

Kamila Migdał-Najman

Uniwersytet Gdański

ZASTOSOWANIE SAMOU CZĄCEJ SIĘ SIECI NEURONOWEJ TYPU SOM W ANALIZIE KOSZYKOWEJ

Streszczenie: Analiza koszykowa jest metodą identyfikacji kombinacji produktów nabywanych razem. Pozwala na wykrywanie powtarzających się, ukrytych powiązań w postaci prostych reguł asocjacyjnych. Jest podstawowym narzędziem analitycznym stosowanym w marketingu detalicznym pomocnym w podejmowaniu wielu decyzji biznesowych. W artykule zaprezentowano możliwości samouczącej się sieci neuronowej typu SOM w poszukiwaniu wzorców zakupowych klientów i podjęto nad nimi dyskusję. Na szczególną uwagę zasługuje również metoda wizualizacji decyzji zakupowych.

1. Wstęp

Nowoczesne systemy rejestracji transakcji handlowych (kody kreskowe, kody paskowe) są cennym źródłem pozyskiwania informacji o zachowaniu konsumentów. Systemy te pozwalają na automatyczne wczytywanie informacji z transakcji i identyfikację produktów w bazach danych. Dostęp do coraz większych baz danych może jednak powodować trudności w bieżącym ich analizowaniu i wyciąganiu wniosków. Ważne staje się więc poszukiwanie nowych technik służących gromadzeniu, przetwarzaniu i analizowaniu takich danych. Istotnym elementem jest również automatyzacja procesu poszukiwania informacji i wizualizacja prezentowanych rozwiązań. Wszystkie produkty, które w drodze transakcji nabył konsument, reprezentują tzw. koszyk (*basket*). Analiza koszykowa (*basket analysis*, *market basket analysis*) to metoda identyfikacji kombinacji produktów nabywanych razem (współkupowanych) i ich sekwencji zakupu. Pozwala na wykrywanie powtarzających się, ukrytych powiązań (*associations*) i korelacji wśród różnych produktów, które znajdują się w bazach danych transakcji i prezentowaniu ich w postaci prostych reguł. Jest jedną z bardziej powszechnych i użytecznych metod analizy danych. Może okazać się pomocna w podejmowaniu wielu decyzji biznesowych, np. takich jak projektowanie katalogów produktów, planowanie strategii cross-marketingu, czy też w analizie zachowań i zwyczajów klientów sklepu [Nong Ye 2003]. Celem artykułu jest próba wykorzystania sieci neuronowej typu SOM w poszukiwaniu zwyczajów i prawidłowości zakupowych klientów. W badaniu wy-

korzystane zostaną dane symulacyjne. Wykryte zależności zostaną zaprezentowane w postaci prostych reguł. Zaproponowana zostanie również wizualizacja uzyskanych wyników decyzji zakupowych.

2. *Self Organizing Map*

Self Organizing Map – SOM, mapa samoorganizująca się, nazywana jest również siecią lub mapą Kohonena (por. [Kohonen 1997; Yin 2002; Deboeck, Kohonen 1998; Migdał-Najman 2009]). Zaproponowana i rozwinięta została ok. 1982 r. przez fińskiego profesora Teuvo Kohonena. Obecnie jest jednym z bardziej znanych nienadzorowanych modeli sztucznych sieci neuronowych, który stosuje się głównie do klasyfikacji, grupowania, redukcji wymiarowości, wyszukiwania anomalii i odchyień od wartości typowych, wizualizacji wielowymiarowych zbiorów danych i badania dynamiki zjawisk (por. [Papadimitriou i in. 2002; Delgado 2000; Migdał-Najman, Najman 2004]). Sieć SOM uporządkowana jest w pewną strukturę nazywaną siecią, która zostaje „rozrzucana” w przestrzeni wejściowej cech. Każdy neuron znajdujący się w takiej sieci lokalnie aproksymuje analizowane obiekty. Każdy taki „lokalny aproksymator” to jeden neuron. Neurony na sieci powiązane są z pozostałymi neuronami pewną relacją sąsiedztwa, która określa topologię (strukturę) sieci. Każdy wektor wejściowy połączony jest równolegle ze wszystkimi neuronami na sieci przez wektor wag (*reference vector*). Uczenie i dopasowanie sieci do przestrzeni wejściowej cech jest adaptacyjnym procesem doboru wag sieci. Proces uczenia sieci rozpoczyna się od inicjalizacji wag, czyli wyznaczenia początkowych wartości współczynników wagowych. Do procedur inicjalizacji wag możemy zaliczyć np. inicjalizację losową lub próbkową. W inicjalizacji losowej (*random initialization*) wartości początkowe współczynników wagowych dobierane są jako liczby losowe. Inicjalizacja losowa stosowana jest głównie w przypadku, gdy niewiele wiemy o zbiorze danych wejściowych. Wadą inicjalizacji losowej jest ryzyko występowania w sieci znacznej liczby „martwych” neuronów. Zdarza się to szczególnie wtedy, gdy obszar losowania nie pokrywa się z obszarem sygnałów wejściowych. W inicjalizacji próbkowej (*sample initialization*) wartości początkowe współczynników wagowych inicjowane są losowo ze zbioru danych wejściowych. Przewagą inicjalizacji próbkowej jest automatyczne zlokalizowanie współczynników wagowych w tej samej części przestrzeni zbioru danych wejściowych [Duch 2000, s. 202; Vesanto 1997]. Neurony, które sąsiadują ze sobą na sieci, wykazują skłonność do rozpoznawania podobnych (bliskich) do siebie obiektów wejściowych i mają podobne wektory wag. Zaproponowany przez T. Kohonena algorytm uczenia tak modyfikuje współczynniki wagowe w neuronach znajdujących się w warstwie stanowiącej mapę topologiczną, aby zbliżyć je do centrów skupień występujących w danych uczących. Każdy neuron staje się pewnym wzorcem grupy bliskich sobie obiektów wejściowych.

3. Eksperyment badawczy

Do weryfikacji możliwości zastosowania sieci SOM w poszukiwaniu wzorców zakupowych wśród klientów przygotowano dwa eksperymenty. W badaniu wykorzystano dwa zbiory danych umownych. Eksperyment pierwszy ma na celu zaprezentowanie krok po kroku sposobu wykorzystania sieci SOM w poszukiwaniu prawidłowości zakupowych (produktów współkupowanych). Wyniki analizy uzyskane dla sieci SOM zaprezentowano wizualnie na wykresach sieciowych oraz porównano z wynikami, które uzyskano by, stosując reguły asocjacyjne¹. W eksperymencie drugim zweryfikowano możliwości sieci SOM w taki sam sposób jak w przypadku eksperymentu pierwszego. Analizę przeprowadzono jednak na dużo większej bazie transakcji zawierającej więcej zarówno przypadków, jak i produktów. Eksperyment drugi rozszerzono o analizę wzorców zakupowych klientów w wyodrębnionych segmentach (jednorodnych grupach klientów). W obu przypadkach analizowano przede wszystkim zależności, które występują najczęściej w zbiorach danych. Budowane sieci SOM dla dwóch eksperymentów testowano, stosując różne topologie sieci. Budowano sieci o heksagonalnej strukturze połączeń o rozmiarach od 2x2 do 4x4 (dla pierwszego eksperymentu) i 5x5 do 12x12 (dla drugiego eksperymentu). Przy budowie sieci uwzględniano cztery funkcje sąsiedztwa: gaussowską, uciętą gaussowską, prostokątną i wykładniczą. W każdym wariancie uczono sieci od 100 do 1000 iteracji (eksperyment pierwszy) i od 1000 do 5000 iteracji (eksperyment drugi). Dla każdej symulacji wyznaczano trzy miary jakości sieci: błąd kwantyzacji, błąd topograficzny i błąd dystorsji. Optymalną sieć SOM w sensie minimalizacji miar błędów sieci poddawano szczegółowej analizie, poszukując wzorców zachowań hipotetycznych klientów.

4. Wyniki badań

Eksperyment pierwszy. W tabeli 1 zaprezentowano 10 hipotetycznych transakcji (koszyków) prezentujących 7 produktów kupowanych przez klientów. Do produktów tych należą: marchew (X_1), sałata (X_2), pomidory (X_3), seler (X_4), jabłka (X_5), pomarańcze (X_6) i śliwki (X_7). Każdy wiersz w tab. 1 prezentuje kombinację produktów kupionych przez poszczególnych klientów w pojedynczej transakcji. Mianowicie np. klient pierwszy, który zrealizował swoje zakupy, miał w koszyku marchew, sałatę i pomidory; klient drugi: marchew, pomidory i seler itd. Każdą transakcję (indywidualny koszyk klienta) przedstawiono w postaci binarnego (zero-jedynkowego) wektora transakcji. Wektor ten ma tyle składowych, ile jest wszyst-

¹ Poszukiwanie ukrytych zależności w postaci prostych reguł asocjacyjnych o postaci: jeżeli [poprzednik], to [następnik]. Wyodrębnianie reguł zawierających istotną informację na podstawie miar jakości uzyskanych reguł: wsparcia, zaufania i korelacji. W badaniu zaprezentowano jedynie reguły najsilniejsze, czyli takie, które informują o najczęściej występujących prawidłowościach (reguły o najwyższym wsparciu).

Tabela 1. Kombinacje artykułów kupionych w 10 transakcjach

Transakcja	Kupione artykuły
1	marchew, sałata, pomidory
2	marchew, pomidory, seler
3	marchew, seler
4	marchew, pomidory, jabłka
5	sałata, pomidory, seler, jabłka
6	pomidory, seler, pomarańcze
7	sałata, pomidory
8	marchew, sałata, pomidory, jabłka
9	marchew, jabłka, śliwki
10	marchew, jabłka

Źródło: opracowanie własne.

kich analizowanych produktów. Jeżeli w wektorze transakcji co najmniej jedna składowa jest równa jeden, to oznacza, że w danym koszyku znajduje się co najmniej jeden z analizowanych produktów. Jeżeli w koszyku klienta nie ma pewnego artykułu, to składowa w wektorze transakcji przypisana do tego artykułu przyjmie wartość równą zero. W tabeli 2 zaprezentowano macierzowy format danych eksperymentu pierwszego.

Tabela 2. Macierzowy format danych

Transakcja	Kupione artykuły						
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
1	1	1	1	0	0	0	0
2	1	0	1	1	0	0	0
3	1	0	0	1	0	0	0
4	1	0	1	0	1	0	0
5	0	1	1	1	1	0	0
6	0	0	1	1	0	1	0
7	0	1	1	0	0	0	0
8	1	1	1	0	1	0	0
9	1	0	0	0	1	0	1
10	1	0	0	0	1	0	0

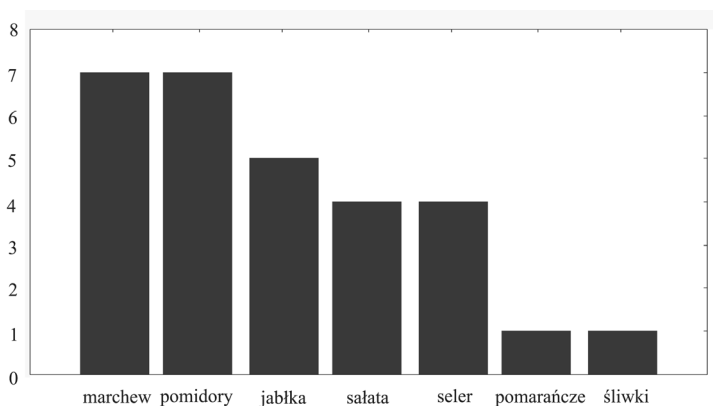
Źródło: opracowanie własne.

Wstępna analiza danych pozwala na zaobserwowanie następujących prawidłowości. Najczęściej klienci kupowali marchew i/lub pomidory (por. rys. 1). Najczęściej współkupowane są ze sobą: {marchew} → {pomidory} (4 koszyki); {jabłka} → {marchew} (4 koszyki) oraz {sałata} → {pomidory} (4 koszyki) (por. tab. 3). Występują 2 koszyki, w których jednocześnie występują: marchew, sałata, pomidory lub marchew, pomidory, jabłka lub sałata, pomidory, jabłka. Liczba wszystkich możliwych reguł asocjacyjnych w prezentowanym eksperymencie wynosi 448.

Tabela 3. Reguły asocjacyjne o najwyższym poziomie wsparcia

Jeżeli [poprzednik], to [następnik]	Wsparcie w %
Jeżeli marchew, to pomidory	40
Jeżeli jabłka, to marchew	40
Jeżeli sałata, to pomidory	40
Jeżeli pomidory, to seler	30
Jeżeli pomidory, to jabłka	30

Źródło: opracowanie własne.

**Rys. 1.** Częstotliwość występowania produktów w 10 transakcjach

Źródło: opracowanie własne.

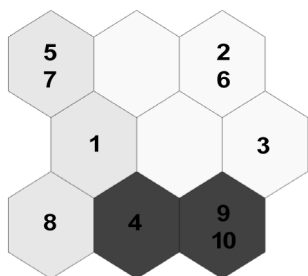
W rzeczywistości w sklepach realizuje się bardzo wiele takich transakcji, jak również sklepy mają bardzo bogaty asortyment. W takich sytuacjach utrudniona staje się analiza tak obszernych zbiorów danych. Utrudniona, a często wręcz niemożliwa staje się również wizualizacja analizowanych reguł. Alternatywą w takich sytuacjach może być zastosowanie sztucznej sieci neuronowej typu SOM. W analizowanym przykładzie optymalna w sensie minimalizacji miar błędów sieci okazała się sieć SOM o następujących parametrach: typ połączeń neuronów – heksagonalny, rozmiar sieci 3×3 neurony, funkcja sąsiedztwa – prostokątna, błąd kwantyzacji – 1,1384, błąd topograficzny – 0, błąd dystorsji – 1,33 i liczba iteracji uczących – 100. Na rysunku 2 zaprezentowano heksagonalną sieć SOM o rozmiarze 3×3 neurony (łącznie 9 neuronów). Poszczególne neurony sieci odpowiadają za pewne obiekty wejściowe (transakcje, koszyki). Mianowicie neuron 1 sieci SOM mapuje transakcję 5 i 7, neuron 2 – transakcję 1, neuron 3 – transakcję 8, neurony 4 i 5 nie odpowiadają za żadną transakcję (są neuronami martwymi), neuron 6 – transakcję 4, neuron 7 – transakcję 2 i 6, neuron 8 – transakcję 3, a neuron 9 – transakcje 9 i 10. W sytuacji, kiedy neuron odpowiedzialny jest tylko za jedną transakcję (np. neuron 2), to wektor wag takiego neuronu jest równy składowym wektora transak-

cji obiektu mapowanego przez ten neuron (np. transakcja 1). W sytuacji, kiedy neuron mapuje większą liczbę obiektów (transakcji) (np. neuron 1), to wektor wag tego neuronu jest punktem ciężkości (centroidem) obiektów mapowanych przez ten neuron (np. transakcje 5 i 7). W tabeli 4 zaprezentowano wektory wag neuronów prezentowanej sieci.

Tabela 4. Wektory wag sieci o rozmiarze 3×3

Neuron	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	0	1	1	0,5	0,5	0	0
2	1	1	1	0	0	0	0
3	1	1	1	0	1	0	0
4	0,4	0,6	1	0,6	0,2	0,2	0
5	0,857	0,143	0,571	0,423	0,429	0,143	0,143
6	1	0	1	0	1	0	0
7	0,5	0	1	1	0	0,5	0
8	1	0	0	1	0	0	0
9	1	0	0	0	1	0	0,5

Źródło: opracowanie własne.



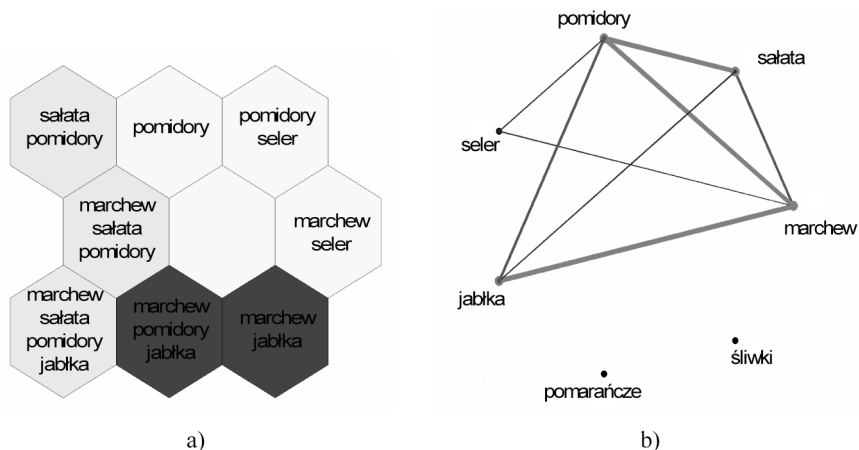
Rys. 2. Heksagonalna sieć SOM o rozmiarze 3×3 neurony i obiekty mapowane przez poszczególne neurony

Źródło: opracowanie własne.

Każdy neuron sieci SOM, a dokładniej jego wagi, odzwierciedla wzorce zachowań grupy bliskich sobie obiektów wejściowych (transakcji). Składowe wektora wag są liczbami z przedziału od 0 do 1. Wartość równa jeden w wektorze wag oznacza, że wśród wszystkich mapowanych przez dany neuron transakcji występuje ten konkretny produkt (100% mapowanych transakcji zawiera ten produkt). Jeżeli np. w wektorze wag występuje wartość równa 0,5, to oznacza to, że wśród mapowanych przez dany neuron transakcji w 50% przypadków występuje ten produkt. Najistotniejsze wzorce postępowania klientów dla wag równych jeden

zaprezentowano na rysunku 3a). Wektor wag neuronu pierwszego przyjmuje wartość równą jeden dla produktów: sałata (X_2) oraz pomidor (X_3). Na sieci SOM trzy neurony odpowiadają za parę produktów: sałata i pomidor (neurony 1, 2, 3), trzy neurony za produkty: pomidor i marchew (neurony 2, 3, 6) oraz trzy neurony za parę: jabłka i marchew (neurony 3, 6, 9). Są to pary produktów, które są najczęściej ze sobą łączone. Reguły asocjacyjne dla tych par produktów przyjmują najwyższy poziom wsparcia (por. tab. 3). Na rysunku 3b) w formie wykresu sieciowego zaprezentowano najczęściej współkupowane produkty. Im prosta łącząca dwa produkty jest „grubsza”, tym dane produkty częściej ze sobą współwystępują (tym jest

więcej neuronów, które odpowiadają za taki wzorec zachowań). W wyniku zastosowania sztucznej sieci neuronowej SOM i reguł asocjacyjnych otrzymano bardzo zbliżone wyniki zależności między produktami.



Rys. 3. a) Wzorce zachowań mapowanych obiektów;
b) Wizualizacja zależności uzyskanych na podstawie sieci SOM

Źródło: opracowanie własne.

Eksperyment drugi. W eksperymencie drugim wykorzystano dane symulacyjne prezentujące 596 transakcji fikcyjnych klientów. Klienci mogli wybierać produkty z następującej listy 22 artykułów: 1 – słodycze, 2 – desery, 3 – kawa i herbata, 4 – dania gotowe, 5 – zdrowa żywność, 6 – ryby, 7 – artykuły sypkie, 8 – dla dzieci, 9 – przetwory, 10 – nabiał, 11 – przyprawy, 12 – wędliny, 13 – piwo, 14 – sosy, 15 – alkohole, 16 – przekąski, 17 – napoje, 18 – woda mineralna, 19 – pieczywo, 20 – mięso, 21 – owoce i 22 – warzywa. Liczbę wystąpień analizowanych artykułów zaprezentowano w tab. 5. Najczęściej w koszyku klienta znalazł się nabiał (439 razy), pieczywo (420 razy), napoje (386 razy), słodycze (377 razy) i warzywa (368 razy). W tabeli 6 zaprezentowano reguły asocjacyjne o najwyższym poziomie wsparcia dla rozpatrywanych produktów.

Tabela 5. Liczba wystąpień poszczególnych artykułów

Artykuły (numer)	10	19	17	1	22	21	7	12	20	16	13
Liczba wystąpień	439	420	386	377	368	336	298	239	214	200	199
Artykuły (numer)	15	9	3	4	2	14	8	18	5	6	11
Liczba wystąpień	168	154	129	118	90	86	74	73	36	35	11

Legenda: numer 10 oznacza nabiał, 19 – pieczywo itd., zob. eksperymnt drugi.

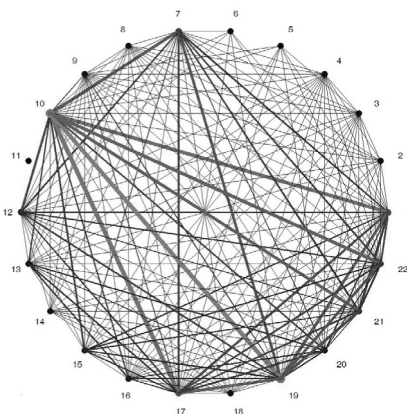
Źródło: opracowanie własne.

W eksperymencie drugim zbiór danych 596 transakcji poddano analizie, wykorzystując sztuczną sieć neuronową typu SOM. Optymalna w sensie minimalizacji miar błędów sieci okazała się sieć SOM o następujących parametrach: typ połączeń neuronów – heksagonalny, rozmiar sieci – 11x11 neuronów, funkcja sąsiedztwa – ucięta gaussowska, błąd kwantyzacji – 1,03, błąd topograficzny – 0,33, błąd dystorsji – 1,21 i liczba iteracji uczących równa 1000. Macierz wag, optymalna w sensie minimalizacji miar błędów sieci, miała wymiar 121x22 (sieć 11x11 neuronów ma łącznie 121 neuronów). Najważniejsze wzorce zachowań klientów zaobserwowano, uwzględniając do dalszej analizy tylko te współczynniki wagowe, które w macierzy wag przyjęły wartość równą jeden. Uwzględniając częstość powtarzania się istniejących kombinacji par produktów prezentowanych przez 121 neuronów sieci, zaproponowano rozwiązanie graficzne w postaci wykresu sieciowego (por. rys. 4).

Tabela 6. Reguły asocjacyjne o najwyższym poziomie wsparcia – eksperyment drugi

Jeżeli [poprzednik], to [następnik]	Wsparcie w %
Jeżeli nabiął, to pieczywo	56,71
Jeżeli nabiął, to napoje	50,17
Jeżeli nabiął, to warzywa	49,83
Jeżeli nabiął, to słodczyce	46,81
Jeżeli pieczywo, to napoje	46,31
Jeżeli nabiął, to owoce	46,31
Jeżeli warzywa, to pieczywo	46,14

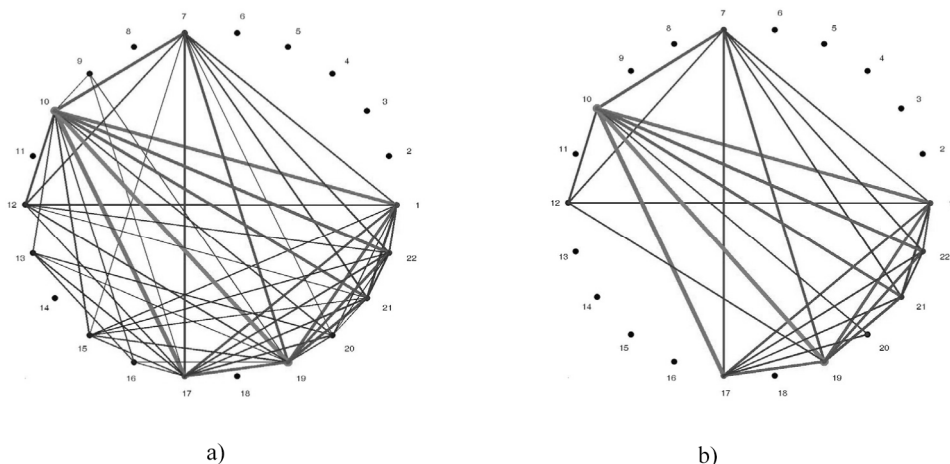
Źródło: opracowanie własne.



Rys. 4. Wizualizacja wzorców zachowań klientów uzyskana na podstawie sieci SOM

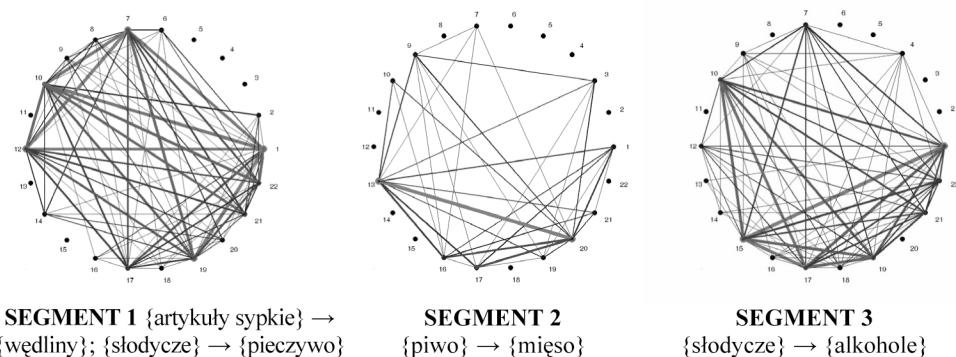
Źródło: opracowanie własne.

Duża liczba istniejących zależności² pomiędzy produktami bardzo często uniemożliwia przejrzystą wizualizację wzorców zachowań konsumentów. Na rysunkach 5a) i 5b) zaprezentowano najważniejsze zależności istniejące w zbiorze danych, redukując te pary produktów, które wśród najistotniejszych wykrytych wzorców występowały najrzadziej³. Zastosowanie sieci neuronowej typu SOM



Rys. 5. Wizualizacja najistotniejszych wzorców zakupowych klientów

Źródło: opracowanie własne.



Rys. 6. Wizualizacja zależności w wyodrębnionych segmentach

Źródło: opracowanie własne.

² Liczba wszystkich możliwych reguł asocjacyjnych w eksperymencie drugim wynosi 46 137 344.

³ Na rysunku 4 zaprezentowano wszystkie istniejące zależności między produktami (waga równa jeden). Na rysunku 5a) dokonano redukcji istniejących zależności (par produktów), które reprezentowane były tylko przez jeden neuron. Na rysunku 5b) dokonano redukcji istniejących zależności, które reprezentowane były przez jeden neuron lub dwa neurony. Liczbę produktów na wykresie można także zredukować przez eliminację tych, które są najsłabiej związane z innymi produktami.

pozwoiliło na wyciągnięcie następujących reguł: artykuły najczęściej współkupo- wane to {nabiał} → {pieczywo} oraz {nabiał} → {napoje}. Zgodność uzyskanych wyników podejścia opartego na sieci SOM i regułach asocjacyjnych w eksperymencie drugim jest bardzo wysoka. Niemal wszystkie reguły asocjacyjne o naj- wyższym poziomie wsparcia znajdują odzwierciedlenie na sieci SOM.

W eksperymencie drugim dokonano segmentacji klientów na podstawie uzy- skanych charakterystyk zakupowych. Na uzyskanej sieci SOM wyodrębniono osiem jednorodnych grup klientów. Uzyskane informacje z przeprowadzonej seg- mentacji klientów mogą zostać wykorzystane do dywersyfikacji oferty na podsta- wie preferencji klientów do określonych produktów. Na rysunku 6 zaprezentowano wzorce zakupowe (wyróżnione reguły) klientów z trzech wybranych segmentów.

5. Wnioski

Sieć SOM ma duży potencjał w rozpoznawaniu ukrytych zależności w postaci pro- stych reguł. Występuje bardzo wysoka zgodność wyników podejścia opartego na sieci SOM i regułach asocjacyjnych. Małe różnice uzyskanych zależności z sieci SOM w porównaniu z regułami asocjacyjnymi wynikają z tego, że neuron pełni funkcję „metaskupienia; środka ciężkości” obiektów grupujących się przy po- szczególnych neuronach. Skala różnic ma jednak niewielkie znaczenie praktyczne. Sieć SOM pozwala na większą klarowność uzyskanych wyników, jak również na przyspieszenie prowadzonej analizy. Zysk ze stosowania prezentowanej metody jest tym większy, im więcej obiektów (transakcji) i produktów (atrybutów) podlega badaniu. Uzyskane wyniki z sieci SOM pozwalają na wizualizację zależności w postaci wykresów sieciowych. Za pomocą sieci SOM można ustalić zależności, ale nie uzyskamy informacji o prawdopodobieństwie realizacji transakcji, które zawie- rają zadane produkty. Sieć SOM może okazać się ciekawą alternatywą w stosunku do „klasycznej” analizy koszykowej, nie ograniczając jej zastosowania jedynie do problemu koszyka sklepowego.

Literatura

- 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.
- Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R. (red.), *Sieci neuronowe*, PAN, Akademicka Oficyna Wydawnicza Exit, Warszawa 2000.
- Kohonen T., *Self-Organizing Maps*, Springer Series in Information Sciences, Springer-Verlag, Berlin Heidelberg 1997.
- Migdał-Najman K., *Analiza porównawcza własności nienadzorowanych sieci neuronowych typu Self Organizing Map i Growing Neural Gas w analizie skupień*, [w:] *Klasyfikacja i analiza danych –*

- teoria i zastosowania*, Taksonomia 16, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 47, UE, Wrocław 2009, s. 205-213.
- 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, s. 507-519.
- Nong Ye (red.), *The Handbook of Data Mining*, Lawrence Erlbaum Associates, Publishers, London 2003.
- 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, s. 187.
- 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, s. 4.
- Yin H., *Data visualization and manifold mapping using the ViSOM*, „Neural Network” 2002, 15.

THE APPLICATION OF NEURAL NETWORK SELF ORGANIZING MAP (SOM) IN MARKET BASKET ANALYSIS

Summary: Market basket analysis seeks to identify combinations of items in the form of association rules. It is a basic analytical tool applied in retail marketing which plays a crucial role in decision support and is regarded as a typical field of application for data mining. In this article the neural networks such as Self Organizing Map (SOM) are presented and discussed with respect to their suitability for the discovery of purchase patterns. Particular attention is paid to the visualization of purchase decisions.