

Sieci neuronowe i ich wykorzystanie w analizie danych ekonomicznych na przykładzie prognozowania sprzedaży energii elektrycznej

1. Wstęp

Sztuczne sieci neuronowe (nazywane prościej sieciami neuronowymi, w skrócie SN) stanowią współcześnie dynamicznie rozwijającą się dziedzinę metod analizy danych, w tym również danych ekonomicznych. Badania nad konstrukcją i zastosowaniem sieci neuronowych w procesach szeroko pojętego przetwarzania informacji zostały zainspirowane przez koncepcje naśladowania funkcjonowania naturalnych struktur komórek nerwowych zawartych w mózgu, przy jednoczesnym wykorzystaniu dużej szybkości i mocy obliczeniowej komputera. W istocie sposoby analizy informacji przez sztuczną sieć neuronową są bardzo podobne do procesów zachodzących w rzeczywistych mózgowych strukturach neuronów, jednak elementarne operacje przetwarzania sygnałów mogą odbywać się w układach elektronicznych komputera z dużo większą szybkością.

Generalnie istota efektywnego funkcjonowania SN polega na jej wcześniejszym „nauczeniu”, przy wykorzystaniu odpowiedniego zbioru danych związanych z analizowanym zagadnieniem, zawartych w tzw. ciągu uczącym. Wiedza (np. o określonych zależnościach występujących na rynku finansowym) nie jest tu przekazywana sieci *a priori* (jak to ma miejsce np. w przypadku konstrukcji systemów ekspertowych), ale w trakcie iteracyjnego procesu uczenia, polegającego na wielokrotnej prezentacji poszczególnych wzorców wchodzących w skład ciągu uczącego. W wyniku tego procesu następuje „dostrojanie” dużej liczby adaptowalnych parametrów sieci — wag połączeń pomiędzy jej poszczególnymi komórkami — w taki sposób, że sieć potrafi prawidłowo reagować na wzorce, których się „nauczyła”, oraz — co więcej — na inne wzorce, które nie występowały podczas fazy uczenia. Zwrócić należy tu uwagę na istotne podobieństwo tego procesu do uczenia się i gromadzenia doświadczeń przez człowieka, a następnie wpływu nabytej wiedzy na podejmowane decyzje.

Poprawnie skonstruowana i nauczona sieć, przechowująca określoną wiedzę w postaci odpowiednio dostrojonych współczynników wag połączeń neuronowych, może być następnie wykorzystywana do generowania właściwych prognoz lub do sugerowania poprawnych reakcji na aktualne dane pochodzące z analizowanego systemu (np. rynku finansowego).

2. Rys historyczny i aktualne kierunki badań nad sieciami neuronowymi

Pionierskie prace nad poznaniem zasad funkcjonowania mózgu prowadził Ramón y Cajál [1911], który zaprezentował koncepcję neuronów jako elementów strukturalnych mózgu, a następnie inni pionierzy neurofizjologii i bioniki (zob. [R. Tadeusiewicz 1994]). Jednak właściwy początek istnienia dziedziny sztucznych sieci neuronowych nastąpił wraz z opublikowaniem historycznej pracy McCullocha i Pittsa [W. S. McCulloch, W. Pitts 1943], w której autorzy przedstawili po raz pierwszy matematyczny model neuronu, stosowany (z niewielkimi modyfikacjami) do dziś. Kolejne fundamentalne opracowania dotyczące tej problematyki stanowią książki Hebba [D. O. Hebb 1949], von Neumana [J. Neuman 1958], oraz Taylora [W. K. Taylor 1960], omawiające zagadnienie funkcjonowania sieci neuronowych z trzech różnych punktów widzenia.

Pierwszą sprzętową implementację sieci neuronowej (tzw. neurokomputer) stanowiło urządzenie o nazwie Mark I Perceptron, zbudowane w roku 1957 przez Rosenblatta i Wightmana. Układ ten, zawierający 512 losowo zadanych połączeń pomiędzy neuronami, służył do rozpoznawania znaków przedstawianych na macierzy o wymiarach 20×20 pikseli. Kolejną interesującą konstrukcją neurokomputera była zbudowana przez Widrowa w roku 1960 sieć o nazwie Madaline, służąca do adaptacyjnego przetwarzania sygnałów.

Zarówno układy sprzętowe, jak i pojawiające się coraz liczniej prace teoretyczne zmierzające do skonstruowania modeli biologicznych struktur neuronowych, zaowocowały w latach sześćdziesiątych intensywnym postępowaniem badań nad problematyką sieci neuronowych. Rozwój ten został jednak istotnie zahamowany na około 15 lat wskutek opublikowania przez Minsky'ego i Paperta książki *Perceptrons*, w której autorzy dowodzili bardzo ograniczonej przydatności sieci jednowarstwowych. W okresie tym prowadzono jednak szereg prac w tej dziedzinie, spośród których na uwagę zasługują m.in. prace Grossberga, Kohonena, Werbosa, Andersona, Fukushima i Hopfielda. W roku 1974 Werbos opracował po raz pierwszy skuteczny algorytm uczenia wielowarstwowych nieliniowych sieci neuronowych [P. J. Werbos 1974] — algorytm wstecznej propagacji błędu (*backpropagation*) — stanowiący obecnie podstawową metodę uczenia tego typu sieci. Ponowne niezależne odkrycie tego algorytmu w połowie lat osiemdziesiątych przez Rumelharta, Hintona i Williama przyczyniło się do przełamania istniejącego impasu i gwałtownego rozwoju badań nad konstrukcją, metodami uczenia i zastosowaniami różnych typów sieci neuronowych (zob. [R. Tadeusiewicz 1993]).

Obecnie obserwuje się olbrzymie zainteresowanie problematyką sieci neuronowych i ich zastosowań, owocujące lawinowo rosnącą liczbą książek oraz publikacji w licznych periodykach poświęconych tej tematyce (należą do nich m.in. *Neural Networks*, *IEEE Transactions on Neural Networks*, *Neural Computing & Ap-*

plications, Neurocomputing, Engineering Application of Artificial Intelligence, i inne), a także w różnych wydawnictwach związanych z pokrewnymi, a często odległymi dziedzinami wiedzy. Organizowana jest też na całym świecie ogromna liczba konferencji naukowych dotyczących tej problematyki. Na uwagę zasługuje również powstanie dużej ilości firm oferujących pakiety oprogramowania służące do komputerowej implementacji sieci neuronowych pod kątem ogólnym, bądź do specjalistycznych zastosowań (m.in. finansowych), a także posiadających w swojej ofercie rozwiązania sprzętowe (neurokomputery) (zob. [R. Tadeusiewicz 1993; D. Rutkowska i in. 1996; *Selecting...* 1996]).

Badania nad praktycznymi zastosowaniami sieci neuronowych obejmują m.in. następujące dziedziny [R. Tadeusiewicz 1993; S. Osowski 1996]:

— Szeroko pojęte rozpoznawanie obrazów, prowadzące się do wyłowienia istotnych cech danego obiektu lub sygnału i przydzielenia go do odpowiedniej klasy. Przykłady stanowią tu m.in. rozpoznawanie i analiza obrazów wizyjnych, rozpoznawanie dźwięków (np. mowy lub sygnałów sonarowych), rozpoznawanie ręcznie pisanych znaków (np. kodów pocztowych), weryfikacja podpisów, wykrywanie podejrzanych obiektów w bagażach lotniczych i inne.

— Klasyfikację bezwzorcową, prowadzącą się do określenia struktury grupowej obiektów w przestrzeni cech (zob. też [T. Kohonen 1995; M. Grabowski 1997]).

— Sterowanie obiektów dynamicznych (np. automatów, robotów, procesów przemysłowych).

— Szeroko pojęte prognozowanie, głównie w zastosowaniach technicznych lub ekonomicznych (np. predykcja obciążeń systemu elektroenergetycznego, prognozy kursów walut, itp.).

— Przetwarzanie sygnałów, obejmujące m.in. filtrację sygnałów, kompresję, transformację, kodowanie i dekodowanie informacji, redukcję danych (np. analizę głównych składowych).

— Badania dotyczące nowych rozwiązań pamięci komputerowych (pamięci asocjacyjne, pamięci rozproszone), a także algorytmów automatycznego uczenia się maszyn.

Większość z powyższych kierunków badań zaowocowała również efektywnymi zastosowaniami sieci neuronowych na gruncie ekonomii i zarządzania (zob. np. [R. Tadeusiewicz 1993; R. Tadeusiewicz 1995; A. P. Refenes 1995; J. Morajda 1997; P. Lula 1999; J. Morajda 2000; J. Morajda 2003]). Przykładowe zagadnienia, w których wykorzystywane są te narzędzia, to:

- ocena zdolności kredytowej podmiotów gospodarczych,
- prognozowanie stóp procentowych i kursów walut,
- prognozowanie wielkości sprzedaży w przedsiębiorstwie,
- prognozy ekonomiczne rozwoju przedsiębiorstw, kondycji finansowej banków, itp.
- prognozowanie dynamiki indeksów giełdowych,

- selekcja papierów wartościowych w portfelu inwestycyjnym,
- wybór potencjalnych klientów w zagadnieniu marketingu bezpośredniego,
- modelowanie zjawisk finansowych w przedsiębiorstwie (np. analiza kosztów, zmiany należności na rachunkach, inwestycje kapitałowe, itp.),
- zarządzanie stanem magazynu przedsiębiorstwa,
- prognozy rynków (np. pieniężnych, długów, metali szlachetnych itd.),
- klasyfikacja klientów i kontrahentów firmy,
- wspomaganie kontroli podatkowej,
- kojarzenie danych i automatyzacja procesów wnioskowania w zarządzaniu,
- filtracja sygnałów i uzupełnianie danych niekompletnych,
- optymalizacja decyzji gospodarczych.

3. Ogólna charakterystyka sieci neuronowych

Sztuczne sieci neuronowe, jako modele rzeczywistych struktur komórek nerwowych, posiadają szereg istotnych cech charakterystycznych m.in. dla sposobu przetwarzania informacji przez biologiczny mózg. Cechy te niejednokrotnie decydują o efektywności wykorzystania sieci neuronowych w konkretnych zastosowaniach, a także o ogólnych właściwościach metod opartych na tych narzędziach. Najistotniejsze z nich zostały wymienione poniżej [zob. S. Haykin 1994; R. Tadeusiewicz 1993].

Zdolność do nabywania wiedzy poprzez uczenie. W przeciwieństwie do klasycznych modeli komputerowych, których budowa opiera się na algorytmie realizującym określoną metodę rozwiązania danego problemu, sieć neuronowa nie wymaga wcześniejszego zdefiniowania sposobu przetwarzania przez nią informacji, nie wymaga zatem procesu programowania, wykorzystującego wiedzę aprioryczną. Proces ten zastępowany jest przez uczenie sieci, w trakcie którego sieć „zdobywa” wiedzę na podstawie prezentowanych jej danych uczących. Fakt ten pozwala na efektywne zastosowanie SN także w problemach, gdzie posiadamy niewielką wiedzę na temat analizowanego zagadnienia, albo też w sytuacjach, gdy stworzenie stosownego modelu komputerowego w postaci programu byłoby bardzo trudne, natomiast dysponujemy odpowiednią ilością danych mogących posłużyć do uczenia sieci.

Równoległy charakter przetwarzania informacji. Informacja w SN jest przetwarzana jednocześnie we wszystkich komórkach sieci¹, co zapewnia równoległość obliczeń i w konsekwencji daje możliwość znacznego przyspieszenia procesu przetwarzania informacji; fakt ten jest istotny zwłaszcza w fazie uczenia sieci².

¹ Oczywiście w implementacjach sieci na komputerach sekwencyjnych równoległość ta jest symulowana i nie przynosi widocznych korzyści.

² Interesujące są tu porównania szybkości pracy maszyny cyfrowej i mózgu człowieka: superkomputery pracują z szybkością rzędu 10^{11} operacji na sekundę, neurony zaś mogą przyjmować i wysyłać impulsy o częstotliwości zaledwie do 1000 Hz, jednak szybkość pracy całej kory mózgo-

Zdolność do generalizacji. Prawidłowo skonstruowana i nauczona sieć neuronowa potrafi dawać rozsądne odpowiedzi nie tylko na sygnały wejściowe wchodzące w skład ciągu uczącego, ale również na inne wzorce wejściowe, które nie były sieci wcześniej pokazywane (a należy zakładać, że w większości praktycznych zastosowań z takimi sygnałami sieć będzie miała do czynienia w fazie realizacji określonego zadania, np. predykcji). Jednak uzyskanie tej zdolności wymaga należytej staranności na etapie budowy i uczenia sieci (dobór architektury sieci, przeprowadzenie niezależnej walidacji).

Nieliniowość³ i nieparametryczność. Przyjęcie nieliniowego modelu neuronu prowadzi do uzyskania nieliniowej zależności między wejściem a wyjściem dla całej sieci. Przy odpowiedniej architekturze sieci może ona zrealizować dowolne odwzorowanie wiążące wartości wejściowe i wyjściowe sieci⁴. Metodę opartą na sieciach neuronowych uznaje się również za metodę nieparametryczną w sensie braku konieczności przyjmowania *a priori* postaci funkcyjnej modelu (zob. też [A. P. Refenes 1995; M. Grabowski 1997]). Cechy te mają istotne znaczenie w aspekcie konieczności modelowania zjawisk nieliniowych, występujących m.in. na rynkach finansowych. Jakkolwiek dla tego typu problemów mogą być wykorzystane inne nieliniowe techniki, to SN ze względu na łatwość implementacji i możliwość realizacji dowolnych odwzorowań wejście — wyjście mogą okazać się narzędziem najbardziej efektywnym (zob. [E. M. Azoff 1994]).

Zdolność do adaptacji. Wagi połączeń w nauczonej sieci neuronowej realizującej określony model, jako parametry sieci podlegające modyfikacji w procesie uczenia, mogą podlegać dalszym adaptacyjnym zmianom w odpowiedzi na zmieniające się warunki zewnętrzne środowiska. W praktyce może to być realizowane w procesie „douczenia” sieci, polegającym na kontynuacji uczenia przy zastosowaniu zmodyfikowanego ciągu uczącego zawierającego „bardziej aktualne” wzorce. Tego typu adaptacja wag sieci może mieć istotne znaczenie w przypadku, gdy sieć ma za zadanie realizować model niestacjonarny, przystosowujący się do zmian zachodzących w systemie rzeczywistym. Powstaje jednak w tym przypadku dylemat określenia wielkości ciągu uczącego i ustalenia kryterium „aktualności” danych⁵.

Spśród innych właściwości SN, eksponowanych przez wielu badaczy i mających istotne znaczenie dla praktycznych implementacji, wymienić można także:

wej, zawierającej ok. 10^{10} neuronów i ok. 10^{15} połączeń między nimi, ocenia się na 10^{18} operacji na sekundę [R. Tadeusiewicz 1993].

³ Pomijamy tu klasę liniowych SN, mających dość ograniczone pole zastosowań.

⁴ Dotyczy to tylko sieci wielowarstwowych, podstawę matematyczną stanowi tu twierdzenie Kołmogorowa (zob. np. [Hecht-Nielsen 1990]).

⁵ Dylemat ten jest dobrze widoczny w zagadnieniu analizy szeregów czasowych w przypadku użycia do uczenia i walidacji sieci danych pochodzących z ostatniego okresu: okres ten powinien być na tyle długi, aby zapewnić właściwą zdolność sieci do generalizacji i uchronić ją przed uczeniem się szumu, z drugiej zaś strony na tyle krótki, aby sieć potrafiła wychwycić i zaadaptować się do istotnych zmian w systemie rzeczywistym.

- uniwersalność modeli opartych na SN w sensie jednolitości metod konstrukcji, uczenia i analizy SN dla różnych zastosowań [S. Haykin 1994],
- odporność SN na uszkodzenia struktury⁶,
- możliwość implementacji SN w postaci układów elektronicznych VLSI.

Praktycznie wszystkie SN, będące przedmiotem badań i zastosowań praktycznych, mają budowę warstwową. Oznacza to, że w strukturze sieci można wyróżnić pewną liczbę warstw, kolejno przetwarzających informację, z których każda zawiera określoną liczbę neuronów. Z reguły sposób funkcjonowania neuronów tej samej warstwy jest identyczny. W każdej sieci możemy wyodrębnić warstwę wejściową (zwaną również warstwą zerową), której zadaniem jest jedynie dystrybuowanie w sieci sygnałów wejściowych (nie przetwarza ona informacji), oraz warstwę wyjściową generującą sygnały wyjściowe sieci. Warstwy pośrednie (jeżeli występują), położone pomiędzy warstwą wejściową a wyjściową, nazywamy warstwami ukrytymi (nie wymieniają one informacji bezpośrednio z otoczeniem sieci).

Klasyfikacji sieci neuronowych można dokonywać według różnych kryteriów. Poniżej wymieniono najistotniejsze metody podziału SN (szczegółową analizę można znaleźć w literaturze, np. [R. Tadeusiewicz 1993; J. Hertz i in. 1993; S. Haykin 1994]).

Podział SN według charakteru połączeń w sieci:

- sieci z połączeniami jednokierunkowymi⁷ (*feedforward*),
- sieci ze sprzężeniami zwrotnymi⁸.

Podział według ilości warstw:

- sieci jednowarstwowe⁹,
- sieci wielowarstwowe (zawierające co najmniej jedną warstwę ukrytą).

Podział według charakteru realizowanego odwzorowania wejście — wyjście:

- sieci liniowe¹⁰,

⁶ Cecha ta ma niebagatelne znaczenie w biologicznych systemach nerwowych, np. mózg człowieka tracąc codziennie od kilku do kilkudziesięciu tysięcy neuronów nie traci praktycznie swoich funkcjonalnych zdolności; właściwość ta została także zauważona już w okresie początkowych badań nad SN, np. perceptron Rosenblatta zachowywał zdolność rozpoznawania znaków nawet po przerwaniu pewnej części połączeń pomiędzy neuronami; wydaje się jednak, że cecha ta ma obecnie mniejsze znaczenie w praktyce [R. Tadeusiewicz 1993].

⁷ W takich sieciach przepływ informacji odbywa się tylko w jednym kierunku, tj. od warstwy wejściowej do wyjściowej, poprzez warstwy pośrednie (ukryte), a zatem sygnały wyjściowe neuronów danej warstwy są przekazywane tylko do warstw następnych. W tego typu sieciach nie występują zatem sprzężenia zwrotne.

⁸ Najczęściej analizowanymi SN ze sprzężeniami zwrotnymi są tzw. sieci Hopfielda, których nazwa pochodzi od nazwiska badacza, który zaproponował tego typu struktury.

⁹ Sieci jednowarstwowe zawierają warstwę wejściową i warstwę wyjściową, jednak ponieważ przy określaniu liczby warstw sieci nie uwzględniamy warstwy wejściowej jako nie przetwarzającej informacji (zob. np. [S. Haykin 1994, s. 18]), tak więc według poprawnej terminologii sieć dwuwarstwowa to sieć z jedną warstwą ukrytą.

¹⁰ Inne nazwy tych sieci to ADALINE i MADALINE. Zawierają one tylko neurony posiadające

— sieci nieliniowe.

Podział według metody uczenia sieci:

— sieci uczone w sposób nadzorowany (tzw. uczenie z nauczycielem),

— sieci samouczące się (uczenie bez nauczyciela, tj. w sposób nienadzorowany).

Jednym z ważniejszych i częściej stosowanych rodzajów SN są sieci typu perceptron. Ten rodzaj sieci jest wysoce efektywny w różnych zastosowaniach praktycznych i dlatego jest najczęściej używany w zagadnieniach prognozowania szeregów czasowych [E. M. Azoff 1994].

4. Struktura i funkcjonowanie sieci neuronowej typu perceptron

Sieci neuronowe typu perceptron stanowią zasadniczą klasę SN uczonych w sposób nadzorowany. Są to sieci o strukturze warstwowej, przy czym praktyczne znaczenie mają w zasadzie tylko perceptrony wielowarstwowe, tzn. posiadające co najmniej jedną (i zwykle tylko jedną) warstwę ukrytą. Charakteryzują się one połączeniami jednokierunkowymi typu „każdy z każdym” (każdy neuron kolejnej warstwy jest połączony z wszystkimi neuronami warstwy poprzedniej, por. rysunek 1). Neurony warstw ukrytych i warstwy wyjściowej oparte są na klasycznym modelu McCullocha-Pittsa i mogą posiadać liniową lub nieliniową funkcję aktywacji [R. Tadeusiewicz 1993; S. Haykin 1994; S. Osowski 1996]. W analizie złożonych problemów nieliniowych zastosowanie znajdują jedynie perceptrony nieliniowe, tzn. złożone z neuronów o nieliniowych funkcjach aktywacji.

Na rysunku 1 zaprezentowano przykładowy schemat wielowarstwowej sieci neuronowej typu perceptron. Sieć w tym przykładzie posiada jedną warstwę ukrytą, oraz dwuelementową warstwę wyjściową. Neurony określonej warstwy odbierają sygnały tylko z warstwy poprzedniej.

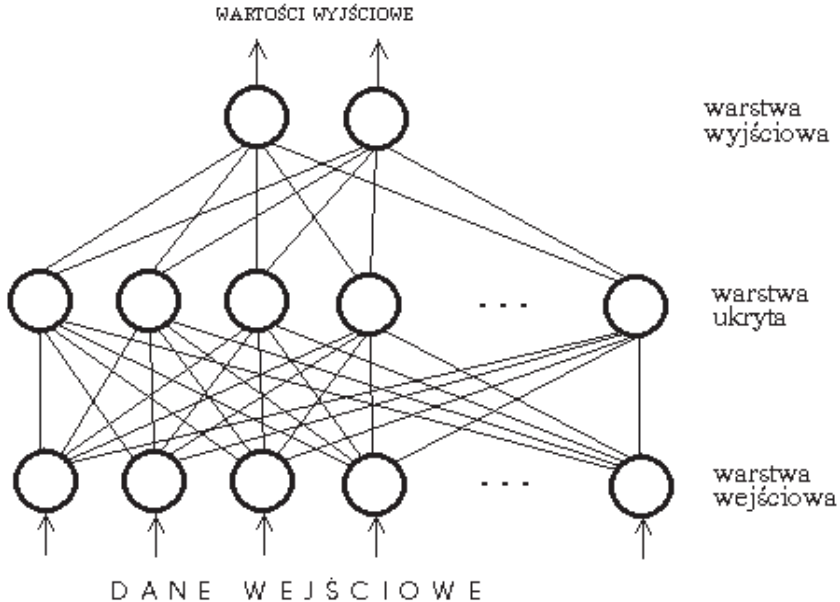
Liczba neuronów w warstwach wejściowej i wyjściowej jest ściśle zdeterminowana odpowiednio przez liczbę zmiennych objaśniających i objaśnianych w analizowanym problemie. Wielkość i ilość warstw ukrytych jest na ogół arbitralnie dobierana przez projektanta sieci w oparciu o pewne przesłanki wpływające zarówno z natury problemu, jak i z pewnych ogólnych wskazań dostępnych w literaturze; często jednak niezbędny jest dobór tych parametrów na drodze eksperymentalnej. Ogólnie liczba neuronów ukrytych nie może być zbyt mała (aby sieć była w stanie zidentyfikować i zapamiętać relacje pomiędzy zmiennymi, występujące w badanym zjawisku). Zbyt duża liczba neuronów ukrytych może z kolei prowadzić do podatności na tzw. efekt przeuczenia (utrata zdolności do generalizacji wiedzy).

Wykorzystywany w warstwach ukrytych i wyjściowej model neuronu bazuje na sposobie przetwarzania informacji w rzeczywistych biologicznych komórkach

liniową funkcję przejścia i wobec tego realizują liniowe odwzorowanie wejście — wyjście. Wyczerpujące omówienie tego typu sieci przedstawione jest w [R. Tadeusiewicz 1993].

Rysunek 1

Przykładowa architektura warstwowej sieci neuronowej typu perceptron z jedną warstwą ukrytą. Każdy neuron warstwy ukrytej lub wyjściowej jest połączony z wszystkimi neuronami należącymi do warstwy poprzedniej



Źródło: opracowanie własne.

nerwowych (oczywiście po przyjęciu wielu uproszczeń). Model taki można zapisać w formie następującej zależności:

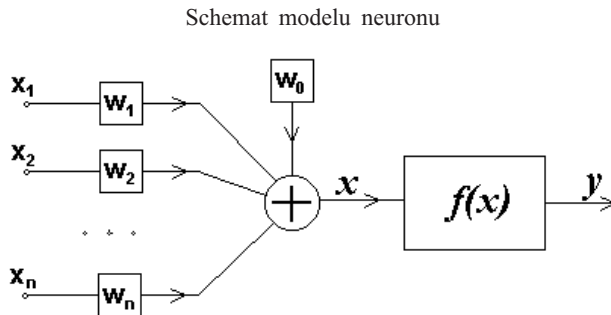
$$y = f\left(\sum_{i=0}^n x_i w_i\right) \quad (1)$$

gdzie:

- y — wartość wyjściowa neuronu,
- n — liczba wejść,
- x_1, x_2, \dots, x_n — wartości sygnałów wejściowych dla neuronu,
- $x_0 = \text{const.} = 1$
- w_1, w_2, \dots, w_n — wartości wag połączeń wejściowych dla danego neuronu, określające znaczenie poszczególnych wejść,
- w_0 — składnik stały (ang. *bias*), reprezentujący wartość progową (z przeciwnym znakiem),
- f — funkcja określająca zależność wyjścia od ważonej sumy wejść, zwana funkcją przejścia lub funkcją aktywacji (ang. *transfer function*).

Graficzny schemat tak skonstruowanego modelu prezentuje rysunek 2.

Rysunek 2



Źródło: opracowanie własne na podstawie [S. Haykin 1994].

Poszczególne wersje modelu neuronu, stosowane w SN typu perceptron, różnią się przede wszystkim doбором funkcji przejścia f . Funkcja ta w dużym stopniu decyduje o właściwościach pojedynczego neuronu, a tym samym całej sieci. Do najczęściej stosowanych funkcji należy funkcja progowa, progowo-liniowa, sigmoidalna, oraz tangens hiperboliczny [R. Tadeusiewicz 1993; S. Haykin 1994, S. Osowski 1996; D. Rutkowska i in. 1997].

Proces uczenia sieci neuronowej perceptron wykorzystuje zbiór przykładów — wzorców — stanowiących tzw. ciąg (zbiór) uczący. Każdy z tych przykładów zawiera wektor wartości wejściowych dla sieci, oraz poprawny (oczekiwany na wyjściu) wektor wzorcowych sygnałów wyjściowych. Sam proces uczenia sprowadza się do wyznaczenia takiego zbioru parametrów połączeń (wag) sieci W_0 , dla którego określona w przestrzeni wag tzw. funkcja błędu $F(W)$ (ang. *error function*, *cost function*) osiąga minimum. Wartość funkcji błędu F dla określonego zestawu wag W stanowi sumaryczną miarę błędów sieci, czyli różnic pomiędzy wartością wzorcową, a faktyczną wartością wyjściową wygenerowaną przez sieć. Wartość funkcji błędu zazwyczaj określa się i kumuluje dla wszystkich elementów ciągu uczącego. Funkcja ta jest najczęściej definiowana za pomocą zależności¹¹ [S. Haykin 1994; E. M. Azoff 1994]:

$$F = \frac{1}{2} \sum_i (y_i - z_i)^2 \quad (2)$$

gdzie:

- z_i — wartość wyjściowa (wzorcowa) i -tego elementu ciągu uczącego,
- y_i — otrzymana faktyczna wartość wyjścia SN dla i -tego elementu ciągu uczącego.

¹¹ Zakładamy tutaj, że sieć posiada jeden element wyjściowy; gdyby neuronów wyjściowych było więcej, należałoby w celu znalezienia wartości funkcji błędu przeprowadzić także skalanie — na przykład sumowanie błędów obliczanych dla poszczególnych wyjść.

Klasyką i jednocześnie najpowszechniej używaną metodą nadzorowanego uczenia sieci neuronowych (minimalizacji funkcji $F(W)$) dla SN typu *perceptron* jest algorytm wstecznej propagacji błędów (ang. *backpropagation*) [P. J. Werbos 1974; R. Tadeusiewicz 1993; S. Haykin 1994; S. Osowski 1996]. Idea tego algorytmu sprowadza się do iteracyjnej¹² modyfikacji wektora wag sieci W w oparciu o gradient funkcji F , po rozpoczęciu tego procesu od (z reguły) losowo określonego wektora początkowego $W^{(1)}$. W danej i -tej iteracji (oznaczymy przez $W^{(i)}$ wartość wektora wag podczas tej iteracji) najpierw określony zostaje gradient $\nabla F(W^{(i)})$ funkcji błędu F w punkcie $W^{(i)}$ przestrzeni wag. Z kolei wyznaczony zostaje wektor korekty wag (tzn. wektor $W^{(i+1)} - W^{(i)}$), którego podstawowa składowa posiada kierunek największego spadku funkcji F w punkcie $W^{(i)}$ (przeciwny do gradientu), a druga składowa jest zgodna z wypadkowym kierunkiem zmiany wag w całej poprzedniej epoce uczenia (reprezentuje ona swoistą „bezwładność” procesu uczenia). Bardzo ważną rolę w procesie modyfikacji wag odgrywają dwa parametry¹³: współczynnik uczenia η , decydujący o wielkości kroku korekty wag w kierunku największego spadku funkcji F , oraz współczynnik bezwładności (*momentum*) μ , określający długość składowej wektora korekty wag równoległej do kierunku zmiany wag dokonanej w poprzedniej epoce.

Oprócz metody wstecznej propagacji opracowano szereg innych metod uczenia SN, zarówno gradientowych (np. algorytm zmiennej metryki, algorytm gradientów sprzężonych) [zob. np. S. Osowski 1996], jak i bezgradientowych (np. metoda wektora losowego).

Funkcjonowanie nauczonej już sieci polega na podaniu na wejścia sieci zestawu wartości wejściowych (cech) opisujących analizowany obiekt (wzorec) i zaobserwowaniu oraz zinterpretowaniu sygnału wyjściowego wygenerowanego przez sieć (może to być np. prognoza określonej wielkości ekonomicznej, np. kursu akcji, lub sygnał wskazujący przynależność rozważanego wzorca do określonej klasy).

5. Zastosowanie sieci neuronowych w procesie prognozowania sprzedaży energii elektrycznej

W niniejszej części opracowania zamieszczono przykład zastosowania sieci neuronowych typu perceptron w ważnym z praktycznego punktu widzenia problemie ekonomicznym dotyczącym prognozowania krótkoterminowego zapotrzebowania na energię elektryczną. Prognozy takie mają fundamentalne znaczenie dla spółek zajmujących się dystrybucją energii elektrycznej (zakładów energetycz-

¹² Iteracja polega na jednokrotnej prezentacji pojedynczego elementu ciągu uczącego oraz na dokonaniu odpowiedniej korekty wag na podstawie tej prezentacji. Sumę iteracji po wszystkich elementach ciągu uczącego (tzn. pojedynczą prezentację całego ciągu uczącego) nazywamy epoką uczenia.

¹³ Parametry te są na ogół odpowiednio modyfikowane w trakcie procesu uczenia.

nych), gdyż są one wykorzystywane do planowania harmonogramów i wielkości przesyłów mocy w systemie energetycznym [W. Bartkiewicz 1998; J. S. Zieliński 2000]. Ponadto zakłady energetyczne, funkcjonując na wolnym rynku energii, zawierają kontrakty na dostawy określonej ilości energii, a odchylenia od zamówień (powstałe np. wskutek niedokładnych prognoz) obciążone są wysokimi karami umownymi (nie istnieją zaś możliwości magazynowania energii).

W badaniach wykorzystano dane historyczne pochodzące z jednej ze spółek dystrybucyjnych z okresu 6 lat: od 1 stycznia 1996 do 1 stycznia 2002. Ogólnie zbiór danych zawierał 2196 rekordów charakteryzujących dzienne zapotrzebowania na energię. Dane z lat 1996—2000 wykorzystano w procesie uczenia i walidacji (bieżącej weryfikacji zapobiegającej przeuczeniu) sieci, natomiast 365 rekordów dotyczących roku 2001 zastosowano w celu testowania (końcowej oceny) sieci.

Zastosowano 27 zmiennych wejściowych sieci, obejmujących:

- 24 zapotrzebowania godzinne na energię, w poszczególnych godzinach w danym dniu,
- prognozy minimalnej i maksymalnej temperatury dla następnego dnia,
- binarną informację wskazującą czy dzień następny jest niedzielą.

Zastosowana zmienna wyjściowa dotyczy dobowego zapotrzebowania na energię elektryczną w dniu następnym.

W modelu prognostycznym wykorzystano sieć neuronową typu nieliniowy perceptron wielowarstwowy, z jedną warstwą ukrytą zawierającą 15 neuronów (tzn. model neuronowy o wielkości poszczególnych warstw: 27—15—1). Zastosowano funkcję tangens hiperboliczny jako funkcję aktywacji neuronów warstwy ukrytej. Sieć uczono przy użyciu algorytmu gradientów sprzężonych. W celach porównawczych skonstruowano również dla tego zagadnienia drugi model oparty na klasycznej wielorakiej regresji liniowej¹⁴.

Tabela 1

Miary błędów dla zbioru testowego

	Model neuronowy	Regresja liniowa
Pierwiastek z błędu średniokwadratowego RMSE	138594,8	392035,0
Błąd średni MAE	104602,0	331450,2

Źródło: obliczenia własne.

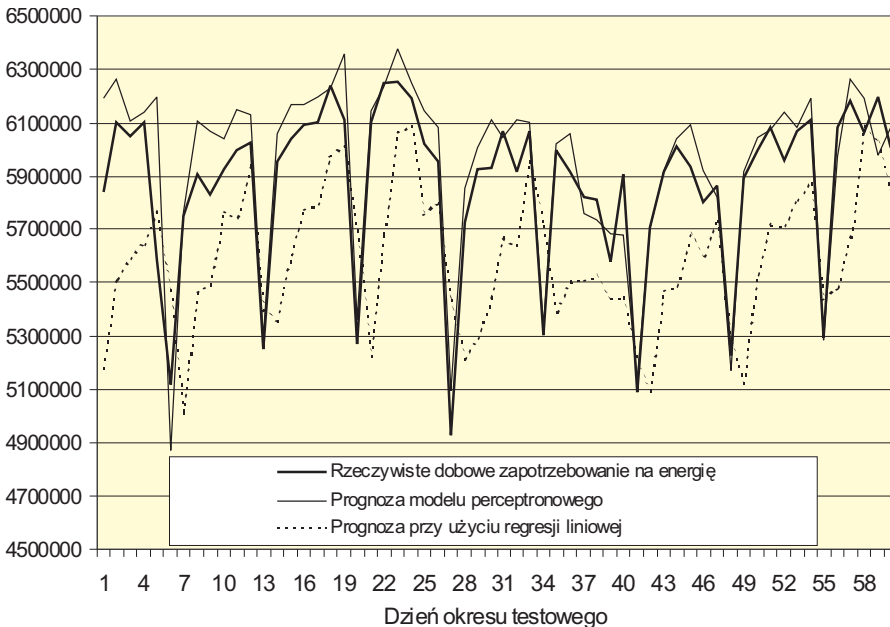
¹⁴ Potencjalna analiza porównawcza innych typów modeli wykracza poza założone ramy niniejszego opracowania, mającego na celu jedynie pokazanie przykładu efektywnego wykorzystania sieci neuronowych.

Po zakończeniu procesu uczenia otrzymano miary efektywności obu modeli w postaci błędów dla 365-elementowego zbioru testowego, przedstawionych w tabeli 1 (oczywiście im mniejszy błąd tym większa efektywność modelu).

Porównanie liczbowych wartości błędów dla zbioru testowego wskazuje na zdecydowanie wyższą efektywność modelu neuronowego w analizowanym problemie. Fakt ten jest również łatwo zauważalny w wyniku wizualnej analizy wykresów zaprezentowanych na rysunku 3, przedstawiających przebiegi szeregów czasowych: rzeczywistego zapotrzebowania na energię, jednodniowej prognozy modelu neuronowego (perceptron), oraz jednodniowej prognozy modelu opartego na regresji liniowej, dla pierwszych 60 dni z okresu testowego.

Rysunek 3

Rzeczywiste i prognozowane zapotrzebowanie dobowe na energię elektryczną dla pierwszych 60 dni testowych. Widoczne wyraźne spadki sprzedaży energii dla niedziel



Źródło: opracowanie własne.

W istocie sieci neuronowe stanowią na tyle efektywne narzędzia prognozowania nieliniowego, że są również w praktyce wykorzystywane przez zakłady energetyczne do konstrukcji modeli prognostycznych w zagadnieniach predykcji sprzedaży energii.

6. Podsumowanie

Sieci neuronowe wykazują wysoką przydatność w modelowaniu nieliniowych zjawisk ekonomicznych. Jako metody nieliniowe i nieparametryczne, stanowią silną alternatywę (lub narzędzie uzupełniające) dla klasycznych modeli statystycznych. O efektywności metod neuronowych świadczą nie tylko wyniki licznych badań, ale również rezultaty ich praktycznych zastosowań (np. w przedstawionym tu problemie prognozowania sprzedaży energii, czy też w predykcji kursów walut albo akcji). Fakt ten wskazuje na celowość badań dotyczących zastosowania sieci neuronowych w różnych problemach ekonomicznych, głównie związanych z analizą złożonych, nieliniowych i słabo zidentyfikowanych systemów.

Bibliografia

- Azoff E.M., *Neural network time series forecasting of financial markets*, Wiley, New York 1994.
- Bartkiewicz W., *Metody sztucznej inteligencji w prognozowaniu obciążenia sieci elektroenergetycznej*, Rozprawa doktorska, Łódź 1998.
- Grabowski M., *Sieci neuronowe w analizie danych społeczno-ekonomicznych*. Praca doktorska, Akademia Ekonomiczna w Krakowie 1997.
- Haykin S., *Neural networks. A comprehensive foundation*, Macmillan College Publishing Company, New York 1994.
- Hebb D.O., *The organization of behavior: A neuropsychological theory*, Wiley, New York 1949.
- Hecht-Nielsen R., *Neurocomputing*, Reading, MA. Addison-Wesley 1990.
- Hertz J., Krogh A., Palmer R. G., *Wstęp do teorii obliczeń neuronowych*, WNT, Warszawa 1993.
- Kohonen T., *Self-organizing maps*, Springer-Verlag, Berlin 1995.
- Lula P., *Jednokierunkowe sieci neuronowe w modelowaniu zjawisk ekonomicznych*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 1999.
- McCulloch W. S., Pitts W., *A logical calculus of the ideas immanent in nervous activity*, „Bulletin of Mathematical Biophysics” 1943, No 5, pp. 115—133.
- Morajda J., *Wybrane możliwości zastosowań sieci neuronowych w ekonomii i zarządzaniu*, Zeszyty Naukowe AE w Krakowie nr 493, Kraków 1997.
- Morajda J., *Neural Networks as Predictive Models in Financial Futures Trading*, Proc. of the 5-th Conference „Neural Networks and Soft Computing”, Zakopane 2000.
- Morajda J., *Neural Networks and Their Economic Applications [w:] Artificial Intelligence and Security in Computing Systems — Proc. of the 9th International Conference ACS'2002*, Kluwer Academic Publishers, Boston, Dordrecht, London 2003.
- Neuman von J., *The computer and the brain*, Yale Univ. Press., New Haven 1958.
- Osowski S., *Sieci neuronowe w ujęciu algorytmicznym*, WNT, Warszawa 1996.
- Refenes A. P. (ed.), *Neural Networks in the Capital Markets*, Wiley, Chichester 1995.
- Rutkowska D., Piliński M., Rutkowski L., *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*, PWN, Warszawa 1997.
- Selecting the right neural network tool, third edition*, „Neurovest Journal”, September/October 1996.
- Tadeusiewicz R., *Sieci neuronowe*, Akademicka Oficyna Wydawnicza RM, Warszawa 1993.
- Tadeusiewicz R., *Problemy biocybernetyki*, PWN Warszawa, 1994.
- Tadeusiewicz R., *Sieci neuronowe w prognozowaniu procesów gospodarczych [w:] Mat. konferencyjne „Sztuczna inteligencja i infrastruktura informatyczna*, Siedlce 1995, s. 325—342.

-
- Taylor W. K., *Computers and the nervous system. Models and analogues in biology*, Cambridge Univ. Press., Cambridge 1960.
- Werbos P. J., *Beyond regression: new tools for prediction and analysis in the behavioral sciences*, Ph.D. thesis, Harvard University, Cambridge, MA, 1974.
- Zieliński J. S. (red.), *Inteligentne systemy w zarządzaniu — teoria i praktyka*, PWN, Warszawa 2000.