

Tomasz Jasiński, Anna Bochenek

Prognozowanie cen nieruchomości lokalowych za pomocą sztucznych sieci neuronowych

Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania 45/1, 317-328

2016

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.



DOI: 10.18276/sjp.2016.45/1-25

Tomasz Jasiński*

Anna Bochenek**

Politechnika Łódzka

PROGNOZOWANIE CEN NIERUCHOMOŚCI LOKALOWYCH ZA POMOCĄ SZTUCZNYCH SIECI NEURONOWYCH

Streszczenie

Opracowanie porusza problematykę modelowania cen nieruchomości lokalowych na rynku wtórnym za pomocą sztucznych sieci neuronowych. Przeprowadzone przez autorów badania mają charakter zarówno literaturowy, jak i empiryczny. W procesie budowy modeli testowano możliwość użycia licznych zmiennych wejściowych. Badania objęły swoim zasięgiem strefę śródmiejską Łodzi. Wśród sieci poddanych testom znalazły się: sieć jednokierunkowa wielowarstwowa o jednej i dwóch warstwach ukrytych oraz sieć o radialnych funkcjach bazowych.

Słowa kluczowe: sztuczna inteligencja, sieci neuronowe, prognozowanie, ceny, nieruchomości lokalowe

Wstęp

Prognozowanie cen nieruchomości za pomocą narzędzi z obszaru sztucznej inteligencji (*artificial intelligence* – AI) stanowi stosunkowo nowy i nie w pełni zbadany obszar analizy. W literaturze przedmiotu można odnaleźć zarówno wyniki badań prowadzonych przy użyciu metod AI bazujących na zasadzie ewolucji Darwina, na

* Adres e-mail: tomasz.jasinski@p.lodz.pl.

** Adres e-mail: An.Bochenek@wp.pl.

przykład algorytmów genetycznych i struktur ewolucyjnych, jak i modeli sztucznych sieci neuronowych (*artificial neural networks* – ANN). AI często stanowią jeden z elementów kompleksowego systemu prognostycznego, w którego skład zazwyczaj wchodzi także narzędzie Systemu Informacji Geograficznej (Geographic Information System – GIS). Takie podejście zaprezentowali między innymi Garcia, Gamez i Alfaro (2008), Liu, Deng i Wang (2011). W niniejszym opracowaniu modelowanie cen sprzedaży zostało przeprowadzone za pomocą symulatora ANN działającego na bazie zmiennych objaśniających stworzonych przy użyciu nie tylko danych pochodzących z systemu GIS oraz archiwalnych cen sprzedaży nieruchomości (pozyskanych z Łódzkiego Ośrodka Geodezji), lecz także z zastosowaniem dodatkowych informacji wyodrębnionych między innymi w procesie analizy ksiąg wieczystych. Celem badań było sprawdzenie możliwości przewidywania m^2 oraz wartości transakcji kupna-sprzedaży nieruchomości lokalowej przy użyciu wybranych modeli ANN.

1. Budowa i zasada działania sztucznych sieci neuronowych

ANN stanowią połączone i komunikujące się ze sobą pojedyncze komórki nerwowe. Budowa sztucznego neuronu nawiązuje do swojego biologicznego pierwowzoru. Odpowiednikami dendrytów są tak zwane wejścia. Służą one do wprowadzania sygnałów do komórki. Należy przy tym zauważyć, że z każdym wejściem jest związana pewna liczba rzeczywista zwana wagą. Sygnał wprowadzany do komórki jest mnożony przez przypisany do danego wejścia współczynnik wagowy. Następnie tak powstałe iloczyny są sumowane, stanowiąc tak zwane całkowite pobudzenie neuronu. Opisane wagi decydują o „inteligencji” danej komórki, ponieważ w procesie uczenia modelu to właśnie one ulegają modyfikacji, tak by sieć jako całość mogła zyskać i zapamiętać nową wiedzę. W zależności od struktury neuronów oraz połączeń między nimi wyróżnia się wiele typów ANN. Do najbardziej popularnych należą stworzone w 1958 roku sieci jednokierunkowe wielowarstwowe (zwane również perceptronem wielowarstwowym, w skrócie MLP od ang. *multilayer perceptron*). Podobnie jak w pozostałych rodzajach sieci, także i w MLP neurony pogrupowane są w trzy rodzaje warstw –zawsze pojedyncze wejściową i wyjściową oraz umiejscowione między nimi opcjonalne warstwy ukryte. Cechą charakterystyczną MLP jest przepływ informacji tylko w jednym kierunku, to jest od warstwy wejściowej, poprzez kolejne warstwy ukryte, aż do wyjściowej.

Neurony umieszczone w warstwie wejściowej nie biorą bezpośrednio udziału w procesie nauki. Ich zadaniem jest rozpowszechnienie danych wejściowych modelu wśród neuronów dalszych struktur ANN. „Inteligencja” jest zatem umieszczona w opisanych uprzednio wagach neuronów wchodzących w skład warstw ukrytych oraz warstwy wyjściowej. Liczba neuronów w warstwie wejściowej jest zawsze równa liczbie zmiennych objaśniających modelu. Analogicznie liczba komórek nerwowych w ostatniej, wyjściowej warstwie sieci pokrywa się z liczbą zmiennych objaśnianych. W przypadku modelowania cen nieruchomości typowo stosuje się modele wyposażone w pojedynczy neuron na wyjściu ANN. Dzięki temu model prognostyczny generuje na swoim wyjściu pojedynczą odpowiedź. Takie rozwiązanie wymusza zastosowanie MLP zawierających między innymi jedną warstwę ukrytą, gdyż model z tylko jednym uczącym się neuronem na wyjściu nie generowałby oczekiwanych wyników. Kwestia doboru liczby warstw ukrytych¹ oraz liczby komórek nerwowych w każdej z nich zazwyczaj jest rozwiązywana eksperymentalnie przez testowanie kolejnych wersji modelu prognostycznego².

Jeżeli w modelu występują połączenia zwrotne przesyłające sygnały z neuronów dalszych, licząc od wejścia, warstw modelu do komórek warstw wcześniejszych (lub neuronów tej samej warstwy), wówczas ANN jest siecią rekurencyjną (zwaną także RNN – *recurrent neural network*) (Tadeusiewicz, 1998). Ich użycie jest uzasadnione szczególnie w przypadku przewidywania szeregów czasowych. Zatem jeśli analiza dotyczyłaby modelowania zmienności cen nieruchomości w czasie, użycie RNN mogłoby doprowadzić do powstania modelu o wyższej jakości prognoz niż te generowane przez MLP. Ponieważ niniejsze opracowanie stanowi analizę cen nieruchomości lokalowych bez badania aspektu czasowego, zatem wykorzystanie RNN nie tylko byłoby nieuzasadnione, lecz wręcz stanowiłoby błędne podejście do zagadnienia modelowania przy użyciu ANN. Z tego względu w niniejszym artykule pominięte zostaną dalsze zasady budowy i funkcjonowania RNN.

Oprócz opisanych powyżej dwóch głównych rodzajów ANN wyróżnia się wiele innych ich podtypów. Ze względu na ich mnogość nie będą one przedmiotem dal-

¹ Literatura przedmiotu, jak i doświadczenia autorów wskazują, iż optymalne są zazwyczaj modele wyposażone w jedną lub dwie warstwy ukryte. W rzadszych sytuacjach warto rozważyć ANN zawierające trzy takie warstwy. Dalsze zwiększanie ich liczby prowadzi w praktyce do pogorszenia jakości prognoz.

² Liczba neuronów jest swego rodzaju kompromisem między dużą liczbą neuronów oznaczającą w teorii większe, potencjalne możliwości modelu a małą liczbą komórek nerwowych oznaczającą mniejszą liczbę błędów, które są związane z każdym neuronem. W praktyce poszukuje się możliwie prostego modelu, który jest w stanie realizować postawione cele.

szych rozważań w tym opracowaniu. Więcej informacji o ich budowie oraz zasadzie funkcjonowania można odnaleźć między innymi w (Jasiński, 2003).

2. Dobór parametrów modelu ANN prognozującego ceny nieruchomości

Analiza literatury przedmiotu wykazała zastosowanie w procesie predykcji cen nieruchomości wielu zmiennych objaśniających, które można podzielić na dwie grupy danych określających cechy (Antipov, Pokryshevskaya, 2012; Chiarazzo, Caggiani, Marinelli, Ottomanelli, 2014; Garcia, Gamez, Alfaro, 2008; Liu, Deng, Wang, 2011; Selim, 2009):

1. endogeniczne nieruchomości:
 - typ nieruchomości (np. lokalowa, budynkowa³),
 - materiał użyty do budowy (np. drewno, cegła),
 - wiek⁴,
 - powierzchnia⁵,
 - liczba łazienek⁶,
 - liczba pokoi,
 - występowanie windy,
 - liczba kondygnacji budynku,
 - numer kondygnacji, na której znajduje się nieruchomość (w przypadku nieruchomości lokalowych),
 - występowanie balkonu⁷,
 - istnienie lub rodzaj ogrzewania,
 - stan nieruchomości⁸,
 - rodzaj podłogi w salonie, sypialni i łazience,

³ W przypadku nieruchomości budynkowej można także uściślać ich rodzaj (Selim, 2009).

⁴ Może być on wyrażony w latach (Garcia, i in., 2008) albo jako jedna z grup wiekowych nieruchomości (Selim, 2009).

⁵ Zarówno całkowita, jak i uzupełniona np. przez powierzchnie pierwszego i drugiego pokoju (w przypadku badań nad nieruchomościami lokalowymi o m.in. dwóch pokojach, np. Antipov, Pokryshevskaya, 2012).

⁶ Garcia i in. (2008) zastosowali podejście polegające na przyznaniu jednego punktu za łazienkę i połowy za tzw. „niepełną” łazienkę, tj. niewyposażoną w wannę lub prysznic.

⁷ Garcia i in. (2008) przyznawali punkt w tej kategorii, tylko jeżeli powierzchnia balkonu przekraczała 15 m².

⁸ Zazwyczaj wyróżnia się kilka kategorii, np. zły, standardowy, bardzo dobry.

- występowanie pomieszczenia przynależnego typu komórka/piwnica,
 - wielkość powierzchni parkingowej lub przynależność garażu,
 - występowanie ogrodu na terenie nieruchomości,
 - występowanie innych udogodnień, na przykład sauny, jacuzzi, telewizji kablowej,
 - dostępność gazu ziemnego (instalacja gazowa),
 - cena wywoławcza nieruchomości,
2. egzogeniczne nieruchomości:
- rodzaj obszaru, na którym znajduje się nieruchomość (np. miejski, wiejski),
 - odległość nieruchomości od centrum miasta⁹,
 - dostępność w okolicy innych udogodnień, takich jak środki transportu, miejsca edukacji i biznesu, telefon¹⁰,
 - dane dotyczące zatrudnienia w najbliższej okolicy (np. liczba pracodawców i pracowników),
 - dane demograficzne (np. gęstość zaludnienia, liczba mieszkańców),
 - stopień zanieczyszczenia powietrza (zarówno przeciętny, jak i wartości maksymalne)¹¹,
 - współrzędne geograficzne¹²,
 - inne zmienne charakterystyczne dla otoczenia badanych nieruchomości, na przykład odległość od plaży w przypadku miejscowości nadmorskich.

Badania literaturowe wykazały, że poprawnie mogą funkcjonować zarówno modele wyposażone w stosunkowo niewielką liczbę zmiennych objaśniających, jak i te zbudowane przy użyciu dużej ich liczby.

⁹ Centrum miasta jest w tym przypadku rozumiane jako miejsce powszechnie uznawane za centrum „życia”. Często taką funkcję pełnią duże obiekty handlowo-rozrywkowe. Tym samym niejednokrotnie nie pokrywa się ono z historycznym centrum miasta. Takie podejście zostało zastosowane m.in. w badaniach autorstwa Garcii i in. (2008).

¹⁰ Część z tych danych może zostać pobrana z systemu GIS (Liu i in., 2011). Pozostałe, np. czas dotarcia do centrum podczas porannych godzin szczytu (Chiarazzo i in., 2014), wymagają dostępności innych źródeł.

¹¹ Chiarazzo i in. (2014) wykorzystali dane z jedenastu stacji pomiarowych dostarczających informacji o wielkości zanieczyszczenia w postaci wartości SO_2 , NO_x , NO , NO_2 , CO , PM_{10} .

¹² Co prawda stanowią one *de facto* cechę endogeniczną nieruchomości, to jednak w tym przypadku zostały one zakwalifikowane do zmiennych typu egzogenicznego, ponieważ celem ich zastosowania jest określenie położenia nieruchomości w stosunku do innych, których dane zostały poddane analizie, a tym samym niejako automatyczne ustalenie poziomu cen obowiązujących w najbliższym sąsiedztwie.

Ze względu na opisaną w poprzednim rozdziale zasadność ograniczenia liczby zmiennych wejściowych warto, o ile istnieje taka możliwość, dokonać analizy stopnia zależności między zmiennymi objaśniającymi a zmienną objaśnianą. Przykładem takich badań jest publikacja autorstwa Antipova i Pokryshevskiy (2012). Zmienne wejściowe zostały podzielone na cztery grupy w zależności od stopnia ich istotności. Największy związek z cenami m² nieruchomości zaobserwowano przy zmiennych objaśniających w postaci: dzielnicy, czasu dotarcia do centrum miasta metrem, typu domu. Do czynników, których istotność była mniejsza, ale wciąż wysoka, zaklasyfikowano: całkowitą powierzchnię nieruchomości, typ (wyposażenie) łazienki. Czynniki o przeciętnym poziomie wpływu na zmienną wyjściową stanowiły: powierzchnia kuchni oraz salonu, odległość od najbliższej stacji metra, liczba kondygnacji w budynku. Jako elementy mało istotne sklasyfikowano: bezwzględną różnicę procentową pomiędzy powierzchniami pokoi, wskaźnik Herfindahla obliczony dla powierzchni pokoi, kondygnację, na której znajduje się nieruchomość lokalowa, oraz dostępność telefonu.

Należy zauważyć, że poszukując optymalnego zestawu zmiennych wejściowych modelu, trzeba uwzględnić zarówno typ nieruchomości, których ceny są modelowane, jak i, co wydaje się równie istotne, specyfikę lokalnego rynku nieruchomości. Trudno oczekiwać, aby na przykład w przypadku nieruchomości lokalowych w dużych polskich miastach czynnik w postaci dostępności telefonu stacjonarnego miał w dobie komunikacji bezprzewodowej zauważalne znaczenie. Powoduje to, że powstałe modele posiadają ograniczoną uniwersalność, zatem ich przeniesienie na zgoła inny rynek nieruchomości musi być związane z dokonaniem co najmniej ich adaptacji do nowych warunków, a w skrajnym przypadku nawet ich całkowitą przebudową.

Przeprowadzenie analizy istotności zmiennych wejściowych modelu może być dokonane na wiele sposobów. Jednym z najbardziej popularnych jest podejście polegające na budowie licznych modeli różniących się zmiennymi objaśniającymi i poszukiwanie wśród nich optymalnego zestawu zmiennych. Inną metodą jest badanie wrażliwości zmiennej wyjściowej na dane wejściowe. Ponieważ stworzony przez ANN model nie ma charakteru liniowego, a dokładnie jest wielomianem dowolnego stopnia, zatem tradycyjne współczynniki korelacji okazują się być trudne do zastosowania. Możliwe jest natomiast wprowadzanie losowych (w danym przedziale) zaburzeń zmiennej wyjściowej i obserwacja zmian na wyjściu modelu. Umożliwia to przede wszystkim wyodrębnienie tych zmiennych modelu, których wpływ na prognozowaną zmienną jest najmniejszy, a następnie ich usunięcie.

3. Modelowanie ceny nieruchomości lokalowej na przykładzie Łodzi

Analizie poddano transakcje kupna-sprzedaży obejmujące nieruchomości lokalowe dostępne na rynku wtórnym w 2015 roku. Wszystkie z wymienionych obiektów znajdowały się w centrum miasta, to jest w obrębach „S”. W badaniach empirycznych zostały użyte przez autorów następujące zmienne objaśniające: 1) powierzchnia nieruchomości lokalowej; 2) powierzchnia nieruchomości wraz z pomieszczeniami przynależnymi; 3) zmienna dwustanowa wskazująca na przynależność piwnicy; 4) liczba izb; 5) liczba pokoi; 6) kondygnacja, na której znajduje się nieruchomość lokalowa; 7) liczba kondygnacji w budynku; 8) stan budynku (oceny w trzy-stopniowej skali); 9) długość trasy samochodowej do kompleksu Manufaktura¹³; 10) szacowany czas przejazdu autem do kompleksu Manufaktura¹⁴; 11) minimum odległości nieruchomości od jednego z dwóch głównych centrów handlowo-rozrywkowych Łodzi, to jest kompleksu Manufaktura oraz Galerii Łódzkiej¹⁵; 12) odległość od placu Wolności; 13) odległość od najbliższego obszaru zieleni.

Zmienne objaśniane stanowiły: 1) wartość transakcji; 2) cena m² nieruchomości lokalowej; 3) cztery zmienne binarne określające przynależność ceny m² do określonego przedziału cenowego.

Modelowanie zostało przeprowadzone z użyciem sieci jednokierunkowych dwu- oraz trójwarstwowych oraz o radialnych funkcjach bazowych (*radial basis function* – RBF). Z przeprowadzonych analiz wynika, że dwie pierwsze wymienione architektury cechowały najlepsze wyniki prognoz. Z tego powodu dalszy opis wyników badań będzie się odnosił jedynie do wspomnianego perceptronu wielowarstwowego. Jako algorytm uczący została zastosowana momentowa metoda wstecznej propagacji błędów z różnymi wartościami współczynnika nauki oraz momentum.

¹³ Długość trasy została wyznaczona przy użyciu serwisu internetowego maps.google.com. Wybrano najkrótszą z proponowanych dróg przejazdu.

¹⁴ Użyty czas przejazdu został wyznaczony przez serwis maps.google.com. Oszacowanie obejmowało przejazd samochodem w dzień roboczy o godz. 17.00. Dla najszybszej trasy wybrano czas średni z podanego przedziału możliwych wartości (w zależności od natężenia ruchu drogowego w podanej godzinie).

¹⁵ Wszystkie odległości zostały obliczone w linii prostej między badaną nieruchomością lokalową a punktem docelowym. Należy zaznaczyć, że takie podejście stanowi w pewnym sensie uproszczenie zagadnienia odległości od siebie opisanych miejsc, ponieważ należy oczekiwać, że potencjalnego nabywcę w większym stopniu interesuje dystans potrzebny do pokonania pieszo lub środkiem komunikacji (np. samochodem) niż odległość w metryce euklidesowej. Pewnym rozwiązaniem na dalszych etapach badań mogłoby być zastosowanie metryki taksówkowej, wydaje się jednak, że oszacowanie odległości (lub czasu) dojścia i/lub dojazdu z wykorzystaniem dostępnych w internecie map jest rozwiązaniem o większym stopniu precyzji.

Zbiór danych zawierał 130 transakcji¹⁶. Został on podzielony w sposób losowy na trzy podzbiory, których liczebność została podana w nawiasach: uczący (78), walidacyjny (21), testowy (31). Jako miernik jakości prognoz został przyjęty średni błąd bezwzględny (*mean absolute error* – MAE).

Najlepsze wyniki zostały osiągnięte dla sieci typu MLP o jednej warstwie ukrytej (sieć dwuwarstwowa). Prognoza obejmowała wartość pojedynczej transakcji¹⁷, zatem model posiadał jeden neuron w warstwie wyjściowej. Liczba komórek w pierwszej warstwie – wejściowej – była pochodną liczby użytych zmiennych objaśniających i dla opisanego modelu wyniosła dziewięć. Na wejściu wprowadzono takie zmienne, jak: powierzchnia nieruchomości lokalowej, powierzchnia wraz z pomieszczeniami przynależnymi, liczba pokoi, liczba izb, kondygnacja, stan budynku, przynależność piwnicy, odległości od placu Wolności i najbliższego terenu zieleni. Liczba neuronów w warstwie ukrytej została dobrana doświadczalnie. Badania empiryczne wskazały na minimalizację wartości miernika błęd przy zastosowaniu pięciu komórek nerwowych w opisaną warstwę.

Średnia wartość transakcji dla zbioru testowego wyniosła 170 631 zł (i była zbliżona do przeciętnej dla zbioru uczącego, która wyniosła 176 604,40 zł). Najmniejszy błąd prognozy wyniósł 23 715,44 zł, czyli poniżej 13,90% średniej wartości transakcji. Model ten cechował się jednak stosunkowo wysokim odchyleniem standardowym wynoszącym 38 369,54 zł.

Kolejnym wartym opisanym modelem był MLP również o architekturze trójwarstwowej (czyli złożony z czterech warstw), zbudowany z dziesięciu komórek w warstwie wejściowej, trzech w obu warstwach ukrytych oraz pojedynczego neuronu wyjściowego. Co prawda błąd prognozy okazał się być większy niż w poprzednio opisaną sieć i wyniósł 24 732,38 zł (ok. 14,49% średniej wartości transakcji), to jednocześnie uległa obniżeniu wartość odchylenia standardowego i wyniosła ona 35 012,13 zł. Model jako zmienne objaśniające zawierał: powierzchnię nieruchomości lokalowej, powierzchnię wraz z pomieszczeniami przynależnymi, liczbę pokoi, liczbę izb, kondygnację, stan budynku, fakt przynależności piwnicy, odległość od najbliższego terenu zieleni, czas dojazdu oraz długość trasy przejazdu samochodem do centrum handlowo-rozrywkowego Manufaktura.

¹⁶ Liczba ta jest stosunkowo mała. Zwiększenie ilości danych mogłoby się przyczynić do poprawy uzyskanych wyników. Należy jednak zauważyć, że obecna ilość danych umożliwia już stworzenie działającego modelu. W sposób empiryczny zostało wykazane, że graniczna ilość danych w zbiorze uczącym, poniżej której budowa modelu jest w praktyce niemożliwa, wynosi ok. 70 transakcji.

¹⁷ Modele prognozujące pozostałe zmienne wyjściowe – w tym wartość m² – cechowały się gorszymi wynikami. Cenę sprzedaży 1 m² nieruchomości można jednak łatwo obliczyć, dzieląc prognozowaną wartość transakcji przez powierzchnię nieruchomości.

Sieci jednokierunkowe trójwarstwowe wyposażone w większą liczbę neuronów mimo potencjalnie większych możliwości prognostycznych okazały się generować wyniki obarczone większymi wartościami miernika błędów. Taki stan rzeczy może wynikać z wielu powodów. Jednym z nich wydaje się być liczebność zbioru uczącego. Sieć o bardziej rozbudowanej architekturze – jak wykazują badania własne autorów, a także analiza literatury przedmiotu – wymaga w celu wytrenowania większej liczby komórek liczniejszych zbiorów danych. Niewystarczająca ilość informacji nie pozwala w takiej sytuacji na wydobycie potencjału tkwiącego w modelu o większym stopniu komplikacji.

W przypadku prognozy ceny m² nieruchomości lokalowej najlepszy z modeli charakteryzował się MAE w wysokości 534,19 zł (odchylenie standardowe 735,71). Jako zmienne wejściowe zostały użyte: powierzchnia nieruchomości lokalowej, powierzchnia wraz z pomieszczeniami przynależnymi, liczba pokoi, liczba izb, kondygnacja, fakt istnienia piwnicy, stan budynku, odległość od najbliższego terenu zieleni, czas dojazdu oraz długość trasy przejazdu samochodem do centrum handlowo-rozrywkowego Manufaktura. Oznacza to, że model dysponuje statystycznie niższą jakością prognozy od ANN szacujących wartość transakcji.

Podsumowanie

W procesie testowania wpływu zmiennych wejściowych modelu na jakość uzyskiwanej przez ANN prognozy wykazano, że:

1. Wpływ odległości nieruchomości lokalowej od „centralnego” punktu miasta, jakim jest plac Wolności, na wartość transakcji kupna-sprzedaży jest minimalny. Zauważalnie lepszymi zmiennymi są dane określające odległość od centrów handlowych lub handlowo-rozrywkowych.
2. Szacowany czas dotarcia do centrum rozrywkowo-handlowego stanowi parametr o podobnym stopniu istotności w procesie prognozy jak wyrażona w kilometrach odległość, jaką należy pokonać samochodem.
3. Odległości reprezentujące faktyczne czasy lub dystans dotarcia do centrum handlowo-rozrywkowego mają większy wpływ na uzyskane wyniki niż odległości obliczone w linii prostej.
4. Przynależność piwnicy do nieruchomości lokalowej stanowi jeden z najmniej znaczących spośród wszystkich testowanych parametrów¹⁸.

¹⁸ Istotność parametrów zmienia się w zależności od modelu, jednak fakt przynależności piwnicy do nieruchomości lokalowej w większości z nich stanowił element w najmniejszym stopniu wpływający na uzyskiwany wynik. W niewielkim procencie ANN czynnik ten stanowił drugi z najmniej istotnych parametrów.

5. Takie parametry, jak liczba kondygnacji oraz ogólny stan budynku, mają ograniczony wpływ na wartość transakcji oraz cenę m² nieruchomości lokalowej.
6. Potwierdzone zostało kluczowe znaczenie takich czynników, jak liczba pokoi, liczba izb¹⁹, powierzchnia nieruchomości (także w przypadku wyznaczania ceny m²), kondygnacja, na której znajduje się nieruchomość.

Badania empiryczne nad wyborem optymalnej struktury modelu ANN wskazują, iż:

1. Przy ograniczonej ilości danych uczących bardziej zasadne wydaje się być użycie sieci jednokierunkowych dwuwarstwowych lub sieci trójwarstwowej o niewielkiej liczbie neuronów w obu warstwach ukrytych.
2. W przypadku dokonywania klasyfikacji ceny m² nieruchomości do jednego z przedziałów większą precyzją cechowały się ANN o jednej zmiennej wyjściowej niż systemy o liczbie neuronów w warstwie ukrytej odpowiadającej liczbie przedziałów.
3. Wyższą jakość wyceny uzyskują modele szacujące wartość całej nieruchomości niż te, których zmienne wyjściowe opisują cenę m².

Badania potwierdziły opisywaną, szczególnie w literaturze zagranicznej, możliwość wykorzystania ANN do modelowania cen na rynku nieruchomości lokalowych. Należy przy tym nadmienić, że ze względu na zastosowane dane wejściowe model miał ograniczone możliwości precyzyjnego wyznaczenia ceny nieruchomości. Należy zatem opisane rozwiązanie rozumieć jako system szacowania cen nieruchomości, który stanowi dopiero punkt wyjścia w procesie określania ceny konkretnej nieruchomości lokalowej. Konieczność wprowadzania indywidualnej korekty w procesie wyceny bezpośrednio wynika z braku informacji o stanie, standardzie wyposażenia i rozkładzie pomieszczeń analizowanych nieruchomości lokalowych. Dane te mogą być – przynajmniej częściowo – uzyskane z lokalnych biur pośrednictwa sprzedaży nieruchomości (dla rynku wtórnego) lub z bazy BaRN (o ile dany rynek objęty jest zasięgiem bazy), co jednak nie było celem niniejszego opracowania.

¹⁹ W niemalże wszystkich modelach liczba pokoi miała większe znaczenie od liczby izb.

Literatura

- Antipov, E.A., Pokryshevskaya, E.B. (2012). Mass Appraisal of Residential Apartments: An Application of Random Forest for Valuation and a CART-based Approach for Model Diagnostics. *Expert Systems with Applications*, 39, 1772–1778. DOI: 10.1016/j.eswa.2011.08.077.
- Chiarazzo, V., Caggiani, L., Marinelli, M., Ottomanelli, M. (2014). A Neural Network Based Model for Real Estate Price Estimation Considering Environmental Quality of Property Location. *Transportation Research Procedia*, 3, 810–817. DOI: 10.1016/j.trpro.2014.10.067.
- Garcia, N., Gamez, M., Alfaro, E. (2008). ANN+GIS: An Automated System for Property Valuation. *Neurocomputing*, 71, 733–742. DOI: 10.1016/j.neucom.2007.07.031.
- Jasiński, T. (2003). Przegląd architektur sztucznych sieci neuronowych wykorzystywanych w ekonomii do przewidywania szeregów czasowych. *Studia i Prace Kolegium Zarządzania i Finansów*, 35, 166–177.
- Liu, X., Deng, Z., Wang, T. (2011). Real Estate Appraisal System Based on GIS and BP Neural Network. *Transactions of Nonferrous Metals Society of China*, 21, 626–630. DOI: 10.1016/S1003-6326(12)61652-5.
- Selim, H. (2009). Determinants of House Prices in Turkey: Hedonic Regression Versus Artificial Neural Network. *Expert Systems with Applications*, 36, 2843–2852. DOI: 10.1016/j.eswa.2008.01.044.
- Tadeusiewicz, R. (1998). *Elementarne wprowadzenie do techniki sieci neuronowych z przykładowymi programami*. Warszawa: Akademicka Oficyna Wydawnicza PLJ.

APARTMENT PRICES FORECASTING BY THE ARTIFICIAL NEURAL NETWORKS

Abstract

The paper describes apartment prices forecasting on the secondary market with the use of the artificial neural networks. The research done by authors is of the theoretical and empirical value. During the construction of the models have been tested numerous input variables. The study was conducted in the centre of Łódź. Among tested networks there were multilayer perceptron with one and two hidden layers as well as radial basis function networks.

Translated by Tomasz Jasiński

Keywords: artificial intelligence, neural networks, forecasting, prices, apartments
JEL Code: R32

