

Wykorzystanie perceptronowych sieci neuronowych w zagadnieniu wyceny nieruchomości

1. Wprowadzenie

W wielu zagadnieniach szeroko rozumianej analizy danych ekonomicznych, oprócz klasycznych metod statystycznych, ważne zastosowanie znajdują także narzędzia oparte na sztucznej inteligencji, zwłaszcza sieci neuronowe. Problematyka sieci neuronowych i ich zastosowań została rozwinięta w bogatej literaturze, np. [S. Haykin 1994; P. Lula 1999; S. Osowski 1996; A. P. Refenes (ed.) 1995; R. Tadeusiewicz 1993; J. S. Zieliński (red.) 2000]. Metody wywodzące się z tego obszaru mogą okazać się szczególnie przydatne w analizie problemów, w których istnieją trudne do zidentyfikowania, nieliniowe relacje pomiędzy zmiennymi w analizowanym systemie ekonomicznym, lub gdy posiadamy niewielką wiedzę aprioryczną dotyczącą badanego obiektu, ale dysponujemy odpowiednią ilością danych mogących posłużyć do uczenia sieci. Tego typu zagadnienia, w których efektywność sieci neuronowych została teoretycznie i empirycznie potwierdzona [A. Beltratti, S. Margarita, P. Terna 1996; P. Lula 1999; J. Morajda 1997; J. Morajda — w druku; J. Morajda 2003; A. P. Refenes (red.) 1995; J. S. Zieliński (red.) 2000], często pojawiają się w różnych dziedzinach ekonomii i zarządzania, np. w prognozowaniu rynków finansowych, marketingu, ocenie ryzyka kredytowego itp.

Zamieszczone w niniejszym opracowaniu badania dotyczą wykorzystania sieci neuronowych typu perceptron do szacowania wartości nieruchomości metodą porównawczą, tzn. w oparciu o dane dotyczące innych, wycenionych wcześniej obiektów. Narzędziami wspomagającymi taką wycenę mogą być modele formalne wykorzystujące funkcje regresji, jednak w związku z nieliniowym i nie w pełni zidentyfikowanym charakterem powiązań pomiędzy czynnikami wpływającymi na wartość nieruchomości a jej ceną, istnieją przesłanki do zastosowania sieci neuronowych w tym zagadnieniu.

Typowa odpowiedź sieci neuronowej, podobnie jak w przypadku modelu regresyjnego, polega na wygenerowaniu jednej „punktowej” wartości liczbowej, stanowiącej szacowaną wycenę. W niniejszej pracy podjęto jednak próbę wyko-

rzystania zbioru sieci typu perceptron do aproksymacji funkcji gęstości prawdopodobieństwa dla ceny sprzedaży danej nieruchomości, co wnosi więcej informacji dotyczącej wartości obiektu i możliwości negocjacji ostatecznej ceny.

2. Opis problemu i zastosowana metoda jego rozwiązania

Podejście porównawcze do wyceny nieruchomości polega na szacowaniu wartości obiektu poprzez jego porównanie z charakterystyką podobnych, wcześniej sprzedanych (wycenionych) obiektów na podstawie określonego zestawu cech. Do konstrukcji modelu służącego do realizacji takiej wyceny można zastosować sieć neuronową, uczoną (i testowaną) przy użyciu wzorców opisujących nieruchomości, dla których wcześniej zawarto transakcje kupna po określonej cenie. Występujące we wzorcach uczących cechy obiektów stanowią zmienne wejściowe (objaśniające) dla modelu neuronowego, a cena transakcji — zmienną wyjściową (objaśnianą). Tak nauczona sieć, po podaniu na wejście opisu wycenianego obiektu, wygeneruje jako sygnał wyjściowy jego wycenę. Procedura ta stanowi typowe wykorzystanie realizujących nieliniową regresję sieci neuronowych do prognozowania określonej wartości.

Jednakże sieci typu perceptron, konstruowane dla realizacji procesu klasyfikacji (a nie w celu obliczania „punktowej” prognozy), można wykorzystać do aproksymacji funkcji gęstości prawdopodobieństwa dla prognozowanej zmiennej. Podstawy teoretyczne tej metodyki zawarte zostały w pracy [D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley, B. W. Suter 1990], w której m.in. pokazano, iż w zagadnieniu klasyfikacji sygnały wyjściowe perceptronu wielowarstwowego odpowiednio uczonego przy użyciu algorytmu wstecznej propagacji błędów¹ reprezentują² prawdopodobieństwa *a posteriori* przynależności danego wektora wejściowego x do poszczególnych klas. Ten fakt został wykorzystany m.in. w pracy [J. Morajda – w druku] do zaproponowania neuronowego systemu aproksymacji funkcji gęstości prawdopodobieństwa w procesie prognozowania szeregów czasowych. Podobna koncepcja może posłużyć w rozważanym tu zagadnieniu wyceny nieruchomości. Opiera się ona na podziale zbioru wartości wyjściowych (cen nieruchomości w ciągu uczącym) na k rozłącznych, uporządkowanych przedziałów (determinujących jednocześnie klasy wektorów wejściowych x), a następnie na przeprowadzeniu uczenia k jednowyjściowych (realizujących zadanie klasyfikacji) sieci perceptronowych przyporządkowanych do poszczególnych

¹ Jeżeli istnieje k klas wektorów wejściowych x , perceptron posiada k wyjść przyporządkowanych odpowiednio do kolejnych klas; w procesie uczenia pożądane wartości wyjściowe przyjmują wielkość 1 na wyjściu przypisanym właściwej klasie oraz 0 na pozostałych wyjściach. Zamiast k -wyjściowego perceptronu można też zastosować k osobnych sieci perceptronowych z jednym wyjściem każda; uczenie takich sieci jest szybsze i przynosi lepsze efekty.

² Dokładność tej reprezentacji zależy jednak m.in. od właściwie dobranej struktury sieci (wielkości warstwy ukrytej, funkcji aktywacji neuronów itp.) i parametrów algorytmu uczenia.

klas (przedziałów cen). Podczas funkcjonowania nauczonych sieci, po zaprezentowaniu określonego wzorca wejściowego charakteryzującego wycenianą nieruchomość, wygenerowane wartości wyjściowe (zinterpretowane jako prawdopodobieństwa przynależności wektora x do poszczególnych klas, a wycenianego obiektu do poszczególnych przedziałów cenowych) służą do realizacji aproksymacji funkcji gęstości prawdopodobieństwa dla wartości (ceny) nieruchomości.

W przeprowadzonych badaniach wykorzystano dane pochodzące ze spisu powszechnego przeprowadzonego w Bostonie (USA) w roku 1970 i zawierające 400 przypadków dotyczących charakterystyki nieruchomości w poszczególnych 400 obszarach spisowych.

Każdy z tych przypadków jest opisany przez 13 następujących cech (potraktowanych tutaj jako zmienne wejściowe dla modeli neuronowych):

- średnia liczba pokoi w nieruchomościach w danym obszarze,
- odsetek obiektów wybudowanych przed 1940 rokiem,
- odsetek ludności murzyńskiej żyjącej w danym obszarze spisowym,
- odsetek ludności o niskim statusie społecznym,
- liczba popełnianych przestępstw w przeliczeniu na 1 mieszkańca,
- odsetek posiadłości zajmujących ponad 25 000 stóp kwadratowych,
- udział stref przemysłowych w danym obszarze,
- stopa podatku od nieruchomości,
- liczba uczniów przypadająca na 1 nauczyciela,
- wskaźnik binarny równy 1 dla obszarów graniczących z rzeką Charles,
- ważona odległość do 5 podstawowych stref zatrudnienia w Bostonie,
- wskaźnik wyrażający łatwość dostępu do obwodnic Bostonu,
- wskaźnik zanieczyszczenia powietrza (stężenie tlenu azotu).

Jako zmienną objaśnianą w modelu przyjęto medianę wartości nieruchomości w danym obszarze spisowym.

Przed wykorzystaniem danych do budowy modelu neuronowego wszystkie zmienne (wejściowe i wyjściową) poddano standaryzacji przez odjęcie wartości średniej dla każdej zmiennej i następnie wykonanie dzielenia przez odchylenie standardowe.

Neuronowy system analizy danych został skonstruowany dla parametru $k = 21$, tzn. przy wykorzystaniu 21 sieci typu perceptron wielowarstwowy, realizujących zadanie klasyfikacji i trenowanych przy zastosowaniu pożądaných wartości wyjściowych typu binarnego (0 i 1). Każda z tych sieci realizuje zadanie rozpoznawania określonej klasy nieruchomości o wartościach z danego zakresu cen o rozpiętości 2250 \$ (kolejne sieci są związane z kolejnymi przedziałami cenowymi).

Każda sieć wchodząca w skład systemu zawierała 5 neuronów ukrytych, posiadających funkcję aktywacji typu tangens hiperboliczny i była trenowana przy wykorzystaniu algorytmu wstecznej propagacji błędów (w wersji z modyfikacją wag po demonstracji każdego wzorca uczącego). Zastosowana procedura tzw. multi-startu (wielokrotnego startu procesu uczenia od różnych zestawów wag początko-

wych) pozwoliła uniknąć przypadkowego przedwczesnego zatrzymania uczenia w lokalnym minimum funkcji błędu.

Niezależnie od opisanego wyżej systemu, analizie poddano (głównie w celach porównawczych) efektywność wykorzystania klasycznej sieci perceptron, realizującej zadanie nieliniowej regresji i trenowanej przy użyciu rzeczywistych wartości cen nieruchomości (zmiennej wyjściowej). Sieć taka, po nauczaniu, generuje jednak tylko „punktowe” rezultaty wyceny bez informacji o kształcie rozkładu prawdopodobieństwa cen. Dla tej sieci zastosowano taką samą strukturę i parametry algorytmu uczenia, jak w przypadku perceptronów klasyfikujących.

Dwadzieścia spośród 400 przypadków zakwalifikowano jako zbiór testowy (nie biorący udziału w uczeniu sieci), natomiast pozostałe 380 rekordów zastosowano do uczenia i walidacji sieci. Na rysunkach 1, 2 i 3 zademonstrowano m.in. rozkład cen nieruchomości (aprioryczny rozkład prawdopodobieństwa), którego podstawowe parametry posiadają następujące wartości (w tysiącach \$): *minimum* = 5,00; *maksimum* = 50,00; *średnia* = 23,02; *odchylenie standardowe* = 9,30.

3. Rezultaty badań

W tabeli 1 przedstawiono rezultaty wyceny nieruchomości dla 20-elementowego zbioru testowego, zrealizowanej najpierw przez klasyczny perceptron realizujący zadanie nieliniowej regresji (są to „punktowe” wartości wycen, nie wnoszące informacji o szacunkowym rozkładzie prawdopodobieństwa).

Tabela 1

Rzeczywiste i oszacowane przez regresyjny perceptron wartości nieruchomości (mediany dla obszaru spisowego) dla 20 przypadków zawartych w zbiorze testowym. Odpowiadające sobie wartości (rzeczywista i oszacowana) dla pojedynczego przypadku są umieszczone w jednej kolumnie jedna pod drugą

Wartości nieruchomości w tys. \$ dla 20 przypadków ze zbioru testowego										
Rzeczywista	18,9	24,8	24,5	18,3	42,3	19,3	19,3	12,7	36,1	23,2
Oszacowana	15,6	27,6	26,2	19,7	44,6	19,2	20,7	12,2	32,7	21,5
Rzeczywista	21,2	15,7	18,2	18,8	11,9	14,3	28,4	43,5	21,2	20,6
Oszacowana	26,5	16,9	16,6	21,0	22,5	16,0	26,7	46,7	21,7	22,8

Źródło: obliczenia własne.

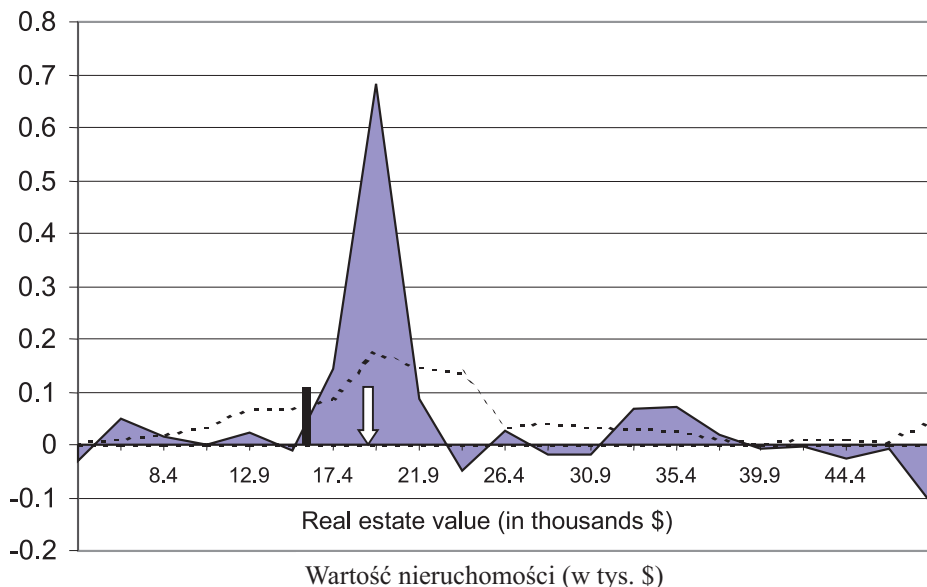
Otrzymano następujące miary efektywności sieci perceptron dla zbioru testowego: pierwiastek z błędu średniokwadratowego $RMSE = 3,286$ oraz błąd średni $MAE = 2,440$.

Z kolei na rysunkach 1, 2 i 3, dla trzech pierwszych przypadków ze zbioru testowego, zaprezentowano w postaci graficznej rezultaty funkcjonowania systemu neuronowego złożonego z 21 sieci perceptron, realizujących zadanie klasyfikacji.

Wykresy przedstawiają rozkłady wygenerowanych przez sieci perceptronowe sygnałów wyjściowych, które tworzą aproksymacje rozkładów gęstości prawdopodobieństwa dla cen nieruchomości w poszczególnych przypadkach. Aby na ich podstawie uzyskać właściwe rozkłady prawdopodobieństwa poprawne w sensie matematycznym, wystarczy przeprowadzić prostą transformację sygnałów, która sprowadza ujemne wartości do zera i następnie przeskalowuje wszystkie wartości sygnałów tak, aby ich suma wynosiła dokładnie 1. Na rysunkach 1, 2, 3 zamieszczono jednak (celowo) nie przekształcone rozkłady sygnałów sieci, aby zobrazować rzeczywiste funkcjonowanie systemu neuronowego. Na wykresach oznaczono też rzeczywiste wartości nieruchomości oraz wielkości wyceny wygenerowane przez klasyczny perceptron realizujący operację regresji.

Rysunek 1

Oszacowany przez klasyfikujący system neuronowy rozkład cen nieruchomości dla przypadku testowego nr 1 (*linia ciągła z zacieniowanym obszarem*). Linia przerywana ukazuje rozkład empiryczny dla wszystkich 400 przypadków. Biała strzałka wskazuje rzeczywistą wartość, natomiast czarny słupek reprezentuje sygnał wygenerowany przez klasyczny perceptron regresyjny

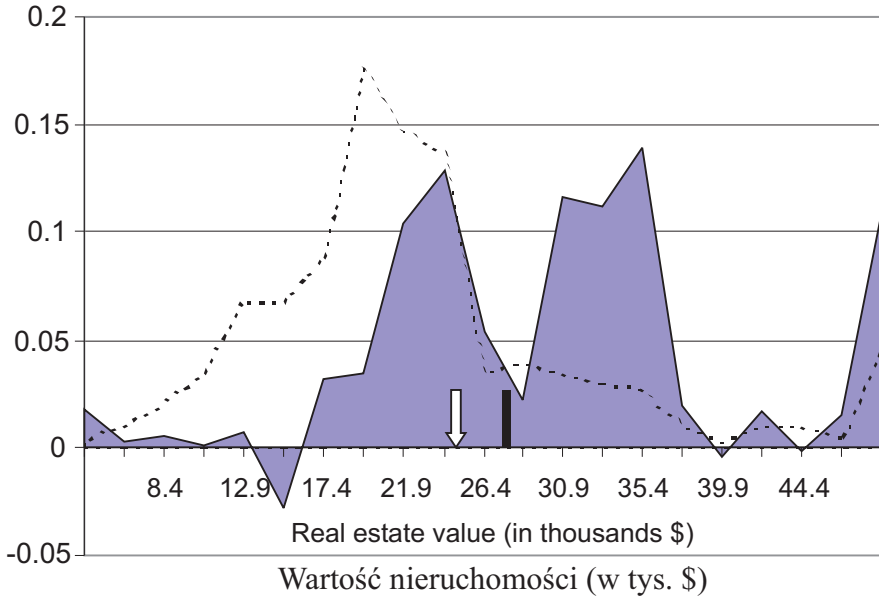


Należy podkreślić, iż otrzymane aproksymacje rozkładów prawdopodobieństwa dla cen nieruchomości dostarczają dużo więcej użytecznej informacji niż zwykła „punktowa” ewaluacja zawierająca pojedynczą wartość.

Rysunek 1 przedstawia sytuację, w której wartość nieruchomości może być dość jednoznacznie i precyzyjnie określona. Wycena wygenerowana przez klasyfikujący system neuronowy wyznacza to oszacowanie na około 19 600 \$ z dość wysokim prawdopodobieństwem (prawie 0,7), co jest wielkością zlokalizowaną

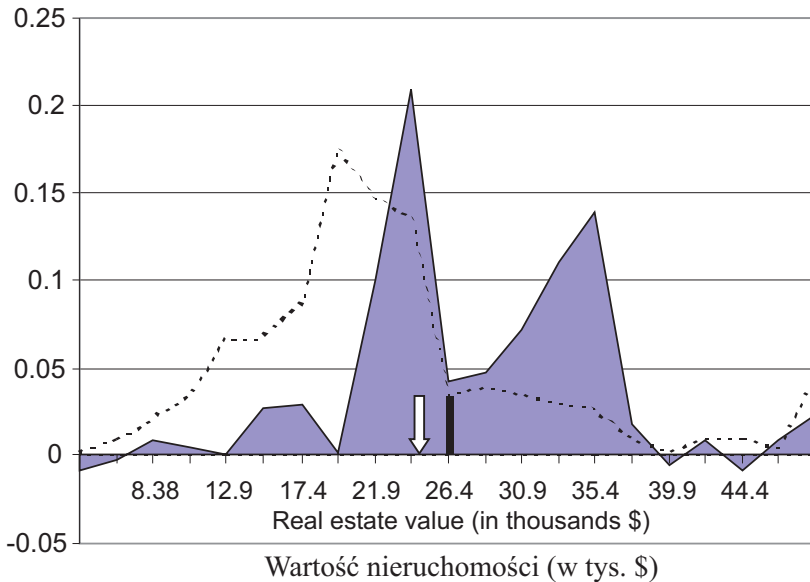
Rysunek 2

Oszacowany przez klasyfikujący system neuronowy rozkład cen nieruchomości dla przypadku testowego nr 2 (oznaczenia jak na rys. 1)



Rysunek 3

Oszacowany przez klasyfikujący system neuronowy rozkład cen nieruchomości dla przypadku testowego nr 3 (oznaczenia jak na rys. 1)



bardzo blisko rzeczywistej wartości wynoszącej w tym przypadku 18 900 \$. Regresyjna sieć perceptronowa popełniła tutaj wyraźnie większy błąd szacowania.

Analiza rozkładu wartości wycen zaprezentowana na rysunku 2 wskazuje na niejednoznaczność oszacowania nieruchomości dla drugiego przypadku testowego, gdyż na wykresie rozkładu występują 3 wyraźne i odrębne „szczyty”, reprezentujące kolejno ceny: 24 100 \$ (wycena zlokalizowana bardzo blisko wartości rzeczywistej wynoszącej 24 800 \$), 35 400 \$, oraz ok. 50 000 \$. Fakt ten pokazuje, iż prawdopodobnie istnieją jeszcze dodatkowe (nie uwzględnione w analizie) czynniki (stanowiące potencjalne dodatkowe zmienne objaśniające dla systemu), które mogą mieć istotny wpływ na cenę danej nieruchomości. Ponadto informacja pochodząca z kształtu tego typu rozkładu może być niezwykle istotna, np. dla sprzedawcy nieruchomości, który wykorzystując znaczną potencjalną przestrzeń do negocjacji cen (wyznaczoną przez prawą część rozkładu) może w efekcie uzyskać dużo większą cenę sprzedawanego obiektu niż oszacowanie wygenerowane przez regresyjny perceptron lub rzeczywista wartość mediany cen w danym obszarze.

Przypadek testowy nr 3 (zobrazowany na rysunku 3), charakteryzujący się dwoma wyraźnymi wierzchołkami na wykresie rozkładu cen, jest w zasadzie podobny w sensie interpretacji do omówionego wyżej przypadku 2.

4. Wnioski końcowe

Zaprezentowana w niniejszym artykule metoda funkcjonowania systemu neuronowego złożonego z pewnej liczby sieci typu perceptron, realizujących zadania klasyfikacji, okazała się użyteczna i efektywna w zagadnieniach wyceny nieruchomości metodą porównawczą. Metoda pozwala na oszacowanie rozkładu prawdopodobieństwa dla wartości badanego obiektu. Analiza takiego rozkładu w określonym przypadku generuje znacznie więcej użytecznej informacji niż oszacowanie pojedynczej wartości, informacja ta zaś może zostać wykorzystana zarówno w celach badawczych (udoskonalenie modelu służącego do wyceny, np. poprzez identyfikację nowych zmiennych objaśniających), jak i w zastosowaniach praktycznych (negocjacja cen). Potencjalne kierunki dalszych badań to analiza doboru parametrów struktury i uczenia sieci neuronowych w omawianym systemie, a także badania efektywności wykorzystania metody w innych, podobnych problemach ekonomicznych.

Bibliografia

- Beltratti A., Margarita S., Terna P. *Neural Networks for Economic and Financial Modeling*, „International Thomson Computer Press”, London 1996.
- Haykin S., *Neural networks. A comprehensive foundation*, Macmillan College Publishing Company, New York 1994.

- Lula P., *Jednokierunkowe sieci neuronowe w modelowaniu zjawisk ekonomicznych*. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 1999.
- Morajda J., *Wybrane możliwości zastosowań sieci neuronowych w ekonomii i zarządzaniu*, Zeszyty Naukowe AE w Krakowie nr 493, Kraków 1997.
- Morajda J., *Multilayer Perceptrons as Approximations to Probability Density Functions in Time Series Forecasting*, „Przegląd Statystyczny”, Warszawa — w druku.
- 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.
- Osowski S., *Sieci neuronowe w ujęciu algorytmicznym*, WNT, Warszawa 1996.
- Refenes A. P. (ed.), *Neural Networks in the Capital Markets*, Wiley, Chichester 1995.
- Ruck D. W., Rogers S. K., Kabrisky M., Oxley M. E., Suter B. W., *The Multilayer Perceptron as an Approximation to a Bayes Optimal Discriminant Function*, „IEEE Transactions on Neural Networks” 1990, vol. 1, no. 4.
- Tadeusiewicz R., *Sieci neuronowe*, Akademicka Oficyna Wydawnicza RM, Warszawa 1993.
- Zieliński J. S. (red.), *Inteligentne systemy w zarządzaniu — teoria i praktyka*, PWN, Warszawa 2000.