

Sieci neuronowe jako narzędzie do predykcji zachowań giełdy papierów wartościowych

Neural networks as a tool to predict the behaviours of exchange stock markets

Karolina Plawgo, Marian Czerwiński

Wrocławska Wyższa Szkoła Informatyki
Stosowanej
ul. Wejherowska 28, 54-239 Wrocław

Treść. Celem badań było znalezienie odpowiedzi na pytanie czy rynek akcji zachowuje się w przypadkowy sposób, nie posiadając właściwie przewidywalnych trendów, czy też trendy te mogą być prognozowane. W celu uzyskania odpowiedzi na to pytanie zostały skonstruowane odpowiednie sieci neuronowe w oparciu o pakiet programu Qnet 2000 [1], które następnie poddano trenowaniu i testowaniu. Z przeprowadzonych analiz wynika, iż relatywnie proste modele mogą dawać dobre wyniki w prognozowaniu indeksu wszystkich spółek WIG-u. Interesującą informacją jest brak możliwości predykcji dla indeksu 20-stu spółek (WIG20) oraz wniosek, że nie można oczekiwać odpowiedzi poprawnej od sieci co do zachowania się cen dla wybranych spółek na podstawie zachowania się cen pozostałych.

Słowa kluczowe: sieci neuronowe, sztuczna inteligencja, prognozowanie wskaźników finansowych, Warszawska Giełda Papierów Wartościowych.

Abstract. The purpose of the researches was to find the reply to a question whether the stock market behaves in an incidental way, without any predictable trends or whether these trends might be predictable. In order to find the reply for these questions, the proper neural networks were build based on Qnet 2000 programme [1], and they were subsequently trained and tested. The analysis which were carried out revealed that the relatively simple models might give good results in forecasting Warsaw Stock Exchange Market Index (WIG). Interesting information is the fact, that there is no possibility to predict the twenty biggest companies' index (WIG20) and the conclusion that we cannot expect from the neural networks the right answer concerning changes of a company's share price based on changes in share prices of the rest of the companies.

Keywords: neural networks, artificial intelligence,

forecasting financial index, Warsaw Exchange Stock Market

1. Wprowadzenie

Obserwuje się coraz większą różnorodność metod, które są stosowane do analizy danych finansowych, a w szczególności do analizy finansowych szeregów czasowych. Zaczyna się sięgać po bardziej skomplikowane (niż analiza techniczna) metody prognozowania finansowego. Z reguły wymagają one stosowania szybkich komputerów i specjalistycznego oprogramowania. Jedną z przyczyn tego faktu jest niewątpliwie to, że rozwój technologii komputerowej umożliwia implementację nawet bardzo skomplikowanych metod matematycznych.

Sztuczne sieci neuronowe (dalej SN) stanowią jedną z najbardziej dynamicznie rozwijających się obecnie gałęzi sztucznej inteligencji. Obserwowane jest, w prowadzonych badaniach naukowych, przesuwanie akcentu z badań podstawowych w kierunku badań związanymi z konkretnymi zastosowaniami. Sieci neuronowe okazały się wygodnym narzędziem, przydatnym do realizacji bardzo wielu różnych praktycznych zadań. W istocie są one z powodzeniem stosowane w niezwykle szerokim zakresie problemów, w tak różniących się od siebie dziedzinach jak finanse, medycyna, zastosowania inżynierskie, geologia czy fizyka. Szeroki zakres zastosowań sieci neuronowych obejmuje również zagadnienia z zakresu nauk ekonomicznych, które do tej pory badano głównie za pomocą modeli statystycznych, ekonometrycznych, czy optymalizacyjnych. Należy mieć na uwadze fakt, iż krytycznymi parametrami przy rozwiązywaniu skomplikowanych problemów optymalizacyjnych są często ograniczenia czasowe bądź ograniczenia sprzętowe. Sieci neuronowe stosowane do rozwiązywania tego typu problemów osiągają bardzo dobre rezultaty w krótkim czasie i przy ograniczonych wymaganiach sprzętowych. Dodatkowo, bardzo istotną zaletą sieci neuronowych jest uniwersalność stosowanej metodologii rozumiana jako niezależność od szczegółowej definicji oraz danych problemu (w ramach określonej grupy problemów). Sieci neuronowe posiadają również zdolność uogólniania zdobytej wiedzy oraz ewolucyjnej auto-poprawy efektywności swojego działania. Konsekwencją takiego stanu rzeczy jest możliwość rozwiązywania przez sieć neuronową po okresie nauki nie tylko problemów treningowych ale również – co istotne – nieznanymi problemami pokrewnymi do tych na których była trenowana [3].

Efektywność sieci neuronowych wynika ze stosowania masowego przetwarzania równoległego w oparciu o bardzo dużą liczbę nieskomplikowanych elementów przetwarzających nazwanych neuronami.

Zastosowanie wielu prostych elementów przetwarzających funkcjonujących w trybie równoległym w miejsce jednego czy kilku wyspecjalizowanych procesorów powoduje, że obok zwiększania szybkości przetwarzania uzyskuje się efekt odporności na ewentualne błędy czy zakłócenia w funkcjonowaniu poszczególnych neuronów. Odporność na błędy wynika z zamierzonej redundancji elementów przetwarzających, z implementacji mechanizmów kontrolnych i autokorekcyjnych w sieciach neuronowych [7].

Uniwersalność metodologii, zdolność generalizacji i autokorekcji, relatywnie małe wymagania czasowe i sprzętowe czynią z sieci neuronowych atrakcyjne i efektywne narzędzie do rozwiązywania wielu problemów np.: przewidywania zachowań (predykcji), grupowania danych, rozpoznawania obrazów, optymalizacji i innych. Metody analizy finansowych szeregów czasowych przy użyciu sieci neuronowych należy zaliczyć do grupy metod ilościowych wywodzących się z szeroko rozumianego pojęcia dochodu. Siłą napędową, która spowodowała ich rozwój, była chęć stworzenia metody prognozowania wskaźników finansowych (w szczególności kursów akcji), których stosowanie na rynku przynosiłoby ponadprzeciętne dochody [5].

Wspólną cechą metod wykorzystywanych do prognozowania finansowego jest założenie braku efektywności rynku, nawet w słabej formie, tzn. zakłada się, że informacje o przeszłych cenach finansowych (np. kursach akcji) nie są odzwierciedlone w cenie, czyli ma sens określenie prognozy ceny, której wykorzystanie prowadziło do uzyskania ponadprzeciętnych dochodów [2].

W tej pracy opisane zostaną wyniki zastosowania SN do badania predykcji indeksów giełdowych na przykładzie Warszawskiej Giełdy Papierów wartościowych (WGPW). Efektywne narzędzia do przewidywania indeksów giełdy a w konsekwencji kursu akcji są z pewnością informacją wielce pożądaną przez inwestorów. Indeks giełdowy jest to podstawowa charakterystyka giełdy. Wartość indeksu giełdowego jest najprostszą odpowiedzią na podstawowe pytanie inwestora: „co się dzieje na giełdzie?”.

Wykorzystanie odpowiednio zbudowanych i wuczonych sieci neuronowych może przyczynić się do usprawnienia kontrolowania i sterowania złożonymi procesami występującymi także w innych rodzajach działalności gospodarczej.

2. Konstrukcja sieci neuronowej do problemu predykcji

Rozpoczynając budowę sieci musimy określić kilka istotnych jej parametrów. Zasadniczą kwestią jest do-

branie odpowiedniej ilości neuronów. Zbyt mała ilość neuronów doprowadzi do braku zbieżności sieci. W zasadzie dla większości problemów nie ma jednoznacznych metod wyznaczenia minimalnej ilości neuronów. Ilość neuronów wejściowych i wyjściowych jest determinowana przez rozwiązywany problem. Ustala się ją doświadczalnie [6].

Sieci neuronowe mają budowę warstwową, która daje się łatwo zdefiniować i szczegółowo opisać nawet w tym przypadku, gdy ilość neuronów wchodzących w jej skład jest bardzo duża. Do pełnego zdefiniowania sieci warstwowej wystarczy podać ilość neuronów w każdej warstwie. Przy innej, w szczególności tzw. dowolnej architekturze sieci - wysiłek związany z definicją tej topologii może być nieakceptowalnie duży. Określenie liczby warstw ukrytych i ilości neuronów znajdujących się w tych warstwach nie jest prostym zadaniem. Jako punkt wyjścia można przyjąć sieć z jedną warstwą ukrytą, zawierającą taką ilość neuronów, która jest równa połowie sumy ilości neuronów wejściowych i ilości neuronów wyjściowych, jednak na ogół najlepsze wyniki otrzymuje się, wybierając te ilości w sposób empiryczny [8].

Kolejnym problemem wymagającym wyjaśnienia jest kwestia struktury połączeń między neuronami. Jak wiadomo, struktura połączeń ma wpływ na działanie sieci, więc jej racjonalny wybór może znacznie przyspieszyć proces jej uczenia.

W celu znalezienia odpowiedniej struktury, można odwołać się do następującego rozumowania. Jednym z ważnych powodów stosowania sieci neuronowych jest fakt, że dla zadań, które chcemy powierzyć do rozwiązania budowanej sieci nie znamy dobrej algorytmicznej metody ich rozwiązania. Jednak skoro nie wiemy, jak postawione zadanie trzeba rozwiązać (chcemy, żeby sieć sama to ustaliła na podstawie przykładów podawanych w trakcie uczenia), to zwykle nie wiemy również z góry, jakie drogi przesyłania sygnałów będą przy tym rozwiązaniu potrzebne, a jakie nie. Nie potrafimy zatem a priori powiedzieć, które połączenia w sieci będą potrzebne, a które nie. Takim rozwiązaniem jest połączenie typu „każdy z każdym” [8]. Przy spełnieniu tej zasady każdy neuron ukryty i każdy neuron wyjściowy jest połączony z każdym neuronem z warstwy poprzedniej. Taka metoda jest kosztowna (czas obliczeń i konieczność posiadania dużego zapasu pamięci).

Dzięki zdefiniowaniu połączeń według zasady „każdy z każdym” niczego z góry nie wyklucza się i dopiero proces uczenia formuje ostateczną strukturę sieci, ustalając niezerowe wartości współczynników wagowych tylko na niektórych (na ogół relatywnie nielicznych) drogach przepływu sygnałów. Te połączenia można potraktować jako nie istniejące i usunąć z sieci.

Musimy rozstrzygnąć także, czy badany problem

jest o charakterze liniowym czy nieliniowym. Stosujemy inne podejście w konstrukcji sieci do problemów liniowych, a inne do nieliniowych. Popęlenie błędu w tym zakresie jest niebezpieczne. Może bowiem okazać się, że badany system jest nieliniowy, jednak okresowo zachowuje się liniowo. Jeżeli analityk zna jedynie okresy liniowe to zastosuje sieć liniową, której dopasowanie będzie doskonałe do rozwiązywanego problemu. Przez pewien okres prognozy będą także bardzo dobre. Po pewnym czasie okaże się jednak, że prognozy pogarszają się – a wkrótce, że nie nadają się do dalszych analiz.

Stosunkowo bezpiecznie jest założyć, że procesy na rynku kapitałowym są nieliniowe. W zdecydowanej większości przypadków jest to prawda. W pozostałych przypadkach sieć nieliniowa i tak dobrze poradzi sobie z liniowymi problemami. Kosztem będzie jedynie dłuższy czas analiz [9].

Pojawia się tutaj problem zbieżności sieci – czyli możliwości nauczenia się problemu przez sieć. Podejście liniowe dla problemów nieliniowych nie doprowadzi do osiągnięcia zbieżności. Sieć nie będzie w stanie się go nauczyć. Sieć nieliniowa nauczy się bez trudu problemu liniowego – jednak najczęściej kosztem dłuższego czasu pracy i (czasami) nieznacznego pogorszenia zdolności prognostycznych. Oczywiście nie jest to jedyny warunek osiągnięcia zbieżności przez sieć.

Budując sieć musimy sprawdzić obszar zmienności analizowanego szeregu. Z niego wynikać będzie postać funkcji aktywacji. Jeżeli analizujemy problem liniowy wybierzemy liniową funkcję aktywacji (purelin). Dla problemów nieliniowych wybierzemy jedną z nieliniowych: sigmoidalną, tangens hiperboliczny lub inną. Sigmoidę wybierzemy, gdy wartości badanego szeregu są wyłącznie dodatnie. Tangens hiperboliczny będzie lepszy, gdy wartości są ujemne i dodatnie. Błędny wybór funkcji aktywacji będzie powodował problemy ze zbieżnością i generalizacją sieci.

Przechodząc do problemu wyboru algorytmu uczenia sieci neuronowej należy odnotować fakt, że najczęściej używanym algorytmem uczenia jest metoda wstecznej propagacji błędu (BP). Po drugie BP jest metoda powolną, ale dość pewną - to znaczy za pomocą tej metody można rozwiązać praktycznie każde zadanie, podczas gdy metody alternatywne w niektórych przypadkach gwarantują bardzo dużą szybkość uczenia jednakże niską praktyczną użyteczność [4].

Podczas eksploatacji sieci neuronowej największe znaczenie ma to, aby sieć osiągała jak najmniejszy błąd związany z prezentacją nowego przypadku. Inaczej mówiąc, najbardziej pożądaną cechą sieci jest zdolność do generalizacji wiedzy na nowe przypadki. Tymczasem w rzeczywistości sieć uczona jest w sposób zapewniający minimalizację błędu wyłącznie dla zbioru

uczącego, co nie jest tym samym co minimalizacja rzeczywistego błędu. Związane jest to ze zjawiskiem tzw. przeuczenia sieci (problem nadmiernego dopasowania). Sieć z większą liczbą wag może modelować bardziej złożone funkcje i z tego powodu ma większą skłonność zbytniego dopasowania się do danych. Sieć z mniejszą liczbą wag może z kolei nie być dostatecznie dobrym narzędziem do opisu występującej w rzeczywistości zależności.

Na przykład, sieć nie posiadająca warstw ukrytych może modelować wyłącznie proste zależności liniowe. Natomiast sieć o zbyt dużej liczbie warstw i zbyt dużej ilości neuronów w warstwach ukrytych będzie miała skłonność do uczenia się „na pamięć” całego zbioru uczącego. W związku z tym pojawia się pytanie dotyczące sposobu wyboru sieci o właściwej złożoności.

Rozwiązaniem powyżej zasygnalizowanego problemu może być użycie procesu walidacji. Polega on na tym, że pewna liczba przypadków uczących jest zaliczana do oddzielnej grupy. Dane znajdujące się w tej oddzielnej grupie nie są bezpośrednio stosowane w trakcie uczenia sieci. Natomiast są one wykorzystane do przeprowadzenia niezależnej kontroli postępów algorytmu uczenia [8].

Jeśli jakość odpowiedzi sieci na dane uczące i na dane walidacyjne nie jest przynajmniej w przybliżeniu identyczna, to najprawdopodobniej podział przypadków między dwa zbiory był obciążony jakąś ukrytą tendencją - zaleca się w takim przypadku przerwanie uczenia i ponowny (losowy) podział posiadanych danych na część uczącą i część walidacyjną. W oparciu o ten sposób można ujawnić w trakcie procesu uczenia tendencje sieci do przeuczenia. W takim przypadku wskazane jest, aby zmniejszyć ilość neuronów ukrytych i/lub liczbę warstw ukrytych. W sytuacji przeciwnej, gdy sieć nie posiada dostatecznych możliwości do modelowania rzeczywistej funkcji, przeuczenie nie pojawia się, ale wtedy mimo długiego czasu uczenia błąd uczenia, ani błąd walidacyjny nie spadnie do satysfakcjonującego poziomu (Hertz, Krogh, Palmer, 1995).

W celu zwiększenia poziomu zaufania do ostatecznego modelu zwykle praktykuje się wydzielenie trzeciego zbioru przypadków - tak zwanego zbioru testowego. Ostateczna postać modelu (sieci) sprawdzana jest za pomocą zbioru testowego (zbiór ten używany jest tylko raz).

Kolejnym istotnym parametrem jest liczba „epok” uczenia sieci (jedna prezentacja podzbioru próbek uczących wraz z odpowiednią korekcją wag). Liczba próbek podzbioru jest nazywana rozmiarem epoki. Jeżeli będzie to liczba zbyt mała to sieć nie zdąży się nauczyć problemu. Zbyt wielka liczba epok z kolei doprowadzi do znacznego wydłużenia czasu uczenia się sieci i zwiększy ryzyko jej przeuczenia. Parametr ten

stosunkowo łatwo ustalić eksperymentalnie. Obserwując wykres uczenia się sieci widzimy, że po przekroczeniu pewnej liczby epok proces ten praktycznie nie postępuje [8].

Następnym parametrem jest określenie dopuszczalnego błędu, czyli żądanej jakości dopasowania sieci do danych empirycznych. Najczęściej jakość mierzy się przy pomocy średniego błędu kwadratowego (RMS). W niektórych typach sieci stosuje się również inne miary np.: SSE (Sum squared error performance function), MAE (Mean absolute error) i inne.

Po osiągnięciu tej wartości algorytm uczenia sieci zostanie przerwany. Gdy ustalimy ten parametr na zbyt niskim poziomie, sieć zostanie słabo wyuczona, natomiast gdy na zbyt wysokim – może go nigdy nie osiągnąć. Nawet gdybyśmy kontynuowali naukę sieci, zwiększając liczbę epok, to nie ma ona szans osiągnąć założonego pułapu. Często w analizie szeregów czasowych MSE ustala się na poziomie 1% przeciętnej wartości badanej cechy.

Kolejnym parametrem jest krok uczenia, który dotyczy wielkości zmian wag neuronów w każdym kolejnym cyklu uczenia. Zbyt duży krok spowoduje problemy ze zbieżnością sieci. Zbyt mały krok spowoduje olbrzymie straty czasu i nadwrażliwość sieci [6].

Jeszcze jednym ważnym parametrem sieci jest czas jej uczenia. Czas potrzebny na uczenie sieci rośnie wykładniczo wraz ze wzrostem liczby danych wejściowych. Przy stałej liczbie wejść rośnie wykładniczo wraz z ze wzrostem ilości neuronów. Natomiast rośnie liniowo wraz ze wzrostem liczby epok uczenia. Bardzo często przy obecnych algorytmach i ograniczonej mocy komputerów, czas budowy rozsądnej sieci może przewyższyć wielokrotnie cały dostępny czas, po upływie którego, sieć jest już niepotrzebna. Znany jest klasyczny paradoks, że „zrobienie prognozy na jutro może potrwać trzy dni”.

Kluczowym problemem we wszystkich badaniach ilościowych jest dostępność i jakość danych. W zasadzie nie ma możliwości zdobycia wszystkich potrzebnych danych o odpowiedniej jakości. Na problem ilości danych prawie nigdy nie mamy wpływu, gdyż jest on ograniczony kosztami, możliwościami technicznymi.

Innym problemem przy konstrukcji sieci neuronowej jest to, że sieci neuronowe przetwarzają wyłącznie dane numeryczne należące do ściśle określonego przedziału. Stwarza to problemy w tych sytuacjach, gdy dane należą do innego przedziału, jeśli występują braki danych, lub też, gdy dane mają charakter nienumeryczny. Dane numeryczne są przeskalowywane do właściwego dla sieci przedziału, przy czym może to być wykonane automatycznie lub sterowane przez użytkownika.

Określenie rozmiaru zbioru uczącego, czyli właściwej liczby przypadków wymaganych do nauczenia

sieci neuronowej, stwarza w ogólności istotne problemy. Istnieją pewne reguły heurystyczne, które uzależniają liczbę wymaganych przypadków od rozmiaru sieci. Najprostsza z nich mówi, że liczba przypadków powinna być dziesięciokrotnie większa od liczby połączeń występujących w sieci. W rzeczywistości liczba potrzebnych przypadków jest również uzależniona od złożoności funkcyjnej poddanej modelowaniu i praktycznie rośnie ona nieliniowo, co powoduje, że nawet przy dość małej liczbie zmiennych liczba ta jest ogromna. Problem ten w literaturze nazywany jest jako „przekleństwo wielowymiarowości”. W przypadku posiadania mniejszych zbiorów danych należy zdawać sobie sprawę z faktu, że zmuszeni jesteśmy używać sieci niedostatecznie nauczonej. Zazwyczaj w takich przypadkach najlepszym sposobem postępowania będzie dopasowanie modelu liniowego [8].

Rozwiązanie wielu rzeczywistych problemów bywa także utrudnione z uwagi na niepełnowartościowość danych tworzących zbiór uczący. Ta niepełnowartościowość przejawia się tym, że odpowiednie dane są w odpowiednim zbiorze, ale nie wnoszą tak wiele informacji, jak powinny. Przyczyny tego mogą być różne. Na przykład wartości pewnych zmiennych mogą być zniekształcone przez występujące szumy albo też niektóre zestawy danych mogą być niekompletne. Powszechnie sądzi się również, że sieci neuronowe są odporne na szumy. Oczywiście jest tak w rzeczywistości, jednak odporność ta ma swoje granice.

Dane można skalować, standaryzować lub normalizować według wielu algorytmów. Techniki te stanowią problem sam w sobie z punktu widzenia zmiany parametrów statystycznych danych przetworzonych. Najczęściej stosowane jest zwykle przeskalowanie danych do zadanego przedziału.

3. Omówienie wyników predykcji sieci neuronowych dla wybranych wskaźników na przykładzie Warszawskiej Giełdy Papierów Wartościowych

Jednym z istotnych zastosowań sieci neuronowych jest predykcja zachowań giełdy papierów wartościowych.

Celem badań było znalezienie odpowiedzi na pytanie czy rynek akcji zachowuje się w przypadkowy sposób, nie posiadając właściwie przewidywalnych trendów, czy też trendy te mogą być prognozowane.

Zachowanie giełdy ma charakter prognozowalny, jeśli istnieje funkcja, która opisuje badany rynek i która przy pomocy sieci może zostać zidentyfikowana.

W celu uzyskania odpowiedzi na to pytanie zostały skonstruowane odpowiednie sieci neuronowe w oparciu

o pakiet programu Qnet 2000 [1], które następnie poddano trenowaniu i testowaniu. Wyniki zostały przedstawione i omówione w bieżącym rozdziale.

Budując model do celów prognostycznych możemy zastosować dwa podejścia. Pierwsze to podejście autoregresyjne, które oznacza, że badany proces jest modelowany na podstawie jego własnych opóźnień. W tym przypadku prognozujemy np. wskaźnik WIG jedynie na podstawie jego przeszłych wartości. Drugie podejście to wprowadzenie do modelowania danych przekształconych np. wartości wskaźników technicznych. W pracy zastosowano drugą metodologię.

Horyzont czasowy prognozy może być dowolny i różnie określony. Możemy robić prognozy wartości np. WIG'u na następny tydzień lub na następne 5 sesji. Różnica jest istotna. W pierwszym przypadku wyznaczamy średnie wartości WIG dla każdego tygodnia. W drugim przypadku wyznaczamy średnie 5-cio sesyjne. Należy zwrócić uwagę, że posługując się średnimi możemy znacznie skrócić liczbę dostępnych próbek uczących i testowych. Gdybyśmy szacowali średnie tygodniowe to próbek będzie tyle co tygodni a więc liczba danych zmniejszy się pięciokrotnie. Przy średnich 5-cio sesyjnych strata będzie równa czterem sesjom, a więc nieistotnie mało.

Drugim, obok prognoz kierunku zmian, jest problem prognozowania wartości badanego waloru. Jest to uzupełnienie powyższych analiz o wskaźnik skali zmian. Staramy się nie tylko ustalić kierunek zmian ale i ich wielkość. Jest to zadanie dużo trudniejsze ze względu na wymaganą wysoką precyzję.

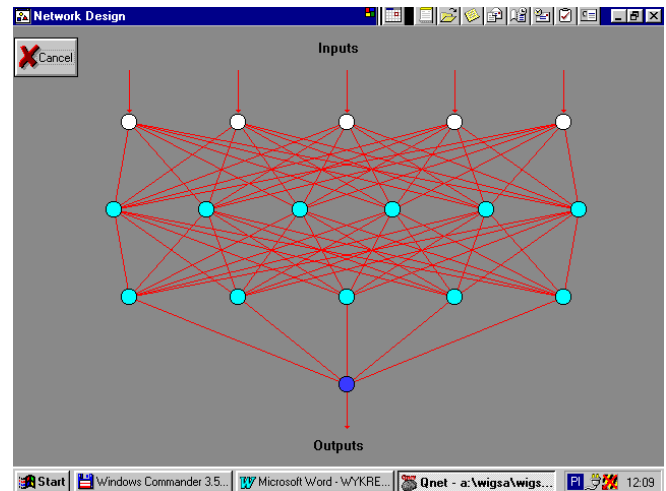
Z oczywistych względów chcielibyśmy robić prognozy na jak najdalszy moment w przyszłości. Trzeba sobie jednak zdawać sprawę z tego, że błędy prognoz rosną gwałtownie wraz ze zwiększaniem horyzontu prognozy.

Klasyczne podejście do budowy sieci nie pozwala zbyt dokładnie kontrolować zdolności generalizacyjnych sieci. Trenujemy sieć na próbie uczącej, następnie sprawdzamy jej zdolności na małej próbie testowej. Wynik tego sprawdzianu nas zadowala lub nie. Jeżeli nie, to zaczynamy budowę i uczenie sieci od nowa. Dla małych sieci jest to strategia do przyjęcia - dla dużych nie.

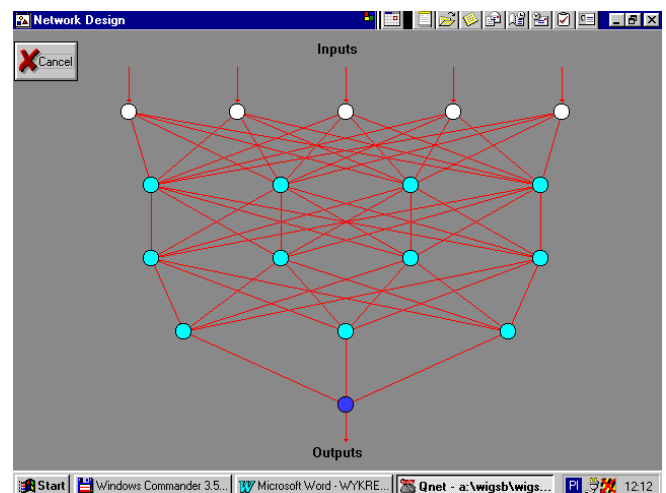
Konstrukcję sieci rozpoczęto od zbadania wpływu wartości wag na starcie na zdolności generalizacyjne sieci. Uruchamiano sieć z ustaloną topologią z różnymi wagami (wyznaczonymi z funkcji Random). Uzyskane rezultaty pozwalają stwierdzić, że początkowe wagi nie wpływają w sposób istotny na końcowy wynik działania sieci. Następnie zbadano wpływ ilości neuronów oraz ilości warstw na jakość pracy sieci. Wyniki dla odmiennie skonstruowanych sieci, trenowanych na przekształconych danych historycznych WIG, zebrano

w tabeli 1.

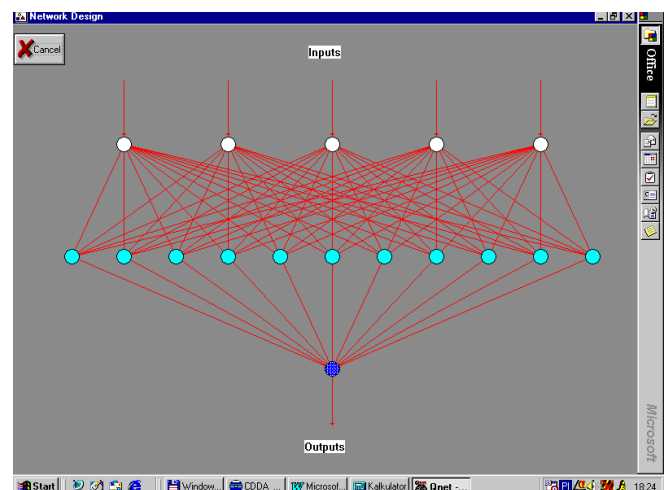
Sieć A to sieć z dwoma warstwami ukrytymi (Rys.1) z łączną ilością 11 neuronów w tych warstwach.



Rys.1 Topologia sieci A.

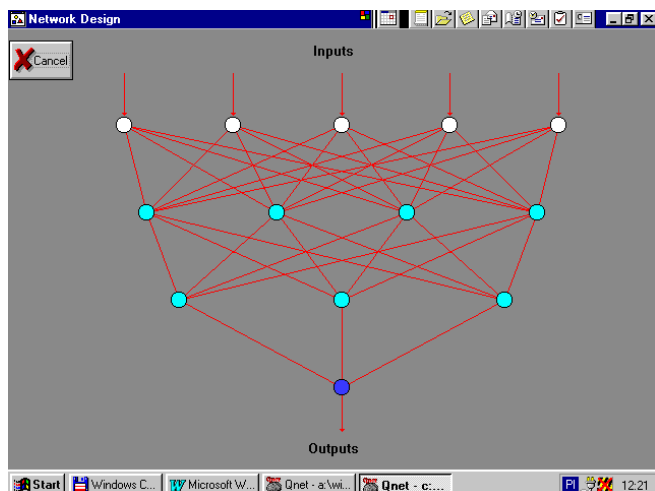


Rys.2 Topologia sieci B.



Rys.3 Topologia sieci C.

Podobną ilość neuronów (Rys.2,3) w warstwach ukrytych ma sieć B i C, odpowiednio z trzema i z jedną warstwą ukrytą.



Rys.4 Topologia sieci D.

Sieć D jest siecią z dwoma warstwami ukrytymi z łączną liczbą trzynastu neuronów. Porównując otrzymane predykcje dla różnych sieci, uwzględniając współczynnik tolerancji 1% oraz 0.5% (dopuszczalna różnica pomiędzy prognozowaną względną wartością a jej rzeczywistą względną wielkością) można zauważyć, że przy większym współczynniku tolerancji najlepszą siecią jest sieć z dwoma warstwami ukrytymi, z trzema i czterema neuronami w tych warstwach (sieć D). Z przeprowadzonych testów wynika, iż na każde pięć przypadków testowych, sieć udzieliła pięciu poprawnych odpowiedzi (mieszczących się w granicach założonego błędu).

RODZAJ SIECI	LICZBA DOBRYCH / ZŁYCH PREDYKCJI	
	WSPÓLCZYNNIK TOLERANCJI 0.01	WSPÓLCZYNNIK TOLERANCJI 0.005
A	4/1	2/3
B	5/0	2/3
C	4/1	2/3
D	5/0	2/3

Tab.1. Wyniki prognozy dla wskaźnika WIG przy różnej topologii sieci.

Podobną liczbę poprawnych predykcji można uzyskać przy większej ilości neuronów i większej liczbie warstw ukrytych (sieć B), lecz ze względu na bardziej złożoną strukturę, a tym samym wydłużony czas uczenia, jest to sieć mniej atrakcyjna. Sieć A daje gorsze rezultaty dla danych testujących, natomiast C jest gorsza w procesie uczenia (wolniej dochodzi do stabilnych wyników pracy).

Przy współczynniku tolerancji 0.5%, stawiającym w stosunku do sieci dużo większe wymagania (problem ten będzie jeszcze dokładniej omawiany), wyniki uzyskane przy pomocy analizowanych sieci są identyczne, jednakże procent poprawnych odpowiedzi obniżył się i wynosi 40% w każdym przypadku.

Analiza sieci A z większą ilością neuronów w warstwach ukrytych w stosunku do sieci D wykazuje, że dla dodatkowych neuronów zarówno w warstwie pierwszej ukrytej jak i drugiej, procentowy udział ich wag jest dużo niższy niż pozostałych neuronów wbudowanych w te warstwy.

SIEĆ	UDZIAŁ PROCENTOWY POSZCZEGÓLNYCH NEURONÓW W WARSTWIE WEJŚCIOWEJ					UDZIAŁ PROCENTOWY WAG POSZCZEGÓLNYCH NEURONÓW W KOLEJNYCH WARSTWACH UKRYTYCH													
						1W					2W				3W				
	1N	2N	3N	4N	5N	1N	2N	3N	4N	5N	1N	2N	3N	4N	5N	6N	1N	2N	3N
A	43	46	10	1	0	18	40	2	18	22	16	13	9	16	29	17	-	-	-
B	44	45	10	1	0	24	30	16	30	-	13	35	19	33	-	-	48	44	8
C	45	50	10	5	0	10	5	6	12	20	3	8	7	18	5	6	-	-	-
D	47	43	10	0	0	21	21	33	25	-	20	70	10	-	-	-	47	45	8

Tab.2. Udział procentowy poszczególnych neuronów w warstwie wejściowej oraz w kolejnych warstwach ukrytych w zależności od rodzaju sieci.

Należy zauważyć, że udział procentowy wag poszczególnych neuronów w warstwie wejściowej jest bardzo podobny dla wszystkich wariantów sieci. Istotnym jest spostrzeżenie, iż decydujące znaczenie mają średnie kroczące względne, 5 i 10 –sesyjnej (zawarte w dwóch pierwszych kolumnach danych wejściowych), natomiast niewielki jest wpływ średniej kroczącej względnej 200-sesyjnej (piąta kolumna danych).

Proces uczenia sieci jak i testowania kończony był w momencie uzyskania rezultatów, które były stabilne w stosunkowo długim okresie pracy sieci. Współczynniki korelacji jak i błędy RMS są bardzo zbliżone dla wszystkich typów sieci.

Na podstawie powyżej zamieszczonej analizy wybrano do dalszych badań sieć o strukturze D.

Wybór ten jednocześnie jest potwierdzeniem reguły, że ostateczna postać sieci posiada dwie warstwy ukryte, które zawierają w każdej warstwie po około połowie sumy neuronów warstwy wejściowej i wyjściowej.

W ten sposób skonstruowana sieć posłużyła do predykcji trzech odmiennie wyznaczonych indeksów Warszawskiej Giełdy Papierów Wartościowych (WGPW). Do analizy wybrano dane dotyczące wartości WIG'u (parametru opisującego zachowanie wszystkich spółek notowanych na WGPW) oraz WIG20 dla 20 największych spółek notowanych na giełdzie z ostatnich sześciu lat, a także ceny wszystkich spółek (138) wchodzących w skład WGPW (dane oznaczono jako WIGS).

Dane te zostały przekształcone w serię prostych wskaźników technicznych wykorzystujących cenę zamknięcia. Zostało wprowadzonych pięć wskaźników dla każdej spółki giełdowej, WIG'u oraz WIG20 - czyli pięć neuronów wejściowych, średnie kroczące 5, 10, 20, 50 i 200-sesyjne. Średnie te zostały następnie podzielone przez odpowiednie ceny (wartości) zamknięcia na dany dzień. Dane te (pięć względnych średnich kroczących, jako dane wejściowe oraz wskaźnik ROC, jako wyjście sieci). Zweryfikowane zostało założenie, iż dane te pozwalają na określenie przyszłych wartości badanych zmiennych, a więc i trendów (rosnących lub malejących).

W procesie przygotowania danych były wykorzystane pojęcia: średniej ruchomej i impetu.

Średnie ruchome (zwane również średnimi kroczącymi) należą do klasycznych narzędzi stosowanych do analizy technicznej. Ich popularność znacznie wykracza poza zagadnienia analizy szeregów finansowych, gdyż są one z powodzeniem stosowane w różnorodnych analizach danych. Średnie tego typu ułatwiają identyfikację trendów, pozwalają na filtrację danych, wskazują na punkty zwrotne w analizowanych szeregach.

Średnie ruchome są szeregiem uśrednionych wartości szeregu pierwotnego, przy czym proces uśredniania realizowany jest na podstawie danych mieszczą-

cych się we fragmencie szeregu o określonej długości (podszereg uwzględniony w czasie obliczeń nazywany jest oknem). Określenie średnia ruchoma lub krocząca związane jest z tym, że okno wyznaczające uśredniane wartości podąża za bieżącą wartością szeregu.

Wyznaczanie szeregu wartości średnich ruchomych jest formą filtracji danych mającej na celu usunięcie bądź zmniejszenie wpływu wahań krótkookresowych (średnia ruchoma jest więc rodzajem filtru). Uzyskany szereg wartości uśrednionych reprezentuje zmiany będące wynikiem wahań o dłuższych okresach.

Wraz z wydłużaniem szerokości okna zastosowanego do obliczania średniej wpływ wygładzania (eliminacji wahań krótkookresowych) jest coraz mocniejszy i uzyskane rezultaty obliczeń informują w coraz większym stopniu o tendencjach długookresowych. Dobór szerokości okna jest uzależniony od celu analizy.

Ze względu na sposób obliczania średniej możliwe jest wyróżnienie różnych typów średnich ruchomych. W pracy zostanie zastosowana prosta średnia ruchoma.

Prosta średnia ruchoma jest średnią arytmetyczną ze wszystkich wartości wchodzących w skład przyjętego okna.

Sposób jej wyznaczania przedstawia wzór

$$S^pr_t = 1/n \sum_{i=t-n+1}^t x_i$$

gdzie x_i oznaczają ceny w czasie i .

Pojęcie impetu definiuje się jako tempo zmiany cen. Analiza impetu nie uwzględnia bezwzględnych wartości cen, lecz zajmuje się wyłącznie zmianami zachodzącymi w określonym okresie czasu. Impet (zwany również pędem lub momentum) wyrażany jest w sposób ilościowy przy pomocy wskaźników impetu. Pojęcie impetu jest zbliżone do pojęcia szybkości i wyrazić można przy pomocy ogólnej formuły:

$$M = \Delta p / t$$

gdzie:

M - impet, Δp zmiana ceny, t - czas.

Istnieje cały szereg miar impetu do najpopularniejszych można zaliczyć wskaźnik zmian (ROC - Rate of Change):

$$ROC = x_t / x_{t-n}$$

Dane dla WIG oraz WIG20 zostały przekształcone zgodnie ze wzorami podanymi powyżej :

Wejścia:

5-cio sesyjna średnia ruchoma/aktualna wartość za-

mknięcia

10-cio sesyjna średnia ruchoma/aktualna wartość zamknięcia

20-sto sesyjna średnia ruchoma/aktualna wartość zamknięcia

50-cio sesyjna średnia ruchoma/aktualna wartość zamknięcia

200-tu sesyjna średnia ruchoma/aktualna wartość zamknięcia

Wyjście:

Zmiana względna wskaźnika po 5 sesjach od aktualnej wartości zamknięcia - wskaźnik zmian ROC (Rate of Change). W przypadku WIGS wejścia sieci stanowiły 5,10,20,50,200-sesyjne średnie obliczone względem cen akcji spółek z dnia 08-04-2002 roku dla wszystkich 138 spółek notowanych na WGPW. Wyjście stanowiły ROC cen poszczególnych spółek po 5-ciu sesjach od daty podanej wyżej.

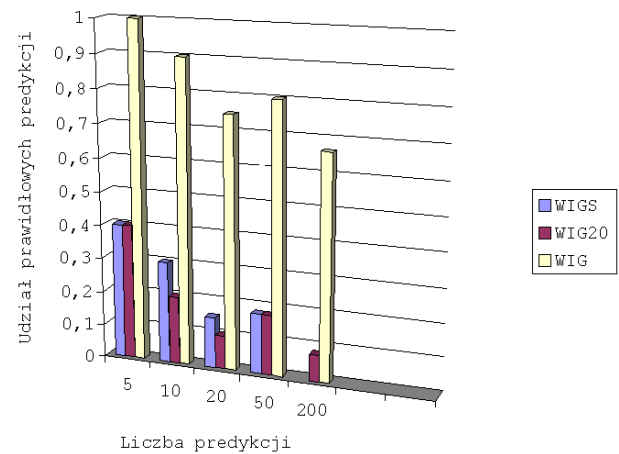
Nie korzystając z żadnego modelu można stwierdzić, że poziom wskaźnika WIG'u może zmienić swoją wartość co najwyżej o 10% w stosunku do obecnej wielkości. Ponieważ WIG, najczęściej, nie zmieniał się więcej niż o 4% to prognozę a priori można jeszcze zawęzić. Przeciętne zmiany WIG w badanym okresie wynosiły średnio 200 punktów z sesji na sesję (co do wartości bezwzględnej). Gdyby wyrazić te wartości w procentach byłoby to : 1.3%. Aby prognozy miały istotną wartość to błędy w dłuższych okresach powinny być mniejsze niż 1%. Nakładając na to doświadczenia inwestorów, można powiedzieć, że przyszła wartość WIG jest dość dobrze znana. Zadaniem sieci jest poprawianie tych przewidywań. W praktyce oznacza to, że prognozy nie powinny różnić się od rzeczywistości o więcej niż połowę tego, co i tak już wiemy. Wtedy trud włożony w budowę sieci może się opłacić. W praktyce oznacza to, że sieć powinna dawać prognozy ze średnim błędem nie przekraczającym, dla WIG-u, 1% bieżącej jego wartości.

Zaprojektowana sieć pozwala na uzyskanie lepszych rezultatów. Otrzymane wyniki zostały zaprezentowane dla współczynnika tolerancji 0.01 na Rys.5., a dla współczynnika tolerancji 0.005 na Rys.6

Analiza poniżej zamieszczonego wykresu wskazuje na doskonałą zdolność sieci do prognozowania kierunku i wartości względnej zmiany wskaźnika WIG'u (przy współczynniku tolerancji 0,01) w stosunku do pozostałych badanych wskaźników (WIG20 i WIGS).

Wysoką pewnością prognozy wskaźnika WIG otrzymuje się dla predykcji pięciosesyjnych, maleje ona prawie monotonicznie i dla prognoz długoterminowych spada, jednak nadal jest ona duża (66% dla prognozy 200-tusecyjnej).

Generalizacja prognozy długoterminowej przy tym stopniu tolerancji jest możliwa jedynie dla WIG'u.

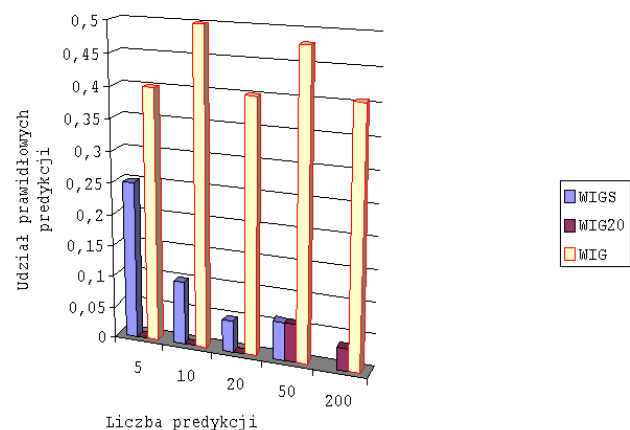


Rys.5 Udział prawidłowych predykcji w całkowitej liczbie predykcji (współczynnik tolerancji 0,01)

Prognozy wskaźników WIG20 i WIGS nie są zadowalające. Predykcje krótkoterminowe wyznaczone zostały przez sieć poniżej 50%, natomiast w długim horyzoncie czasowym spadają nawet poniżej 20%. W przypadku tych wskaźników wystąpił słaby proces generalizacji podczas procesu uczenia sieci. Przy założonym poziomie tolerancji (0.01) można stwierdzić, że użyteczność tych modeli jest niewielka.

Ceny akcji spółek wchodzących w skład WIG-u nie podlegają predykcji (WIGS), oznacza to, że rzeczywistość są one kształtowane losowo. Model neuronowy był w stanie zapamiętać zbiór uczący, ale nie potrafił reagować właściwie w stosunku do przypadków pochodzących ze zbioru testowego.

Prognozy krótko i długookresowe dla wskaźnika WIG20 przy współczynniku tolerancji 0.5%, (wypełnienie którego pozwala na dokonywanie korzystnych inwestycji), są bliskie zeru.

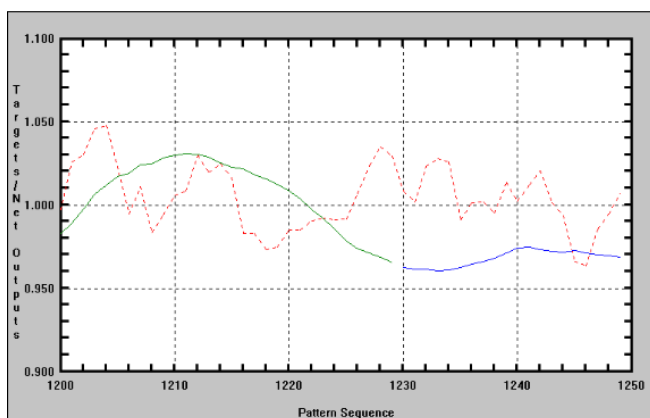


Rys.6 Udział prawidłowych predykcji w całkowitej liczbie predykcji (współczynnik tolerancji 0,005)

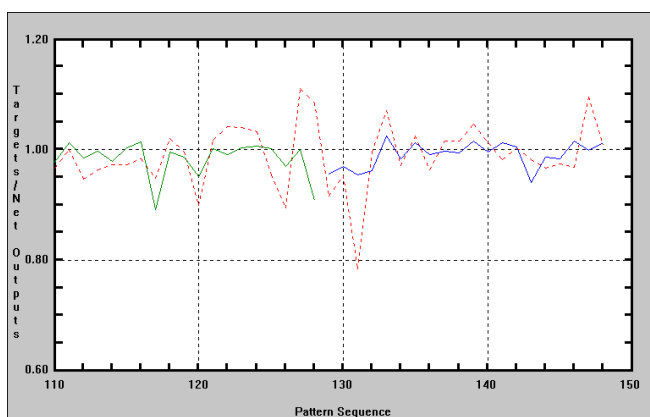
Udział udanych prognoz dla WIGS jest niewiele większy niż w przypadku WIG20, jedynie dla testu wykorzystującego pięć cen akcji spółek wynosi on nieco ponad 20%.

W przypadku WIG'u dobre predykcje utrzymują się na stałym, dość wysokim poziomie (od 40 - 50%) dla każdego z wybranych zbiorów testowych.

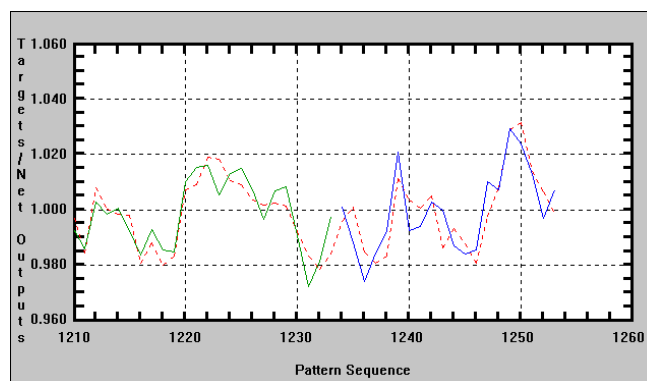
Z przeprowadzonych analiz wynika, iż relatywnie proste modele mogą dawać dobre wyniki w prognozowaniu indeksu wszystkich spółek WIG-u (Rys.9). Badania opublikowane wcześniej przeprowadzone przy użyciu odmiennie skonstruowanych sieci neuronowych potwierdzają otrzymane w pracy wyniki. Interesującą informacją jest brak możliwości predykcji dla indeksu 20-stu spółek (Rys.8) oraz wniossek, że nie można oczekiwać odpowiedzi poprawnej od sieci co do zachowania się cen (wzrostu lub spadku) dla wybranych spółek na podstawie zachowania się cen pozostałych (Rys.7).



Rys.7 Rysunek ilustrujący zależność względnego indeksu WIG20 dla predykcji 20-sto sesyjnej w procesie trenowania sieci (zielony), testowania sieci (niebieski) w stosunku do rzeczywistej względnej wartości (czewony).



Rys.8 Rysunek ilustrujący zależność względnego indeksu WIGS dla predykcji 20-sto sesyjnej w procesie trenowania sieci (zielony), testowania sieci (niebieski) w stosunku do rzeczywistej względnej wartości (czerwony).



Rys.9 Rysunek ilustrujący zależność względnego indeksu WIG dla predykcji 20-sto sesyjnej w procesie trenowania sieci (zielony), testowania sieci (niebieski) w stosunku do rzeczywistej względnej wartości (czerwony).

4. Podsumowanie

Sieci neuronowe są nowym narzędziem informatycznym umożliwiającym konstrukcję modeli nieliniowych rozwiązujących złożone i trudne do identyfikacji zadania klasyfikacyjne i regresyjne. Używanie sieci neuronowych w analizie złożonych danych jest szczególnie korzystne, ponieważ prowadzi do konstrukcji modeli opartych na właściwościach samych danych, a nie na arbitralnych hipotezach tworzącego model badacza.

Do reprezentatywnych przykładów problemów rozwiązywanych za pomocą sieci neuronowych należy prognozowanie giełdowe. Zmiany cen akcji są szeroko znanym przykładem zjawisk podlegających klasyfikacji i rozpoznawaniu. Zjawiska te charakteryzują się z jednej strony złożonością i wielowymiarowością, zaś z drugiej strony występowaniem, w niektórych okolicznościach, składowych mających przynajmniej charakter deterministyczny.

Wielu analityków stosujących analizę techniczną wykorzystuje w związku z tym sieci neuronowe do wyznaczania na przykład prognoz cen akcji na podstawie dużej liczby czynników, takich jak kształtowanie się w przeszłości cen innych akcji i różnych wskaźników ekonomicznych.

Sieci posiadają szereg cech, dzięki którym mogą stanowić przydatne narzędzie modelowania i prognozowania zjawisk społeczno - ekonomicznych. Celowość ich zastosowania wynika z pewnych własności charakteryzujących wspomniany typ zjawisk, jak również ze sposobu budowy i funkcjonowania modeli neuronowych. Jako uzasadnienie stosowania modeli neuronowych w modelowaniu prawidłowości występujących na rynkach finansowych można przyjąć następujące fakty [7]:

- znaczna ilość zjawisk rozpatrywanych na gruncie

finansów ma charakter nieliniowy, co stanowi podstawową przesłankę do tego, aby do ich modelowania stosować narzędzia przystosowane do opisu zależności nieliniowych. Jednym z narzędzi spełniających ten warunek są jednokierunkowe sieci neuronowe. Posiadają one zdolności do aproksymacji dowolnych zależności nieliniowych jak również charakteryzują się zdolnościami generalizującymi,

- proces budowy modelu neuronowego polega na eksploracji dostępnych zbiorów danych i oszacowaniu na tej podstawie modelu opisującego stwierdzone prawidłowości. Stosowanie modeli tego typu nie wymaga znajomości postaci funkcji opisującej istniejącą prawidłowość. W związku z tym modele neuronowe mogą znaleźć zastosowanie wszędzie tam, gdzie nie jest znane dokładne prawo opisujące kształtowanie się badanych zależności. Nie wyklucza to możliwości stosowania sieci w przypadkach, gdy znana jest postać formuł matematycznych opisujących badany aspekt rzeczywistości, ale wówczas nakłady związane z oszacowaniem modelu neuronowego mogą być wyższe niż nakłady niezbędne do obliczenia parametrów danego w postaci równania prawa,

- modele neuronowe mają charakter adaptacyjny. Mogą służyć do opisu zależności zmieniających się w czasie. W chwili pojawienia się nowych danych przeprowadzony może zostać proces douczenia sieci, co umożliwia uwzględnienie w tworzonym modelu informacji zawartych w najnowszych obserwacjach, sieć neuronowa może być traktowana nie tylko jako mechanizm opisujący przebieg zjawiska i generujący przyszłe jego wartości. Daje ona możliwości przeprowadzania wszechstronnej analizy badanego fragmentu rzeczywistości. Podstawowe informacje o systemie uzyskać można poprzez zastosowanie analizy wrażliwości modelu. Pozwala ona na przedstawienie charakteru związku pomiędzy badaną wielkością o poszczególnymi wpływającymi na nią czynnikami,

- proces szacowania i wykorzystania modeli neuronowych może być realizowany współbieżnie w systemach wieloprocesorowych lub przez szereg komputerów połączonych w sieć komputerową. Taki sposób realizacji obliczeń neuronowych pozwala na znaczne skrócenie czasu potrzebnego na realizację niezbędnych działań.

Wyniki uzyskane przy pomocy sieci neuronowych mogą być wykorzystywane samodzielnie lub też mogą stanowić uzupełnienie rezultatów uzyskanych przy pomocy innych technik – na przykład klasycznych metod statystycznych.

W pracy przedstawiono wyniki badań nad zdolnościami sieci w prognozowaniu wybranych wskaźników dotyczących funkcjonowania Warszawskiej Giełdy Papierów Wartościowych. Dostarczają one dość obiecujących informacji. Udało się zbudować prognozy war-

tości zmian WIG'u o 100 % skuteczności w krótkim i o 66% skuteczności w długim horyzoncie czasu (dla współczynnika tolerancji 0,01). Prognozy wartości tego indeksu są więc zadowalające.

Oczywiście należy zdawać sobie sprawę, że pokazane tu ujęcie jest tylko jednym z bardzo wielu możliwych. Wyniki te nie są ostateczne, mają przede wszystkim zachęcić do dalszych badań. Istnieje wiele innych typów sieci, które należało by przetestować. Powinno się optymalizować także inne parametry sieci niż te wzięte pod uwagę w pracy. Należało by zoptymalizować wartości współczynników technicznych, sprawdzić wpływ informacji fundamentalnych na możliwości prognostyczne. Można by zbudować sieć rozpoznającą formacje techniczne – do wspomagania prognoz długoterminowych. Dróg poprawy prezentowanych wyników jest wiele.

Literatura (References)

- [1] Qnet 2000 V2K build 721 firmy Neural Network Modeling 2002, www.horyzont.eu/wydawnictwo/biuletyn/informatyka/zalacznik/qnet2000
- [2] E. M. Azoff, *Neural Network Time Series Forecasting of Financial Market*. Wiley & Sons, Chichester 1994.
- [3] A. Beltratti, S. Margarita, P. Terna, *Neural Networks for Economic and Financial Modeling*. International Thompson Computer Press, London 1996.
- [4] S. T. Gallant, *Neural Network Learning and Expert Systems*. MIT Press. Cambridge 1993.
- [5] S. Goonatilake, P. Treleven (red.), *Intelligent Systems for Finance and Business*. Wiley & Sons, Chichester. 1995.
- [6] J. Hertz, A. Krogh, R. G. Palmer, *Wstęp do teorii obliczeń neuronowych*. Wydawnictwo Naukowo-Techniczne, Warszawa 1995.
- [7] S. Shochen, G. Ariav, *Neural Networks for Decision Support: Problems and Opportunities*, Vol.11, Decision Support Systems 1994.
- [8] R. Tadeusiewicz, *Elementarne wprowadzenie do sieci neuronowych z przykładowymi programami*. Akademicka Oficyna Wydawnicza, Warszawa 1998.
- [9] J. S. Zirilli, *Financial Prediction Using Neural Networks*. International Thompson Computer Press, London 1996.