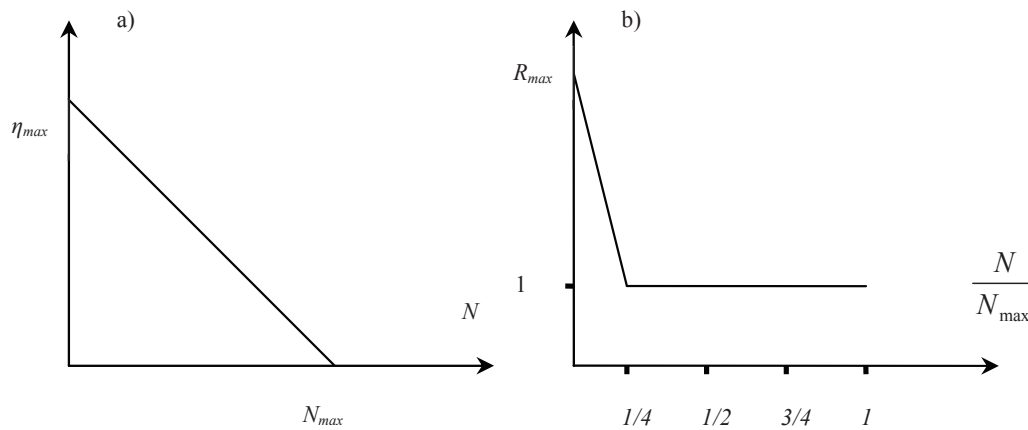
$d(i, j)$ – oznacza odległość euklidesową między neuronem j (zwycięzcą) a i -tym neuronem z sąsiedztwa G lub też odległość mierzona liczbą neuronów [Rutkowski 2006, s. 206–215].

Zmiana wektora wag tego neuronu powoduje, że neuron jako wzorzec staje się bardziej podobny do prezentowanego przypadku. Oprócz zwycięskiego neuronu modyfikowane są również neurony sąsiednie, zgodnie z przyjętą topologią oraz promieniem sąsiedztwa¹.

Dla przykładu, jeżeli parametr sąsiedztwa $R=2$ (topologia kwadratu, odległość mierzona liczbą neuronów), to sąsiedzi tworzą kwadrat 3×3 neuronów (cała mapa z Rys. 2.). Jeżeli zwycięski neuron znajduje się przy brzegu mapy, to sąsiedztwo kończy się na jej brzegu. Promień sąsiedztwa zmieniany jest w trakcie procesu uczenia od ustalonej przez badacza wartości początkowej do końcowej. Promień sąsiedztwa R jest przechowywany i skalowany jako liczba rzeczywista. Jednakże w celu określenia konkretnych sąsiadów zwycięskiego neuronu, brana jest pod uwagę najbliższa wartość całkowita [StatSoft 2006].

Duży wpływ na proces uczenia ma sposób zmiany wielkości parametrów η oraz R . Najczęściej stosuje się strategię pokazaną na Rys. 4.

¹ Przykład działania algorytmu Kohonena można przestudiować w: [Żurada 1992, s. 247–253].



Rys. 4. Strategia doboru parametrów w trakcie uczenia algorytmem Kohonena

a) zmiana stałej uczenia w funkcji cykli uczących,

b) zmiana wielkości sąsiedztwa w funkcji cykli uczących

Źródło: [Osowski 1994, s. 192]

Do symulacji użyto oprogramowanie **Statistica 7.1** firmy StatSoft. Proces uczenia sieci przeprowadzono przy następujących parametrach:

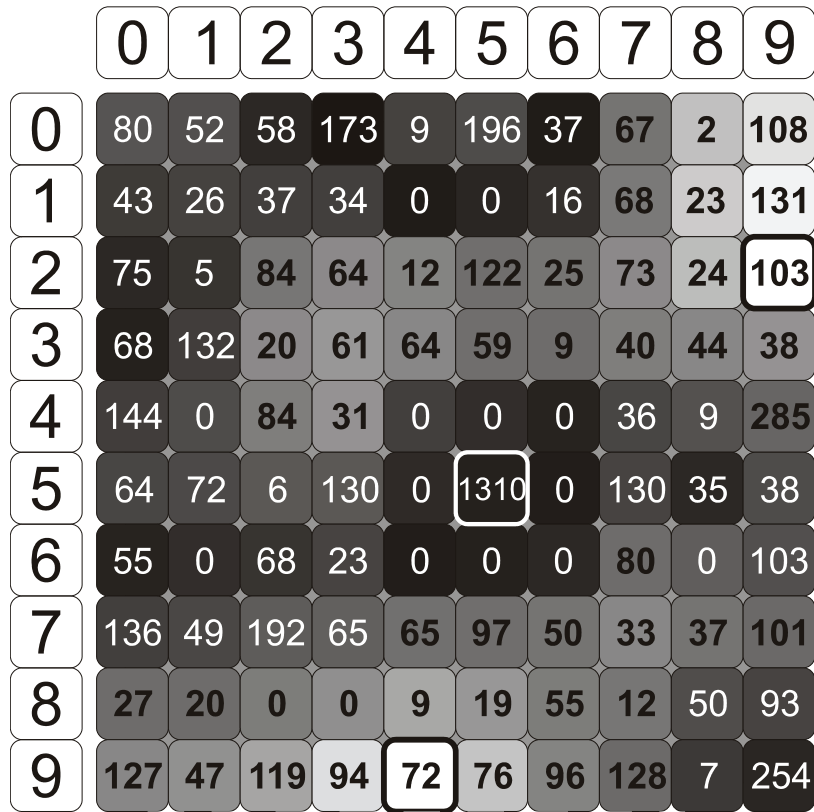
- wymiar mapy topologicznej 10×10 ,
- współczynnik uczenia i parametr sąsiedztwa:
 - faza I (100 epok): $\eta_{max}=0,1$ i $\eta_{min}=0,02$ $R_{max}=3$ i $R_{min}=1$
 - faza II (1000 epok): $\eta_{max}=0,1$ i $\eta_{min}=0,01$ $R_{max}=1$ i $R_{min}=1$
- inicjacja wag losowa (zgodna z rozkładem Gaussa).

Wyniki badań

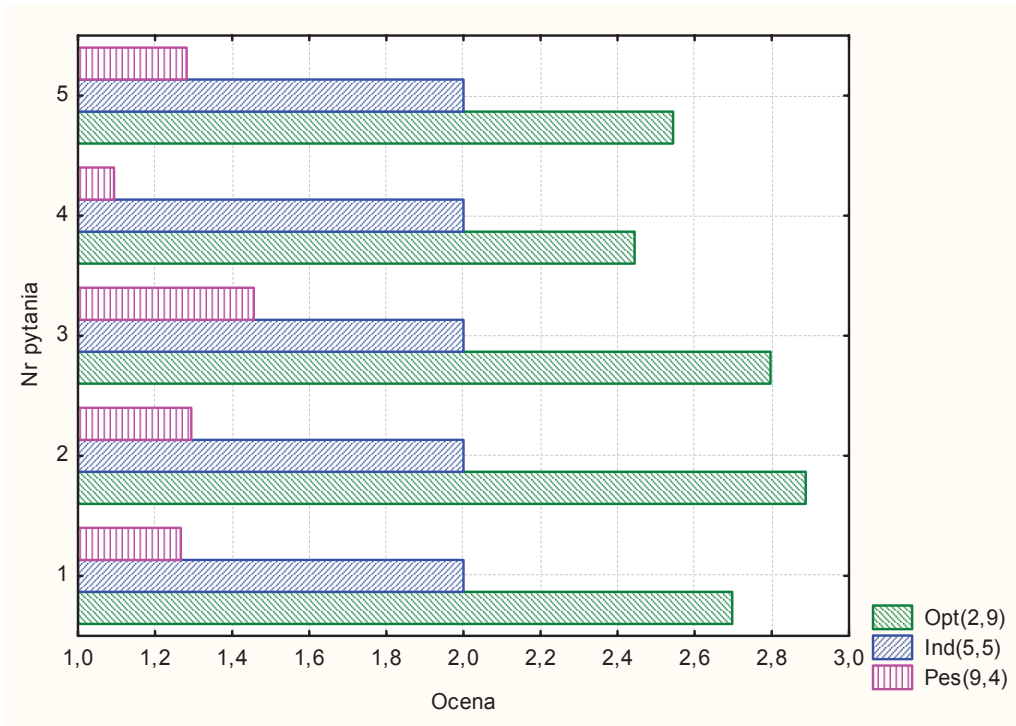
Realizowane zgodnie z opisaną metodologią badania doprowadziły do wyodrębnienia na mapie (SOM) obszarów dla trzech arbitralnie przyjętych przez autora grup profili respondentów: optymistów (Opt), indyferentnych (Ind) i pesymistów (Pes). Wygląd mapy pokazano na Rys. 5.

Zmiana intensywności zabarwienia neuronu oznacza stopień przynależności do odpowiedniej grupy. Im mniejsze nasycenie barwy, w przypadku prawego, górnego obszaru mapy, tym bardziej optymistyczne nastroje w stosunku do oczekiwanych zmian, określonych zakresem pytań zawartych w ankiecie, a odnoszących się do najbliższego kwartału. Analogicznie, w lewym, dolnym obszarze, wzrost pesymistycznych. W ten sposób na mapie pokazano 3 obszary: złożony z 25 profili (neuronów) obszar optymistów, złożony z 23 profili obszar pesymistów i pomiędzy nimi, 40 profili respondentów indyferentnych. Osoby indyferentne na potrzeby tej analizy to te, które pasywnie podchodzą do odpowiedzi na pytania lub w najbliższym okresie nie spodziewają się większych zmian. Każdy neuron mapy reprezentuje odrębny klaster. Wartości wagowe jego połączeń z wektorem wejściowym pokazują stopień przynależności do określonej grupy.

Symbol każdego neuronu mapy dodatkowo opisany został liczbą respondentów, których profil najbardziej pasuje do cech przez ten neuron reprezentowanych. Celem ułatwienia percepcji, neurony w obszarach Opt i Pes wartości liczbowe mają zapisane czarną czcionką, a w graniczącym z nimi obszarze Ind, białą. Jak wynika z Rys. 5., neuron mapy o współrzędnych (5,5), znajdujący się w obszarze Ind, reprezentuje aż 1310 ankiet poddanych badaniu. Trzy neurony, które w dalszej części badań poddano głębszej analizie, wyróżniono poprzez pogrubienie ich konturu. W obszarze Opt neuron (2,9), o najmniejszej intensywności zabarwienia, to klaster skrajnych optymistów. W obszarze Pes neuron (9,4) to klaster skrajnych pesymistów. W obszarze Ind wybrano neuron (5,5), którego cechy odpowiadają wystawieniu ocen neutralnych na wszystkie pytania ankiety dot. prognozy. Te trzy charakterystyczne typy profili pokazane zostały na Rys. 6.



Rys. 5. Mapa Kohonena



Rys. 6. Charakterystyczne profile ankietowanych

Korzystając z modelu SOM, dokonano analizy liczebności w klastrach w rozbiciu na kwartalne okresy badawczych w latach 2003-07. Wyniki analizy, dla wybranych profili, zebrano w Tabelicy 1.

Tab. 1. Rozkład liczebności wybranych, charakterystycznych profili w latach 2003–2007

Okres	Liczba			Udział ^{a)} [%]		
	Opt	Ind	Pes	Opt	Ind	Pes
I kw. 2003	0	47	15	0,00	12,77	4,08
II kw. 2003	6	70	2	1,68	19,55	0,56
III kw. 2003	5	69	10	1,21	16,71	2,42
IV kw. 2003	11	68	2	3,03	18,73	0,55
I kw. 2004	7	49	5	1,94	13,57	1,39
II kw. 2004	14	58	5	3,98	16,48	1,42
III kw. 2004	10	63	4	2,86	18,00	1,14
IV kw. 2004	15	69	12	4,29	19,71	3,43
I kw. 2005	6	71	5	1,71	20,29	1,43
II kw. 2005	1	82	1	0,29	23,43	0,29
III kw. 2005	3	75	0	0,86	21,43	0,00
IV kw. 2005	4	69	3	1,14	19,71	0,86
I kw. 2006	4	83	1	1,14	23,71	0,29
II kw. 2006	0	86	2	0,00	24,57	0,57
III kw. 2006	6	45	1	1,71	12,86	0,29
IV kw. 2006	2	61	0	0,57	17,43	0,00
I kw. 2007	2	49	1	0,57	14,00	0,29
II kw. 2007	1	80	1	0,29	22,86	0,29
III kw. 2007	4	64	1	1,14	18,29	0,29
IV kw. 2007	2	52	1	0,57	14,86	0,29

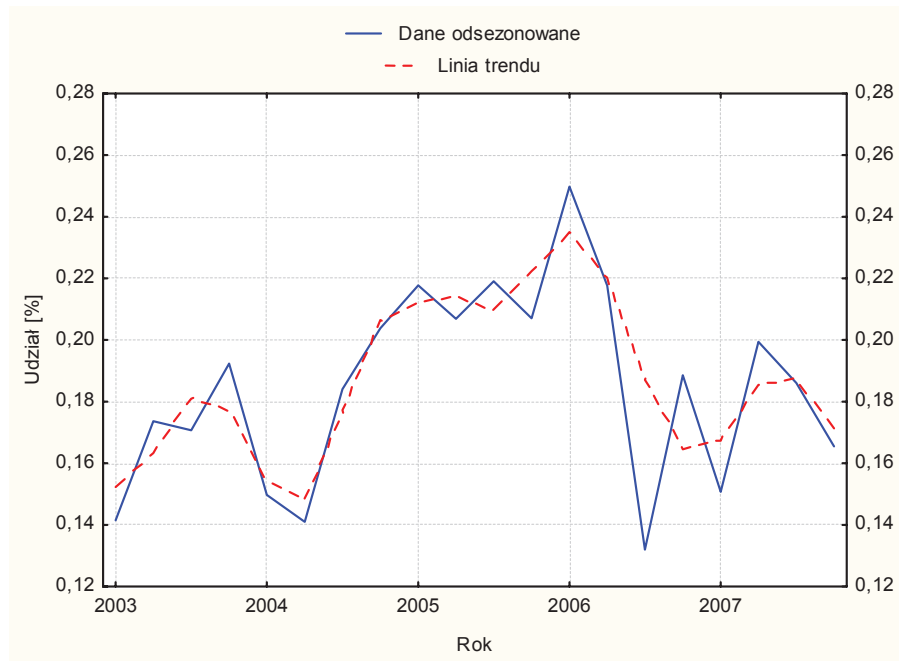
^{a)} Liczba ankiet reprezentowanych przez neuron, np. Opt(2,9), do wszystkich ankiet w określonym okresie.

Dyskusja wyników

Utworzony w trakcie badań model neuronowy, ze względu na opisane założenia, posłużył do bezwzorcowej klasyfikacji respondentów badań nastrojów gospodarczych na terenie woj. lubelskiego w latach 2003–07. Badaneankiety dotyczyły nastrojów gospodarczych w gospodarstwach domowych. Jak wynika z analizy Rys. 5., w trakcie badania sieć wyodrębniła 85 niezależnych profili respondentów spośród 100 możliwych, które, ze względu na podobieństwo (topologiczne sąsiedztwo), można zaliczyć do 3 specyficznych grup: optymistów, indyferentnych i pesymistów. Liczba ta potwierdza prawidłowość arbitralnego wyboru wymiaru mapy.

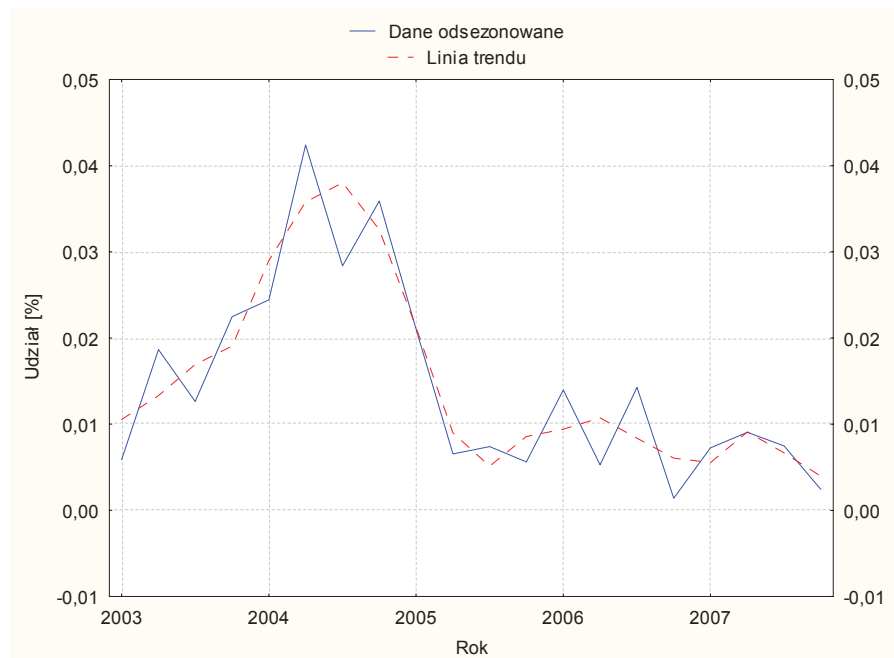
Już pobieżna analiza wykazuje, że klaster (5,5) zdecydowanie przewyższa liczebnością wszystkie pozostałe. Ich udział procentowy w całym zbiorze wynosi ponad 18%. Klaster ten reprezentuje najbardziej typowych respondentów o profilu Ind (Rys. 6.). Jak widać, odpowiedzi udzielane na wszystkie pytania mają wartość 2 (bez zmian). Liczebność oraz udział procentowy ankiet dla tego profilu można odczytać z Tabelicy 1., a dane po odsezonowaniu i linię trendu² zaobserwować na Rys. 7.

² Korekcja kwartalna addytywna prowadzona metodą Census II. Szereg wygładzony średnią ruchomą Hendersona.



Rys. 7. Zmiany udziału respondentów o profilu Ind(5,5) w ogólnej liczbie respondentów w latach 2003–2007

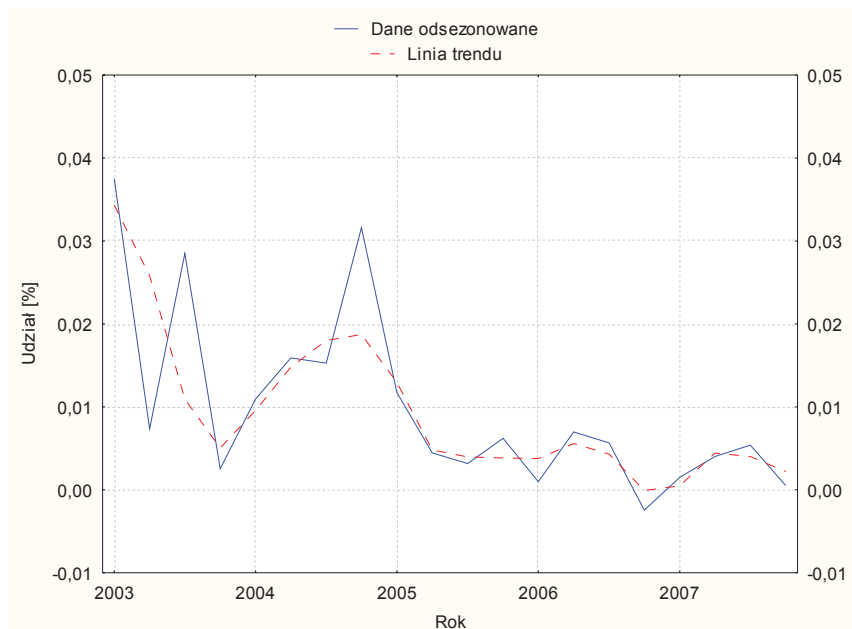
Jak widać, wyraźny wzrost udziału osób, które biernie podchodziły do wypełnienia ankiety lub przewidywały stabilizację w obszarze podlegającym ocenie, wystąpił w latach 2004–2006. Po 2006 roku udział ten znacznie spadł i od III kw. 2006 r. wyraźnie się ustabilizował. Należy jednak zauważyć, że ostatni kwartał 2007 roku być może rozpoczyna kolejną fazę trendu spadkowego.



Rys. 8. Zmiany udziału respondentów o profilu Opt(2,9) w ogólnej liczbie respondentów w latach 2003–2007

Analogiczny przebieg zmienności dla profilu Opt(2,9) pokazany jest na Rys. 8. Na wykresie daje się zaobserwować, że po początkowym wyraźnym wzroście udziału optymistów od początku 2003 r. do połowy 2004 r., nastąpił gwałtowny spadek, który został zahamowany w połowie 2005 r. Od tego momentu daje się zauważyć bardzo ustabilizowany trend boczny.

Wyraźny trend spadkowy wykazuje natomiast przebieg zmienności udziału respondentów o profilu Pes(9,4). Po gwałtownym spadku w 2003 roku i korekcie wzrostowej, przez kolejne cztery kwartały, od IV kw. 2004 r., nastąpił początkowo bardzo szybki spadek, trwający trzy kwartały. Od połowy 2005 roku daje się zaobserwować trend boczny z minimalną tendencją spadkową.



Rys. 9. Zmiany udziału respondentów o profilu Pes(9,4) w ogólnej liczbie respondentów w latach 2003–2007

Podsumowanie

Wyniki przeprowadzonych badań sugerują, że Sieci Kohonena mogą być dobrym narzędziem wspomaganego analizy badań nastrojów gospodarczych, komplementarnym do analizy zmian indeksów. Przedstawiony na Rys.1. indeks nastrojów gospodarczych dla gospodarstw domowych woj. lubelskiego w latach 2001-2008, wykazuje w całym okresie badawczym wyraźny, stały wzrost. Od początku 2006 roku możemy mówić o podlegającym okresowym wahaniom, optymizmie respondentów (wartość indeksu jest większa od 0). Zestawiając tę obserwację z wykresami z Rys.8 i 9. można zauważyć, że stało się to bardziej za przyczyną zmniejszającej się liczby pesymistów niż przyrostowi w grupie optymistów.

Analiza linii trendów profili optymistów i pesymistów pokazuje dodatkowo wyraźny wzrost polaryzacji nastrojów od III kw. 2003 do II kw. 2004 roku, czego z oczywistych względów nie może pokazać indeks. Bardzo interesującym byłoby nałożenie tej obserwacji na sytuację społeczno-gospodarczą. Równie interesującym mogłoby też być zestawienie tych zmian z kampaniami politycznymi i medialnymi.

Charakterystyczne profile ankietowanych, przedstawione na rys. 6., ukazują, że skrajni optymiści lepiej oceniają sytuację swojego gospodarstwa domowego, w perspektywie najbliższego kwartału, niż sytuację w regionie. W grupie skrajnych pesymistów, zaobserwować można wyraźną obawę o stan gospodarki w regionie, przy ok. 3-krotnie mniejszych obawach o wzrost poziomu bezrobocia. Na uwagę zasługuje również fakt, że ta grupa respondentów stosunkowo dobrze ocenia przewidywaną zmianę oszczędności we własnym gospodarstwie domowym.

Oprócz opisanych w pracy profili, pozostałe 82 wyodrębnione na mapie³, nie odbiegają od pewnego schematu. Daje się zauważyć tendencję, że respondenci w swojej większości podobnie (lepiej lub gorzej) oceniają wszystkie aspekty sytuacji społeczno-gospodarczej zawartej w pytaniach. Należy w tym miejscu zauważyć, że korzystając ze standardowego narzędzia analizy, jakim jest np. arkusz kalkulacyjny, bardzo szybko można na wykresie przeanalizować wszystkie profile w poszukiwaniu specyficznych.

Przeprowadzone badania można poszerzyć o analizę zmian w całych wyodrębnionych obszarach, a na jej podstawie sformułować analityczną postać, alternatywnego indeksu nastrojów.

³ Dla wyodrębnionych na mapie 15 profili nie ma reprezentantów w zbiorze ankiet poddanych badaniom.

Literatura

- Komisja Wspólnot Europejskich 2006: *Wspólny Unijny program sondaży wśród przedsiębiorstw i konsumentów*. (Komunikat Komisji) KOM, Bruksela.
- OSOWSKI S. 1994: *Sieci neuronowe*. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
- RUTKOWSKI L. 2006: *Metody i techniki sztucznej inteligencji. Inteligencja obliczeniowa*. Wydawnictwo Naukowe PWN, Warszawa.
- StatSoft 2006: *Elektroniczny Podręcznik Statystyki PL*, Kraków, [[:@:]] <http://www.statsoft.pl/textbook/stathome.html>.
- VARMUZA K. 1998: *Chemometrics: Multivariate View on Chemical Problems*. [w:] Schleyer P. v. R., Allinger N. L., Clarc T., Gasteiger J., Kollman P. A., Schafer III H. F., Schreiner P. R. (Eds.): *The Encyclopedia of Computational Chemistry*, J. Wiley & Sons Ltd, Chichester 1998, Vol. 1.
- WSZiA2008: *System przeciwdziałania powstawaniu bezrobocia na terenach słabo zurbanizowanych*. Zamość, [[:@:]] http://ebarmetr.pl/mezo_scan_economical_atmosphere.php?province_id=2&r_type=prognoza.
- ŻURADA J. 1992: *Sztuczne sieci neuronowe*. PWN, Warszawa.