

Anna Kasznia

Uniwersytet Ekonomiczny w Katowicach

PORÓWNANIE PODEJŚCIA APROKSYMUJĄCEGO I KLASYFIKUJĄCEGO W PROGNOZOWANIU KURSÓW WYBRANYCH AKCJI NA GPW W WARSZAWIE S.A. Z UŻYCIEM JEDNOKIERUNKOWYCH SIECI NEURONOWYCH

Wykorzystywanie sztucznych sieci neuronowych do prognozowania szeregów czasowych ma długą historię. Już w 1964 roku Hu [10, s. 36] zastosował prostą sieć ADALINE do prognozowania pogody. Możliwości sieci bez warstw ukrytych były jednak mocno ograniczone. Wierzono, że sieci wielowarstwowe są w stanie pokonać ograniczenia perceptronów prostych, nie znano jednak skutecznych metod ich uczenia. Dopiero publikacja w 1986 roku dwutomowej monografii *Learning Internal Representation by Error Propagation* opracowanej przez McClellanda i Rumelharta na temat równoległego przetwarzania rozproszonego spowodowała gwałtowny przyływ zainteresowania techniką wstecznej propagacji błędów, podstawowego algorytmu uczenia sieci wielowarstwowych. W ten sposób rozpoczęła się nowa era dla niegdyś niedocenionych sieci wielowarstwowych, w tym także dla badań nad zdolnościami prognostycznymi modeli sztucznych sieci neuronowych. Już w 1987 roku Lapedes i Farber [10, s. 36] przeprowadzili badania, w których doszli do wniosku, że modele te mogą być zastosowane do modelowania i prognozowania nieliniowych szeregów czasowych. Jako że wiele procesów na gruncie ekonomii jest opisywanych za pomocą szeregów czasowych, a ich charakter jest nierzadko nieliniowy, sztuczne sieci neuronowe szybko znalazły zastosowanie także w tej dziedzinie. Od początku lat 90. z powodzeniem prowadzi się badania dotyczące ich użycia na rynku papierów wartościowych*.

Najpowszechniej wykorzystywane w prognozowaniu są sieci działające jako aproksymator funkcji [5, s. 174]. W podejściu tym badacz zmierza do tego, by nauczyć sieć z jak największą precyzją prognozować przyszłe wartości bada-

* Obszerne opisy zastosowań modeli sztucznych sieci neuronowych podano między innymi w pracach [8; 4; 7].

nego szeregu. W prognozach gospodarczych wartości te są używane do podejmowania decyzji (budowy, optymalizacji strategii). Powodzenie finansowych strategii inwestycyjnych zależy jednak w dużej mierze nie tyle od precyzji oszacowania, ile od sygnałów, jakie na podstawie odpowiedzi sieci są generowane. Sygnały te są zazwyczaj generowane w zależności od spodziewanego kierunku zmian prognozowanych wartości. Nasuwa się pytanie, czy nie lepiej starać się nauczyć sieć prognozowania kierunku zmian i podejść do problemu prognozowania jak do zadania klasyfikacji.

W artykule tym podjęto próbę odpowiedzi na pytanie, czy modele klasyfikujące są w stanie osiągnąć lepsze rezultaty w sensie prawidłowego prognozowania kierunków zmian szeregu i osiągniętego z wykorzystaniem wygenerowanych prognoz zysku. Do badań wykorzystano kursy akcji 3 wybranych spółek notowanych na Gieldzie Papierów Wartościowych w Warszawie. Dane pobrano ze strony bossa.pl, natomiast badania przeprowadzono w środowisku obliczeniowym MATLAB.

1. Podejście aproksymujące a klasyfikujące

W podejściu aproksymującym badacz, zaopatrując sieć w zbiór zmiennych wejściowych i odpowiadający tym próbkom zbiór pożądanych wyjść (celi), liczy na to, że sieć w trakcie procesu uczenia odnajdzie postać funkcji opisującej zależność zmiennej prognozowanej od wybranych zmiennych wejściowych. Tym samym nauczy się z jak największą precyzją prognozować przyszłe wartości szeregu. Podejście to oczywiście zakłada, że istnieje jakaś funkcja opisująca badaną zależność, dzięki której można wyznaczyć konkretne przyszłe wartości kursów, cen, stóp zwrotu itd. Estymacja parametrów takiego modelu dokonuje się poprzez minimalizację funkcji błędu* między prognozami sieci a wartościami rzeczywistymi. Wypracowane w toku uczenia rozwiązania sieci są przybliżone. Sam algorytm uczenia wymusza uśrednione odpowiedzi sieci, co powoduje, że dokładne przewidywanie poziomu dziennych cen, kursów itd. jest bardzo trudne.

W przypadku ekonomicznych systemów prognostycznych model szeregu czasowego stanowi element pomocniczy w procesie podejmowania decyzji. Stąd ważna jest nie tyle dokładność liczbowa prognoz, ile prawidłowe prognozowanie kierunków zmian i skuteczność strategii podjętej z użyciem sygnałów wygenerowanych przez sieć [7, s. 69]. To ostatecznie decyduje o użyteczności modelu, a nie niska wartość błędu prognoz. Jako alternatywne dla modeli aproksymujących można więc wykorzystać sieci realizujące zadanie klasyfikacji. W takim podej-

* Najczęściej jest to błąd średniokwadratowy MSE.

ściu w trakcie treningu nie zaopatruje się sieci w rzeczywiste wartości zmiennej jako celi, ale daje jej zestaw wzorców przeszłych zachowań zmiennej i trenuje tak, by prognozowała wystąpienie któregoś z tych wzorców, klasyfikując próbkę do jednej z arbitralnie określonych klas. W przypadku prognozowania finansowych szeregów czasowych klasy te mogą reprezentować sesje wzrostowe i spadkowe. Gdyby miał nastąpić wzrost prognozowanej zmiennej, wartość wyjścia sieci wynosiłaby 1, w przeciwnym wypadku 0.

Warto w tym momencie wspomnieć pracę Leunga [5], który wykazał, że modele prognostyczne oparte na kierunku zmian indeksów giełdowych (S&P 500, FTSE 100 i Nikkei 225) przewyższają wynikami modele bazujące na poziomie zmian i pozwalają osiągnąć większe zyski ze strategii inwestycyjnych.

1.1. Architektura sieci użytych do badań

W zasadzie dowolny typ sieci może zostać użyty do prognozowania szeregów czasowych [6, s. 55]. Dokonując wyboru typu sieci, należy uwzględnić te topologie, które będą w stanie wymodelować zależności występujące pomiędzy zestawem danych wejściowych i prognozowaną zmienną. Typ sieci musi być odpowiedni do realizacji odpowiedniej formy przetwarzania sygnałów wejściowych regresji lub klasyfikacji w zależności od rodzaju rozważanego problemu. W przeważającej liczbie eksperymentów dotyczących prognozowania finansowych szeregów czasowych wykorzystuje się sieci jednokierunkowe wielowarstwowe typu perceptron (MLP) [10, s. 37], dlatego ten typ sieci został użyty do badań. W zależności od architektury i użytych funkcji aktywacji w poszczególnych warstwach sieć ta może realizować zadanie zarówno aproksymacji, jak i klasyfikacji, a także, w obu formach, zostanie użyta do badań.

Drugi typ sieci użyty w eksperymentach to probabilistyczna sieć neuronowa (PNN). W przeciwieństwie do sieci typu MLP probabilistyczne sieci neuronowe nie generują ciągłych wartości wyjściowych, a dzielą zbiór danych na ustaloną liczbę kategorii wyjściowych. Stąd też sieci te stosuje się w problemach klasyfikacyjnych. Wykazano, że sieci te z powodzeniem radzą sobie w przewidywaniu kierunku zmian finansowych szeregów czasowych. Kamruzzaman i in. [4] podają przykłady badań, w których zastosowano sieci typu PNN do budowy strategii inwestycyjnych. Wymieniają między innymi pracę Aggarwala i Demaskeya, których efektywność strategii hedgingowej znacznie się poprawiła po tym, jak zaimplementowano w niej wyniki predykcji kierunku zmian kursów wymiany dokonanych z użyciem PNN. W innym przytoczonym badaniu Marbely, wykorzystując PNN, odkrył związek między kierunkiem zmian cen typu intraday i interday na rynku Forex.

Sieć typu PNN uczy się obliczania prawdopodobieństw warunkowych a posteriori przynależności wektora wejściowego do każdej z wydzielonych klas. W sieciach PNN wektory uczące są zakodowane w wagach pomiędzy warstwą wejściową a wyjściową jako wektory-wzorce. Tak naprawdę więc w sieciach typu PNN nie ma procesu uczenia, a jedynie zapamiętywanie wzorców poprzez ich zakodowanie w wagach pomiędzy warstwą wejściową a ukrytą, a następnie ważenie [3, s. 104].

Dla każdego wektora wejściowego oblicza się odległość (euklidesową lub inną) między tym przypadkiem a wektorem wag (w charakterze wag wykorzystuje się wektory wejściowe poszczególnych wzorców). Tak wyznaczona wartość staje się parametrem radialnej funkcji aktywacji neuronów warstwy ukrytej (najczęściej jest wykorzystywana funkcja gaussowska). Następnie w neuronach warstwy wyjściowej oblicza się sumę tych wartości dla wszystkich elementów zbioru uczącego każdej klasy. Warstwa wyjściowa jest konkurencyjna, tzn. funkcja aktywacji neuronów wyjściowych przyporządkowuje 1 największej sumie i 0 pozostałym, tym samym klasyfikuje przypadek do klasy o największej sumie, gdyż prawdopodobieństwo, że to właśnie do tej klasy należy przypadek, jest największe. Tabela 1 przedstawia porównanie architektur sieci użytych do eksperymentów.

Tabela 1

Podstawowe cechy modeli użytych w badaniu

	MLP – aproksymująca	MLP – klasyfikująca	PNN
Ilość warstw	2	2	2
Funkcja aktywacji w warstwie ukrytej	tangens hiperboliczny	tangens hiperboliczny	radialna
Funkcja w warstwie wyjściowej	liniowa	tangens hiperboliczny	konkurencyjna
Liczba wyjść	1	2	2
Zmienne wyjściowe	prognoza kursu na 1 sesję do przodu	sesja wzrostowa/ /sesja spadkowa (0-1)	sesja wzrostowa/ /sesja spadkowa (0-1)

1.2. Dane uczące

Do badań wykorzystano szeregi kursów zamknięcia dla akcji 3 spółek notowanych na GPW w Warszawie należących do indeksu WIG 20, takich jak: Telekomunikacja Polska SA (TPS), PKO Bank Polski (PKO) oraz KGHM Polska Miedź SA (KGH), odpowiednio z sektora usług, finansów i przemysłu.

W literaturze przedmiotu nie ma zgodności co do optymalnej wielkości zbioru uczącego. Zdecydowano się więc dla każdej ze spółek wyznaczyć wielkość zbioru treningowego, posługując się algorytmem Mean-Change-Point Test

zaproponowanym przez Huanga^{*}. Polega on na wyznaczeniu optymalnego rozmiaru zbioru uczącego poprzez znalezienie punktu w szeregu czasowym, który dzieli ten szereg na dwie części o możliwie najbardziej różnych średnich. Minimalną wielkość zbioru ustalono na dwa lata [9, s. 203-222]. W badaniach użyto danych od momentu wejścia spółki na giełdę do końca 2010 roku. Otrzymane w ten sposób wielkości zbiorów uczących przedstawia tabela 2.

Tabela 2

Zakres czasowy zbiorów uczących

KGH	14.12.2005-31.12.2010
PKO	7.10.2008-31.12.2010
TPS	3.02.2009-31.12.2010

W trakcie treningu sieci typu MLP zostało wybranych losowo po 15% próbek ze zbioru uczącego do zbioru testowego i walidacyjnego. W uczeniu sieci typu PNN brał udział cały zbiór uczący. Wyniki symulacji sieci na danych z pierwszej połowy 2011 roku zostały wykorzystane do ostatecznej weryfikacji jakości działania opracowanych modeli.

Zmienne wejściowe

Rolę zmiennych wejściowych pełniły:

- opóźnione realizacje szeregu, tj. kursy akcji z ostatnich 20 sesji,
- trzy średnie ruchome, 4-, 9- oraz 12-dniowe wraz z ostatnią dzienną stopą zwrotu.

Początkowo rolę zmiennych wejściowych pełniły tylko opóźnione realizacje kursów akcji, jednak modele prognozujące wartość kursu akcji, bazując na takich zmiennych, nie były w stanie wygenerować dobrych prognoz. Wynika to z tego, że sieci prognozujące przyszłe wartości szeregu na podstawie wcześniejszych jego realizacji mają tendencję do ustalania wag w trakcie uczenia na takim poziomie, że wygenerowane prognozy są praktycznie przesuniętym w czasie o jeden okres szeregiem. Nie chcąc więc skazywać tych modeli z góry na niepowodzenie, zdecydowano się na przeprowadzenie symulacji także z wykorzystaniem wskaźników analizy technicznej. Dane zostały przeskalowane, tak by zawierały się w przedziale $[-1;1]$. Wszystkich symulacji dokonano w programie Matlab, wykorzystując bibliotekę Neural Networks Toolbox. Jakość modelu została oceniona na podstawie dwóch podstawowych kryteriów:

- zgodności kierunków zmian (DS),
- hipotetycznego zysku wygenerowanego dzięki strategii opartej na wynikach sieci.

^{*} Algorytm postępowania szczegółowo przedstawiono w [2].

2. Symulacje

Liczbę neuronów ukrytych w warstwach sieci typu MLP ustalono metodą prób i błędów, rozpoczynając w każdym przypadku od minimalnej [6, s. 170] liczby neuronów, jednocześnie obserwując błąd na zbiorze walidacyjnym i testowym, zwiększając ich liczbę aż do uzyskania satysfakcjonującego poziomu. Po wyborze najlepszych modeli przeprowadzono symulacje dla danych, które nie uczestniczyły w uczeniu, tj. sesji od 3.01.2011-30.06.2011. Dla uzyskanych prognoz obliczono miarę zgodności kierunków zmian (w ilu procentach przypadków prognozowany kierunek zmian był taki jak w rzeczywistości). Wyniki przedstawia tabela 3.

Dla każdego ze współczynników zgodności zmian sprawdzono efektywność predykcji, testując przy poziomie istotności 0,05 hipotezę, czy proporcja trafnie zakwalifikowanych prognoz jest znacząco większa od 50%, czyli wyniku, jaki można by uzyskać przez tzw. ślepy traf [5, s. 184]. W tabeli 3 wartości miary istotnie większe od 50% zostały oznaczone gwiazdką. Dodatkowo pogrubioną czcionką wyróżniono największe wartości współczynnika DS dla każdej ze spółek.

Tabela 3

Wartości współczynników zgodności kierunków zmian dla danych od 3.01 do 30.06.2011

		KGH	PKO	TPS
1.	PNN (zm. opóźnione)	0,62*	0,58*	0,56
2.	PNN (zm. analizy tech.)	0,52	0,65*	0,50
3.	MLP – klasyfikująca (zm. opóźnione)	0,54	0,62*	0,58*
4.	MLP – klasyfikująca (zm. analizy tech.)	0,58*	0,54	0,52
5.	MLP – aproksymująca (zm. opóźnione)	0,52	0,55	0,49
6.	MLP – aproksymująca (zm. analizy tech.)	0,53	0,60*	0,58*

Jak widać, w grupie modeli opartych na opóźnionych zmiennych tylko modele klasyfikujące osiągnęły wyniki znacząco lepsze od 50% (cztery z sześciu modeli). W grupie modeli opartych na wskaźnikach analizy technicznej były to natomiast aż dwa z trzech modeli aproksymujących. Co więcej, w grupie tej tylko dwa z czterech modeli klasyfikujących osiągnęły takie wyniki (PKO 0,65 oraz KGH 0,58). Nie zmienia to faktu, że dla każdej spółki modelem o najwyższej wartości współczynnika DS był model klasyfikujący. Chcąc zbadać, czy wielkość osiągniętych miar przekłada się na zysk, jaki można osiągnąć na podstawie wygenerowanych prognoz, zaimplementowano prostą strategię inwestycyjną.

W przypadku modeli aproksymujących, jeśli sieć prognozuje jutrzejszy kurs na poziomie wyższym niż dzisiejszy, oznacza to sygnał kupna, w przeciwnym wypadku – sprzedaży. Dla modeli klasyfikujących decyzje kupna-sprzedaży

są dokonywane na podstawie prognoz, czy następna sesja będzie wzrostowa, czy spadkowa. Transakcja jest dokonywana w momencie, gdy rekomendacja sugeruje zajęcie pozycji odwrotnej do dotychczasowej. Wartości zysku zostały odniesione do zysku, jaki zostałby otrzymany z zastosowaniem strategii kup i trzymaj, tj. gdyby dokonano kupna w pierwszym dniu badanego okresu i sprzedano posiadane akcje po kursie z ostatniego dnia. Tak skonstruowany benchmark pozwala ocenić, czy osiągnięty zysk jest zasługą zastosowanego modelu, czy też panującego na rynku trendu. Hipotetyczny kapitał początkowy wyniósł 100.

Wyniki przedstawiające kapitał końcowy uzyskany z zastosowanej strategii na dzień 30.06.2011 przedstawiono w tabeli 4. Kwoty ujęte w nawiasach to wartości zysku uwzględniające prowizję maklerską w wysokości 0,4% wartości transakcji. Wartości wyróżnione pogrubioną czcionką oznaczają zysk, jaki osiągnięto dla modeli wykazujących najwyższe wartości miary DS. Modele te w przypadku dwóch spółek – KGHM oraz PKO BP – wykazują również największy zysk. Dla spółki TP SA wartość zysku jest minimalnie niższa od najlepszego pod względem zysku modelu (106,55 w porównaniu do 108,20). Nie oznacza to jednak, że wysoka wartość miary DS zawsze przekłada się na osiągnięty zysk. Dla przykładu największą wartością zysku dla spółki TP SA wykazał się model 2, przewidujący prawidłowo kierunek zmian tylko w 50% przypadków. Natomiast model 6, dla którego miara DS wynosiła 58%, nie tylko przyniósł stratę, ale także nie udało mu się osiągnąć zysków większych niż dla strategii „kup i trzymaj”.

Tabela 4

Wartości zysku dla danych od 3.01 do 30.06.2011

		KGH	PKO	TPS
1.	PNN (zm. opóźnione)	148,29 (125,49)	122,53 (101,73)	106,45 (80,85)
2.	PNN (zm. analizy tech.)	116,99 (102,99)	135,61 (116,81)	112,60 (108,20)
3.	MLP – klasyfikująca (zm. opóźnione)	130,42 (125,62)	113,59 (105,19)	119,75 (106,55)
4.	MLP – klasyfikująca (zm. analizy tech.)	131,70 (113,30)	119,25 (114,05)	112,20 (106,60)
5.	MLP – aproksymująca (zm. opóźnione)	116,19 (112,19)	111,68 (98,88)	101,39 (82,86)
6.	MLP – aproksymująca (zm. analizy tech.)	113,60 (105,60)	120,71 (105,91)	118,77 (97,57)
7.	Strategia „kup i trzymaj”	115,03	96,77	99,88

Na podstawie przeprowadzonych symulacji nie da się jednoznacznie określić, który model daje znacząco lepsze rezultaty. Dla spółki KHGM były to modele klasyfikujące bazujące na opóźnionych zmiennych (model 1 i 3), dla spółki PKO BP – oparte na wskaźnikach analizy technicznej (model 2 i 4). W przypad-

ku modeli dla TP SA oba podejścia dały podobne rezultaty (modele 2, 3 i 4). Niemniej jednak to strategie wspierane sygnałami generowanymi przez modele klasyfikujące okazały się bardziej zyskowne od tych, które wykorzystywały prognozy modeli aproksymujących.

Wnioski

W artykule podjęto próbę odpowiedzi na pytanie, czy użycie do prognozowania finansowych szeregów czasowych sieci realizujących zadanie klasyfikacji przynosi lepsze rezultaty w sensie prawidłowego prognozowania kierunku zmian i osiągniętego zysku niż klasyczne podejście polegające na prognozowaniu konkretnej przyszłej wartości szeregu. Do badań wykorzystano kursy akcji 3 wybranych spółek notowanych na GPW w Warszawie. Przeprowadzone symulacje pokazują, że choć niektóre z modeli aproksymujących osiągnęły satysfakcjonujące rezultaty pod względem miary DS, to najlepsze wyniki dla każdej spółki wykazały sieci klasyfikujące. Również w kategoriach osiągniętego zysku modele te przyniosły lepsze rezultaty. Odnotowano jednak, że wysoki współczynnik zgodności kierunków zmian nie zawsze przekłada się na osiągnięty zysk. Przeprowadzone badania wskazują również na konieczność poświęcenia większej uwagi doborowi odpowiednich zmiennych wejściowych.

Literatura

1. Gately E.J., *Sieci neuronowe. Prognozowanie finansowe i projektowanie systemów transakcyjnych*, WIG-Press, Warszawa 1999.
2. Huang W., Nakamori Y., Wang S., Zhang H., *Select the Size of Training Set for Financial Forecasting with Neural Networks*, Proceedings of the International Symposium on Neural Networks, Springer-Verlag, Berlin-Heidelberg 2005, s. 879-884.
3. Jankowski N., *Ontogeniczne sieci neuronowe*, EXIT, Warszawa 2003.
4. Kamruzzaman J., Begg R., Sarker R., *Artificial Neural Networks in Finance and Manufacturing*, Idea Group Inc., 2006.
5. Leung M.T., Daouk H., Chen A., *Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models*, „International Journal of Forecasting” 2000, Vol. 16, s. 173-190.
6. Masters T., *Sieci neuronowe w praktyce. Programowanie w języku C++*, WNT, Warszawa 1996.

7. *Neural Networks in the Capital Markets*, red. A.-P. Refenes, John Wiley & Sons Ltd, 1995.
8. Trippi R., Lee J.K., *Artificial Intelligence in Finance and Investments*, IRWIN, 1996.
9. Walczak S., *An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks*, „Journal of Management Information Systems” 2001, Vol. 17, No. 4, Spring, s. 203-222.
10. Zhang G., Eddy Patuwo B., Hu M.Y., *Forecasting with Artificial Neural Networks: The State of Art*, „International Journal of Forecasting” 1998, Vol. 14, s. 35-62.

FORECASTING STOCK PRICES USING FEED-FORWARD NEURAL NETWORK – A COMPARISON OF APPROXIMATION AND CLASSIFICATION APPROACHES

Summary

In this paper two approaches to financial time series forecasting using neural networks were compared. First one, the function approximation approach, in which neural networks are trained to forecast the exact one day ahead value of stock price. And the second one, classification approach, in which the output variable is the direction of future stock price movements. The aim of this work was to check if using the classification models can lead to better results in terms of direction of change forecasting and profits generated by their forecasts. This research was conducted on the basis of the time series of daily closing stock prices for three companies listed on the Warsaw Stock Exchange. Simulations show that some of the approximating models achieved satisfactory results in terms of the directional symmetry measure, although the best results for each of the analyzed company have been achieved for classification models.