

## **Ocena ilościowego spożycia żywności przez rodziny polskie z wykorzystaniem ich klasyfikacji przy użyciu sieci neuronowych**

O ilości spożywanej żywności wiemy z danych GUS, tj. z danych bilansowych oraz z tzw. badań budżetów gospodarstw domowych, prowadzonych corocznie już od ponad 40. lat na losowanej corocznie próbie gospodarstw domowych. Konfrontacja znanych z tych źródeł przeciętnych wielkości spożycia z naukowymi zaleceniami żywnościowymi prowadzi do wniosku o dość zasadniczych między nimi rozbieżnościach, a to oznacza ryzyko chorób na tle wadliwego żywienia (Kowrygo 2000).

Dane z badań budżetów pozwalają dodatkowo stwierdzić, że istnieje duża indywidualna zmienność w spożyciu poszczególnych produktów, a nawet ich grup, przy tym skądinąd uzasadniona jest teza o istnieniu w odżywianiu się ludności względnie jednolitych stylów (wzorców) respektowanych konsekwentnie przez konkretne rodziny. Wzorce te warto by ocenić precyzyjnie ze zdrowotnego punktu widzenia, a w tym określić negatywne następstwa stosowanych racji pokarmowych. W tym celu przydatne byłoby wyłonienie takich grup gospodarstw domowych, które byłyby pod względem spożycia żywności względnie jednolite, a jednocześnie różniły się między sobą możliwie dużo, co upoważniałoby do uznania ich za reprezentujące odrębne style żywnościowe. Wydaje się, że sieci neuronowe mogą być bardzo pomocne w takiej klasyfikacji gospodarstw domowych.

**Sieci neuronowe** to w istocie urządzenia techniczne zbudowane na podstawie teorii formalnej o tej samej nazwie, której rozwój – zapoczątkowany oficjalnie w 1943 roku – był i jest nadal inspirowany biologiczną wiedzą człowieka o układach nerwowych [Hertz i inni 1995].

Sztuczne modele mózgu tworzone są ze znacznie mniejszej liczby neuronów jako podstawowych elementów składowych niż prawdziwy mózg. Sposób działania sieci neuronowych cechuje przede wszystkim to, że nie wymagają one algorytmu wykonania powierzanego im zadania. W to miejsce sieci wymagają wyuczenia (wytrenowania) przeprowadzanego na bazie przykładów. Proces uczenia sieci prowadzi do kształtowania wag połączeń pomiędzy poszcze-

gólnymi neuronami, które w ten sposób „przejmują” strukturę danych (sygnałów) i nabierają zdolności do generalizacji.

W koncepcjach neuronowych dostrzega się obiecujące możliwości zastosowań w wielu dziedzinach, a szczególnie w ekonomii, finansach, medycynie, geologii i fizyce, do realizacji różnych celów, a zwłaszcza tych, w których mamy do czynienia z klasyfikacją, przewidywaniem, optymalizacją, sterowaniem, rozpoznawaniem, filtrowaniem lub kojarzeniem danych (sygnałów) [Tadeusiewicz 1999].

Zasadniczym przełomem w rozwoju sieci neuronowych stała się możliwość ich symulowania przy użyciu współczesnych komputerów. Sprzyjająca jest duża ich szybkość i zasobność w pamięć operacyjną.

Jednocześnie coraz wyraźniej dokuczliwe stają się ograniczenia metod algorytmicznych i opartych na nich programach komputerowych. Dla fragmentarycznej tylko ilustracji uwarunkowań algorytmów można przywołać klasyczne procedury statystycznego opracowywania danych, które bardzo często wymagają spełnienia wielu założeń, np. co do normalności rozkładu przypadków i liniowości związków pomiędzy analizowanymi cechami. Założenia te nie zawsze możliwe są do spełnienia, a przynajmniej wymagają sporej wiedzy już na wstępnym etapie analiz. Jest więc zapotrzebowanie na narzędzia, które byłyby bardziej praktyczne i uniwersalne.

Wiele wskazuje, że narzędziem takim mogą być sieci neuronowe. Ocenia się, że przy ich stosowaniu poziom wymaganej wiedzy teoretycznej do skutecznego zbudowania modelu aplikacyjnego jest znacznie mniejszy niż w przypadku tradycyjnych metod. Nie znaczy to oczywiście, że sieci są niezwykle prostym i zawsze skutecznym narzędziem, użytkownik musi bowiem dokonywać wyboru najlepszego typu neuronu, jak i określać architekturę (ich ułożenie) do zadania, jakie ma wykonać, wreszcie musi dysponować pewną wiedzą empiryczną pozwalającą na „wyuczenie sieci” [Tadeusiewicz 1999].

Sieć neuronowa typu Kohonena – wykorzystana przy prezentowanych analizach – jest siecią dwuwarstwową, w której podczas trenowania – w tym przypadku na całej zbiorowości – modelowane są wagi neuronów w taki sposób, by bardzo zbliżone do siebie przypadki reprezentował ten sam neuron, przy czym możliwe jest wprowadzenie zasady sąsiedztwa, pozwalającej na przejmowanie przez sąsiednie neurony przypadków podobnych. Zadaniem sieci Kohonena jest więc wykrycie „naturalnych” skupień przypadków ze względu na przyjęte cechy, co skłania do użycia określenia, że prowadzą one do podziału optymalnego.

**Przedmiotem analizy** są dane jednostkowe o spożyciu żywności z uwzględnieniem jej podziału na 13 produktów żywnościowych lub ich grup, spożyciu badanym przez GUS w 1998 roku w 31 756 gospodarstwach domowych traktowanych jako przypadki. Okresem uczestnictwa w badaniu jest jeden miesiąc, przy czym cała wylosowana próba składa się z podpróbek uczestniczących w badaniu w kolejnych miesiącach.

Analizuje się zatem charakterystyki poszczególnych gospodarstw domowych w postaci wielkości spożycia żywności przypadających (przeliczonych) na każdą osobę danego gospodarstwa. Są to wielkości w kg na miesiąc, z wyjątkiem jaj, które ewidencjonuje się w sztukach.

**Metoda analizy** wykorzystuje program Statistica z modułem Sieci Neuronowe formy StatSoft. Tworzono za pomocą tego modułu sieci neuronowe typu Kohonena o liczbie neuronów w warstwie wyjściowej od 40 do 150 przy zawsze 13 neuronach w warstwie wejściowej (tyle, ile zmiennych wejściowych). Trenowano z wykluczeniem sąsiedztwa, jak i z sąsiedztwem w promieniu 2 neuronów, na wszystkich danych; zwykle 30 epok (przebiegów) dawało najlepsze z możliwych ukształtowanie wag.

Wytrenowanymi sieciami klasyfikowano gospodarstwa domowe, czyli oznaczano je numerami neuronów zwyciężających w projekcji danego przypadku. Następnie oceniano, na ile powstałe skupienia gospodarstw przy tych samych neuronach wyjaśniają zmienność spożycia wyróżnionych produktów żywnościowych. Oceniano to przy użyciu stosunku korelacyjnego<sup>1</sup> ustalanego osobno dla każdego produktu, przy czym jako syntetyczne miary przyjęto stosunek średni, różnice korelacji poszczególnych produktów oraz liczbę uzyskanych przez daną sieć skupień (grup) i ich wielkości (tab. 1).

Sprawdzono też, w jakim stopniu wyjaśniana jest zmienność sumarycznego spożycia żywności. Ponieważ zmienna ta nie uczestniczy w modelu, to jej wyjaśnienie należy traktować jako efekt wtórny. Mimo to jej zmienność wyjaśniana jest też na zaskakująco wysokim poziomie (do 0,75). Wszystkie sieciowe grupowania okazały się dużo lepsze od klasycznego podziału na grupy społeczno-ekonomiczne, który daje wyjaśnienie zmienności spożycia przy wskaźniku średnim na poziomie 0,19 (tab. 1).

---

<sup>1</sup>Przyjmuje wielkości od 0 (brak korelacji) do 1 (związek funkcyjny).

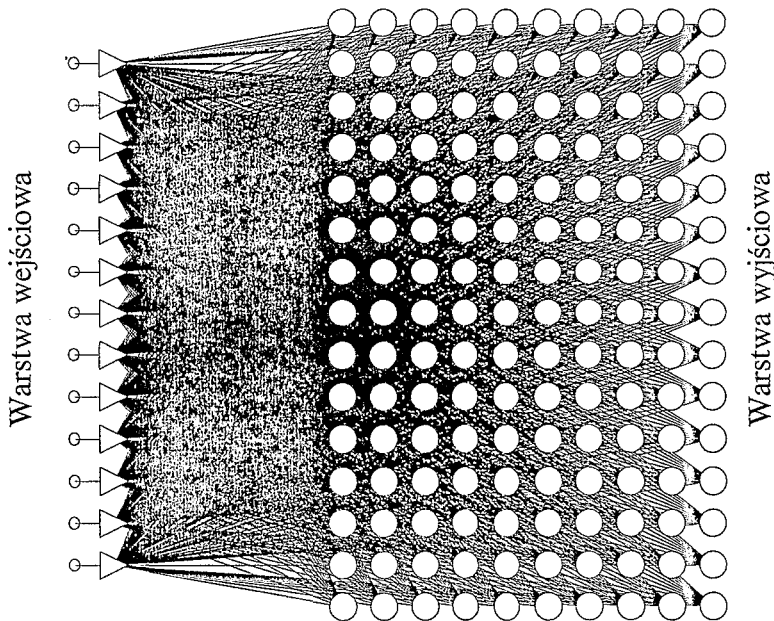
**Tabela 1**

Korelacyjne charakterystyki klasyfikacji zaproponowanych przez różne sieci neuronowe i porównawczo klasyfikacji społeczno-ekonomicznej

Wyszczególnienie	Klasyfikacje gospodarstw domowych dokonywane przez sieci o liczbie neuronów:					Przynależność społeczno-ekonomiczna
	40	60	70	90 i sąsiedztwo	150	
Stosunki korelacji spożycia produktu w danych klasyfikacjach						
Zbożowe	0,50	0,51	0,52	0,60	0,54	0,30
Mięso i przetwory	0,53	0,65	0,65	0,77	0,63	0,17
Ryby i przetwory	0,34	0,67	0,65	0,76	0,65	0,13
Mleko	0,38	0,46	0,47	0,48	0,49	0,27
Jaja	0,42	0,59	0,61	0,77	0,64	0,20
Masło	0,17	0,25	0,41	0,35	0,45	0,16
Tłuszcze i oleje roślinne	0,78	0,85	0,79	0,89	0,85	0,21
Smalec, słonina	0,17	0,69	0,68	0,80	0,67	0,18
Owoce	0,66	0,66	0,64	0,79	0,65	0,22
Warzywa	0,43	0,68	0,49	0,78	0,69	0,17
Strączkowe	0,12	0,14	0,36	0,38	0,39	0,08
Ziemniaki	0,19	0,27	0,75	0,75	0,74	0,11
Cukier (też w wyrobach)	0,35	0,38	0,35	0,61	0,46	0,20
ŚREDNIO	0,39	0,52	0,57	0,67	0,60	0,19
Spożycie razem	0,58	0,64	0,73	0,79	0,75	0,29
Współczynnik zmienności korelacji (w %)	39	33	22	22	18	24
Liczba skupień obejmujących po co najmniej:						
10% zbiorowości	2	4	3	2	4	4
5% zbiorowości	4	5	5	5	4	7
1% zbiorowości	4	9	9	23	11	8
Łączna liczba grup	4	12	12	37	16	8

Źródło: Obliczenia własne na podstawie indywidualnych wyników badań budżetów gospodarstw domowych za 1998 rok oraz badań prowadzonych przez GUS.

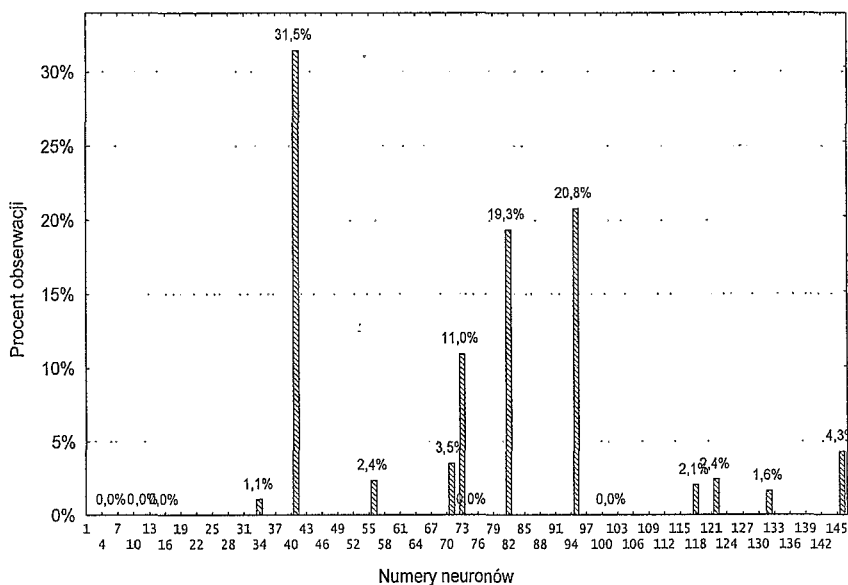
Z punktu widzenia wygody dalszych analiz przydatniejsze okazało się korzystanie z klasyfikacji sieci Kohonena trenowanej bez sąsiedztwa. Sieci te dają mniej grup. Sieci o większej liczbie neuronów dawały rezultaty bardziej zachęcające. Do dalszych prezentacji wybrano zatem klasyfikację dokonaną przez sieć ze 150 neuronami, która dawała uśrednioną korelację na poziomie 0,60 przy najmniejszej skali zmienności tego wskaźnika w przekroju poszczególnych produktów, bo przy współczynniku tylko 18%, czyli przy stosunku korelacji od 0,39 w przypadku strączkowych do 0,85 w przypadku tłuszczów roślinnych. Korelacje te uznawane są jako wysokie. Za klasyfikacją tą przemawia też liczba uzyskanych grup i ich wielkości. Zachęcający jest podział z czterema względnie dużymi grupami obejmującymi po więcej niż 10% zbiorowości i siedmioma już wyraźnie mniejszymi, obejmującymi po od 1 do 5% zbiorowości (rys. 1 i 2).



**Rysunek 1**

Schemat sieci neuronowej Kohonena ze 150 neuronami

Od klasyfikacji wykonanej przez sieć ze 150 neuronami lepsza pod względem stopnia wyjaśniania ogólnej zmienności spożycia jest sieć z 90 neuronami trenowana z uwzględnieniem niewielkiego sąsiedztwa. Jednakże duża liczba zaproponowanych grup zasadniczo utrudnia i rozprasza ich głębszą analizę, należałoby je łączyć, co jednak prowadzi nieuchronnie do utraty siły wyjaśniania (tab. 1).



**Rysunek 2**

Histogram wielkości skupień utworzonych przez sieć Kohonena ze 150 neuronami

Podstawowe charakterystyki skupień gospodarstw domowych wyłonionych przez sieć Kohonena ze 150 neuronami zestawiono w tabeli 2. Poszczególne grupy rozmieszczono według malejącej ich wielkości, przy czym pominięto 5 bardzo małych grup obejmujących łącznie tylko 11 gospodarstw domowych. O tych gospodarstwach domowych powiedzieć można to, że odbiegają w sposób zasadniczy swymi charakterystykami od pozostałych. Wielkości ich dotyczące wzbudzają nawet podejrzenie o błąd albo są to bardzo szczególne przypadki, np. badane w okresie wesela, stąd o bardzo wysokich wielkościach spożycia. Z uwagi na to, że jest ich bardzo niewiele, pomija się je w dalszych prezentacjach. Niemniej jednak warto dostrzec zdolności sieci neuronowej do samodzielnego wyłonienia tego typu przypadków.

Pierwsza z prezentowanych grup jest najliczniejsza, obejmuje 31,5% zbiorowości, druga 20,8%, trzecia 19,3%, czwarta 11%, natomiast grupy od piątej do jedenastej obejmują już tylko po od niewiele ponad jeden do nieco ponad cztery procent, w sumie te ostatnie grupy obejmują 17,4% ogółu badanych. Można więc powiedzieć, że grupy od I do IV to grupy reprezentujące cztery względnie często spotykane style żywieniowe, pośród nich grupa I reprezentuje styl najczęstszy, dotyczący praktycznie co trzeciego gospodarstwa domowego. Pozostałe grupy reprezentują odmienne style, spotykane już rzadziej, a niektóre bardzo rzadko (tab. 2).

**Tabela 2**

Podstawowe charakterystyki grup gospodarstw domowych wg malejących ich wielkości i stosunku korelacji, będące rezultatem klasyfikacji wykonanej przez sieć Kohonena ze 150 neuronami

Wyszczególnienie	Razem	W tym skupienia (grupy gospodarstw domowych):											
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	
Liczebność →	31 756	<b>9 992</b>	6 594	6 140	3 485	1 373	1 121	768	755	652	522	343	
Liczebność w % →	100,0	<b>31,5</b>	20,8	19,3	11,0	4,3	3,5	2,4	2,4	2,1	1,6	1,1	
Produkty spożywcze	Stosunek korelacji	Spożycie na 1 osobę, na miesiąc w kg (jaja w sztukach)											
Tłuszcze roślinne	0,85	1,3	<b>0,6</b>	1,5	1,0	2,5	1,1	1,9	1,7	1,4	1,7	5,0	1,6
Ziemniaki	0,74	9,2	<b>5,7</b>	6,7	8,9	9,2	7,4	11,6	18,8	11,8	8,2	14,5	132,0
Warzywa	0,69	6,5	<b>4,0</b>	4,3	7,2	7,4	8,8	9,1	31,2	7,7	6,9	11,7	12,2
Smalec, słonina	0,67	0,27	<b>0,15</b>	0,16	0,29	0,23	0,10	0,42	0,34	2,53	0,27	0,50	0,40
Ryby	0,65	0,44	<b>0,25</b>	0,36	0,40	0,51	0,59	0,62	0,43	0,54	3,11	0,96	0,41
Owoce	0,65	5,5	<b>3,8</b>	3,5	5,1	5,7	19,4	8,5	11,0	4,7	8,4	11,2	5,5
Jaja	0,64	17,4	<b>10,7</b>	12,9	22,8	20,7	16,6	47,6	23,0	21,4	21,6	36,1	18,1
Mięso i wyroby	0,63	6,0	<b>4,1</b>	4,8	7,5	7,1	5,6	12,2	7,7	10,3	8,5	10,9	7,0
Zbożowe	0,54	8,4	<b>6,3</b>	7,7	8,8	10,4	8,3	13,2	10,8	11,3	10,0	16,6	9,6
Mleko (też w wyrobach)	0,49	19,9	<b>15,1</b>	16,1	22,2	23,6	26,9	33,5	27,9	23,1	23,5	37,8	20,6
Cukier (też w wyrob.)	0,46	2,4	<b>1,6</b>	1,9	2,7	3,2	2,7	4,4	4,5	3,2	2,9	5,9	2,7
Masło	0,45	0,41	<b>0,36</b>	0,22	0,55	0,35	0,57	0,79	0,55	0,37	0,59	0,61	0,46
Strączkowe	0,39	0,11	<b>0,07</b>	0,08	0,13	0,13	0,13	0,16	0,27	0,20	0,18	0,20	0,15
Spożycie razem	0,75	61,4	<b>42,6</b>	48,0	66,1	71,4	82,5	99,1	116,4	78,5	75,5	117,9	193,5

Źródło: Jak w tabeli 1.

Zestawione w analizowanej tabeli przeciętne informacje o spożyciu poszczególnych produktów w poszczególnych grupach pozwalają owe style opisać i ocenić. Na samym początku zauważyć trzeba, że poszczególne grupy jaskrawo odróżnia sumaryczna wielkość spożywanej żywności na 1 osobę na miesiąc: od 42,6 kg w grupie I do 193,5 kg w grupie XI, przy czym ilość spożytej żywności systematycznie wzrasta w kolejnych grupach, z wyjątkiem VIII i IX. Prawdopodobność dotycząca sumy spożytej żywności w dużym zakresie przenosi się na poszczególne produkty, zwłaszcza jeśli analizujemy grupy od I do IV, a więc te, które reprezentują najczęstsze style żywnościowe. W pozostałych grupach obserwujemy już częstsze zmiany proporcji w żywieniu lub pewne upodobania, np. do spożywania w większych ilościach masła zamiast tłuszczów roślinnych czy wyraźnie większe spożycie owoców i warzyw, ryb, smalcu albo ziemniaków.

Poszczególne grupy, które z uwagi na wysoki stopień wyjaśniania zmienności można uznać za reprezentujące występujące w rzeczywistości style żywnościowe, przedstawić można krótko w następujący sposób:

- **grupa I**, najliczniejsza, to osoby spożywające najmniej żywności; obserwujemy w tej grupie najniższe w przekroju wszystkich grup spożycie poszczególnych produktów, z wyjątkiem masła i owoców, których nieznacznie mniejsze spożycie jest tylko w grupie II,
- **grupa II**, tylko nieco wyższe niż w grupie I poziomy spożycia poszczególnych produktów, choć niektórych nieco wyraźniejsze (zbożowych, ryb, tłuszczów roślinnych), z wyjątkiem jednak jeszcze niższego niż w grupie I spożycia owoców i masła, przy czym to drugie jest kompensowane dosyć wysokim spożyciem tłuszczów roślinnych,
- **grupa III**, obserwujemy tutaj już wyraźnie wyższe poziomy spożycia wszystkich produktów, przy czym dotyczy to najbardziej spożycia jaj, smalcu, strączkowych, mięsa i warzyw; wielkości spożycia w tej grupie są najczęściej dosyć zbliżone do wielkości średnich w całej zbiorowości; warto ponadto zwrócić uwagę na umiarkowane spożycie tłuszczów roślinnych w tej grupie przy relatywnie dosyć wysokim poziomie spożycia masła, co jest odwrotnością sytuacji w II grupie,
- **grupa IV**, jest już stosunkowo mniej liczna, niemniej ponad 10% udziału świadczy o jeszcze znacznej popularności reprezentowanego przez nią stylu; wyróżnia go jeszcze wyższe niż we wszystkich poprzednich grupach spożycie żywności, na co składa się przede wszystkim wysokie spożycie tłuszczów roślinnych, a po nich wyrobów z dużą zawartością cukru, a także ryb, jaj i warzyw; ponadto, obserwujemy tutaj niskie spożycie masła przy bardzo wysokim poziomie spożycia tłuszczów roślinnych – można to zaliczyć do specyfiki tej grupy,



- **pozostałe grupy, czyli skupienia od V do XI**, są już wyraźnie mniejsze, niemniej jednak wyraźnie dystansują się od siebie i pozostałych, są grupami osób o wysokim odnotowanym poziomie spożywania żywności, przy czym w grupie V mamy osoby spożywające szczególnie dużo owoców, a także ryb; w grupie VI mamy osoby szczególnie obficie jedzące zbożowe, mięso, mleko, jaja, masło i cukier; w grupie VII mamy z kolei osoby spożywające najwięcej warzyw i strączkowych, a także dużo owoców, ziemniaków i cukru (są to więc osoby jedzące obficie, ale starające się uwzględniać zalecenia); w grupie VIII mamy osoby spożywające najwięcej smalcu oraz dużo strączkowych i mięsa; w grupie IX spotykamy osoby spożywające najwięcej ryb, przy jednocześnie sporym spożyciu owoców; w X grupie mamy osoby, które jedzą wyjątkowo obficie – podobnie jak w grupie VII – ale ze szczególną koncentracją na zbożowych, mleku, tłuszczach i cukrze; wreszcie XI grupę charakteryzują poziomy spożycia podobne do obserwowanych w grupie III, z wyjątkiem ziemniaków i warzyw, których wypada tu na 1 osobę bardzo dużo (można podejrzewać, że są to osoby badane w miesiącu gromadzenia zapasów, zwłaszcza ziemniaków).

Uzyskanie w miarę jednorodnych grup gospodarstw domowych pod względem spożycia żywności jest korzystne z punktu widzenia prowadzenia dalszej oceny, a przede wszystkim oceny zdrowotnej rejestrowanych u nich racji pokarmowych. Można bowiem stosunkowo precyzyjnie określać między innymi najbardziej niedoborowe w danej grupie składniki odżywcze i wskazywać negatywne tego następstwa. Wyniki takiej analizy mogą mieć duże znaczenie w edukacji żywieniowej i profilaktyce zdrowotnej. Ramy niniejszego artykułu nie pozwalają takiej oceny kontynuować.

## Podsumowanie

Sieci neuronowe stanowią ciekawe i frapujące narzędzia badawcze. Mogą być bardzo przydatne w rozpoznawaniu danych, ich klasyfikacji, a także – jak wskazuje literatura – w wielu innych zastosowaniach. Nie wymagają rzeczywiście od badacza głębokiej wiedzy statystycznej czy ekonometrycznej, ale by rezultaty ich zastosowania były wiarygodne, trzeba mieć pewną wprawę w posługiwaniu się nimi. Teoria – jak na razie – nie dostarcza wystarczających reguł na potrzeby warsztatu badawczego.

Sieci neuronowe Kohonena otwierają drogę do klasyfikacji optymalnych, gdyż dają szansę na podziały przypadków przy bardzo wysokich współczynnikach wyjaśniania ogólnej zmienności uwzględnianych w modelu cech. W ten

sposób zwracają uwagę na ograniczoność poznawczą klasycznych podziałów. Analiza wyżej prezentowana wykazała, że możliwe jest uzyskanie tą drogą klasyfikacji gospodarstw domowych przy korelacji bardzo wysokiej i jednocześnie znacznie przewyższającej korelacje uzyskiwane przy typowych podziałach.

Warto ponadto zauważyć zdolności sieci neuronowych Kohonena do wyłaniania przypadków nietypowych, bardzo odbiegających. Może to mieć duże znaczenie w odnajdywaniu błędów lub eliminacji przypadków szczególnych, incydentalnych.

Przedstawiona analiza pozwoliła wyłonić grupy gospodarstw domowych, za którymi kryją się określone i zasadniczo odmienne style żywnościowe, które można poddać dalszej ocenie, a w tym ocenie zdrowotnej.

## Literatura

1. HERTZ J., KROGH A., PALMER R., 1995: *Wstęp do teorii obliczeń neuronowych*, Wydawnictwa Naukowo-Techniczne, Warszawa.
2. KOWRYGO B.. 2000: *Studium wpływu gospodarki rynkowej na sferę żywności i żywienia w Polsce*, Wydawnictwo SGGW, Warszawa.
3. LASKOWSKI W., 2000: Sieci neuronowe w analizach statystycznych; *Wiadomości Statystyczne* 7, GUS, Warszawa.
4. OSOWSKI S., 1996: *Sieci neuronowe*, Wydawnictwa Naukowo-Techniczne, Warszawa.
5. TADEUSIEWICZ R., 1993: *Sieci neuronowe*, Akademicka Oficyna Wydawnicza RM, Warszawa.

## **Quantitative analysis of food consumption by Polish families with the utilization of their classification with the use of neural networks**

### **Abstract**

Data from surveys carried out by GUS permit to state that there is large individual variability in consumption of food by the Polish population. However, traditional partitioning of households, for example, according to income or dwelling place does not fully explain this variability. It was assumed that neural networks as a new research tools can shade more light is respect to consumption of food in concentrated households.

The use of neural networks of the Kohonen type enabled the separation of 4 basic concentrations of households differing in the level of food consumption as well as in the contribution of various products. This partitioning explained at a high level the variability in food consumption (the average correlations coefficient of groups of products examined reached the level of 0.6).

Moreover this partitioning gave the opportunity to further analyze relatively homogenous households that clarity differ with each other.