

Akademia Ekonomiczna w Krakowie
Wydział Zarządzania

Janusz Morajda

**Metody sztucznej inteligencji w zarządzaniu
portfelem inwestycyjnym**

(praca doktorska)

Promotor:

Prof. zw. dr hab. inż. Ryszard Tadeusiewicz

Kraków 1999

Spis treści

| | |
|--|-----------|
| Wstęp | 5 |
| 1. ZAGADNIENIE WSPOMAGANIA DECYZJI INWESTYCYJNYCH. | 9 |
| 1.1. Klasyczne metody wykorzystywane w działalności inwestycyjnej. | 9 |
| 1.2. Aktywne zarządzanie portfelem inwestycyjnym..... | 12 |
| 1.3. Systemy wspomaganie decyzji inwestycyjnych. | 14 |
| 1.3.1. Problemy modelowania rynków finansowych. | 15 |
| 1.3.2. Wykorzystanie narzędzi alternatywnych - przesłanki metodologiczne. | 17 |
| 1.3.3. Wykorzystanie narzędzi alternatywnych - przesłanki informatyczne. | 19 |
| 2. WYBRANE ASPEKTY ZASTOSOWANIA SIECI NEURONOWYCH W ANALIZIE RYNKÓW FINANSOWYCH..... | 22 |
| 2.1. Rys historyczny i aktualne kierunki badań nad sieciami neuronowymi. | 23 |
| 2.2. Ogólna charakterystyka sieci neuronowych. | 26 |
| 2.3. Model pojedynczego neuronu. | 30 |
| 2.4. Model warstwowej sieci neuronowej typu <i>perceptron</i> | 38 |
| 2.5. Metody uczenia sieci neuronowych. | 41 |
| 2.6. Przykłady wybranych z literatury zastosowań sieci neuronowych na rynkach finansowych. | 46 |
| 2.7. Zasadnicze aspekty efektywnego wykorzystania sieci neuronowych w analizie i predykcji rynków finansowych. | 49 |
| 2.7.1. Zdolność sieci do generalizacji - unikanie efektu przeuczenia. | 49 |
| 2.7.2. Zagadnienie niezależnej walidacji sieci. | 50 |
| 2.7.3. Dobór architektury sieci. | 51 |
| 2.7.4. Problem właściwego przygotowania danych wejściowych. | 52 |
| 2.7.5. Wstępna eliminacja szumów - wykorzystanie średnich kroczących. | 53 |
| 2.7.6. Zagadnienie optymalnego doboru (selekcji) danych wejściowych. | 56 |
| 2.7.7. Redukcja ilości oraz dekorelacja zmiennych wejściowych. | 57 |
| 2.7.8. Normalizacja zmiennych wejściowych. | 59 |
| 2.7.9. Konstrukcja zbiorów: uczącego, walidacyjnego i testującego. | 60 |
| 2.7.10. Dobór i normalizacja zmiennej wyjściowej. | 62 |
| 2.7.11. Ocena funkcjonowania sieci neuronowej. | 63 |
| 3. ALGORYTMY GENETYCZNE I MOŻLIWOŚCI ICH ZASTOSOWAŃ NA RYNKACH FINANSOWYCH. | 66 |
| 3.1. Idea algorytmu genetycznego. | 66 |
| 3.1.1. Dobór genetycznej reprezentacji potencjalnych rozwiązań oraz funkcji celu i przystosowania. | 67 |

| | |
|--|----|
| 3.1.2. Operatory genetyczne. | 68 |
| 3.1.3. Schemat oraz wybrane aspekty stosowania algorytmu genetycznego. | 70 |
| 3.2. Możliwości i przykłady bezpośredniego wykorzystania algorytmów genetycznych w procesie optymalizacji strategii inwestycyjnych. | 73 |
| 3.2.1. Uwagi wstępne. | 73 |
| 3.2.2. Przykłady zastosowania algorytmów genetycznych na rynkach finansowych. ... | 74 |
| 3.2.3. Inne aspekty zastosowań algorytmów genetycznych w ekonomii. | 75 |
| 3.3. Możliwości zastosowania algorytmów genetycznych w procesie konstrukcji sieci neuronowych. | 76 |

4. PROPOZYCJA NOWEJ METODOLOGII KONSTRUOWANIA I REALIZACJI AKTYWNYCH STRATEGII INWESTYCYJNYCH. 79

| | |
|---|-----|
| 4.1. Ogólna struktura systemu decyzyjnego i postulowany sposób jego wykorzystania. | 80 |
| 4.2. Moduł przetwarzania informacji wejściowej i wyboru danych wejściowych. | 82 |
| 4.3. Moduł preprocessingu danych wejściowych. | 96 |
| 4.3.1. Trymowanie (obcinanie) wartości nietypowych. | 96 |
| 4.3.2. Normalizacja zmiennych wejściowych. | 97 |
| 4.3.3. Skokowe rozdzielanie wartości dodatnich i ujemnych. | 98 |
| 4.3.4. Zastosowanie antysymetrycznej funkcji potęgowej. | 99 |
| 4.3.5. Analiza głównych składowych. | 99 |
| 4.4. Przyjęta architektura sieci neuronowych typu <i>perceptron</i> | 103 |
| 4.5. Modele i normalizacja zmiennej wyjściowej dla sieci <i>perceptron</i> | 103 |
| 4.5.1. Względna zmiana kursu. | 104 |
| 4.5.2. Trend względnych zmian kursu. | 104 |
| 4.5.3. Najbliższa istotna względna zmiana średniej adaptacyjnej z kursu. | 105 |
| 4.5.4. Trend wielkości różnic logarytmów kursu i indeksu. | 108 |
| 4.5.5. Normalizacja wartości wzorcowych zmiennej wyjściowej. | 108 |
| 4.6. Uczenie i testowanie sieci neuronowej. | 109 |
| 4.7. Koncepcja algorytmu genetycznego. | 116 |
| 4.7.1. Zastosowany sposób genetycznego kodowania rozwiązań. | 116 |
| 4.7.2. Dobór funkcji celu i funkcji przystosowania. | 118 |
| 4.7.3. Propozycje zmodyfikowanych operatorów genetycznych i innych parametrów algorytmu. | 119 |
| 4.8. Moduł realizacji strategii decyzyjnej. | 123 |

5. ANALIZA REZULTATÓW ZASTOSOWANIA PRZEDSTAWIONEJ METODOLOGII W STRATEGII OPARTEJ O PROGNOZY INDEKSU WIG 20. 125

| | |
|---|-----|
| 5.1. Dobór parametrów systemu. | 125 |
| 5.2. Analiza otrzymanych wyników oraz efektywności metody. | 128 |

| | |
|---|-----|
| 5.3. Analiza wzajemnych relacji pomiędzy ilościowymi miarami efektywności sieci..... | 137 |
| 5.4. Badania doboru wybranych parametrów systemu i ich wpływu na efektywność metody..... | 141 |
| 5.4.1. Zależność efektywności sieci od wartości wag początkowych..... | 141 |
| 5.4.2. Analiza przykładowego procesu uczenia sieci neuronowej..... | 149 |
| 5.4.3. Dylemat doboru optymalnej wielkości warstwy ukrytej..... | 153 |
| 5.4.4. Celowość zastosowania analizy głównych składowych jako elementu wstępnego przetwarzania danych. | 157 |
| 5.4.4.1. Badanie wpływu liczby zastosowanych głównych składowych na efektywność sieci | 157 |
| 5.4.4.2. Celowość proporcjonalnej redukcji warstwy ukrytej sieci. | 165 |
| 5.4.4.3. Dobór liczby epok uczenia sieci PCA. | 167 |
| 5.4.4.4. Metoda neuronowa a metoda klasyczna przekształcenia PCA. | 169 |
| 5.4.5. Dobór funkcji błędu w procesie uczenia sieci..... | 170 |
| 5.4.6. Dobór parametrów funkcji aktywacji neuronu..... | 173 |
| 5.4.7. Analiza czasowych zmian efektywności aktywnej strategii opartej o wybraną sieć..... | 179 |
| 5.4.8. Badanie przydatności douczania i ponownego uczenia wybranej efektywnej sieci. | 182 |

6. KONCEPCJA AKTYWNEJ STRATEGII ZARZĄDZANIA PORTFELEM AKCJI W OPARCIU O ZAPROPONOWANĄ METODOLOGIĘ ORAZ DISKUSJA UZYSKANYCH EFEKTÓW..... 186

| | |
|--|-----|
| 6.1. Metodologia aktywnego zarządzania portfelem inwestycyjnym..... | 187 |
| 6.2. Algorytm symulacji procesu aktywnego zarządzania portfelem. | 192 |
| 6.3. Badania efektywności metody dla wybranego zestawu akcji. | 195 |
| 6.3.1. Zaproponowane parametry systemu decyzyjnego..... | 196 |
| 6.3.2. Wyniki badań efektywności zaproponowanej metody zarządzania portfelem. | 199 |

7. PODSUMOWANIE OTRZYMANYCH WYNIKÓW I WNIOSKI KOŃCOWE.....206

Zakończenie 202

Literatura 205

Wstęp.

Obserwowany obecnie dynamiczny rozwój nowoczesnych metod informatyki, stymulowany szybkim postępem technologii komputerowych, stwarza nowe obszary potencjalnych badań naukowych związanych z wykorzystaniem tych metod w różnych dyscyplinach wiedzy. Szczególnie obiecującą tematykę badań stanowią tu metody oparte na tak zwanej sztucznej inteligencji - dziedzinie czerpiącej inspirację z obserwacji naturalnych systemów biologicznych i zachodzących w nich procesów przetwarzania i wykorzystywania informacji. Metody te mogą zostać zastosowane między innymi do tworzenia modeli różnych złożonych zjawisk, także tych obserwowanych w ekonomii. Ponieważ - jak dowodzą studia literaturowe - narzędzia modelowania oparte na sztucznej inteligencji często stanowią korzystną alternatywę dla metod klasycznych, a niejednokrotnie okazują się od nich lepsze i bardziej skuteczne, za wysoce celowe należy uznać prowadzenie prac badawczych nad ich efektywnością, rozwojem i zastosowaniami.

W niniejszej pracy postanowiono zbadać możliwości wykorzystania dwóch podstawowych narzędzi sztucznej inteligencji - sieci neuronowych i algorytmów genetycznych - w modelowaniu dynamiki rynków finansowych. Badania oparto na polskich uwarunkowaniach, przy czym ukierunkowane zostały one na zagadnienia konstrukcji systemów decyzyjnych, generujących strategie możliwe do wykorzystania w procesie zarządzania portfelem inwestycyjnym.

Cele pracy można zdefiniować w następujących punktach:

- ◆ zaprezentowanie sieci neuronowych i algorytmów genetycznych jako efektywnych narzędzi w modelowaniu zjawisk zachodzących na rynku finansowym oraz w konstrukcji systemów decyzyjnych dla tego rynku,

- ◆ zaproponowanie pewnej nowej metodologii konstrukcji aktywnych strategii inwestycyjnych, wykorzystującej oryginalne połączenie obu prezentowanych narzędzi,
- ◆ przeprowadzenie badań efektywności zaproponowanej metodologii, badań skutków doboru wartości jej wybranych, istotnych parametrów, a także badań uwarunkowań jej wykorzystania.

Zasadnicza teza niniejszej pracy oparta jest na stwierdzeniu, iż **metody sztucznej inteligencji, takie jak sieci neuronowe i algorytmy genetyczne, stanowią mogą efektywne narzędzia w procesie modelowania dynamiki rynków finansowych dla potrzeb generowania decyzji inwestycyjnych, jednak efektywność ta może być uzależniona od prawidłowego doboru szeregu parametrów stosowanej metody.** Dodatkowo autor pragnie sformułować pogląd, iż generalnie efektywność systemów decyzyjnych wykorzystujących te narzędzia będzie wzrastać w miarę dalszego rozwoju techniki komputerowej oraz postępu badań w dziedzinie sztucznej inteligencji.

Autor pracy - dostrzegając sceptycyzm części badaczy i analityków rynków finansowych preferujących raczej klasyczne (niewątpliwie dobrze rozwinięte) metody analizy tych rynków - pragnie zwrócić uwagę na fakt, iż w czasie gdy powstawały klasyczne teorie funkcjonowania rynków finansowych, możliwości wykorzystania omawianych tutaj metod sztucznej inteligencji były (z uwagi na dostępną moc obliczeniową) praktycznie znikome. Skoro obecnie te możliwości się pojawiły i stale wzrastają, obowiązkiem badacza jest eksploracja nowych metodologii i obszarów potencjalnych ich zastosowań. Stwierdzenie to - zdaniem autora - w pełni uzasadnienia celowość podjęcia przedmiotowej tematyki badawczej w niniejszej pracy, a także jej kontynuację w trakcie dalszych badań. Z kolei rozważając potencjalny aplikacyjny aspekt badań nie można wykluczyć faktu, iż w niedalekiej przyszłości inwestor nie stosujący aktywnych, komputerowych, opartych na nowoczesnych metodologiach sposobów analizy rynku i realizacji strategii inwestycyjnych, nie będzie w stanie podejmować efektywnych decyzji inwestycyjnych.

Praca składa się z siedmiu rozdziałów, zakończenia oraz spisu literatury.

Rozdział pierwszy przedstawia ogólną charakterystykę rynków finansowych oraz skrótowo omawia podstawowe metody (określane w pracy jako klasyczne), wykorzystywane w analizie tych rynków i w działalności inwestycyjnej. Następnie przedstawiono rozważania świadczące o celowości stosowania aktywnych metod inwestycyjnych wspieranych przez odpowiednio skonstruowane systemy decyzyjne, oraz omówiono przesłanki przemawiające za zastosowaniem sieci neuronowych i algorytmów genetycznych w tym zagadnieniu.

W rozdziale drugim dokonano charakterystyki sieci neuronowych jako przydatnych narzędzi do modelowania zjawisk zachodzących na rynkach finansowych. Omówiono przy tym szereg aspektów tych narzędzi, takich jak model pojedynczego neuronu, architektura sieci, metody uczenia, cechy modeli opartych na sieciach, kierunki badań i zastosowań. Szczególną uwagę zwrócono na problematykę zastosowania sieci neuronowych do analizy rynków finansowych, prezentując szereg zagadnień i uwarunkowań efektywnej realizacji tego typu celów.

Rozdział trzeci poświęcony został prezentacji algorytmów genetycznych jako efektywnych narzędzi w zagadnieniach globalnej optymalizacji modeli dla złożonych systemów, w tym także w problemach optymalizacji strategii inwestycyjnych na rynkach finansowych. Omówiono też możliwości wykorzystania tych metod w procesach konstrukcji sieci neuronowych.

W rozdziale czwartym zaproponowano nową metodologię generowania i realizacji aktywnych strategii decyzyjnych. Metodologia ta zakłada konstrukcję odpowiedniego systemu decyzyjnego opartego o sieci neuronowe, przy czym podczas procesu tej konstrukcji wykorzystuje się pewną (zaproponowaną przez autora) modyfikację algorytmu genetycznego dla doboru zmiennych wejściowych i wybranych parametrów sieci. Szczegółowo omówiono poszczególne elementy tej metodologii oraz problematykę doboru jej parametrów.

Rozdział piąty zawiera badania dotyczące zastosowania zaprezentowanej metodologii w procesie realizacji aktywnej strategii inwestycyjnej opartej o prognozy indeksu giełdowego WIG 20. Badania te dotyczą zarówno analizy efektywności

samej metody, jak i zagadnień doboru jej określonych elementów, takich jak wykorzystanie analizy głównych składowych do wstępnej transformacji danych wejściowych, dobór optymalnej wielkości warstwy ukrytej dla sieci, dobór parametrów funkcji aktywacji i funkcji błędu, analiza procesów uczenia i douczania sieci, itp. Oprócz wyników badań zamieszczono szczegółową dyskusję otrzymanych rezultatów.

Rozdział szósty zawiera propozycję metody aktywnego zarządzania portfelem złożonym z wybranego zestawu instrumentów finansowych (np. akcji), w oparciu o przedstawioną metodologię. Zaprezentowano algorytm generujący odpowiednie decyzje inwestycyjne na podstawie sygnałów sieci neuronowych, oraz algorytm symulacji procesu zarządzania portfelem w określonym przedziale czasu. Następnie przeprowadzono badania efektywności tej metody dla wybranego pakietu akcji notowanych na giełdzie warszawskiej.

W rozdziale siódmym dokonano podsumowania wyników badań oraz zawarto syntetyczne wnioski dotyczące efektywności i uwarunkowań zastosowania zaproponowanej metodologii.

Zawarte w niniejszej pracy badania przeprowadzono w oparciu o napisany przez autora zestaw programów komputerowych, zaimplementowanych w języku programowania C.

Pragnę wyrazić serdeczne podziękowania Panu prof. zw. dr hab. inż. Ryszardowi Tadeusiewiczowi za niezwykle cenną pomoc merytoryczną podczas pisania pracy oraz za wnikliwe uwagi, bardzo pomocne podczas redakcji rozprawy.

1. Zagadnienie wspomaganie decyzji inwestycyjnych.

Istotnym zagadnieniem w działalności każdego przedsiębiorstwa lub instytucji finansowej jest efektywne zarządzanie finansami, obejmujące m.in. dokonywanie inwestycji w określone instrumenty finansowe. Do podstawowych instrumentów, stanowiących przedmiot tych inwestycji, zaliczyć można papiery wartościowe (np. akcje, obligacje, bony skarbowe). Niezwykle ważnym elementem działalności inwestycyjnej staje się zatem właściwe zarządzanie portfelem inwestycyjnym, tj. zestawem posiadanych papierów wartościowych, a często także zestawem instrumentów pochodnych (takich jak np. opcje czy kontrakty terminowe).

Metody konstrukcji i zarządzania portfelem inwestycyjnym zależą w głównej mierze od przyjętych celów i warunków działalności inwestycyjnej, a te z kolei są zdeterminowane przez strategiczne cele oraz preferencje inwestora. Niewątpliwie jednak metody te, aby były dostatecznie efektywne, powinny być oparte na szeroko pojętej, systematycznej analizie rynku finansowego oraz jego instrumentów, realizowanej przy wykorzystaniu najlepszych dostępnych narzędzi. Wynika stąd potrzeba stałego rozwoju i doskonalenia (a także oceny i weryfikacji) narzędzi dla takiej analizy, służących m.in. do generowania prognoz i/lub do wspomaganie decyzji.

1.1. Klasyczne metody wykorzystywane w działalności inwestycyjnej.

Naukowe teorie dotyczące inwestowania w papiery wartościowe wywodzą się głównie z połączenia statystyki oraz nauki o finansach. Stanowią one podstawy

podejmowania decyzji inwestycyjnych w warunkach niepewności i ryzyka, a więc w sytuacji stale panującej na rynkach finansowych. Teorie te są również bazą dla wielu stosowanych obecnie w praktyce strategii inwestycyjnych. Najbardziej znaczące osiągnięcia w tej dziedzinie, mające decydujący wpływ na rozwój nauki o inwestowaniu oraz praktycznych metodologii analizy instrumentów finansowych i budowy portfela inwestycyjnego, zostały poniżej skrótowo wymienione (wyczerpujący opis tej problematyki można znaleźć np. w Bodie, Kane, Marcus, 1993; Haugen, 1996).

Teoria portfela określa metody tworzenia tzw. portfeli efektywnych, tj. optymalnych z punktu widzenia maksymalizacji zysku i minimalizacji ryzyka, oraz portfela rynkowego uwzględniającego też instrumenty pozbawione ryzyka. Teoria ta jest oparta na analizie stóp zysku i ryzyka (rozumianego jako odchylenie standardowe stóp zysku), oraz współczynników korelacji papierów wartościowych. Podstawy tej teorii stworzył Markowitz (1952). Uproszczoną metodę konstrukcji portfela zaproponował Sharpe (1963). Kolejnym ważnym uzupełnieniem tej teorii był opracowany przez Sharpe'a, Lintnera i Mossina model równowagi rynku papierów wartościowych CAPM (*Capital Asset Pricing Model*). Metody konstrukcji portfela papierów wartościowych zostały omówione i rozwinięte w dość bogatej literaturze (zob. np. Elton & Gruber, 1991; Levy & Sarnat, 1972; Bodie et al., 1993; Jajuga & Jajuga, 1996; Haugen, 1996) i stanowią podstawę powszechnie stosowanej w praktyce tzw. analizy portfelowej.

Analiza papierów wartościowych (*security analysis*) stanowi metodologię określania wartości papierów wartościowych (głównie akcji) w oparciu o prognozy przyszłych dywidend i zysków danej firmy. Obejmuje szeroko pojętą analizę wskaźników makroekonomicznych i branżowych¹, analizę sprawozdań finansowych i wybranych wskaźników finansowych dla danej spółki, oraz wykorzystanie określonych modeli wyceny akcji, jak np. model zdyskontowanych dywidend (*dividend discount model*) w powiązaniu z określonymi modelami wzrostu

¹ Zyski firmy zależą w dużej mierze od otoczenia ekonomicznego w którym ona funkcjonuje.

dywidendy². Działalność ta jest podstawą tzw. analizy fundamentalnej, której wynikiem może być wyszukanie papierów niedoszacowanych (dla których wartość rynkowa jest niższa od wartości otrzymanej wyceny) oraz analogicznie papierów przeszacowanych, i w konsekwencji podjęcie odpowiednich decyzji inwestycyjnych.

Analiza pochodnych papierów wartościowych obejmuje zagadnienia wyceny instrumentów pochodnych (derywatów), przede wszystkim opcji. Powszechnie stosowanym modelem służącym do określania wartości tych papierów jest model Blacka-Scholesa (Black & Scholes, 1973).

Duże znaczenie w problematyce analizy rynków finansowych i wykorzystania dostępnej informacji do podejmowania decyzji inwestycyjnych wniosła **hipoteza efektywności rynków finansowych** (*Efficient Market Hypothesis – EMH*) (zob. Fama, 1970; Fama, 1991). Opierając się na pewnych ogólnych założeniach stwierdza ona, iż cała dostępna informacja o rynku jest już odzwierciedlona w aktualnych cenach i w związku z tym przyszłe ruchy cen akcji są nieprzewidywalne i podlegają one jedynie przypadkowym zmianom (*random walk*). W zależności od kontekstu w jakim rozumiemy pojęcie informacji o rynku, wyróżniamy następujące pojęcia efektywności rynków:

- słaba efektywność (*weak-form efficiency*) - dotyczy rynków, na których informacja o przeszłych ruchach cen i wolumenie obrotu jest już uwzględniona w aktualnym poziomie cen, informacje te nie przyniosą zatem ponadprzeciętnych zysków;
- średnia efektywność (*semistrong-form efficiency*) - jest pojęciem szerszym, zakładającym że cała publicznie dostępna informacja o rynku i o perspektywach działalności poszczególnych firm znalazła już odzwierciedlenie w cenach akcji;
- mocna efektywność (*strong-form efficiency*) - wiąże się z hipotezą, że cała informacja, również aktualnie niedostępna publicznie, jest już odzwierciedlona w aktualnych cenach akcji, a więc nawet wykorzystanie informacji poufnych (*insider trading*) nie przyniesie ponadprzeciętnych zysków.

² Por. (Bodie et al., 1993, Jajuga & Jajuga, 1996).

Fakt braku efektywności danego rynku nawet w słabej postaci³, który może być potwierdzony odpowiednimi badaniami statystycznymi, lub też zakwestionowanie pewnych założeń, na których oparta jest hipoteza efektywności rynków, i tym samym wniosków z tej teorii płynących, stanowią przesłanki do wykorzystania metod **analizy technicznej** w celu wspomagania decyzji inwestycyjnych⁴. Analiza techniczna zajmuje się próbą prognozowania przyszłych ruchów cen akcji na podstawie dotychczasowych przebiegów ich kursów oraz wolumenu obrotów. Jej aspektami są analiza trendów, średnich kroczących, wielkości obrotów, formacji rysujących się na wykresach cenowych, oraz wskaźników technicznych. Opiera się nie tyle na naukowych podstawach matematycznych, co na różnych hipotezach wysnutych przez analityków na podstawie wieloletnich obserwacji rynków finansowych (należą do nich np. teoria Dow'a, teoria Elliotta, i szereg innych).

1.2. Aktywne zarządzanie portfelem inwestycyjnym.

Dwa podstawowe zagadnienia związane z działalnością inwestycyjną opartą na teorii portfela to analiza poszczególnych papierów wartościowych pod kątem określenia ich oczekiwanych stóp zysku, ryzyka i korelacji pomiędzy nimi, oraz konstruowanie portfela w oparciu o wyniki tej analizy i odpowiednie podstawy teoretyczne. Przy szacowaniu oczekiwanej stopy zysku często stosowaną metodą jest przyjęcie średniej stopy zysku z pewnego okresu z przeszłości, a zatem przyjęcie założenia, że zysk z danego papieru oraz związane z nim ryzyko (a także korelacja z innymi papierami) będą się kształtowały podobnie jak w przeszłości. W istocie teoria EMH wyklucza przydatność prognoz rynkowych przyjmując, że wszelkie informacje mogące być podstawą tego typu prognoz zostały już odzwierciedlone w bieżących cenach. Przy takich założeniach celowe jest przyjęcie pasywnej metody zarządzania portfelem inwestycyjnym, polegającej na konstrukcji zdywersyfikowanego portfela w oparciu o podstawy teoretyczne, a następnie na jego konsekwentnym trzymaniu z możliwością

³ Dotyczy to zwłaszcza rynków stosunkowo słabo rozwiniętych, jak np. w Polsce.

⁴ Analogiczne rozumowanie z uwzględnieniem średniej efektywności rynku można przeprowadzić w celu uzasadnienia możliwości wykorzystania analizy fundamentalnej.

okresowych korekt w przypadku zmian charakterystyk poszczególnych jego składników.

Bardziej wnikliwe podejście do zagadnienia zarządzania portfelem inwestycyjnym może wyłonić pewne przesłanki, stanowiące uzasadnienie dla prób dokonywania prognoz przyszłego zachowania się rynku oraz poszczególnych jego instrumentów. W teorii EMH można dopatrzeć się bowiem pewnych wad, które mogą spowodować jej zakwestionowanie przez część inwestorów (poddane w wątpliwość może być m.in. jej istotne założenie przyjmujące pełną racjonalność zachowań uczestników rynku i nie uwzględniające tym samym ewidentnie występującego czynnika psychologicznego⁵). Innym aspektem jest potwierdzony przez szereg badań fakt, iż nawet rynki rozwinięte (jak np. rynek w USA) charakteryzują się występowaniem pewnych zjawisk zaprzeczających pełnej efektywności tych rynków nawet w słabej postaci (zob. np. Martinez Abascal & Pregel, 1994); tego typu zjawiska, jak np. efekt styczniowy, efekt końca tygodnia, czy też efekt małych firm, mogą być znacznie bardziej wyraźne na słabiej rozwiniętych rynkach.

Innego typu postulaty, stojące w sprzeczności do hipotezy rynków efektywnych (EMH), wywodzą się z nabierającej coraz większej popularności teorii chaosu deterministycznego. W jej ujęciu rynki kapitałowe stanowią nieliniowe systemy dynamiczne, posiadające naturę fraktalną (Peters, 1997), przy czym charakteryzują się one występowaniem tzw. efektów długotrwałej pamięci⁶. Teorię tą potwierdzają liczne badania finansowych szeregów czasowych, wykorzystujące tzw. analizę R/S (statystykę Hursta) (np. Peters, 1997; Jaśiń, 1998).

Przedstawione wyżej fakty mogą uzasadniać celowość wykorzystywania prognoz zachowania się rynku i/lub prognoz stóp zysku poszczególnych papierów wartościowych, a także analiz poszczególnych papierów pod kątem ich niedoszacowania lub przeszacowania. Tego typu prognozy, generowane np. na podstawie wskazań analizy technicznej, analizy fundamentalnej, lub odpowiednio

⁵ Zob. (Plummer, 1995).

⁶ Efekty te są związane z wpływem poprzednich wartości szeregu czasowego z pewnego (dość długiego) okresu przeszłości na bieżące jego realizacje i pojawiają się w tzw. persystentnych szeregach czasowych, charakteryzujących się występowaniem wyraźnych trendów.

skonstruowanych modeli statystycznych, mogą być szczególnie przydatne na rynkach instrumentów o dużym stopniu ryzyka, jak np. rynek akcji i ich pochodnych. Związane z tymi prognozami strategie inwestycyjne zakładać mogą kupno lub sprzedaż całego portfela papierów wartościowych (ew. relokację środków finansowych w inny typ papierów, np. sprzedaż akcji i kupno obligacji) w przypadku wykorzystywania prognoz rynku (*ang. market timing*), lub też kupno walorów niedoszacowanych i sprzedaż przeszacowanych w przypadku wykorzystywania ocen poszczególnych papierów (*ang. security selection*). Działalność taką, polegającą na dokonywaniu permanentnych korekt składu portfela na podstawie bieżących prognoz i analiz, określa się mianem **aktywnego zarządzania portfelem** (*ang. active portfolio management*), zob. np. (Bodie et al., 1993).

Zasadniczym elementem warunkującym efektywność aktywnych metod zarządzania portfelem inwestycyjnym jest źródło możliwie wiarygodnych i precyzyjnych prognoz rynkowych i/lub właściwych sygnałów kupna lub sprzedaży poszczególnych walorów lub grup walorów, tylko bowiem wtedy strategia aktywna może okazać się lepsza od pasywnej. Szereg badań statystycznych potwierdza w pewnych przypadkach możliwość uzyskiwania efektywnych prognoz i tym samym skuteczność aktywnych strategii inwestycyjnych (zob. np. Taylor & Yoder, 1994; Lee & Rahman, 1990; Kester, 1990). Powstaje zatem dylemat skonstruowania odpowiednich narzędzi służących do generowania jak najlepszych prognoz lub sygnałów kupna/sprzedaży. Dylemat ten stanowi niewątpliwą inspirację do prowadzenia badań naukowych nad rozwojem instrumentów oraz doskonaleniem technik służących do predykcji rynków finansowych i generowania sygnałów transakcyjnych, przy wykorzystaniu m.in. najnowszych osiągnięć informatyki, w tym metod sztucznej inteligencji.

1.3. Systemy wspomagania decyzji inwestycyjnych.

Stałym elementem stosowania aktywnych strategii inwestycyjnych jest ciągła analiza informacji napływających z rynku, generowanie bieżących prognoz w oparciu o te informacje i w konsekwencji podejmowanie odpowiednich decyzji inwestycyjnych w oparciu o uzyskane prognozy. Precyzyjne analizy dużych ilości informacji w

połączeniu z koniecznością podejmowania szybkich decyzji inwestycyjnych, pozbawionych dodatkowo czynnika emocjonalnego, mogą stanowić trudność nawet dla doświadczonych analityków. Fakt ten stanowi przesłankę do wykorzystania do tego celu komputerowych systemów wspomagania decyzji inwestycyjnych (systemów decyzyjnych, *ang. trading systems*), których zadaniem jest automatyczne generowanie decyzji dotyczących kupna lub sprzedaży poszczególnych walorów notowanych na rynku i realizacja tym samym określonej aktywnej strategii inwestycyjnej zdeterminowanej przez konstrukcję systemu. Jakkolwiek stosowanie tego typu systemów nie jest warunkiem koniecznym efektywności aktywnych metod zarządzania portfelem inwestycyjnym, może jednak istotnie poprawić skuteczność podejmowanych decyzji.

1.3.1. Problemy modelowania rynków finansowych.

Elementami systemu decyzyjnego mogą być odpowiednie modele reprezentujące wiedzę na temat danego rynku (systemu rzeczywistego) zawartą w postaci równań matematycznych lub reguł logicznych. Konstrukcja modeli (modelowanie), jako powszechnie stosowana metoda opisu rzeczywistych systemów, sprowadza się do stworzenia takiego typu równań lub reguł, które zdolne są opisać system możliwie wiernie z punktu widzenia realizacji założonego celu. Model nie musi zatem odzwierciedlać wszystkich zależności i zjawisk zachodzących wewnątrz rzeczywistego systemu, a jedynie uwzględniać te zmienne wejściowe, zmienne stanu i zmienne wyjściowe oraz powiązania między nimi, które mają istotny wpływ na osiągnięcie przyjętego celu. W szczególności celem takim może być uzyskiwanie wiarygodnych prognoz dynamiki cen akcji lub wybranych wskaźników rynkowych, albo generowanie decyzji kupna lub sprzedaży dla poszczególnych walorów notowanych na rynku. Zadania takie mogą realizować wybrane klasy modeli ekonometrycznych (stanowiące podstawę konstrukcji odpowiedniego predyktora)⁷, albo odpowiednie modele decyzyjne (zob. np. Bierman et al., 1986; Levin et al., 1986). Elementy generujące prognozy lub decyzje niekoniecznie muszą być

⁷ Często wykorzystywane są tu np. modele typu ARIMA, a także ARCH (zob. np. Engle, 1993). Por. też (Pawłowski, 1982; Bortoli, 1996).

zbudowane w postaci równań matematycznych, mogą to być odpowiednio skonstruowane systemy ekspertowe⁸, wykorzystujące przechowywaną w formie reguł logicznych wiedzę o danym rynku, pochodzącą od ekspertów z zakresu inwestowania. Niekiedy systemy ekspertowe mogą być zintegrowane z klasyczną procedurą tworzenia portfela papierów wartościowych (zob. Lee J.K. et al., 1990).

Rozważając problem stworzenia odpowiednich modeli mogących wchodzić w skład systemów decyzyjnych, natrafiamy na bardzo istotną barierę charakterystyczną dla wszystkich rynków finansowych. Rynki te są bowiem niezwykle skomplikowanymi systemami dynamicznymi, charakteryzującymi się m.in.:

- występowaniem olbrzymiej liczby czynników (zmiennych egzogenicznych), oraz zmiennych stanu, mających wpływ na odpowiedź systemu (zachowanie się rynku); zmienne te są często powiązane relacjami trudnymi lub wręcz niemożliwymi do zidentyfikowania,
- występowaniem znacznych szumów i zakłóceń o trudnych do zidentyfikowania parametrach (szumy te mają często podłoże psychologiczne i emocjonalne związane z zachowaniem się inwestorów).

Wymienione czynniki sprawiają, że budowa efektywnych predyktorów lub modeli decyzyjnych (i w konsekwencji całego systemu decyzyjnego) w tradycyjny sposób, a więc na podstawie znajomości praw rządzących rynkiem lub poprzez identyfikację systemu na podstawie obserwacji jego wejść i wyjść, może okazać się zadaniem niezwykle trudnym. Dodatkowy istotny problem stanowi tu trudno przewidywalna zmienność systemu rzeczywistego w czasie, co powoduje konieczność budowy modelu niestacjonarnego lub ciągłej adaptacji przynajmniej parametrów modelu, a być może także jego struktury.

⁸ W ogólnym pojęciu system ekspertowy zawiera zestaw wnioskujących reguł logicznych, stworzonych przez eksperta z danej dziedziny wiedzy, generujących odpowiednie prognozy, diagnozy lub decyzje na podstawie dostarczonych informacji wejściowych.

1.3.2. Wykorzystanie narzędzi alternatywnych - przesłanki metodologiczne.

W świetle omówionych problemów modelowania rynków finansowych, w procesie konstrukcji systemów decyzyjnych lub ich elementów generujących prognozy albo decyzje, celowe wydaje się wykorzystanie pewnych narzędzi alternatywnych w stosunku do klasycznych metod budowy modeli. Narzędzia te, uwzględniając fakt stosunkowo niskiego poziomu wiedzy apriorycznej o systemie rzeczywistym, powinny spełniać następujące postulaty:

- pozyskiwać dodatkową, przydatną do celów generowania prognoz i/lub decyzji wiedzę - podczas etapu konstrukcji elementów predykcyjnych lub decyzyjnych,
- selekcjonować wiedzę istotną dla założonych celów, przy jednoczesnej eliminacji czynników nieistotnych,
- wykorzystywać metody nieliniowe i nieparametryczne, które ani nie zakładają *a priori* postaci funkcji tworzącej model (zależności pomiędzy wejściem i wyjściem), ani nie czynią założeń dotyczących postaci i parametrów rozkładów zmiennych losowych,
- charakteryzować się odpornością na występujące w rzeczywistym systemie szумы.

Przy realizacji powyższych postulatów obiecujące może okazać się wykorzystanie metodologii budowy modułów⁹ systemu decyzyjnego, opartych na następujących narzędziach:

1. **Sztuczne sieci neuronowe** - posiadają one zdolność nabywania wiedzy (a następnie jej przechowywania i późniejszego wykorzystywania) na podstawie otrzymanych dostępnych informacji¹⁰ z otoczenia (rynku finansowego), pochodzących z pewnego okresu przeszłości. Adaptacyjny proces gromadzenia

⁹ Pojęcie „moduł” oznacza tu element systemu decyzyjnego, który na podstawie określonych danych wejściowych generuje na wyjściu wartość prognozy danego wskaźnika rynku lub kursu papieru wartościowego, albo sygnał rekomendujący kupno, sprzedaż lub brak dokonania transakcji dla danego instrumentu finansowego.

¹⁰ Informacje te nazywać będziemy danymi uczącymi.

wiedzy i wykształcania relacji pomiędzy wejściem i wyjściem nazywamy procesem uczenia, w jego trakcie struktura sieci (modułu systemu decyzyjnego) nie podlega zmianom, natomiast modyfikowane są wartości jej parametrów (tzw. wag). Sztuczne sieci neuronowe funkcjonują analogicznie do rzeczywistych struktur neuronowych w biologicznym mózgu, stanowią one przy tym metodę nieliniową i nieparametryczną (zob. Cherkassky & Lari-Najafi, 1991).

2. **Algorytmy genetyczne** - realizują iteracyjną procedurę umożliwiającą konstrukcję i ocenę kolejnych, coraz efektywniejszych modułów (rozwiązań pośrednich), zmierzającą do znalezienia rozwiązania (modułu) optymalnego. Gromadzenie wiedzy następuje w procesie łączenia informacji z najlepiej ocenionych poprzednich rozwiązań i jej wykorzystania do budowy następnych. Ocena jest dokonywana poprzez testowanie modułu w oparciu o rzeczywiste dane z rynku. Jedną ze znaczących zalet tych narzędzi jest ich wysoka efektywność przy braku praktycznie jakichkolwiek wstępnych założeń.

Należy podkreślić, że powyższe dwie metodologie nie tylko nie wykluczają się wzajemnie, ale nawet w pewien sposób uzupełniają. Możliwe jest zatem zastosowanie rozwiązań hybrydowych, wykorzystujących oba podejścia łącznie do budowy elementów systemu decyzyjnego, jak również użycie drugiej metody jako rozszerzenia i uzupełnienia pierwszej.

Zwrócenia uwagi wymagają następujące (niekiedy uznawane za negatywne) aspekty zastosowania powyżej zaproponowanych narzędzi do budowy systemów decyzyjnych lub ich elementów:

- w przypadku zastosowania sieci neuronowych brak jest możliwości określenia *explicite* zależności między wejściem i wyjściem sieci, a tym samym podania uzasadnienia dla tych relacji; jest to wada tego narzędzia eksponowana przez niektórych badaczy i praktyków,
- praktycznie jedynym sposobem walidacji (weryfikacji) modułu stworzonego przy użyciu tych narzędzi jest ocena wyników jego testowania na rzeczywistych

danych¹¹; proces ten należy przeprowadzić w oparciu o dane pochodzące z możliwie długiego i niezbyt dawnego okresu czasu (lub kilku okresów),

- użycie powyższych narzędzi wymaga bardzo dużego nakładu obliczeń komputerowych, niezbędnych do realizacji procesu pozyskiwania wiedzy.

Niewątpliwie w ostatnich latach można jednak zaobserwować intensywny rozwój badań w dziedzinach sieci neuronowych, algorytmów genetycznych, i ich zastosowań na rynkach finansowych. Posiada on odzwierciedlenie w coraz liczniejszych publikacjach na ten temat w prasie i w licznych periodykach naukowych. Publikacje te, oprócz wyników badań stanowiących wkład naukowy w rozwój metod opartych na sieciach neuronowych i algorytmach genetycznych, często zawierają również informacje o efektywnym wykorzystaniu praktycznym tych narzędzi na rynkach finansowych. Fakty te (pomimo pojawiających się niekiedy pewnych głosów sceptycznych) stanowią dodatkową, mocną inspirację do prowadzenia badań związanych z zastosowaniem wyżej przedstawionych metod w systemach wspomagania decyzji inwestycyjnych.

1.3.3. Wykorzystanie narzędzi alternatywnych - przesłanki informatyczne.

Obecnie większość programów komputerowych wykorzystywanych w dziedzinie zarządzania portfelem inwestycyjnym realizuje przetwarzanie i analizę napływających z rynku danych na postać wynikową posiadającą najczęściej formę wykresów lub tabel. Prezentowane w takiej formie informacje stanowią niewątpliwie istotną pomoc dla analityka posługującego się analizą techniczną, fundamentalną lub portfelową i ułatwiają podejmowanie finalnych decyzji. Jednak w dobie gwałtownego rozwoju informatyki i technologii związanej z produkcją komputerów oraz powszechnej dostępności tych narzędzi, powstają coraz większe i stale rosnące rezerwy mocy obliczeniowej komputerów, będących w dyspozycji inwestorów operujących na rynkach finansowych. Fakt ten stanowi inspirację do wykorzystania

¹¹ W przypadku genetycznie skonstruowanych modułów w postaci decyzyjnych reguł logicznych, na ogół możliwa jest także ich dodatkowa interpretacja i weryfikacja przez eksperta.

tych rezerw do realizacji dalszych etapów przetwarzania napływających z rynku informacji, a więc do tworzenia i zastosowania komputerowych implementacji coraz bardziej zaawansowanych systemów decyzyjnych. Postęp w dziedzinie informatyki stwarza więc coraz szersze pole do budowy nowych technik wspomagania decyzji inwestycyjnych w oparciu o komputerową analizę informacji daleko wykraczającą poza formę prostego przetwarzania i zwykłej prezentacji wykresów i wskaźników, a także do prowadzenia badań nad przydatnością tych technik w procesie zarządzania portfelem.

Powyższe czynniki szczególnie sprzyjają rozwojowi metod sztucznej inteligencji (ang. *artificial intelligence*, w skrócie AI) i ich zastosowań do konstrukcji systemów decyzyjnych. Do metod sztucznej inteligencji zaliczamy m.in. sieci neuronowe i algorytmy genetyczne, wymienione już poprzednio jako obiecujące narzędzia w zagadnieniach prognozowania i wspomagania decyzji na rynkach finansowych. Wspólną cechą technik AI jest ich rodowód wywodzący się z biologii i obserwacji natury. Jakkolwiek prace badawcze nad narzędziami opartymi o AI trwają już stosunkowo długo, to właśnie wzrost możliwości obliczeniowych dostępnych komputerów stał się zasadniczym czynnikiem stymulującym obserwowany obecnie intensywny rozwój zarówno badań naukowych jak i zastosowań praktycznych tych metod.

Podkreślenia wymaga jeszcze jeden istotny aspekt przemawiający za rozwojem i szerokim stosowaniem omawianych narzędzi. Zarówno sieci neuronowe jak i algorytmy genetyczne posiadają naturalną właściwość, polegającą na zdolności do przetwarzania informacji w sposób równoległy. Cecha ta wywodzi się z biologicznych, rzeczywistych pierwowzorów tych metod, bowiem zarówno w mózgu, jak i w podlegającym ewolucji systemie przyrodniczym, procesy przetwarzania informacji mają właśnie charakter równoległy. Pomimo faktu, iż większość obecnych implementacji sztucznych sieci neuronowych i algorytmów genetycznych dokonywana jest na komputerach sekwencyjnych, należy liczyć się z szybkim rozwojem komputerowych architektur równoległych a także

specjalizowanych układów elektronicznych¹², umożliwiającą realizację operacji przetwarzania informacji w sposób rzeczywiście równoległy. Fakt ten pozwala na bardzo znaczące przyspieszenie obliczeń związanych z wykorzystaniem metod sztucznej inteligencji, a tym samym na istotną poprawę ich efektywności w zastosowaniach praktycznych.

¹² Tego typu układy (neurokomputery), realizujące sprzętowo implementację sieci neuronowych są już produkowane i są też dostępne w powszechnej sprzedaży.

2. Wybrane aspekty zastosowania sieci neuronowych w analizie rynków finansowych.

Badania nad konstrukcją i zastosowaniem sieci neuronowych¹³ (SN) w procesach przetwarzania informacji zostały zainspirowane przez koncepcje naśladowania funkcjonowania naturalnych struktur komórek nerwowych zawartych w mózgu, przy jednoczesnym wykorzystaniu dużej szybkości i mocy obliczeniowej komputera. W istocie sposoby analizy informacji przez sztuczną sieć neuronową są bardzo podobne do procesów zachodzących w rzeczywistych mózgowych strukturach neuronów, jednak elementarne operacje przetwarzania sygnałów mogą odbywać się w układach elektronicznych komputera z dużo większą szybkością.

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

¹³ Właściwie: sztucznych sieci neuronowych (*ang. artificial neural networks*).

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

2.1. Rys historyczny i aktualne kierunki badań nad sieciami neuronowymi.

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

Pierwszą sprzętową implementację sieci neuronowej (tzw. neurokomputer) stanowiło urządzenie o nazwie *Mark I Perceptron*, zbudowane w 1957 roku przez Rosenblatta i Wightmana (zob. Rosenblatt, 1958). Układ ten, zawierający 512 losowo zadanych połączeń pomiędzy neuronami, służył do rozpoznawania znaków przedstawianych na matrycy o wymiarach 20 x 20 pikseli. Kolejną interesującą konstrukcją neurokomputera była zbudowana przez Widrowa w 1960 roku sieć o nazwie *Madaline*, służąca do adaptacyjnego przetwarzania sygnałów.

Zarówno układy sprzętowe, jak i pojawiające się coraz liczniej prace teoretyczne zmierzające do skonstruowania modeli biologicznych struktur neuronowych, zaowocowały w latach 60-tych intensywnym postępem badań nad problematyką sieci neuronowych. Rozwój ten został jednak istotnie zahamowany na około 15 lat

wskutek opublikowania książki *Perceptrons* (Minsky & Papert, 1969), w której autorzy dowodzili bardzo ograniczonej przydatności sieci jednowarstwowych. W okresie tym prowadzono jednak szereg prac w tej dziedzinie, spośród których na uwagę zasługują m.in. prace Grossberga, Kohonena, Werbosa, Andersona, Fukushima i Hopfielda. W 1974 roku Werbos opracował po raz pierwszy skuteczny algorytm uczenia wielowarstwowych nieliniowych sieci neuronowych (Werbos, 1974) - algorytm wstecznej propagacji błędu (*backpropagation*) - stanowiący obecnie podstawową metodę uczenia tego typu sieci. Ponowne niezależne odkrycie tego algorytmu w połowie lat 80-tych przez Rumelharta, Hintona i Williamsa (Rumelhart et al., 1986) przyczyniło się do przełamania istniejącego impasu i gwałtownego rozwoju badań nad konstrukcją, metodami uczenia i zastosowaniami różnych typów sieci neuronowych (zob. Tadeusiewicz, 1993).

Obecnie obserwuje się olbrzymie zainteresowanie problematyką sieci neuronowych i ich zastosowań, owocujące lawinowo rosnącą liczbą książek oraz publikacji w licznych periodykach poświęconych tej tematyce (należą do nich m.in. *Neural Networks*, *IEEE Transactions on Neural Networks*, *Neural Computing & Applications*, *Journal of Computational Intelligence in Finance* (poprzednio *Neurovest Journal*), *Neurocomputing*, *Engineering Application of Artificial Intelligence*, i inne), a także w różnych wydawnictwach związanych z pokrewnymi, a często odległymi dziedzinami wiedzy. Organizowana jest też na całym świecie ogromna liczba konferencji naukowych dotyczących tej problematyki. Na uwagę zasługuje również powstanie dużej ilości firm oferujących pakiety oprogramowania służące do komputerowej implementacji sieci neuronowych pod kątem ogólnym, bądź do specjalistycznych zastosowań (m.in. finansowych), a także posiadających w swojej ofercie rozwiązania sprzętowe (neurokomputery) (zob. Tadeusiewicz, 1993; Rutkowska et al., 1996; *Selecting...*, 1996).

Badania nad praktycznymi zastosowaniami sieci neuronowych obejmują m.in. następujące dziedziny (Tadeusiewicz, 1993; Osowski, 1996):

- Szeroko pojęte rozpoznawanie obrazów¹⁴, sprowadzające się do wyłowienia istotnych cech danego obiektu lub sygnału i przydzielenia go do odpowiedniej klasy. Przykłady stanowią tu m.in. rozpoznawanie i analiza obrazów wizyjnych, rozpoznawanie dźwięków (np. mowy lub sygnałów sonarowych), rozpoznawanie ręcznie pisanych znaków (np. kodów pocztowych), weryfikacja podpisów, wykrywanie podejrzanych obiektów w bagażach lotniczych i inne.
- Klasyfikację bezwzorcową, sprowadzającą się do określenia struktury grupowej obiektów w przestrzeni cech (zob. też Kohonen, 1995; Grabowski, 1997).
- Sterowanie obiektów dynamicznych (np. automatów, robotów, procesów przemysłowych).
- Szeroko pojęte prognozowanie, głównie w zastosowaniach technicznych lub ekonomicznych (np. predykcja obciążeń systemu elektroenergetycznego, prognozy kursów walut, itp.)
- Przetwarzanie sygnałów, obejmujące m.in. filtrację sygnałów, kompresję, transformację, kodowanie i dekodowanie informacji, redukcję danych (np. analizę głównych składowych).
- Badania dotyczące nowych rozwiązań pamięci komputerowych (pamięci asocjacyjne, pamięci rozproszone), a także algorytmów automatycznego uczenia się maszyn.

Większość z powyższych kierunków badań zaowocowała również efektywnymi zastosowaniami praktycznymi sieci neuronowych na gruncie ekonomii i zarządzania (zob. np. Tadeusiewicz, 1993; Tadeusiewicz, 1995; Morajda, 1997a). Przykładowe zagadnienia, w których wykorzystywane są te narzędzia, to:

- ocena zdolności kredytowej podmiotów gospodarczych,
- prognozowanie stóp procentowych i kursów walut,
- prognozowanie wielkości sprzedaży w przedsiębiorstwie,

¹⁴ Zob. też (Tadeusiewicz & Mikrut, 1994; Tadeusiewicz 1994b; Jajuga, 1995).

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

2.2. Ogólna charakterystyka sieci neuronowych.

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

Zdolność do nabywania wiedzy poprzez uczenie. W przeciwieństwie do klasycznych modeli komputerowych, których budowa opiera się na algorytmie realizującym określoną metodę rozwiązania danego problemu, sieć neuronowa nie wymaga wcześniejszego zdefiniowania sposobu przetwarzania przez nią informacji,

nie wymaga zatem procesu programowania, wykorzystującego wiedzę aprioryczną. Proces ten zastępowany jest przez uczenie sieci, w trakcie którego sieć „zdobywa” wiedzę na podstawie prezentowanych jej danych uczących¹⁵. Fakt ten pozwala na efektywne zastosowanie SN także w problemach, gdzie posiadamy niewielką wiedzę na temat analizowanego zagadnienia, albo też w sytuacjach gdy stworzenie stosownego modelu komputerowego w postaci programu byłoby bardzo trudne, natomiast dysponujemy odpowiednią ilością danych mogących posłużyć do uczenia sieci. Warto przy tym dodać, że ilość danych potrzebna do efektywnego użycia sieci neuronowej (w szczególności niezbędnych do tego, by poprawnie zrealizować proces uczenia) jest znacznie mniejsza, niż ilość danych wymagana przy analizach statystycznych.

Równoległy charakter przetwarzania informacji. Informacja w SN jest przetwarzana jednocześnie we wszystkich komórkach sieci¹⁶, co zapewnia równoległość obliczeń i w konsekwencji daje możliwość znacznego przyspieszenia procesu przetwarzania informacji; fakt ten jest istotny zwłaszcza w fazie uczenia sieci¹⁷.

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

¹⁵ Zagadnienie to jest szerzej potraktowane w rozdziale 2.5.

¹⁶ Oczywiście w implementacjach sieci na komputerach sekwencyjnych równoległość ta jest symulowana i nie przynosi widocznych korzyści.

¹⁷ Interesujące są tu porównania szybkości pracy maszyny cyfrowej i mózgu człowieka: najszybsze w świecie superkomputery pracują z szybkością rzędu 10^{10} operacji na sekundę, zaś neurony mogą przyjmować i wysyłać impulsy o częstotliwości zaledwie do 1000 Hz, jednak szybkość pracy całej kory mózgowej, zawierającej ok. 10^{10} neuronów i ok. 10^{15} połączeń między nimi, ocenia się na 10^{18} operacji na sekundę (Tadeusiewicz, 1993).

Nieliniowość¹⁸ i nieparametryczność. Przyjęcie nieliniowego modelu neuronu prowadzi do uzyskania nieliniowej zależności między wejściem a wyjściem dla całej sieci. Przy odpowiedniej architekturze sieci może ona zrealizować dowolne odwzorowanie wiążące wartości wejściowe i wyjściowe sieci¹⁹. Metodę opartą na sieciach neuronowych uznaje się również za metodę nieparametryczną w sensie braku konieczności przyjmowania a priori postaci funkcyjnej modelu (zob. też Cherkassky & Lari-Najafi, 1991; Refenes, 1995; Grabowski, 1997). Cechy te mają istotne znaczenie w aspekcie konieczności modelowania zjawisk nieliniowych, występujących m.in. na rynkach finansowych. Jakkolwiek dla tego typu problemów mogą być wykorzystane inne nieliniowe techniki, to SN ze względu na łatwość implementacji i możliwość realizacji dowolnych odwzorowań wejście-wyjście mogą okazać się narzędziem najbardziej efektywnym (zob. Azoff, 1994).

Zdolność do adaptacji. Wagi połączeń w nauczonej sieci neuronowej realizującej określony model, jako parametry sieci podlegające modyfikacji w procesie uczenia, mogą podlegać dalszym adaptacyjnym zmianom w odpowiedzi na zmieniające się warunki zewnętrzne środowiska. W praktyce może to być realizowane w procesie „douczenia” sieci, polegającym na kontynuacji uczenia przy zastosowaniu zmodyfikowanego ciągu uczącego zawierającego „bardziej aktualne” wzorce. Tego typu adaptacja wag sieci może mieć istotne znaczenie w przypadku, gdy sieć ma za zadanie realizować model niestacjonarny, przystosowujący się do zmian zachodzących w systemie rzeczywistym. Powstaje jednak w tym przypadku dylemat określenia wielkości ciągu uczącego i ustalenia kryterium „aktualności” danych²⁰.

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

¹⁸ Pomijamy tu klasę liniowych SN, mających dość ograniczone pole zastosowań.

¹⁹ Dotyczy to tylko sieci wielowarstwowych, podstawę matematyczną stanowi tu twierdzenie Kołmogorowa (zob. np. Hecht-Nielsen, 1990).

²⁰ Dylemat ten jest dobrze widoczny w zagadnieniu analizy szeregów czasowych w przypadku użyciu do uczenia i walidacji sieci danych pochodzących z ostatniego okresu czasu: okres ten powinien być na tyle długi, aby zapewnić właściwą zdolność sieci do generalizacji i uchronić ją przed uczeniem się szumu, zaś z drugiej strony na tyle krótki, aby sieć potrafiła wychwycić i zaadoptować się do istotnych zmian w systemie rzeczywistym.

- uniwersalność modeli opartych na SN w sensie jednolitości metod konstrukcji, uczenia i analizy SN dla różnych zastosowań (Haykin, 1994),
- odporność SN na uszkodzenia struktury²¹,
- możliwość implementacji SN w postaci układów elektronicznych VLSI.

Praktycznie wszystkie SN, będące przedmiotem badań i zastosowań praktycznych, posiadają budowę warstwową. Oznacza to, że w strukturze sieci można wyróżnić pewną liczbę warstw, kolejno przetwarzających informację, z których każda zawiera określoną liczbę neuronów²². Z reguły sposób funkcjonowania neuronów tej samej warstwy jest identyczny. W każdej sieci możemy wyodrębnić warstwę wejściową (zwaną również warstwą zerową), której zadaniem jest jedynie dystrybuowanie w sieci sygnałów wejściowych (nie przetwarza ona informacji), oraz warstwę wyjściową generującą sygnały wyjściowe sieci. Warstwy pośrednie (jeżeli występują), położone pomiędzy warstwą wejściową a wyjściową, nazywamy warstwami ukrytymi (nie wymieniają one informacji bezpośrednio z otoczeniem sieci).

Klasyfikacji sieci neuronowych można dokonywać według różnych kryteriów. Poniżej wymieniono najistotniejsze metody podziału SN (szczegółową analizę można znaleźć w literaturze, np. Tadeusiewicz, 1993; Hertz et al., 1993; Haykin, 1994).

Podział SN według charakteru połączeń w sieci:

- sieci z połączeniami jednokierunkowymi²³ (*feedforward*),
- sieci ze sprzężeniami zwrotnymi²⁴.

²¹ Cecha ta ma niebagatelne znaczenie w biologicznych systemach nerwowych, np. mózg człowieka tracąc codziennie od kilku do kilkudziesięciu tysięcy neuronów nie traci praktycznie swoich funkcjonalnych zdolności; właściwość ta została także zauważona już w okresie początkowych badań nad SN, np. perceptron Rosenblatta zachowywał zdolność rozpoznawania znaków nawet po przerwaniu pewnej części połączeń pomiędzy neuronami; wydaje się jednak, że cecha ta ma obecnie mniejsze znaczenie w praktyce (Tadeusiewicz, 1993).

²² Pojęcie „neuron” będzie w tej pracy używane zarówno w odniesieniu do biologicznej komórki nerwowej, jak i do elementu sztucznej sieci neuronowej; ponieważ znaczenie tego pojęcia zawsze jasno wynika z kontekstu, nie powinny powstawać związane z tym dwuznaczności.

Podział według ilości warstw:

- sieci jednowarstwowe²⁵,
- sieci wielowarstwowe (zawierające co najmniej jedną warstwę ukrytą).

Podział według charakteru realizowanego odwzorowania wejście-wyjście:

- sieci liniowe²⁶;
- sieci nieliniowe.

Podział według metody uczenia sieci:

- sieci uczone w sposób nadzorowany (tzw. uczenie z nauczycielem),
- sieci samouczące się (uczenie bez nauczyciela, tj. w sposób nienadzorowany).

Jednym z ważniejszych i częściej stosowanych rodzajów SN są sieci typu *perceptron*. Jest to klasa sieci z połączeniami jednokierunkowymi, nieliniowych, uczonych w sposób nadzorowany (mogą one być jedno- lub wielowarstwowe). Ten rodzaj sieci jest wysoce efektywny w różnych zastosowaniach praktycznych i dlatego jest najczęściej używany w zagadnieniach prognozowania szeregów czasowych (Azoff, 1994). Właśnie tego typu sieci będą przedmiotem dalszych rozważań w niniejszej pracy.

2.3. Model pojedynczego neuronu.

Podstawę do stworzenia matematycznego modelu neuronu²⁷ stanowi opis sposobu przetwarzania informacji przez biologiczną komórkę nerwową²⁸ wchodzącą w skład

²³ W takich sieciach przepływ informacji odbywa się tylko w jednym kierunku, tj. od warstwy wejściowej do wyjściowej, poprzez warstwy pośrednie (ukryte), a zatem sygnały wyjściowe neuronów danej warstwy są przekazywane tylko do warstw następnych. W tego typu sieciach nie występują zatem sprzężenia zwrotne.

²⁴ Najczęściej analizowanymi SN ze sprzężeniami zwrotnymi są tzw. sieci Hopfielda, których nazwa pochodzi od nazwiska badacza, który zaproponował tego typu struktury (zob. Hopfield, 1982).

²⁵ Sieci jednowarstwowe zawierają warstwę wejściową i warstwę wyjściową, jednak ponieważ przy określaniu liczby warstw sieci nie uwzględniamy warstwy wejściowej jako nie przetwarzającej informacji (zob. np. Haykin, 1994, str. 18), tak więc według poprawnej terminologii sieć dwuwarstwowa to sieć z jedną warstwą ukrytą.

²⁶ Inne nazwy tych sieci to ADALINE i MADALINE. Zawierają one tylko neurony posiadające liniową funkcję przejścia (por. podrozdział 2.3) i wobec tego realizują liniowe odwzorowanie wejście-wyjście. Wyczerpujące omówienie tego typu sieci przedstawione jest w (Tadeusiewicz, 1993).

większej struktury systemu nerwowego. Komórka taka posiada dużą liczbę wypustek - tzw. dendrytów - za pośrednictwem których odbiera ona sygnały od innych neuronów, oraz jedną wypustkę zwaną aksonem, za pomocą której przekazuje wygenerowane przez siebie sygnały w formie impulsów elektrycznych do kolejnych komórek (zob. rys. 2-1).

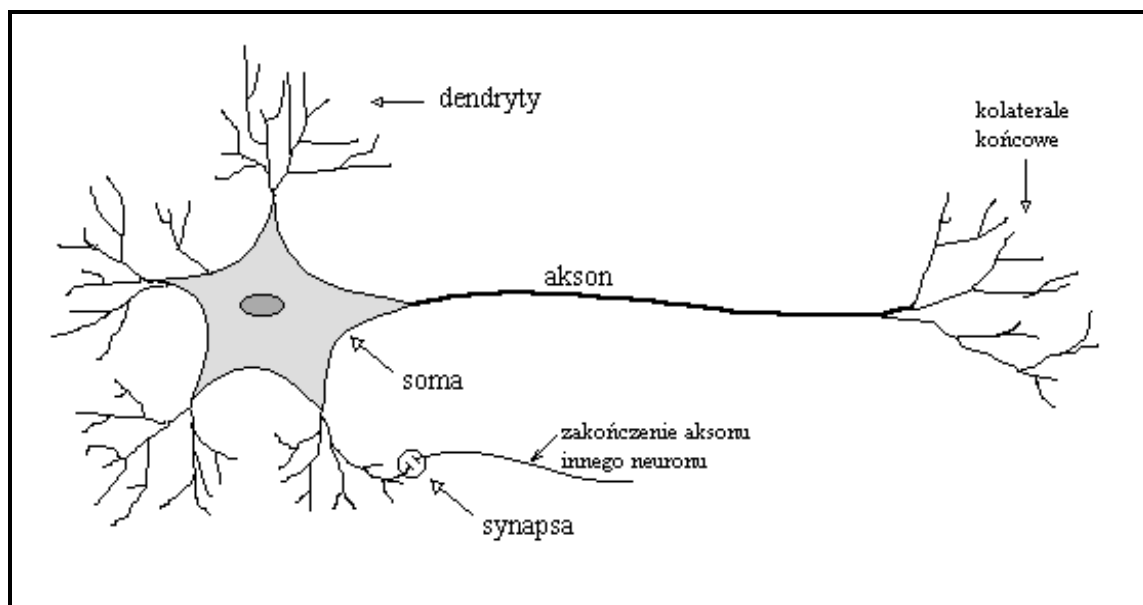
Wartość i charakter (pobudzający lub hamujący) odebranego przez komórkę nerwową sygnału zależy w dużej mierze od elementu stanowiącego bezpośrednie połączenie dwóch neuronów, tzw. synapsy. Przekształca ona w określony sposób docierające do niej impulsy i decyduje o wielkości przekazywanych dalej sygnałów. Charakter tego przekształcenia, a tym samym znaczenie (waga) transmitowanych sygnałów, może ulegać modyfikacjom (stanowi to bardzo istotną cechę synaps). Odebrane i przekazane za pośrednictwem synaps sygnały elektryczne są następnie przez komórkę sumowane i w przypadku gdy sumaryczny potencjał przekroczy pewną wartość progową, generowany jest tzw. impuls czynnościowy, który jest przekazywany (w tej samej postaci) do wszystkich połączeń (synaps) z neuronami odbierającymi informacje od danej komórki²⁹. Generalnie można stwierdzić, że częstotliwość tych impulsów, a w konsekwencji wielkość sygnałów odbieranych przez następne neurony zależy w sposób niemalejący i nieliniowy od wartości owego sumarycznego potencjału³⁰.

²⁷ Pionierskie prace w zakresie modelowania neuronu wykonali McCulloch i Pitts (zob. McCulloch & Pitts, 1943).

²⁸ Celem modelowania sieci neuronowych jest konstrukcja systemów *przetwarzania informacji*, działających w oparciu o podobieństwo do sposobu pracy mózgu. Stąd też wynika celowość analizowania tylko takich modeli neuronu, które opisują zjawiska przetwarzania informacji z pominięciem szeregu innych aspektów budowy i funkcjonowania komórki nerwowej. Pełną analizę zagadnienia modelowania neuronu można znaleźć w pracy (Majewski, 1981).

²⁹ Opisane tu bardzo skrótowo zjawiska są w istocie niezwykle skomplikowanymi procesami fizjologicznymi zachodzącymi w obrębie błony komórkowej neuronu (zob. Tadeusiewicz, 1994a).

³⁰ Konstruując statyczny model neuronu, w którym nie uwzględniamy dynamiki wynikającej z wpływu historii zachowania się neuronu na jego bieżące reakcje, wystarczy wymienić tą właśnie zależność, interpretując ją jako powiązanie aktualnych sygnałów wejściowych z wyjściem w danej chwili. W przypadku modelu dynamicznego należałoby również wziąć pod uwagę tzw. sumowanie czasowe sygnałów odbieranych przez synapsy, wpływ poprzednio wygenerowanych impulsów na zmiany wartości progowej, opóźnienia w transmisji sygnałów przez błonę komórkową i poprzez synapsy, itp. (zob. Tadeusiewicz, 1994a; Morajda, 1991). W tej pracy rozważać będziemy tylko postać statyczną modelu neuronu, ponieważ modele dynamiczne nie zyskały jak dotąd większego zainteresowania badaczy i nie doczekały się w zasadzie zastosowań praktycznych.



Rys. 2-1. Schematyczny rysunek komórki nerwowej. Źródło: opracowanie własne na podstawie (Tadeusiewicz, 1994a).

Z analizy wyżej opisanych procesów wynikają następujące istotne przesłanki, z których wynika kształt biocybernetycznego modelu neuronu:

- konieczne jest uwzględnienie zmiennych wag synaps, decydujących o wielkości (znaczeniu) odbieranych sygnałów,
- należy zastosować sumowanie przekazanych przez synapsy, już zmodyfikowanych sygnałów, z uwzględnieniem (odjęciem) wartości progowej,
- należy określić funkcję, definiującą zależność wartości sygnału wyjściowego neuronu³¹ od obliczonej jak wyżej sumy,
- neuron generuje tylko jeden, przekazywany dalej sygnał wyjściowy.

Model taki można zapisać w formie następującej zależności:

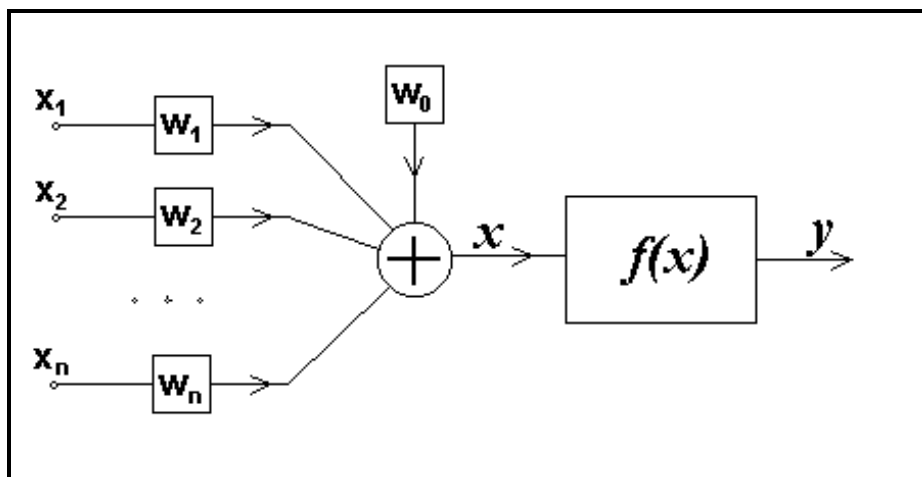
$$y = f\left(\sum_{i=0}^n x_i w_i\right) \quad (2.1)$$

³¹ Nawiązując do biologicznego neuronu, wartość tą można interpretować jako aktualną częstotliwość generowanych przez neuron impulsów czynnościowych.

gdzie:

- y - wartość wyjściowa neuronu,
- n - liczba wejść,
- x_1, x_2, \dots, x_n - wartości sygnałów wejściowych³² dla neuronu,
- $x_0 = \text{const.} = 1$
- w_1, w_2, \dots, w_n - wartości wag połączeń wejściowych dla danego neuronu, określające znaczenie poszczególnych wejść,
- w_0 - składnik stały (*ang. bias*), reprezentujący wartość progową (z przeciwnym znakiem)
- f - funkcja określająca zależność wyjścia od ważonej sumy wejść, zwana funkcją przejścia lub funkcją aktywacji (*ang. transfer function*).

Graficzny schemat tak skonstruowanego modelu prezentuje rys. 2-2.



Rys. 2-2. Schemat modelu neuronu. Źródło: opracowanie własne na podstawie (Haykin, 1994).

Poszczególne wersje modelu neuronu, stosowane w SN, różnią się przede wszystkim doбором funkcji przejścia f . Funkcja ta w dużym stopniu decyduje o właściwościach pojedynczego neuronu, a tym samym całej sieci. W niektórych przypadkach stosuje się funkcje liniowe postaci $f(x) = b \cdot x$, jednak tego typu

modele mają ograniczone zastosowania³³. Spośród nieliniowych funkcji przejścia najczęściej stosowane są funkcje monotoniczne niemalejące, które można podzielić na dwie klasy w zależności od charakteru ich zbioru wartości (zbioru możliwych wartości wyjściowych neuronu):

- Klasa I (zbiór wartości symetryczny względem punktu zero) zawiera funkcje posiadające na ogół formę antysymetryczną, których przeciwdziedzinę mogą stanowić: przedział domknięty $[-a, a]$, przedział otwarty $(-a, a)$, lub zbiór dwuelementowy $\{-a, a\}$;
- Klasa II (zbiór wartości zawierający tylko liczby nieujemne); funkcje te są tworzone przez przesunięcie odpowiednich funkcji z klasy I, polegające na dodaniu do ich wartości stałej a ; przeciwdziedziną dla tych funkcji może być odpowiednio: przedział domknięty $[0, 2a]$, przedział otwarty $(0, 2a)$, lub zbiór dwuelementowy $\{0, 2a\}$;

W obydwu wyżej omówionych przypadkach $a \in \mathbf{R}$, $0 < a < \infty$.

Wybór odpowiedniej klasy funkcji przejścia jest ważnym aspektem konstrukcji modelu neuronu i zależy w dużej mierze od specyfiki problemu, dla którego sieć będzie zastosowana³⁴. Natomiast dobór stałej a wynika z zakresu wartości sygnałów, które ma przetwarzać neuron; zakres ten z kolei jest pochodną zarówno analizowanego zagadnienia, jak i zastosowanego sposobu normalizacji danych

³² Są to sygnały odbierane przez synapsy danego neuronu, a zatem sygnały wyjściowe innych, połączonych z nim neuronów.

³³ Używa się ich w liniowych sieciach typu ADALINE lub MADALINE (gdzie z reguły stosuje się funkcję tożsamościową przyjmując, że $b = 1$, zob. Tadeusiewicz, 1993), a niekiedy również jako elementów warstw wyjściowych sieci nieliniowych (Azoff, 1994).

³⁴ Już porównanie k -wymiarowego sygnału binarnego (wartości 0 lub 1) z bipolarnym (wartości -1 lub 1) prowadzi w pierwszym przypadku do zbioru możliwych wartości w postaci jednostkowego hipersześcianu w przestrzeni k -wymiarowej, zaś w drugim przypadku do jednostkowej k -wymiarowej sfery; wbrew intuicji są to zbiory diametralnie różne (Tadeusiewicz, 1993). Innym ważnym aspektem jest fakt, iż sygnały o wartościach bliskich zera praktycznie nie przyczyniają się do procesu przetwarzania informacji przez sieć neuronową (intuicyjnie można to wyjaśnić tym, że sygnał taki po pomnożeniu przez dowolny współczynnik wagowy dalej będzie miał bliską zera wielkość i wniesie znikomą ilość informacji w operacji sumowania); jest zatem istotną kwestią czy zero ma się znajdować pośrodku „zakresu roboczego” sygnałów neuronów, czy też na jego krańcu.

wyjściowych³⁵. Należy także podkreślić, że w każdym przypadku zbiory wartości funkcji f są tu zbiorami ograniczonymi³⁶.

Analizując zagadnienie doboru funkcji przejścia, należy też zwrócić uwagę na różniczkowalność branej pod uwagę funkcji. Niekiedy wygodniejsze wydaje się użycie funkcji nieróżniczkowalnych (jak np. zaprezentowanych niżej funkcji progowej lub progowo-liniowej), jednak ze względu na fakt występowania w nich punktów, w których pochodna nie istnieje, funkcje te nie nadają się do zastosowania w sieciach uczonych za pomocą algorytmów wykorzystujących pochodne funkcji aktywacji (jak np. powszechnie używany algorytm wstecznej propagacji błędów, omówiony w podrozdziale 2.5).

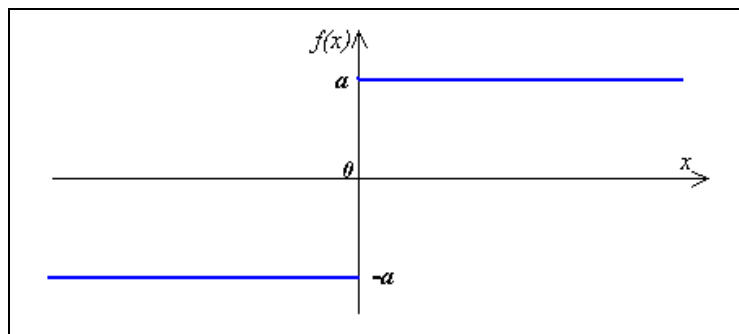
Poniżej przedstawiono kilka najczęściej stosowanych funkcji przejścia (zaprezentowano tylko funkcje należące do klasy I z uwagi na fakt, że odpowiednie funkcje z klasy II można otrzymać przez trywialną transformację, polegającą na przesunięciu i ewentualnym przeskalowaniu wartości funkcji z klasy I). W formułach (2.2) do (2.8) parametr a decyduje o zakresie wartości danej funkcji, natomiast parametr b (jeśli występuje) - o jej nachyleniu. We zależnościach tych przyjęto ogólne założenie, iż $b > 0$.

Funkcja progowa (rys. 2-3).

$$f(x) = \begin{cases} -a & \text{dla } x \leq 0 \\ a & \text{dla } x > 0 \end{cases} \quad (2.2)$$

³⁵ Problem normalizacji zostanie szeroko omówiony w dalszej części pracy.

³⁶ Cecha ta odpowiada właściwościom rzeczywistego neuronu, który może generować impulsy czynnościowe o ograniczonej częstotliwości (od zera do ok. 1000 Hz). W istocie jeżeli sumaryczne pobudzenie neuronu nie przekracza wartości progowej, nie generuje on żadnych impulsów, co można interpretować jako sygnał o wartości zero. Po przekroczeniu tego progu częstotliwość impulsów rośnie nieliniowo wraz ze wzrostem pobudzenia, nigdy jednak, nawet przy ekstremalnych wartościach pobudzenia, nie przekracza pewnej wartości granicznej (ok. 1000 Hz), co wynika ze zjawiska tzw. refrakcji bezwzględnej (zob. Tadeusiewicz, 1994a). Fakt ten znalazł odbicie w zagadnieniu doboru funkcji przejścia modelu neuronu, gdzie często stosuje się funkcje osiągające stan nasycenia dla wartości argumentów spoza określonego przedziału (jak np. zaprezentowane dalej funkcje progowo-liniowa, sigmoidalna, tangens hiperboliczny i inne).



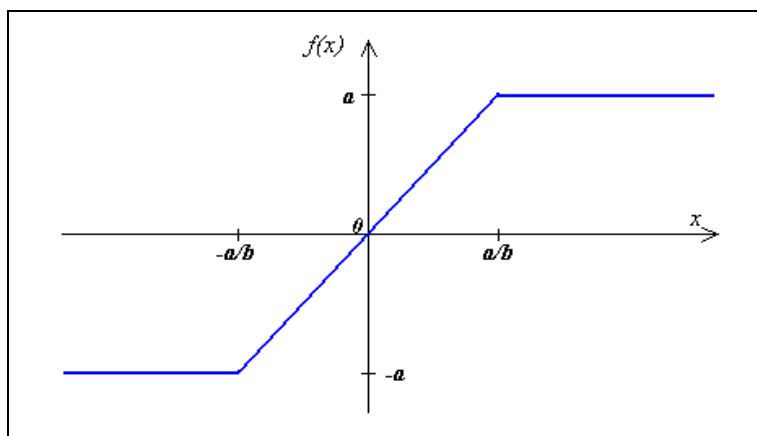
Rys. 2-3. Funkcja progowa. Źródło: opracowanie własne.

Należącym do klasy II odpowiednikiem dla tej funkcji (po przeskalowaniu) jest funkcja skoku jednostkowego (Heaviside'a). Funkcja ta była używana w klasycznych konstrukcjach sieci typu *perceptron* (Rosenblatt, 1958; Tadeusiewicz, 1993). Przy zastosowaniu tego typu funkcji perceptron może być interpretowany jako układ realizujący pewną operację logiczną w oparciu o logikę dwuwartościową (zob. McCulloch & Pitts, 1943; Tadeusiewicz, 1991). Wadą takich funkcji jest ich nieróżniczkowalność.

Należy też zwrócić uwagę, że w przypadku użycia tego rodzaju funkcji składnik stały w_0 (*bias*) we wzorze (2.1) może być rzeczywiście interpretowany jako wartość progowa z przeciwnym znakiem, określająca wartość sumarycznego pobudzenia neuronu, powyżej której przechodzi on w stan aktywny ($+a$).

Funkcja progowo-liniowa (*ang. piece-wise linear*, rys. 2-4).

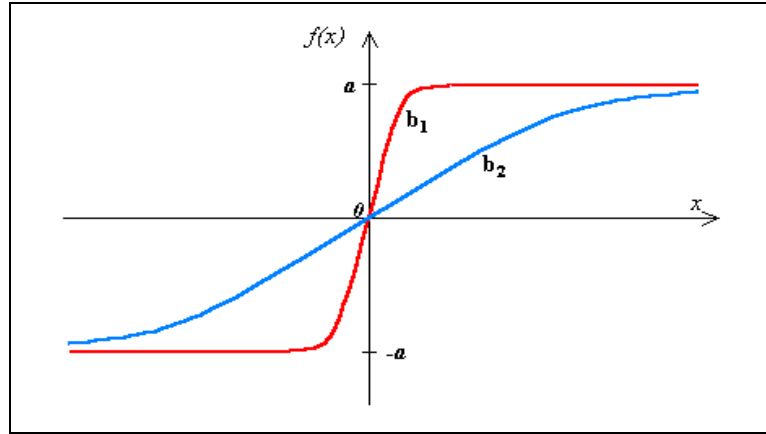
$$(2.3)$$



Rys. 2-4. Funkcja progowo-liniowa. Źródło: opracowanie własne.

Funkcja sigmoidalna³⁷ (inaczej: logistyczna, *ang. sigmoid*, rys. 2-5).

$$f(x) = \frac{2a}{1 + e^{-bx}} - a \quad (2.4)$$



Rys. 2-5. Funkcja sigmoidalna dla dwóch różnych wartości parametru b ($b_1 > b_2$).

Źródło: opracowanie własne.

Zaletą tej funkcji jest nie tylko różniczkowalność, ale także prostota obliczania jej pochodnej, co w przypadku algorytmów uczenia SN wykorzystujących pochodną funkcji aktywacji znacznie poprawia efektywność procesu obliczeń. Pochodna funkcji (2.4) określona jest zależnością (2.5):

$$f'(x) = \frac{b(a^2 - f(x)^2)}{2a} \quad (2.5)$$

Funkcja tangens hiperboliczny.

$$f(x) = a \cdot \tanh(bx) = a \cdot \frac{e^{bx} - e^{-bx}}{e^{bx} + e^{-bx}} \quad (2.6)$$

Funkcja erf (error function).

$$f(x) = \frac{2a}{\sqrt{\pi}} \int_0^{bx} e^{-t^2} dt \quad (2.7)$$

³⁷ Przedstawiono tu funkcję przeskalowaną i przesuniętą o stałą a (zbiorem wartości oryginalnej funkcji

Funkcja typu $x/(1+|x|)$.

$$f(x) = a \cdot \frac{bx}{1 + |bx|} \quad (2.8)$$

Funkcje (2.6), (2.7) i (2.8) są funkcjami antysymetrycznymi i różniczkowalnymi, oraz mają postać bardzo podobną³⁸ do funkcji sigmoidalnej przedstawionej na rys.2-5.

Niekiedy używane są również inne nieliniowe funkcje przejścia nie należące do rozważanej dotychczas klasy funkcji monotonicznych. Można do nich zaliczyć m.in. funkcje sinus, cosinus, oraz funkcję Gaussa (zob. Azoff, 1994). Funkcje te są stosunkowo rzadko używane w SN stosowanych w problemach analizy szeregów czasowych, nie będą też przedmiotem zainteresowania w niniejszej pracy.

2.4. Model warstwowej sieci neuronowej typu *perceptron*.

Konstrukcja modelu warstwowej sieci neuronowej polega na połączeniu szeregu modeli pojedynczych neuronów w strukturę warstwową, z uwzględnieniem następujących aspektów:

- określenie ilości warstw sieci;
- sprecyzowanie liczby neuronów w poszczególnych warstwach;
- dobór modeli neuronów w poszczególnych warstwach³⁹;
- zdefiniowanie charakteru i liczby połączeń w sieci.

Zagadnienia te, wiążące się generalnie z określeniem architektury i parametrów sieci, nie doczekały się dotychczas pełnych i jednoznacznych rozwiązań, jakkolwiek pojawiła się olbrzymia liczba publikacji dotyczących zarówno teoretycznych

sigmoidalnej jest przedział $(0, 1)$.

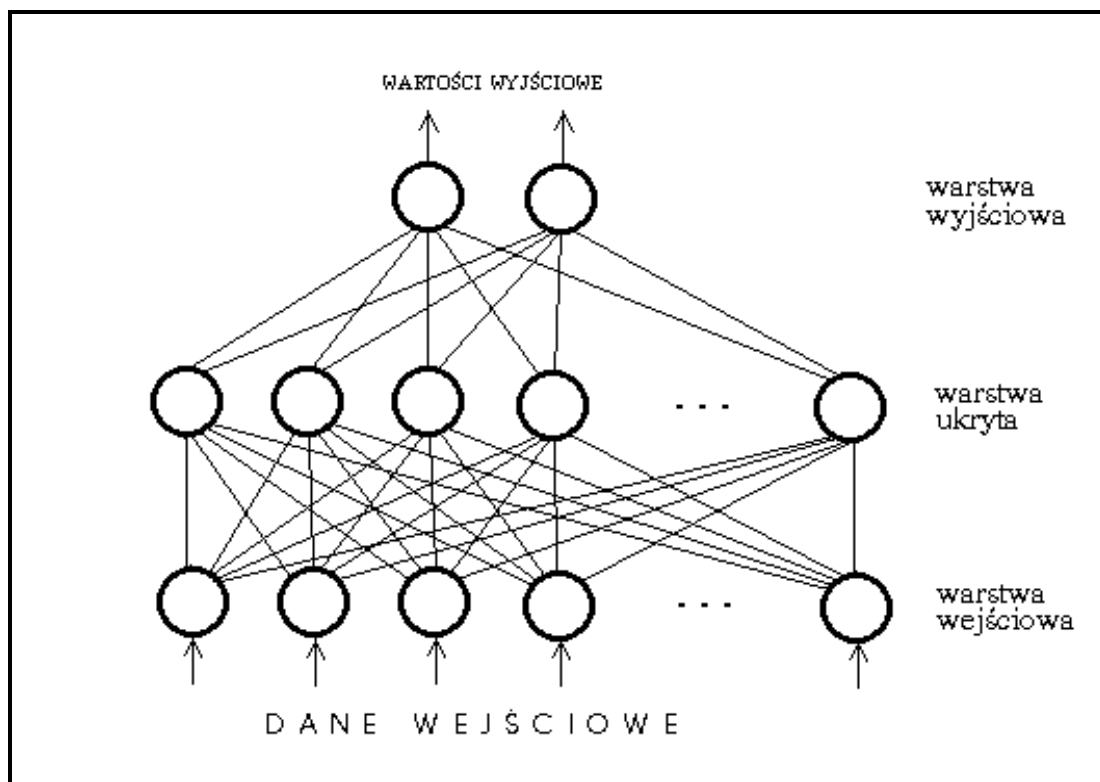
³⁸ Z uwagi na ten fakt pominięto prezentację wykresów tych funkcji.

³⁹ Z reguły przyjmuje się założenie, że postać modelu dla wszystkich neuronów danej warstwy jest identyczna, często założenie to rozszerza się na całą sieć. Założenie to jest wygodne i prowadzi do oszczędności obliczeń podczas uczenia i eksploatacji sieci, nie jest jednak warunkiem koniecznym.

aspektów związanych z tą problematyką, jak i wnoszących liczne przesłanki wynikające z praktycznych eksperymentów. Ogólnie można stwierdzić, że rozwiązanie powyższych zagadnień zależy od specyfiki analizowanego problemu, ale po części również od arbitralnych decyzji badacza dokonującego wyboru właściwego narzędzia. Niewątpliwie oprócz kierowania się wnioskami płynącymi z opublikowanych wyników badań, wymagane jest także przeprowadzenie pewnej ilości własnych eksperymentów zmierzających do wyznaczenia właściwych wartości parametrów sieci dla konkretnego problemu.

Należy ponadto podkreślić, że ostateczna architektura sieci nie musi być zdefiniowana całkowicie a priori, gdyż może ona ulegać modyfikacjom w trakcie procesu uczenia i walidacji sieci lub po zakończeniu wstępnego etapu tego procesu. Zmiany te mogą polegać np. na zwiększaniu lub zmniejszaniu ilości neuronów w warstwach ukrytych (według pewnego algorytmu) w celu otrzymania w efekcie ich optymalnej liczby⁴⁰. Inną metodą optymalizacji struktury sieci jest tzw. obcinanie wag (*weight pruning*), polegające na eliminowaniu w wyuczonej sieci połączeń nieistotnych i tym samym na zmniejszaniu ogólnej liczby wag (Azoff, 1994; Hertz et al., 1993). Ponadto zwrócić należy uwagę na fakt, że architektura sieci nie musi wcale pozostawać niezmienna, w trakcie jej eksploatacji może ona podlegać adaptacyjnym zmianom wraz z analogicznymi modyfikacjami współczynników wagowych.

Ponieważ globalna analiza powyższych zagadnień wykraczałaby daleko poza ramy niniejszej pracy, przeprowadzono jedynie (w następnych rozdziałach) dyskusję tej problematyki w kontekście zastosowania SN do analizy danych pochodzących z rynków finansowych, oraz przedstawiono propozycje pewnych rozwiązań wykorzystanych następnie w przeprowadzonych badaniach. Dość wyczerpujące, ogólniejsze ujęcia tematyki związanej z doбором architektury i parametrów sieci można znaleźć m.in. w książkach: (Haykin, 1994), (Tadeusiewicz, 1993), (Osowski, 1996), (Hertz et al., 1993), (Azoff, 1994), (Hecht-Nielsen, 1990), (Żurada, 1992) i innych.



Rys. 2-6. Przykładowa architektura warstwowej sieci neuronowej z jedną warstwą ukrytą.
Każdy neuron warstwy ukrytej lub wyjściowej jest połączony z wszystkimi neuronami należącymi do warstwy poprzedniej. Źródło: opracowanie własne.

Na rys. 2-6 zaprezentowano przykładowy schemat warstwowej sieci neuronowej z połączeniami jednokierunkowymi (*feedforward*). Sieć ta posiada jedną warstwę ukrytą, oraz dwuelementową warstwę wyjściową. Neurony określonej warstwy odbierają sygnały *tylko* z warstwy poprzedniej, zaś charakter połączeń jest zrealizowany według schematu „każdy z każdym” (oznacza to, że dany neuron warstwy ukrytej lub wyjściowej jest połączony ze *wszystkimi* neuronami należącymi do warstwy bezpośrednio poprzedzającej).

⁴⁰ W literaturze zaproponowano szereg algorytmów realizujących dobór optymalnej architektury sieci, jak np. algorytm upstart, algorytm kaskadowo-korelacyjny, czy też algorytm kafelkowy (zob. Hertz et al., 1993).

2.5. Metody uczenia sieci neuronowych.

Proces uczenia⁴¹ sieci neuronowej sprowadza się do wyznaczenia takiego zbioru parametrów połączeń (wag) sieci W_0 , dla którego określona w przestrzeni wag tzw. funkcja błędu $F(W)$ (ang. *error function, cost function*) osiąga minimum. Wartość funkcji błędu F dla określonego zestawu wag W stanowi sumaryczną miarę błędów sieci, czyli różnic pomiędzy wartością wzorcową, a faktyczną wartością wyjściową wygenerowaną przez sieć. Wartość funkcji błędu zazwyczaj określa się i kumuluje dla wszystkich elementów ciągu uczącego. Funkcja ta jest najczęściej definiowana jedną z następujących zależności⁴² (Azoff, 1994):

$$F = \frac{1}{2} \sum_i (y_i - z_i)^2 \quad (2.9)$$

$$F = \sum_i |y_i - z_i| \quad (2.10)$$

$$F = \sum_i \frac{|y_i - z_i|^u}{u}, \quad 1 \leq u \leq 2 \quad (2.11)$$

$$F = \frac{1}{2} \sum_i \left((1 + z_i) \ln \frac{(1 + z_i)}{(1 + y_i)} + (1 - z_i) \ln \frac{(1 - z_i)}{(1 - y_i)} \right) \quad (2.12)$$

gdzie: z_i - wartość wyjściowa (wzorcowa) i -tego elementu ciągu uczącego
 y_i - otrzymana faktyczna wartość wyjścia SN dla i -tego elementu ciągu uczącego

⁴¹ W niniejszym podrozdziale omówiono jedynie zagadnienie tzw. uczenia nadzorowanego (z nauczycielem), w którym dysponujemy poprawnymi wartościami wzorcowymi (pożądanymi) dla zmiennej wyjściowej sieci i wielkości te wykorzystujemy w procesie uczenia.

⁴² Zakładamy, że sieć posiada jeden element wyjściowy (założenie to przyjęto też w prezentowanej w dalszej części metodologii badań); gdyby neuronów wyjściowych było więcej, należałoby w celu znalezienia wartości funkcji błędu przeprowadzić także scalanie - na przykład sumowanie błędów obliczanych dla poszczególnych wyjść.

Najpowszechniej stosowana jest funkcja kwadratowa, określona zależnością (2.9), w związku z czym omawiana wielkość oznaczana bywa symbolem SSE (*Sum Squared of Errors*).

Poniżej przedstawiono klasyczną i jednocześnie najpowszechniej używaną metodę nadzorowanego uczenia sieci neuronowych (minimalizacji funkcji $F(W)$) dla SN typu *perceptron* - algorytm wstecznej propagacji błędów (ang. *backpropagation*) (Werbos, 1974; Rumelhart et al., 1986, Tadeusiewicz, 1993; Haykin, 1994; Osowski, 1996). Idea tego algorytmu sprowadza się do iteracyjnej⁴³ modyfikacji wektora wag sieci W w oparciu o gradient funkcji F , po rozpoczęciu tego procesu od (z reguły) losowo określonego wektora początkowego $W^{(1)}$. W danej i -tej iteracji (oznaczymy przez $W^{(i)}$ wartość wektora wag podczas tej iteracji) najpierw określony zostaje gradient $\nabla F(W^{(i)})$ funkcji błędu F w punkcie $W^{(i)}$ przestrzeni wag. Z kolei wyznaczony zostaje wektor korekty wag (tzn. wektor $W^{(i+1)} - W^{(i)}$), którego podstawowa składowa posiada kierunek największego spadku funkcji F w punkcie $W^{(i)}$ (przeciwny do gradientu), a druga składowa jest zgodna z wypadkowym kierunkiem zmiany wag w całej poprzedniej epoce uczenia (reprezentuje ona swoistą „bezwładność” procesu uczenia). Bardzo ważną rolę w procesie modyfikacji wag odgrywają dwa parametry⁴⁴: współczynnik uczenia η , decydujący o wielkości kroku korekty wag w kierunku największego spadku funkcji F , oraz współczynnik bezwładności (*momentum*)⁴⁵ μ^p , określający długość składowej wektora korekty wag równoległej do kierunku zmiany wag dokonanej w poprzedniej epoce.

Rozważmy sieć neuronową typu *perceptron*, posiadającą n elementów w warstwie wejściowej, m elementów w warstwie ukrytej (przy założeniu, że istnieje jedna taka

⁴³ Iteracja polega na jednokrotnej prezentacji pojedynczego elementu ciągu uczącego oraz na dokonaniu odpowiedniej korekty wag na podstawie tej prezentacji. Sumę iteracji po wszystkich elementach ciągu uczącego (tzn. pojedynczą prezentację całego ciągu uczącego) nazywamy *epoką uczenia*.

⁴⁴ Parametry te są na ogół odpowiednio modyfikowane w trakcie procesu uczenia.

⁴⁵ Wartość współczynnika bezwładności może zostać określona albo dla pojedynczej iteracji (korekty wag)-parametr ten oznaczać będziemy symbolem μ^p , albo dla całej epoki (odnosi się ona wtedy do sumy korekt wektora wag podczas pojedynczej prezentacji całego ciągu uczącego) - tak określony parametr oznaczmy symbolem μ . Wartości parametru μ stanowią L-krotność wielkości μ^p (gdzie L oznacza liczbę elementów w ciągu uczącym). O ile podczas omawiania algorytmu wstecznej propagacji błędów posługiwać się będziemy parametrem μ^p , o tyle w dalszej części pracy pojęcie współczynnika bezwładności (*momentum*) odnosić się będzie do parametru μ .

warstwa), oraz jeden element wyjściowy. Odnośnie do struktury sieci zakładamy, że każdy neuron warstwy ukrytej lub wyjściowej jest połączony ze wszystkimi neuronami z warstwy bezpośrednio poprzedzającej. Ponadto zakładamy, że neurony warstwy ukrytej oraz neuron wyjściowy są połączone ze źródłem stałego sygnału o wartości 1 - wagi tych połączeń stanowią dla tych neuronów składnik stały (*bias*) (zob. podrozdział 2.3). Przyjmijmy także, że funkcja aktywacji neuronów sieci jest przesuniętą funkcją sigmoidalną, zdefiniowaną zależnością (2.4).

Wprowadźmy dodatkowo następujące oznaczenia zakładając, że dotyczą one określonego *i*-tego elementu ciągu uczącego (dla uproszczenia indeks *i* został w tych oznaczeniach pominięty), przy czym $i = 1, 2, \dots, L$ (gdzie *L* jest ilością elementów w ciągu uczącym):

x_k - *k*-ta wartość wejściowa sieci ($k = 0, 1, 2, \dots, n$), czyli sygnał *k*-tego elementu warstwy zerowej (wejściowej) SN, przy czym z założenia $x_0 \equiv 1$ (sygnał związany z *biasem*).

v_j - wartość wyjściowa (sygnał) *j*-tego neuronu warstwy ukrytej ($j = 0, 1, 2, \dots, m$), przy czym $v_0 \equiv 1$ (sygnał związany z *biasem*).

w_j - wartość wagi połączenia pomiędzy *j*-tym neuronem warstwy ukrytej, a elementem wyjściowym (w_0 - *bias*).

w_{jk} - wartość wagi połączenia pomiędzy *k*-tym elementem wejściowym, a *j*-tym neuronem warstwy ukrytej (w_{j0} - *bias*, $j = 1, 2, \dots, m$).

y - wartość wyjściowa sieci (sygnał neuronu wyjściowego).

z - wartość wzorcowa (pożądana) dla wyjścia sieci.

Przy powyższych założeniach *algorytm wstecznej propagacji błędów* może być opisany następująco:

Krok 1. Określ wartości początkowe współczynników η oraz μ^p ,

Podstaw $e = 1$ (licznik epok uczenia),

Podstaw $\Psi = \mathbf{0}$ (Ψ - wektor równy zmianie wektora wag w poprzedniej epoce),

Określ losowy wektor wag początkowych $W^{(l)}$.

Krok 2. Dokonaj randomizacji (losowego mieszania elementów) ciągu uczącego.

Krok 3. Podstaw $i = 1$ (numer analizowanego elementu ciągu uczącego).

Krok 4. Oblicz wartość wyjściową sieci y dla wektora wejściowego (x_1, x_2, \dots, x_n) , wchodzącego w skład i -tego elementu ciągu uczącego.

Krok 5. Wyznacz wartość ϕ pochodnej funkcji błędu F względem y następująco:

$$\text{dla funkcji kwadratowej (2.9):} \quad \phi = y - z$$

$$\text{dla funkcji (2.10):} \quad \phi = \text{sgn}(y - z)$$

$$\text{dla ogólnej postaci funkcji (2.11):} \quad \phi = \frac{|y - z|^u}{y - z}$$

$$\text{dla funkcji (2.12):} \quad \phi = \frac{y - z}{1 - y^2}$$

Uwagi: Obliczając wartość ϕ uwzględniamy tylko jeden składnik funkcji F związany z rozważanym aktualnie i -tym elementem ciągu uczącego. Problem nieróżniczkowalności funkcji (2.10) i (2.11) dla $y = z$ rozwiązujemy przyjmując w tym punkcie $\phi = 0$. Symbol sgn oznacza funkcję signum.

Krok 6. Oblicz wartości pochodnych funkcji F względem wag połączeń pomiędzy neuronami warstwy ukrytej a elementem wyjściowym:

$$\frac{\partial F}{\partial w_j} = \phi \cdot \frac{b(a^2 - y^2)}{2a} \cdot v_j \quad j = 0, 1, 2, \dots, m \quad (2.13)$$

Ułamek stanowiący drugi czynnik prawej strony równania reprezentuje określoną zależnością (2.5) pochodną funkcji aktywacji $f(x) = \frac{2a}{1 + e^{-bx}} - a$ (zob. wzór (2.4)), przy czym stałe a oraz b (zdefiniowane w podrozdziale 2.3) określają odpowiednio zakres wartości oraz nachylenie funkcji f .

Krok 7. Oblicz wartości pochodnych funkcji F względem wag połączeń pomiędzy elementami wejściowymi SN a neuronami warstwy ukrytej:

$$\frac{\partial F}{\partial w_{jk}} = \phi \cdot \frac{b(a^2 - y^2)}{2a} \cdot w_j \cdot \frac{b(a^2 - v_j^2)}{2a} \cdot x_k \quad (2.14)$$

gdzie: $j = 1, 2, \dots, m$; $k = 0, 1, 2, \dots, n$; $x_0 = 1$

Powyższą zależność uzyskujemy po uwzględnieniu faktu: $v_j = f\left(\sum_{k=0}^n w_{jk} x_k\right)$

Krok 8. W oparciu o uzyskane w **kroku 6** i **kroku 7** elementy gradientu $\nabla F(W^{(i)})$ dokonaj modyfikacji wektora wag według formuły:

$$W^{(i+1)} = W^{(i)} - \eta \cdot \nabla F(W^{(i)}) + \mu^p \cdot \Psi \quad (2.15)$$

Drugi składnik określa zmianę wektora wag w kierunku największego spadku funkcji $F(W)$, natomiast trzeci składnik reprezentuje „bezwładność” procesu uczenia (tutaj przyjęto cząstkowe uwzględnianie tego elementu podczas każdej korekty wag, niektóre wersje algorytmu zakładają jednorazową korektę na końcu danej epoki uczenia o cały wektor $\mu \cdot \Psi$).

Krok 9. Podstaw⁴⁶: $i \leftarrow i + 1$. Jeżeli $i \leq L$ to przejdź do **kroku 4**.

Krok 10. Dokonaj ewentualnej modyfikacji współczynników η oraz μ^p .

Istnieje wiele omawianych w literaturze reguł modyfikacji tych współczynników (zob. np. Haykin, 1994; Osowski, 1996; Tadeusiewicz, 1993); w niniejszej pracy opracowano własną metodę adaptacyjnej korekty parametrów η oraz μ (zob. podrozdział 4.6).

Krok 11. Wyznacz wektor: $\Psi = W^{(i)} - W^{(1)}$

Podstaw: $W^{(1)} \leftarrow W^{(i)}$ oraz $e \leftarrow e + 1$

W kroku tym dysponujemy aktualną wartością funkcji błędu F jako sumarycznego błędu dla wszystkich L wzorców ciągu uczącego; może ona być wykorzystana do realizacji bieżącej oceny sieci i/lub do kontroli spełnienia kryterium stopu algorytmu.

Krok 12. Jeżeli $e > M$ (gdzie M oznacza maksymalną założoną ilość epok uczenia) lub spełnione zostało dodatkowe przyjęte kryterium stopu, to **zakończ** proces uczenia; w przeciwnym przypadku przejdź do **kroku 2**.

Zaprezentowany wyżej algorytm zakłada aktualizację wektora wag po każdorazowej prezentacji pojedynczego wzorca uczącego (ang. *pattern learning*), metoda ta jest oparta o wyznaczanie gradientu dla jednego składnika funkcji F , odpowiadającego temu wzorcowi. Inna wersja algorytmu wstecznej propagacji może zakładać jednorazową korektę wag dopiero na końcu każdej epoki uczenia w oparciu o gradient uwzględniający wszystkie elementy ciągu uczącego (ang. *batch learning*) (zob. np. Haykin, 1994; Morajda, 1996).

Należy też podkreślić, że istnieje cały szereg odmian i modyfikacji algorytmu wstecznej propagacji błędów, charakteryzujących się w niektórych zastosowaniach wyższą efektywnością (szybszą zbieżnością) od algorytmu klasycznego. Ponadto oprócz metody wstecznej propagacji opracowano szereg innych metod uczenia SN, zarówno gradientowych (np. algorytm zmiennej metryki, algorytm gradientów sprzężonych) (zob. np. Osowski, 1996), jak i bezgradientowych (np. metoda wektora losowego).

2.6. Przykłady wybranych z literatury zastosowań sieci neuronowych na rynkach finansowych.

Sieci neuronowe znalazły szereg zastosowań do analizy danych pochodzących z rynków finansowych i stały się obiektem zainteresowań zarówno badaczy rozwijających teoretyczne i praktyczne aspekty wykorzystania SN w tej problematyce, jak i praktyków (np. menedżerów funduszy inwestycyjnych) stosujących te narzędzia do wspomaganie decyzji inwestycyjnych. Ich prace

⁴⁶ Znak „ \leftarrow ” będzie w niniejszej pracy reprezentował operację podstawiania (przypisania), polegającą na obliczeniu wartości wyrażenia znajdującego się po prawej stronie tego symbolu, a następnie na przypisaniu tej wartości - zmiennej stojącej po lewej stronie.

zaowocowały olbrzymią liczbą publikacji związanych z wykorzystaniem SN do analizy i predykcji rynków lub poszczególnych instrumentów finansowych. Spektrum tej literatury rozpoczyna się od krótkich ogólnikowych raportów, zawierających głównie doniesienia o praktycznych rezultatach zastosowań SN (z reguły bez żadnych informacji na temat topologii używanej sieci), a kończy na opracowaniach czysto naukowych o charakterze badawczym. Publikacje poświęcone tej tematyce można znaleźć zarówno w periodykach poświęconych SN (np. *Neural Networks*), jak i w pismach z zakresu finansów (np. *Financial Analyst Journal*). Wydawane są również specjalistyczne czasopisma dotyczące zastosowań SN na rynkach kapitałowych (np. *Journal of Computational Intelligence in Finance* - poprzednio *Neurovest Journal*), a także wydawnictwa książkowe (zob. Azoff, 1994; Refenes, 1995). Poniżej krótko omówiono kilka przykładowych artykułów w celu przybliżenia prezentowanej problematyki.

W pracy (Trippi & DeSieno, 1992) opisano zastosowanie SN do prognozowania rynku akcji. Sześć odpowiednio nauczonych sieci zostało użytych do generowania sygnałów zakupu lub sprzedaży dla stosowanego na giełdzie nowojorskiej indeksu Standard&Poor's 500 (pojęcie kupna/sprzedaży indeksu można odnosić do transakcji dotyczącej pakietu akcji o składzie proporcjonalnym do zestawu papierów tworzących indeks, lub do kontraktu terminowego typu *financial futures* dla indeksu). Wartościami wejściowymi dla sieci są wybrane wielkości i wskaźniki techniczne oparte na danych z okresu ostatnich dwóch tygodni. Każda z sieci jest odpowiednio przygotowywana (douczana) przed kolejną sesją giełdową. W celu wygenerowania rekomendacji zakupu/sprzedaży w danej chwili stworzono prosty system podejmujący decyzje na podstawie wartości wyjściowych wszystkich sieci. Rezultaty funkcjonowania takiego systemu okazały się lepsze niż poszczególnych sieci osobno, jak również były znacznie lepsze od strategii pasywnej opartej o wzrost indeksu S&P 500.

Inny przykład zastosowania SN, tym razem w zagadnieniu wyboru właściwych akcji do portfela inwestycyjnego, opisano w pracy (Kryzanowski et al., 1993). Jako kryterium wyboru przyjęto prognozowany zysk z akcji danej firmy w ciągu najbliższego roku. W celu wykorzystania SN do predykcji zysków, przygotowano

ciąg uczący zawierający 88 zmiennych wejściowych, obejmujących 28 czynników fundamentalnych (związanych z czternastoma wybranymi wskaźnikami finansowymi danej firmy), 25 wielkości (również opartych o wskaźniki finansowe) przedstawiających pozycję danej firmy na tle swojej branży, oraz 35 wielkości stworzonych w oparciu o 7 wybranych wskaźników makroekonomicznych. Wykorzystano dane z 231 firm (przypadków). Uzyskane rezultaty, otrzymane dla niezależnego zbioru testującego, były dość obiecujące, np. w zagadnieniu klasyfikacji akcji na dwie grupy (odpowiednio przynoszące zysk i stratę) średnio w 66,4% przypadków sieć podejmowała prawidłowe decyzje, w 26,2% były to decyzje błędne, zaś w 7,4% przypadków sieć nie wygenerowała decyzji.

W pracy (Siripoulos et al., 1996) zaprezentowano wyniki badań zastosowania SN do prognozy krótkoterminowych trendów rynku akcji, oraz predykcji zmian kursu wybranych akcji w 10-dniowym horyzoncie czasowym. Jako informacje wejściowe dla SN wykorzystano wybrane wskaźniki analizy technicznej, a w przypadku prognoz dla akcji - także pewne dane o charakterze fundamentalnym. Jako typ SN zastosowano perceptron z jedną warstwą ukrytą, uczony za pomocą algorytmu wstecznej propagacji błędów, przeprowadzono także empiryczne badania wielkości warstwy ukrytej. Pokrótce przeanalizowano zagadnienia doboru i wstępnego przetwarzania danych wejściowych, wielkości ciągu uczącego, oraz długości procesu uczenia. Otrzymano obiecujące rezultaty funkcjonowania sieci, np. w przypadku predykcji kursu akcji otrzymano 68% poprawnych prognoz dotyczących wzrostu lub spadku kursu. Następnie zaproponowano aktywną strategię inwestycyjną wykorzystującą regułę logiczną⁴⁷ opartą na sygnałach wyjściowych SN. Ocena zastosowania tej strategii, dokonana w oparciu o szereg kryteriów (np. całkowity zysk, wskaźnik Sharpe'a⁴⁸, itp.) okazała się istotnie lepsza niż podobna ocena

⁴⁷ Decyzja kupna była generowana w sytuacji, gdy pojawiał się dodatni sygnał wyjściowy sieci i był on większy od sygnału z poprzedniego dnia; decyzja sprzedaży była podejmowana jeżeli sygnał wyjściowy sieci miał wartość niższą niż sygnał poprzedni oraz (warunek opcjonalny) gdy od chwili kupna upłynęło więcej niż 4 dni.

⁴⁸ Wskaźnik Sharpe'a (*Sharpe ratio*, *Sharpe measure*) stanowi uwzględniającą zarówno zysk jak i ryzyko miarę

oceny strategii inwestycyjnej lub portfela inwestycyjnego według formuły:
$$S = \frac{E(r) - r_f}{\sigma}$$
, gdzie:

S - wskaźnik Sharpe'a, $E(r)$ - wartość oczekiwana (estymowana najczęściej średnią arytmetyczną) stóp zwrotu z portfela (strategii), σ - odchylenie standardowe stóp zwrotu, r_f - stopa zwrotu wolna od ryzyka. Zob. np. (Bodie et al., 1993).

strategii pasywnej, przy czym wykazano dużą większą skuteczność aktywnej strategii opartej na SN dla rynku greckiego (jest to rynek rozwijający się o stosunkowo niskiej efektywności), niż dla znacznie bardziej rozwiniętego rynku niemieckiego. W pracy przedyskutowano także wybrane aspekty dotyczące ryzyka związanego z aktywną strategią inwestycyjną.

Podane wyżej przykłady stanowią niewielką próbkę pokazującą, jakie prace (w dużych ilościach) pojawiają się ostatnio w literaturze, potwierdzając możliwość i celowość wykorzystania sieci neuronowych jako nowego narzędzia do predykcji kursów akcji oraz do optymalizacji decyzji podejmowanych na rynkach finansowych. Wyniki te wskazują na zdecydowaną celowość podjęcia tych zagadnień także w odniesieniu do polskiego rynku finansowego, ponieważ - co warto podkreślić - specyfika poszczególnych rynków mocno wpływa na strukturę i parametry stosowanych rozwiązań.

2.7. Zasadnicze aspekty efektywnego wykorzystania sieci neuronowych w analizie i predykcji rynków finansowych.

Wyniki dotychczasowych badań teoretycznych i empirycznych, dotyczących konstrukcji i uczenia SN w oparciu o dane finansowe, oraz ich wykorzystania w analizie i predykcji rynków finansowych, wskazują na szereg aspektów związanych z tą problematyką, których uwzględnienie decyduje o efektywności funkcjonowania SN. Poniżej wskazano i pokrótce omówiono najważniejsze z nich.

2.7.1. Zdolność sieci do generalizacji - unikanie efektu przeuczenia.

Efekt przeuczenia SN jest to zjawisko polegające na tym, że sieć bardzo dobrze potrafi się nauczyć wzorców wchodzących w skład ciągu uczącego, podczas gdy praktycznie nie jest w stanie prawidłowo rozpoznawać i reagować na inne wzorce, które nie występowały w fazie uczenia; mówimy wówczas, że sieć traci zdolność do generalizacji (zob. podrozdział 2.2). Występowanie tego efektu eliminuje model oparty na SN z praktycznych zastosowań finansowych, w których oczekujemy przecież od sieci właściwych reakcji na nowe informacje wejściowe, pojawiające się

już po zakończeniu procesu uczenia. Dlatego też zdolność sieci do generalizacji jest w analizowanej problematyce bardzo ważnym, a być może najważniejszym elementem konstrukcji SN.

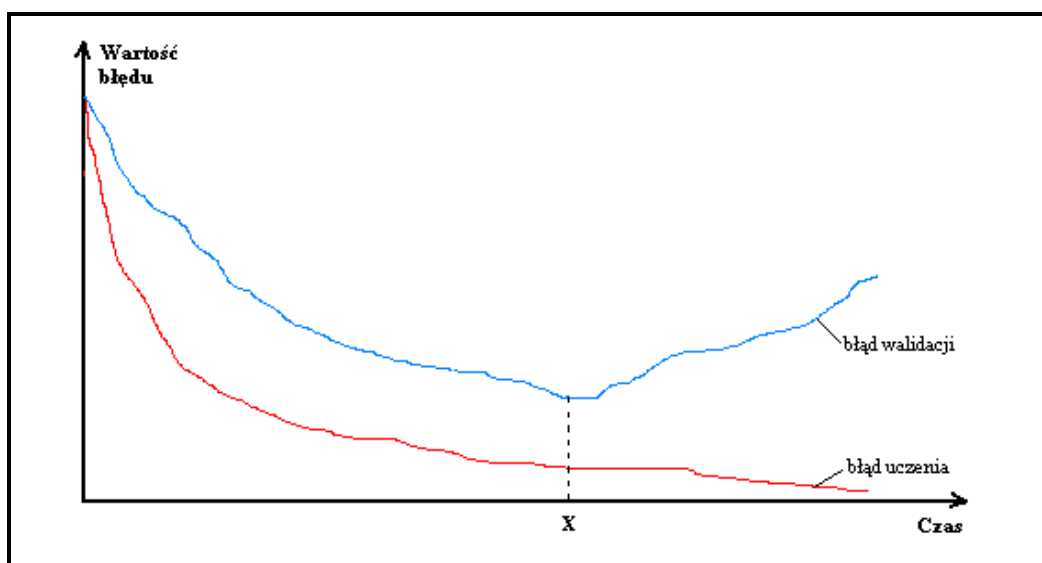
W teorii SN istnieje szereg technik pozwalających na unikanie efektu przeuczenia (np. ograniczenie wielkości warstw ukrytych, stosowanie niezależnej walidacji). Metody te są jednak efektywne w systemach stacjonarnych lub takich, gdzie charakter zmian systemu w czasie jest możliwy do przewidzenia. Rynki finansowe są jednak złożonymi systemami niestacjonarnymi o trudnym do prognozowania charakterze zmienności, dlatego też często okazuje się, iż poprawnie nauczone SN, wykazujące stosunkowo małe błędy dla wzorców użytych do uczenia i testowania sieci, w przypadku danych rzeczywistych - nie użytych w procesie konstrukcji i optymalizacji (uczenia) sieci - mogą funkcjonować niedostatecznie i z tego powodu okazać się nieprzydatne w praktyce. Zagadnienie konstrukcji takich modeli, które wykazywałyby również wysoką skuteczność dla danych spoza ciągu uczącego, jest w problematyce prognozowania rynków finansowych kwestią bardzo istotną, a jednocześnie niezwykle złożoną, wymagającą ciągłych badań związanych m.in. z odpowiednim doбором danych wejściowych dla SN, rzetelnym testowaniem sieci, a także przyjęciem właściwej metody oceny funkcjonowania sieci. Część z tych zagadnień podjęto w ramach realizacji badań, których wyniki referowane są w dalszej części pracy.

2.7.2. Zagadnienie niezależnej walidacji sieci.

Podstawową metodą stosowaną w celu zapobiegania wystąpienia efektu przeuczenia jest prowadzenie niezależnej walidacji sieci w trakcie procesu uczenia. Zasada walidacji oparta jest na wyodrębnieniu ze zbioru danych użytych do konstrukcji SN (do doboru współczynników wagowych w ramach procesu uczenia) podzbioru, zwanego ciągiem walidacyjnym, który nie będzie używany bezpośrednio do uczenia sieci. Ciągi uczący i walidacyjny są wzajemnie rozłączne.

Proces walidacji polega na dokonywaniu bieżącej oceny stopnia generalizacji osiąganego przez sieć (podczas jej uczenia), poprzez obliczanie (po każdej epoce uczenia) błędu walidacji, liczonego identycznie jak błąd uczenia (zob. podrozdział

2.5), ale w oparciu o ciąg walidacyjny. Jeżeli wraz ze spadkiem błędu uczenia spada również błąd walidacji, uczenie powinno być kontynuowane (oraz powinna być dokonywana odpowiednia korekta wag), natomiast wzrost błędu walidacji przy jednoczesnym spadku błędu uczenia świadczy o pojawiającym się niekorzystnym efekcie przeuczenia. Jako końcowe wartości wag, otrzymane w procesie uczenia, przyjmuje się te wielkości, dla których błąd walidacji osiągnął minimum⁴⁹ (rys. 2-7).



Rys. 2-7. Przykładowy przebieg zmian błędu uczenia i błędu walidacji w trakcie procesu uczenia sieci neuronowej. Za optymalny (końcowy) przyjmuje się zestaw wag uzyskany w punkcie X, później pojawia się efekt przeuczenia.

Źródło: opracowanie własne.

2.7.3. Dobór architektury sieci.

W większości zastosowań do analizy danych finansowych wykorzystywana jest sieć typu *perceptron* wielowarstwowy⁵⁰ z połączeniami jednokierunkowymi, z jednym elementem w warstwie wyjściowej. Niektórzy badacze podkreślają jednak w kontekście tych zastosowań zalety sieci posiadającej sprzężenia zwrotne, które mogą

⁴⁹ Niektórzy badacze w kryterium tym uwzględniają również błąd uczenia.

⁵⁰ Niekiedy używane są inne typy SN, np. sieci typu RBF lub tzw. maszyny Boltzmann. W niniejszym opracowaniu badania dotyczą jedynie perceptronu - najpowszechniej stosowanej sieci neuronowej.

zwiększać efektywność SN w niektórych przypadkach (zob. np. Bassi, 1996), na przykład przy eliminacji wpływu nie skompensowanej linii trendu.

Kolejnym zagadnieniem jest dobór ilości warstw ukrytych, oraz charakteru połączeń w sieci. W praktyce najczęściej w zastosowaniach finansowych wykorzystywane są SN z jedną warstwą ukrytą (niekiedy z dwoma), wykazano bowiem, że użycie większej liczby warstw ukrytych rzadko poprawia efektywność działania sieci, przyczynia się natomiast niewątpliwie do wzrostu czasu uczenia i kosztów przetwarzania informacji przez SN. Z kolei powszechnie stosowanym rodzajem połączeń w sieci jest architektura typu *każdy z każdym*, polegająca na połączeniu określonego neuronu z wszystkimi elementami warstwy poprzedzającej⁵¹.

Ważnym problemem pozostaje dobór optymalnej liczby neuronów w warstwie ukrytej (por. podrozdz. 2.4). Z jednej strony zbyt mały rozmiar warstwy ukrytej może przyczynić się do znacznego zmniejszenia efektywności procesu uczenia (mała sieć nie jest w stanie nauczyć się rozpoznawania odpowiednio dużej liczby wzorców), z drugiej strony nadmierna liczba elementów ukrytych, oprócz zwiększenia czasu obliczeń, może prowadzić do powstania efektu „uczenia na pamięć” elementów ciągu uczącego i utraty zdolności uogólniania (generalizacji). W zastosowaniach finansowych dąży się na ogół do określenia możliwie najmniejszej liczby neuronów w warstwie ukrytej - takiej, że jej zwiększanie nie powoduje widocznej poprawy oceny sieci według przyjętego kryterium. Dobór tej wielkości następuje często na drodze eksperymentalnej, ponieważ proponowane w literaturze formuły nie gwarantują optymalnych rezultatów.

2.7.4. Problem właściwego przygotowania danych wejściowych.

W typowych zadaniach prognozowania szeregów czasowych przy użyciu SN, jako informację wejściową wykorzystuje się poprzednie realizacje analizowanego szeregu czasowego, pochodzące z określonego okna (przedziału czasu), oraz ewentualnie

⁵¹ Czasami po wstępnym nauczaniu sieci stosowana jest redukcja połączeń (*weight pruning*), polegająca na eliminowaniu połączeń nieistotnych, to znaczy takich, których wpływ na kształtowanie sygnału wyjściowego sieci jest pomijalnie mały.

wartości innych szeregów czasowych, mogących mieć wpływ na wielkość prognozy. Okazuje się jednak, że w finansowych zastosowaniach SN często lepsze efekty można osiągnąć dokonując odpowiedniego wstępnego przekształcenia oryginalnych wartości określonego szeregu czasowego, zamiast użyć wyłącznie samych tych wartości jako danych wejściowych dla sieci. Przykładowo jeżeli analizujemy szereg czasowy kursów akcji:

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots) \quad \text{gdzie } \mathbf{k} - \text{numer kolejnej sesji giełdowej}$$

zamiast (oprócz) samych wartości \mathbf{x}_k można jako dane wejściowe dla SN wykorzystać:

- zmiany kursu: $\mathbf{x}_k - \mathbf{x}_{k-1}$
- względne zmiany kursu: $(\mathbf{x}_k - \mathbf{x}_{k-1}) / \mathbf{x}_{k-1}$
- względne zmiany w stosunku do kursu sprzed n sesji: $(\mathbf{x}_k - \mathbf{x}_{k-n}) / \mathbf{x}_{k-n}$
- logarytmny iloraz kursu bieżącego i kursu sprzed n sesji: $\ln (\mathbf{x}_k / \mathbf{x}_{k-n})$
- różnice wybranych średnich kroczących obliczonych dla szeregu \mathbf{X} (zob. niżej)
- wartości wybranych wskaźników analizy technicznej⁵²
- wartości innych wskaźników, pochodzących m.in. z analizy fundamentalnej.

2.7.5. Wstępna eliminacja szumów - wykorzystanie średnich kroczących.

Dane pochodzące z rynków finansowych są zakłócone znacznymi szumami o na ogół nieznanym rozkładach i parametrach. Zakłócenia te w dużym stopniu utrudniają wydobycie z tych danych informacji istotnej dla predykcji rynku. Pomimo faktu, iż sieci neuronowe posiadają wysoką zdolność eliminacji szumów (niekiedy są wykorzystywane właśnie głównie do tego celu), to jednak warto te zakłócenia eliminować lub ograniczać ich wpływ. Wykazano, że wstępna redukcja zakłóceń na etapie przygotowania danych uczących często prowadzi do znacznej poprawy efektywności procesu uczenia SN. W celu ograniczenia poziomu szumów w danych,

stosować można szereg narzędzi obliczeniowych funkcjonujących na zasadzie filtrów dolnoprzepustowych; popularną, prostą i skuteczną metodą jest wykorzystanie omówionych niżej średnich kroczących.

Rozważmy szereg czasowy $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k, \dots)$ (może to być np. rozpatrywany powyżej szereg kursów akcji). Dla tego szeregu przytoczymy definicje średnich kroczących: zwykłej, wykładniczej, adaptacyjnej i ważonej. W poniższych definicjach przyjmujemy parametry \mathbf{n} oraz \mathbf{k} , gdzie \mathbf{n} jest okresem średniej⁵³, zaś \mathbf{k} jest momentem czasu (np. numerem sesji) w którym średnia jest obliczana.

Zwykłą średnią kroczącą określamy zależnością:

$$S_{\mathbf{n},\mathbf{k}}(\mathbf{X}) = \frac{1}{\mathbf{n}} \sum_{i=\mathbf{k}-\mathbf{n}+1}^{\mathbf{k}} \mathbf{x}_i \quad (2.16)$$

Wykładniczą średnią kroczącą obliczamy na podstawie zależności rekurencyjnej:

$$E_{\mathbf{n},\mathbf{k}}(\mathbf{X}) = \frac{2}{\mathbf{n}+1} \mathbf{x}_k + \frac{\mathbf{n}-1}{\mathbf{n}+1} E_{\mathbf{n},\mathbf{k}-1}(\mathbf{X}) \quad , \quad E_{\mathbf{n},1}(\mathbf{X}) = \mathbf{x}_1 \quad (2.17)$$

Adaptacyjną średnią kroczącą określamy następująco:

$$A_{\mathbf{n},\mathbf{k}}(\mathbf{X}) = E_{\mathbf{n},\mathbf{k}}(\mathbf{R}) \quad (2.18)$$

gdzie:

$$\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_k, \dots)$$

$$\mathbf{r}_1 = \mathbf{x}_1 \quad , \quad \mathbf{r}_k = \mathbf{n} \cdot E_{\mathbf{n},\mathbf{k}}(\mathbf{X}) - (\mathbf{n}-1) \cdot E_{\mathbf{n},\mathbf{k}-1}(\mathbf{X}) \quad \text{dla } \mathbf{k} \geq 2$$

$E_{\mathbf{n},\mathbf{k}}$ - oznacza średnią wykładniczą zdefiniowaną zależnością (2.17).

⁵² Wskaźniki techniczne są dość powszechnie stosowane jako narzędzia wstępnego przetwarzania informacji dla SN (Azoff, 1994). Szeroki przegląd tych wskaźników można znaleźć np. w (Colby & Meyers, 1988).

⁵³ Jeżeli szereg czasowy dotyczy wartości dziennych (np. z kolejnych sesji giełdowych), taką średnią nazywamy \mathbf{n} -dniową.

Ważoną średnią kroczącą wyznaczamy ze wzoru:

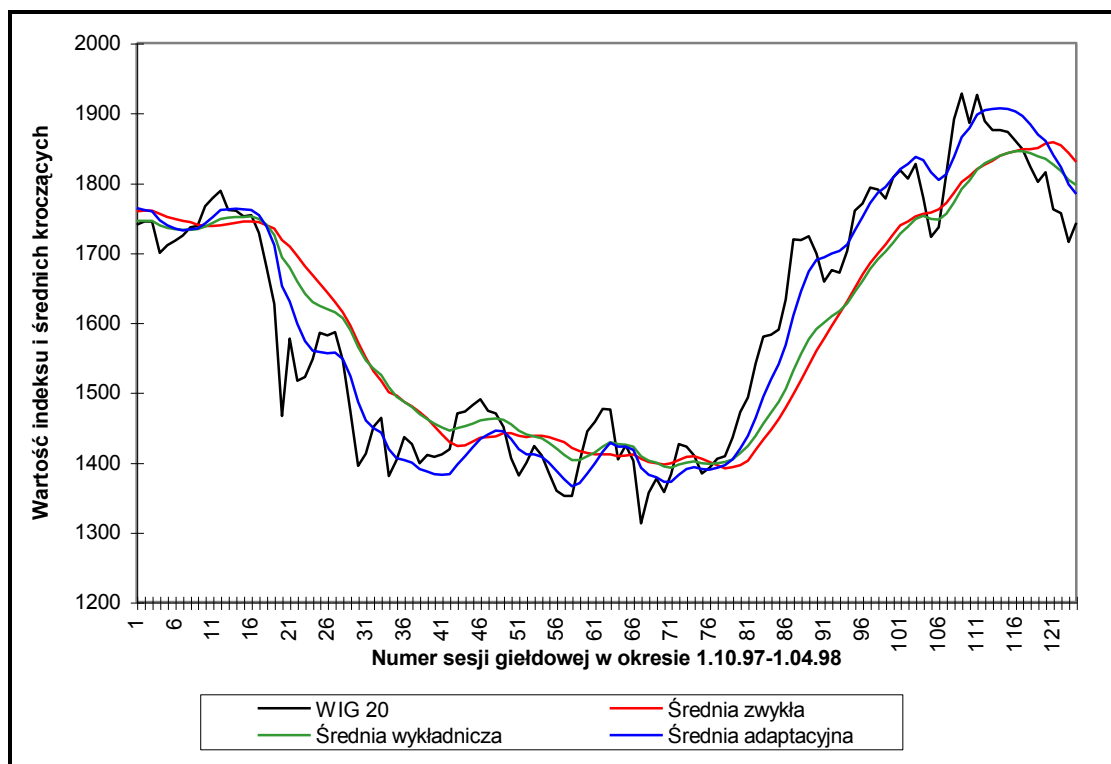
$$W_{n,k}(\mathbf{X}) = \frac{\sum_{i=1}^n w_i X_{k-i+1}}{\sum_{i=1}^n w_i} \quad (2.19)$$

gdzie dobierane do konkretnych danych współczynniki wagowe spełniają zależność: $w_1 > w_2 > \dots > w_n > 0$

Pewną wadą zwykłej średniej kroczącej jest występowanie opóźnienia przebiegu średniej w stosunku do zmian wartości szeregu czasowego. Analogiczna średnia wykładnicza charakteryzuje się mniejszym opóźnieniem, zaś średnia adaptacyjna cechuje się już stosunkowo dobrym dopasowaniem do wartości, z których jest obliczana (spełnia zadanie „wygładzania” szeregu czasowego, unikając przy tym dużych opóźnień przy znaczących zmianach jego wartości). Z kolei średnia ważona nadaje większą wagę ostatnim realizacjom szeregu traktując bardziej odległe w czasie wartości jako mniej znaczące.

Przykładowe przebiegi średnich kroczących: zwykłej, wykładniczej i adaptacyjnej dla indeksu WIG 20 przedstawiono na rys. 2-8.

Dobierając odpowiednio typ (typy) średnich kroczących oraz wielkości parametru n , można dokonać stosownej wstępnej transformacji danych wejściowych dla SN podając na wejścia sieci wartości tych średnich zamiast (oprócz) „surowych” wartości danego szeregu czasowego. Można także dokonywać omówionych w podrozdziale 2.7.4 przekształceń danych (np. obliczania względnych zmian kursu) zastępując w podanych formułach wartości samego kursu - wartościami odpowiednich średnich z kursu; można też obliczać różnice pomiędzy średnimi różnych typów lub o różnych okresach (tzn. porównywać ich aktualne wartości - stanowi to ważny element analizy technicznej), oraz wykonywać szereg innych operacji przekształcania danych z użyciem średnich kroczących.



Rys. 2-8. Przebiegi 15-dniowych średnich kroczących ($n=15$): zwykłej, wykładniczej i adaptacyjnej dla indeksu WIG 20 w 6-miesięcznym okresie 1.10.97-1.04.98. Widoczny jest fakt dobrego dopasowania (niewielkiego opóźnienia) średniej adaptacyjnej w stosunku do wartości szeregu zasadniczego (WIG 20), przy jednoczesnej realizacji wygładzania tego szeregu (eliminacji szumów). Źródło: opracowanie własne.

2.7.6. Zagadnienie optymalnego doboru (selekcji) danych wejściowych.

Analizując bogate możliwości tworzenia różnych typów danych wejściowych dla SN, zwiększone dodatkowo możliwością doboru różnych ich parametrów (np. okresów średnich kroczących), stwierdzić można, iż w następstwie tej obfitości potencjalnych rodzajów informacji powstaje problem identyfikacji danych istotnych oraz doboru optymalnego zestawu zmiennych wejściowych. Wykazano bowiem, że ograniczenie zbioru tych zmiennych, poza redukcją czasu obliczeń w procesie uczenia SN, często poprawia również działanie samej sieci (zob. Rutkowska et al., 1997). Fakt ten jest związany z poprawą efektywności procesu uczenia SN (zbędne lub nadmiarowe wejścia zwiększają poziom zakłóceń nie wprowadzając przy tym istotnej informacji).

W literaturze dotyczącej finansowych zastosowań SN autorzy często prezentują przyjęte zmienne wejściowe dla sieci, nie podając przy tym metodologii ich doboru, można zatem przypuszczać, że zmienne te zostały dobrane arbitralnie lub ustalone metodą prób i błędów. Istnieją także publikacje wskazujące na duże znaczenie i jednocześnie na wysoki stopień skomplikowania tego problemu, oraz analizujące różne metody stosowane przy doborze wejść dla SN (np. Thomason, 1996). Przegląd tych prac pozwala na stwierdzenie, że chociaż istnieją obecnie techniki pozwalające w pewnych przypadkach uzyskać obiecujące rozwiązania tego zagadnienia (np. analiza wrażliwości lub analiza wag połączeń w sieci), to nie istnieje wciąż jeszcze uniwersalna metodologia pozwalająca na optymalny dobór zmiennych wejściowych, zwłaszcza w kontekście niestacjonarności rynków finansowych jako systemów, dla których SN mają być użyte. Problem ten będzie w związku z tym również podjęty jako oddzielne, ważne zagadnienie w niniejszej pracy.

2.7.7. Redukcja ilości oraz dekorelacja zmiennych wejściowych.

Niezależnie od wysiłków zmierzających do rozwiązania problemu efektywnego doboru zmiennych wejściowych, celowa jest często w zastosowaniach sieci neuronowych redukcja wymiarowości przestrzeni wejść, a także transformacja tej przestrzeni w taki sposób, aby stopień korelacji zmiennych wynikowych (po transformacji) był możliwie niski. Takie operacje mogą przyczynić się do wzrostu efektywności procesu uczenia SN (Azoff, 1994). Problem dekorelacji ma szczególne znaczenie w przypadku analizy danych finansowych, gdzie poszczególne szeregi czasowe stosowane jako wejścia SN (np. wartości określonych wskaźników technicznych) mogą być silnie skorelowane.

W celu przeprowadzenia dekorelacji danych wejściowych można zastosować szereg technik analitycznych, zmierzających do ortogonalizacji przestrzeni wejść. Dość powszechnie stosowanym narzędziem, umożliwiającym zarówno ortogonalizację, jak i dalszą redukcję wymiarowości przestrzeni cech, jest analiza głównych składowych (*ang. Principal Component Analysis – PCA*), znana też w teorii przetwarzania sygnałów pod nazwą transformacji Karhunen-Loeve.

Analiza głównych składowych (Jajuga, 1990; Haykin, 1994) oparta jest na liniowym przekształceniu n -wymiarowej przestrzeni skorelowanych cech (zmiennych wejściowych SN) w k -wymiarową przestrzeń cech nieskorelowanych (głównych składowych) ($k \leq n$) według zależności:

$$\mathbf{Y} = \mathbf{V} \cdot \mathbf{X} \quad , \quad \text{gdzie } \mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]^T \text{ jest macierzą } k \times n \quad (2.20)$$

przy czym: \mathbf{X} - wektor zmiennych wejściowych przed transformacją

\mathbf{Y} - wektor zmiennych wejściowych po transformacji (głównych składowych)

$\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ - wektory własne macierzy kowariancji wyznaczonej dla całego zbioru wektorów \mathbf{X} , odpowiadające kolejnym wartościom własnym $\lambda_1, \lambda_2, \dots, \lambda_k$ tej macierzy.

Ponieważ wartości własne $\lambda_1, \lambda_2, \dots, \lambda_k$ stanowią miary wariancji zmiennych po transformacji (głównych składowych), w wyniku przekształcenia PCA otrzymujemy zestaw zdekorelowanych zmiennych, uporządkowanych według ich malejącej wariancji. Traktując wariancję zmiennej jako wskaźnik jej informatywności (a więc znaczenia danej zmiennej⁵⁴), oraz przyjmując $k < n$, możemy dokonać redukcji wymiarowości przestrzeni wejść poprzez odrzucenie pewnej ilości ostatnich głównych składowych (tych o najmniejszej wariancji). W wyniku takiej operacji realizujemy jednoczesną dekorelację zmiennych wejściowych dla SN oraz zmniejszenie ich ilości, stąd też PCA jest powszechnie używanym narzędziem wstępnego przetwarzania danych, szczególnie w zastosowaniach SN do analizy i predykcji finansowych szeregów czasowych (Azoff, 1994).

Należy podkreślić, że oprócz klasycznych metod statystycznych (do których zaliczyć można m.in. analizę głównych składowych) istnieją też rozwiązania problemu wstępnego przetwarzania danych i redukcji ich ilości, oparte na specyficznych konstrukcjach sieci neuronowych. Do nich należy rodzaj wielowarstwowego perceptronu, tzw. autoasociator, który oprócz redukcji wymiaru przestrzeni wejść dla

⁵⁴ Przyjmuje się często iż zmienne o małej wariancji nie wnoszą istotnej informacji.

właściwej SN może spełniać dodatkowo zadanie filtracji szumów (Azoff, 1994). Również omawiana tu transformacja PCA może być realizowana przez pewien typ samouczących się sieci neuronowych (Haykin, 1994; Osowski, 1996; Grabowski, 1997); wykorzystanie tego typu sieci w procesie przetwarzania danych wejściowych zostanie omówione w dalszej części pracy.

2.7.8. Normalizacja zmiennych wejściowych.

Bardzo ważnym etapem wstępnego przetwarzania danych, stosowanym praktycznie we wszystkich aplikacjach SN, jest normalizacja zmiennych wejściowych, polegająca na takim przeskalowaniu wartości każdej zmiennej, aby sprowadzić te wartości do określonego zakresu (przedziału). Głównym celem tej operacji jest nadanie każdej zmiennej wejściowej jednakowego znaczenia (wagi) w stosunku do innych zmiennych. Gdyby bowiem przykładowo zakresem zmiennej wejściowej A był przedział $[-1000, 1000]$, natomiast wartości zmiennej B należałyby do przedziału $[-0,1, 0,1]$, to sieć neuronowa reagowałaby w nieporównywalnie większym stopniu na zmiany wartości A, ignorując przy tym praktycznie wartości zmiennej B. Przyjmując aprioryczne założenie, że obie zmienne A i B posiadają jednakowy wpływ na wartość wyjściową obliczaną przez SN, należy liniowo przeskalować wartości tych zmiennych do tego samego zakresu (np. $[-1, 1]$). Operacja normalizacji zmiennej do przedziału $[-1, 1]$ wyrażona jest przekształceniem:

$$\hat{x} = 2 \frac{x - \min}{\max - \min} - 1 \quad (2.21)$$

gdzie: x - pierwotna wartość zmiennej

\hat{x} - wartość znormalizowana

\min - minimum wszystkich wartości zmiennej przed normalizacją

\max - maksimum wszystkich wartości zmiennej przed normalizacją.

Do innych często stosowanych metod normalizacji zaliczyć można także standaryzację, (polegającą na odjęciu wartości średniej obliczonej dla danej zmiennej i następnie na wykonaniu dzielenia przez odchylenie standardowe tej zmiennej), oraz

tzw. normalizację „bez przesunięcia zera” (ang. *zero offset* - zachowuje ona znak wartości danej zmiennej wejściowej). Metody te opisano w rozdziale 4.

Należy zwrócić uwagę, że niektóre techniki normalizacji (np. operacja określona zależnością (2.21)) są wrażliwe na występowanie danych nietypowych (*outliers*) w normalizowanej serii wartości. Dane nietypowe (znacznie odstające od średniej) decydują o wielkościach minimum i maksimum zmiennej wejściowej i powodują, iż po znormalizowaniu pewne istotne dane osiągną wielkości zbyt bliskie zera, przez co staną się mało znaczące jako sygnały dla SN. Jeżeli na skutek normalizacji dane będą należeć do wąskiego przedziału wartości, to będą zbyt słabo różnicowane przez SN. Celowe jest zatem w takich sytuacjach dokonanie przed normalizacją sprowadzenia wartości nietypowych (odstających) do wielkości bliższych średniej, a więc wykonanie operacji zawężenia pierwotnego zakresu zmiennej poprzez tzw. obcięcie wartości nietypowych (*outliers trimming*).

Normalizacja poszczególnych zmiennych wejściowych według powyższych zasad nie jest jedynym podejściem wykorzystywanym przy analizie i predykcji finansowych szeregów czasowych przez SN. Można także stosować normalizację poszczególnych wzorców (wektorów wejściowych) osobno, normalizować pierwotne (nie przekształcone) szeregi czasowe (np. ceny akcji) sprowadzając ich wartości do określonego przedziału, lub realizować określone kombinacje tych metod (Azoff, 1994).

2.7.9. Konstrukcja zbiorów: uczącego, walidacyjnego i testującego.

Przed realizacją procesu uczenia i testowania SN konieczne jest wyznaczenie:

- okresu czasu, z którego dane zostaną wykorzystane w tym procesie,
- podziału dostępnych danych (wektorów wejściowych) na zbiory: uczący, walidacyjny i ewentualnie dodatkowy zbiór testujący.

Rozwiązanie pierwszego z powyższych problemów jest kwestią kompromisu pomiędzy wykorzystaniem możliwie aktualnych danych z ostatniego okresu czasu, a wymaganą ilością wzorców niezbędnych do właściwego nauczania i oceny SN. Z jednej strony bowiem dane ze zbyt odległych w czasie okresów mogą być

nieadekwatne w stosunku do aktualnej, często innej niż wówczas sytuacji (np. ekonomicznej) wpływającej na funkcjonowanie rynku finansowego, mogą więc one wpływać na pogorszenie funkcjonowania sieci. Z drugiej jednak strony, im większa jest ilość elementów użytych w ciągu uczącym, tym lepsza zdolność nauczonej sieci do generalizacji, efektywnego rozpoznawania wzorców wejściowych i generowania prawidłowych odpowiedzi na te wzorce. Niewątpliwie rozstrzygnięcie powyższej kwestii zależy w znacznej mierze od konkretnego analizowanego problemu i często jest podejmowane arbitralnie. Azoff podaje, że w przypadku analizy finansowych szeregów czasowych wielkość ciągu uczącego powinna być co najmniej 10-krotnie większa od liczby parametrów (wag połączeń) w sieci (Azoff, 1994), chociaż istnieją możliwości efektywnego uczenia sieci również w przypadku znacznie ograniczonej liczby wzorców uczących, co często zdarza się w praktyce.

W celu dokonania dla nauczonej już sieci neuronowej rzetelnej oceny końcowej, której wartość stanowi ważny element analizy efektywności SN, często celowe jest dodatkowe testowanie sieci w oparciu o dane nie użyte w procesie jej uczenia i walidacji (doboru wag). Powstaje w takiej sytuacji konieczność wyodrębnienia (spośród dostępnego zbioru danych) - oprócz ciągu uczącego i walidacyjnego - także niezależnego ciągu testującego. Generalnie powstaje zatem dylemat podziału całego zbioru danych na trzy rozłączne podzbiory: uczący, walidacyjny i testujący. Właściwe rozwiązanie tego zagadnienia nie jest proste i może mieć znaczący wpływ na późniejszą efektywność modelu opartego na SN. Problemy wymagające wówczas rozstrzygnięcia polegają na:

- określeniu liczebności (lub wzajemnych proporcji) podzbiorów uczącego, walidacyjnego i testującego,
- zdefiniowaniu lokalizacji każdego z tych podzbiorów w całym zbiorze⁵⁵.

⁵⁵ Jednym z głównych dylematów jest lokalizacja podzbioru testującego - duża grupa badaczy uważa, że należy go umieścić na samym końcu (w sensie czasowym) zbioru wszystkich dostępnych danych tak, aby zawierał on najbardziej aktualne dane, co powinno zapewnić rzetelną ocenę użyteczności modelu i stworzyć ewentualne przesłanki do wykorzystania sieci w przyszłej praktyce. Rozwiązanie takie ma jednak wady - dane z ostatniego okresu czasowego nie zostaną użyte do uczenia sieci, co pogorszy (niejako z definicji) jakość jej dostosowania do zadań, które będzie spełniała w trakcie normalnej (regularnej) eksploatacji.

Niestandardowym sposobem rozwiązania tych problemów może być dokonanie podziału całego zbioru dostępnych danych na kilka lub kilkanaście bloków, po czym wyodrębnienie w każdym z tych bloków danych uczących, walidacyjnych i testujących – według tych samych zasad alokacji i z zachowaniem przyjętych proporcji.

2.7.10. Dobór i normalizacja zmiennej wyjściowej.

Pomimo faktu, iż w ogólnym przypadku sieć typu *perceptron* może posiadać wiele elementów wyjściowych, w zastosowaniach SN do predykcji rynków finansowych stosuje się na ogół sieci z jednym neuronem w warstwie wyjściowej, gdyż wykazują one największą efektywność⁵⁶ (Azoff, 1994). Bardzo istotnym i nietrywialnym zagadnieniem (zwłaszcza w przypadku prognoz finansowych) jest odpowiedni dobór zmiennej wyjściowej. Właściwego określenia wymaga tu nie tylko charakter tej zmiennej, ale również trafne zdefiniowanie horyzontu predykcji h .

Najczęściej spotykane typy zmiennych wyjściowych to:

- wartość analizowanego szeregu czasowego (np. cena akcji) po upływie okresu h ,
- względna zmiana wartości prognozowanej w okresie h ,
- trend szeregu czasowego (współczynnik kierunkowy prostej regresji) w okresie h ,
- sygnały bezpośrednio rekomendujące kupno lub sprzedaż danego instrumentu finansowego.

Normalizacja zmiennej wyjściowej (uczącej, stanowiącej wzorzec) jest konieczna i dokonywana jest na ogół według zależności (2.21) (w niektórych przypadkach wskazana jest normalizacja „bez przesunięcia zera”). Zakres wartości znormalizowanych przyjętych dla wyjścia SN zależy od przedziału stanowiącego zbiór wartości funkcji aktywacji neuronu wyjściowego – powinien on się mieścić wewnątrz tego przedziału. Podczas operacji normalizacji należy także

⁵⁶ Prognozując np. wartości dwóch różnych finansowych szeregów czasowych lepsze efekty uzyskamy wykorzystując dwie oddzielne sieci z jednym wyjściem każda, niż stosując jedną sieć z dwoma elementami wyjściowymi.

zwrócić uwagę na możliwość występowania wartości nietypowych w serii danych wyjściowych (uczących) i ewentualnie zastosować obcinanie (*trimming*) tych wartości.

2.7.11. Ocena funkcjonowania sieci neuronowej.

Ocena sieci neuronowej według określonego kryterium, związanego z celem zastosowania danej sieci, odbywa się na ogół w dwóch etapach:

- ocena wstępna, która dokonywana jest na bieżąco w trakcie uczenia SN i realizowana jest w oparciu o zbiór walidacyjny (niektórzy badacze uwzględniają tu również zbiór uczący),
- ocena testowa, która odbywa się w oparciu o dane testujące nie używane w procesie uczenia sieci (o ile takie dane wyodrębniono).

W celu dokonania wstępnej oceny SN należy wyznaczyć błąd walidacji SN, błąd ten określamy jako wartość funkcji błędu F , obliczaną według identycznej formuły jak błąd uczenia (zob. podrozdział 2.5, zależności (2.9) - (2.12)), ale w oparciu o dane z ciągu walidacyjnego (a nie uczącego). Najczęściej przyjmuje się, że liczbową miarą oceny sieci jest właśnie wielkość błędu walidacji (obliczana w trakcie każdej epoki uczenia), pamiętając, że im ta miara jest mniejsza, tym ocena sieci jest lepsza (niekiedy jako dodatkowy składnik tej miary uwzględnia się również błąd uczenia).

Podczas procesu uczenia korekta najlepszego (według przyjętego kryterium oceny sieci), dotychczas zapamiętanego zbioru wag następuje zatem wtedy i tylko wtedy, gdy bieżąca wartość oceny sieci osiąga swoje (kolejne) minimum. Wielkość ta po zakończeniu uczenia stanowi także wstępną ocenę nauczonej sieci i w wielu przypadkach może być przyjęta jako ocena **końcowa** (ostateczna). Należy jednak pamiętać, że może ona być nieadekwatna w stosunku do rzeczywistej efektywności sieci, gdyż:

- zbiór walidacyjny, w oparciu o który został wyznaczony błąd walidacji, ma znaczący wpływ na dobór końcowych wartości wag, ponadto zbiór ten może być

niekiedy (z konieczności) zbyt mały, aby dokonana w oparciu o niego ocenę uznać za rzetelną,

- ewentualny składnik wyrażający jakość rozpoznawania wzorców z ciągu uczącego, nie wnosi żadnej informacji odnoszącej się do zdolności sieci do generalizacji.

Pomimo tych potencjalnych zagrożeń w praktyce dość powszechnie wstępną ocenę sieci, uzyskaną po zakończeniu procesu uczenia, uznaje się za ocenę końcową, stanowiącą istotną (a często zasadniczą) miarę efektywności nauczonej sieci.

Jeżeli spośród dostępnych danych wyodrębniono zbiór testujący, dokonywana jest - właśnie w oparciu o ten zbiór - testowa ocena sieci, która w takim przypadku przyjmowana jest jednocześnie jako ocena końcowa (ostateczna). W przypadku wykorzystania SN na rynkach finansowych możliwe są dwa alternatywne podejścia przy wyznaczaniu takiej oceny:

- Obliczenie błędu sieci dla danych testujących w sposób analogiczny jak dla błędu uczącego lub walidacyjnego, tzn. przy użyciu funkcji błędu F , ale w oparciu o zbiór testujący; wielkość tego błędu stanowi liczbową miarę oceny sieci⁵⁷.
- Sprawdzenie rzeczywistego funkcjonowania sieci w przedziale czasu obejmującym dane testujące i ocena jej efektywności poprzez stopień realizacji ostatecznego celu, dla którego sieć została skonstruowana. Jeżeli przykładowo zadaniem sieci jest generowanie sygnałów kupna lub sprzedaży określonego instrumentu finansowego, to ocena końcowa może być dokonywana w oparciu o ocenę efektywności aktywnej strategii wykorzystującej sygnały sieci w okresie testowym⁵⁸. Podstawą oceny efektywności strategii może być wygenerowany przez nią całkowity zysk, albo też inna miara uwzględniająca nie tylko sam zysk (np. wskaźnik Sharpe'a).

⁵⁷ Im mniejsza wartość oceny sieci (błędu), tym ta ocena jest lepsza.

⁵⁸ Tego typu metoda końcowej oceny sieci może być dokonywana również w przypadku, gdy nie wyodrębniono osobnego ciągu testowego - wówczas oceniana (według przyjętego kryterium) strategia inwestycyjna wykorzystująca sygnały sieci realizowana jest dla okresów obejmujących dane walidacyjne. Takie właśnie rozwiązanie zostało zaproponowane i zrealizowane w rozdziale 6 niniejszej pracy.

Należy zwrócić także uwagę, że nawet w przypadku uzyskania zadowalających wyników i podjęcia praktycznego wykorzystania SN do wspomagania decyzji inwestycyjnych, konieczna jest stała bieżąca kontrola efektywności funkcjonowania sieci, gdyż sieć może w trakcie eksploatacji pogarszać swoje działanie np. na skutek dezaktualizacji wiedzy, w oparciu o którą była uczona.

3. Algorytmy genetyczne i możliwości ich zastosowań na rynkach finansowych.

3.1. Idea algorytmu genetycznego.

Idea algorytmu genetycznego (AG) została zaczerpnięta z nauk przyrodniczych opisujących zjawiska doboru naturalnego i dziedziczności. Mechanizmy te (w największym uproszczeniu) polegają na przetrwaniu osobników najlepiej dostosowanych w danym środowisku, podczas gdy osobniki gorzej przystosowane są eliminowane z populacji. Z kolei te osobniki, które przetrwają, przekazują informację genetyczną swoim potomkom, przy czym liczba potomków może być związana ze stopniem dostosowania „rodziców”. Krzyżowanie informacji genetycznej otrzymanej od „rodziców” prowadzi do sytuacji, w której kolejne pokolenia są przeciętnie coraz lepiej dostosowane do warunków środowiska, a jednocześnie proces przypadkowych zmian (mutacji) jest generatorem różnorodności osobników. Mamy więc tu w istocie do czynienia ze swoistym procesem optymalizacji.

Tego typu procesy funkcjonujące w naturze stanowią inspirację do wykorzystania podobnych mechanizmów adaptacyjnych w innych systemach o bardzo dużym stopniu złożoności, m.in. na rynkach finansowych. Algorytmy genetyczne są wówczas w istocie numerycznymi metodami optymalizacji, rozwiązującymi problem właściwego wyznaczenia dużej liczby parametrów systemu (np. decyzyjnego) przy użyciu specyficznych operatorów genetycznych, które generują kolejne punkty w przestrzeni rozwiązań wykorzystując zasady analogiczne jak występujące właśnie w zjawiskach ewolucji i dziedziczności. Obserwowany obecnie rozwój opartych na AG metod rozwiązywania problemów zawdzięczać możemy w dużej mierze szybkiemu

wzrostowi mocy obliczeniowej współczesnych komputerów oraz rozpowszechnianiu się komputerowych architektur równoległych.

Dla problemu rozwiązywanego przy wykorzystaniu AG należy zdefiniować:

- przestrzeń możliwych rozwiązań i sposób ich genetycznego kodowania,
- funkcję celu (kryterium optymalizacji), której maksimum będziemy poszukiwać, a także funkcję przystosowania (będącą z reguły pewną transformacją funkcji celu), według której będziemy oceniać poszczególne rozwiązania; dziedziną tych funkcji musi być cała przestrzeń rozwiązań.

Ponadto należy określić:

- operatory genetyczne (generujące kolejne rozwiązania) i ich parametry,
- parametry samego algorytmu (ilość iteracji, kryterium zatrzymania, itp.).

3.1.1. Dobór genetycznej reprezentacji potencjalnych rozwiązań oraz funkcji celu i przystosowania.

Każde możliwe rozwiązanie zagadnienia z przestrzeni rozwiązań powinno posiadać swój unikalny kod genetyczny, zwany przez odpowiednią analogię chromosomem⁵⁹. Chromosom złożony jest ze skończonej liczby elementów zwanych genami. Geny mogą przybierać wartości z określonego alfabetu, zawierającego skończoną liczbę znaków. Najczęściej stosuje się alfabet binarny (złożony z dwóch znaków) za pomocą którego można łatwo zapisywać zmienne typu logicznego, a także kodować (w przybliżeniu) parametry typu rzeczywistego; niekiedy jednak celowe może okazać się zastosowanie szerszego alfabetu, obejmującego np. pewien podzbiór liczb całkowitych. Należy podkreślić, że stworzenie właściwej reprezentacji genetycznej potencjalnych rozwiązań jest zagadnieniem o kluczowym znaczeniu przy wszystkich zastosowaniach algorytmów genetycznych i w zasadniczy sposób wpływa na efektywność metody. Przykładowy kod genetyczny, złożony z 20 genów binarnych, prezentuje rys. 3-1.

⁵⁹ Pojęcia „chromosom” i „osobnik” traktowane są tu jako synonimy i używane są dla określenia pojedynczego rozwiązania - elementu danej populacji.

W każdym zagadnieniu optymalizacji (także z wykorzystaniem AG) musimy dysponować **funkcją celu** której ekstremum (przyjmijmy tu, że maksimum) poszukujemy. Postać funkcji celu zależy oczywiście od konkretnego zagadnienia, które rozważamy. Dodatkowo wymagane jest zdefiniowanie funkcji przystosowania (poprzez odpowiednie przekształcenie funkcji celu⁶⁰), której wartości stanowią będą liczbowe oceny stopnia przystosowania poszczególnych chromosomów (wzrost wartości funkcji celu powinien powodować jednoczesny wzrost wartości funkcji przystosowania). Musi przy tym istnieć możliwość określenia wartości tych funkcji dla każdego potencjalnego rozwiązania, tzn. dla każdej możliwej kombinacji genów w chromosomie. Celem algorytmu jest znalezienie takiego rozwiązania (lub rozwiązań) problemu, dla których funkcja celu (przystosowania) osiąga maksimum.

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

Rys. 3-1. Przykładowy kod genetyczny pojedynczego chromosomu. Źródło: opracowanie własne.

3.1.2. Operatory genetyczne.

Algorytmy genetyczne przewidują trzy podstawowe operacje, wykonywane na osobnikach (chromosomach) z populacji⁶¹. Są to selekcja, krzyżowanie i mutacja.

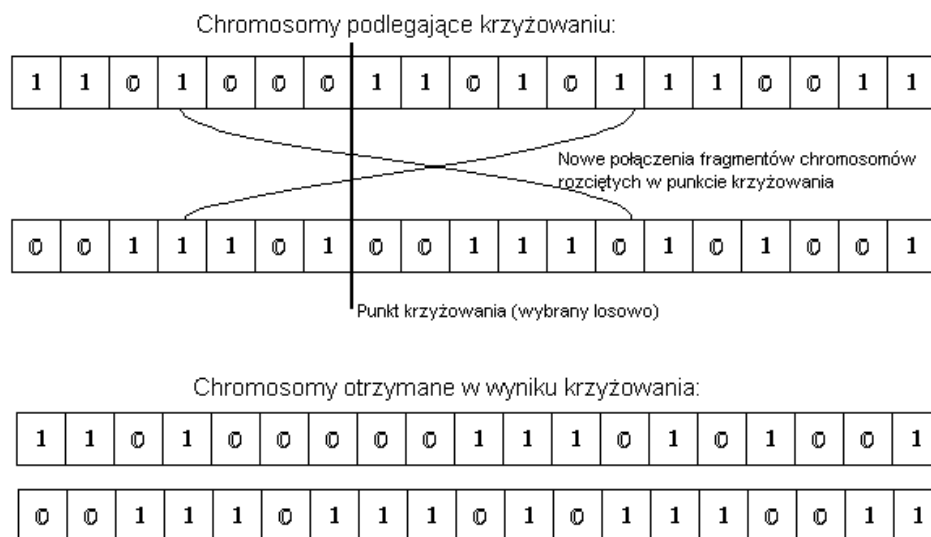
Selekcja polega na wyborze z populacji najlepiej przystosowanych osobników, których materiał genetyczny zostanie poddany operacji krzyżowania i przekazany osobnikom następnego pokolenia⁶². Kryterium wyboru jest tu wartość funkcji przystosowania, przy czym często stosuje się wybór określonej liczby chromosomów w sposób losowy, ale z prawdopodobieństwem proporcjonalnym do wartości tej funkcji (daje to szanse wyboru również dla najsłabszych osobników, ale najlepiej przystosowani mają największe prawdopodobieństwo wyboru).

⁶⁰ W szczególności funkcje celu i przystosowania mogą być tożsamościowo równe, przekształcenie stosuje się jednak w celu poprawy efektywności operatorów genetycznych.

⁶¹ Populacją nazywamy zbiór potencjalnych rozwiązań analizowanych w jednej iteracji algorytmu.

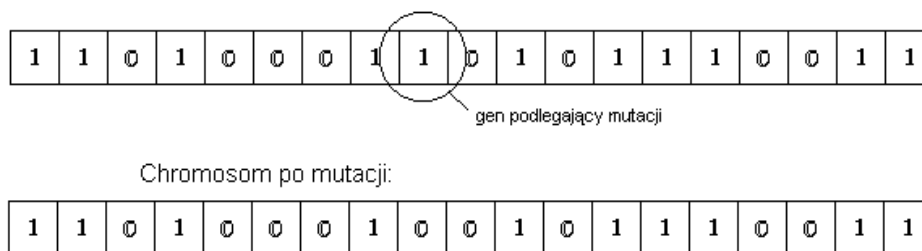
⁶² Pokoleniami będziemy nazywali kolejne populacje, występujące w kolejnych iteracjach algorytmu.

Krzyżowanie polega na wymianie materiału genetycznego pomiędzy losowo dobranymi parami osobników wybranych podczas selekcji. W wyniku krzyżowania powstają nowe chromosomy, które wejdą w skład następnego pokolenia. Okazuje się, że chromosomy powstałe w wyniku krzyżowania często są lepiej przystosowane (mają większą wartość funkcji celu) niż ich „rodzice”. Ponadto oprócz osobników powstałych w wyniku tej operacji, w skład nowego pokolenia może wejść pewna liczba najlepiej dostosowanych chromosomów z pokolenia poprzedniego. Istnieje wiele możliwych odmian operatorów krzyżowania, np. można wybrać więcej niż jeden punkt krzyżowania, przeprowadzać wymianę genów na losowo wybranych pozycjach, itp. Standartową operację krzyżowania obrazuje rys. 3-2.



Rys. 3-2. Schemat operacji krzyżowania. Źródło: opracowanie własne.

Mutacja polega na zmianie wartości losowo wybranego genu. Mutacji może podlegać każdy gen każdego chromosomu z określonym (zwykle niewielkim) prawdopodobieństwem. Niekiedy celowe jest zastrzeżenie, aby geny jednego lub kilku najlepiej dopasowanych chromosomów nie podlegały mutacji. Zadaniem operatora mutacji jest zapewnienie zmienności chromosomów (np. niedopuszczenie do powstania całej populacji identycznych osobników) i tym samym stworzenie możliwości wyjścia procedury optymalizacji z maksimów lokalnych funkcji celu. Działanie operatora mutacji prezentuje rys. 3-3.



Rys. 3-3. Schemat operacji mutacji. Źródło: opracowanie własne.

W różnych aplikacjach AG, opisanych w literaturze, stosowanych bywa wiele odmian i modyfikacji wyżej przedstawionych operatorów; niekiedy celowe okazuje się użycie także innych typów operacji genetycznych, np. inwersji⁶³. Oprócz sprecyzowania operatorów genetycznych w algorytmie należy także określić szereg jego parametrów, np. wielkość (liczebność) populacji, sposób generowania początkowej populacji rozwiązań, warunek zakończenia procesu (np. liczba iteracji), itd.

3.1.3. Schemat oraz wybrane aspekty stosowania algorytmu genetycznego.

Ogólny schemat blokowy podstawowego algorytmu genetycznego prezentuje rys. 3-4. Najlepsze rozwiązanie (rozwiązania) otrzymane w ostatniej populacji (po spełnieniu kryterium stopu) traktujemy jako rozwiązania wynikowe.

Algotymy genetyczne stanowią mocne narzędzie służące do rozwiązania zagadnienia optymalizacji, do którego można sprowadzić również pewną ilość problemów ekonomicznych. Do głównych **zalet** AG można zaliczyć:

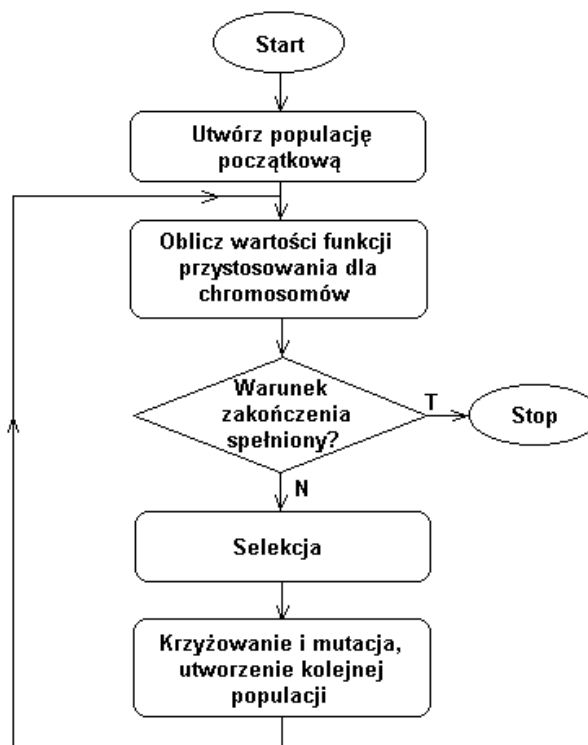
- jednoczesne poszukiwanie rozwiązań w wielu punktach przestrzeni rozwiązań,
- zdolność do łatwego pokonywania ekstremów lokalnych, co może prowadzić do stosunkowo szybkiego znalezienia ekstremum globalnego funkcji celu,
- możliwość szybkiego i efektywnego rozwiązywania nawet bardzo złożonych problemów,

⁶³ Inwersja polega na odwróceniu kolejności genów we fragmencie chromosomu.

- łatwość implementacji na komputerowych architekturach równoległych,
- odporność na szумы,
- prostota łączenia algorytmu z istniejącymi modelami i tym samym łatwość tworzenia systemów hybrydowych,
- brak wymogu posiadania szczegółowej wiedzy o rozwiązywanym problemie (konieczna jest jedynie umiejętność identyfikacji potencjalnych rozwiązań i ich oceny),
- brak matematycznych ograniczeń dotyczących funkcji celu.

Spośród **wad** algorytmów genetycznych wymienić można:

- zależność rozwiązań końcowych od punktów startowych i innych czynników losowych,
- ryzyko nadmiernego dopasowania rozwiązania do danych „uczących” użytych w procesie poszukiwania najlepszych osobników; tym samym funkcjonowanie znalezionej rozwiązania dla innych danych może być niezadowalające,
- wymóg, aby zakodowane dane, oraz funkcja celu miały wartości numeryczne (w niektórych problemach mogą wystąpić trudności ze stworzeniem kodu genetycznego i z doбором funkcji przystosowania).



Rys. 3-4. Schemat podstawowego algorytmu genetycznego. Źródło: opracowanie własne.

Algorytmy genetyczne nadają się szczególnie dobrze do skomplikowanych zagadnień - tam gdzie tradycyjne metody optymalizacji zawodzą. Ich istotną cechą jest fakt, że zachowują całą populację potencjalnych rozwiązań, podczas gdy inne metody analizują w danej chwili tylko jeden punkt przestrzeni rozwiązań. Pozwala to z jednej strony na unikanie ekstremów lokalnych funkcji celu i przyspieszenie poszukiwania suboptymalnych rozwiązań, z drugiej strony zaś na łatwość realizacji algorytmu na komputerowych architekturach równoległych, co daje im dodatkową przewagę nad tradycyjnymi algorytmami.

Problematyka algorytmów genetycznych i ich zastosowań jest wyczerpująco omówiona w bogatej literaturze dotyczącej tej tematyki; należą do niej m.in. następujące książki: (Goldberg, 1995), (Michalewicz, 1996), (Cytowski, 1996), (Davis, 1991), (Holland, 1975), (Whitley, 1992), (Bauer, 1994), (Koza, 1992), (Rutkowska et al., 1997), (De Jong, 1993), (Bäck et al., 1996) i inne. Szereg

informacji, wraz z nieodpłatnie oferowanym do celów badawczych oprogramowaniem, jest również dostępnych w Internecie⁶⁴.

3.2. Możliwości i przykłady bezpośredniego wykorzystania algorytmów genetycznych w procesie optymalizacji strategii inwestycyjnych.

3.2.1. Uwagi wstępne.

Kierując się licznymi przesłankami wskazującymi na celowość zastosowania AG do analizy rynków finansowych (omówiono je tutaj w podrozdziałach 1.3 oraz 3.1), szereg badaczy i praktyków podjęło próby wykorzystania AG w procesie konstrukcji strategii inwestycyjnej dla operacji na poszczególnych instrumentach finansowych. Metodologia takich badań polega na zdefiniowaniu ogólnego schematu strategii generującej sygnały kupna i sprzedaży dla poszczególnych instrumentów, a następnie na realizacji procesu doboru zestawu parametrów lub reguł decyzyjnych dla tej strategii przy wykorzystaniu określonego zbioru danych historycznych pochodzących z rynku, czyli tak zwanych danych uczących. Zadaniem algorytmu genetycznego jest znalezienie optymalnego (lub suboptymalnego) zbioru parametrów i/lub reguł, dla których strategia przynosi najlepsze rezultaty według przyjętego kryterium (tym kryterium może być np. całkowity zysk w okresie, z którego pochodzą dane uczące, wskaźnik Sharpe'a, lub inna miara).

Jak już wspomniano wyżej, w powyższym procesie występuje jednak (podobnie jak w przypadku uczenia sieci neuronowych) ryzyko nadmiernego dopasowania parametrów strategii do danych uczących kosztem utraty zdolności prawidłowego funkcjonowania dla innych danych, nie wchodzących w skład ciągu uczącego. Jest to

⁶⁴ Zob. np.: <http://www.trajan-software.demon.co.uk/sugal.htm>
<http://GARAGe.cps.msu.edu/>
<http://GAL4.ge.uiuc.edu/illigal.home.html>

bardzo poważny problem, ponieważ pojawienie się tej wady eliminuje możliwości praktycznego wykorzystania opisanej strategii. Celowe jest zatem (analogicznie jak dla SN) zastosowanie niezależnej walidacji w trakcie realizacji AG, oraz ewentualnie dodatkowego testowania końcowego otrzymanego systemu decyzyjnego na danych nie wchodzących w skład próbki uczącej.

3.2.2. Przykłady zastosowania algorytmów genetycznych na rynkach finansowych.

Dość szerokie spektrum badań nad optymalizacją strategii inwestycyjnych przy użyciu AG prezentuje Bauer w książce (Bauer R., 1994). W opracowaniu tym omówiono główne przesłanki zastosowania AG na rynkach finansowych oraz przedstawiono pewne koncepcje oraz rezultaty zastosowania tej techniki do konstrukcji reguł decyzyjnych, wchodzących w skład strategii opartych o prognozy zachowania się rynku. We wstępnym etapie badań wybrano 10 (ze 167) szeregów czasowych dotyczących określonych danych makroekonomicznych. Następnie stosowano AG dla wszystkich 120 trzelementowych kombinacji wybranych zmiennych. Zadaniem AG było sprecyzowanie warunku logicznego (dobór operatorów porównania, wartości do porównania dla tych zmiennych, oraz spójników logicznych) generującego decyzję inwestycyjną (np. sprzedaż akcji i kupno bonów skarbowych). Zaprezentowane wyniki badań pozwalają uznać AG za obiecujące narzędzia dla konstrukcji tego typu strategii inwestycyjnych.

Z powodzeniem zastosowano także AG do analizy szeregów czasowych przy konstrukcji systemu predykcyjnego dla rynków walutowych (prognozy kursów walut) (Packard, 1990). W tym przypadku populację rozwiązań stanowił zbiór reguł, prognozujących przyszły trend kursu. Reguły te podlegały adaptacji podczas procesu minimalizacji funkcji celu, uwzględniającej zarówno sumaryczny średniokwadratowy błąd predykcji, jak i ryzyko. Otrzymano w rezultacie zbiór 20 reguł, których zastosowanie pozwoliło osiągnąć efektywność inwestycji porównywalną do rezultatów najlepszych analityków funkcjonujących na tym rynku.

Przykład wykorzystania techniki opartej na AG do zagadnienia wyboru funduszy powierniczych funkcjonujących na rynku USA, które przyniosą największy zwrot z powierzonego kapitału przedstawiony jest w pozycji (Bauer R.J. Jr, 1994). Tu również zagadnienie sprowadzono do znalezienia najlepszej reguły logicznej, według której wyszukiwany będzie właściwy fundusz.

Inny przykład wykorzystania tej metodologii do wspomagania decyzji inwestycyjnych (wraz z szeregiem istotnych sugestii i wskazówek) zaprezentowano w pracy (Leinweber & Arnott, 1995). Autorzy opracowania wykorzystywali narzędzia oparte na AG w zarządzaniu aktywami funduszu inwestycyjnego First Quadrant. Przedstawiono m.in. bardzo dobre rezultaty uzyskane przy wykorzystaniu algorytmów genetycznych w zarządzaniu portfelem akcji i obligacji na rynku amerykańskim i na rynkach innych państw. Przykładowo w sześcioletnim okresie od kwietnia 1987 do kwietnia 1993 roku aktywna strategia inwestycyjna z wykorzystaniem AG przyniosła wzrost wartości portfela funduszu o 155%, w porównaniu do 105% w przypadku stosowania standartowych strategii aktywnych funduszu, oraz do 85% w przypadku strategii pasywnej. Stworzenie strategii opartej o algorytmy genetyczne nastąpiło w oparciu o dane finansowe z okresu 1973 - 1986.

Próbie wykorzystania AG do optymalizacji strategii inwestycyjnej, generującej decyzje kupna/sprzedaży portfela akcji w oparciu o przewidywane zachowanie się rynku (indeksu WIG) zaproponowano w opracowaniu (Morajda, 1997b). Zadaniem algorytmu genetycznego jest tu dobór zestawu wskaźników technicznych oraz parametru reguły, generującej na ich podstawie sygnał kupna, oraz analogicznego zestawu wskaźników i parametru dla sygnału sprzedaży. Przedstawiono także propozycje doboru parametrów samego AG, jak i funkcji celu oraz funkcji przystosowania. W efekcie uzyskano dość obiecujące rezultaty zarówno w kontekście całkowitego zysku dla strategii aktywnej, jak i odpowiednich wskaźników Sharpe'a.

3.2.3. Inne aspekty zastosowań algorytmów genetycznych w ekonomii.

Należy nadmienić, że w ekonomii i biznesie narzędzia oparte o AG znalazły zastosowanie nie tylko na rynkach finansowych. Wykorzystano je np. do

prognozowania rentowności przedsiębiorstw (Thompson & Thompson, 1886), lub też do konstrukcji systemu uczącego się reguł opisujących preferencje konsumentów (Greene & Smith, 1887). Podstawowym obszarem wykorzystania AG w ekonomii i zarządzaniu pozostają jednak finanse. Fakt ten wynika m.in. ze stale wzrastającego potencjału obliczeniowego (związanego z rozwojem technologii komputerowej) dla rozwoju opartych na AG metod ilościowych w zagadnieniach finansowych, a także ze stosunkowej prostoty sformułowania większości problemów finansowych (np. inwestycyjnych) w kategoriach metodologii wykorzystującej algorytmy genetyczne.

Zarówno badania naukowe, jak i praktyczne aplikacje związane z wykorzystaniem AG owocują stale rosnącą liczbą artykułów i raportów dotyczących zastosowania tej metodologii w procesie inwestowania. Prace te zawierają głównie wyniki badań empirycznych dotyczących optymalizacji aktywnych strategii inwestycyjnych przy użyciu AG. Publikowane są one przede wszystkim w specjalistycznych periodykach (np. *Journal of Computational Intelligence in Finance*, poprzednio *Neurovest Journal*), ale także w innych wydawnictwach z dziedziny finansów.

3.3. Możliwości zastosowania algorytmów genetycznych w procesie konstrukcji sieci neuronowych.

Problematyka wzajemnego łączenia algorytmów genetycznych i sieci neuronowych (na zasadach współdziałania obu narzędzi lub wspomagania jednego narzędzia przez drugie - zob. np. (Schaffer et al., 1992)) jest od dawna rozwijana w licznych pracach badawczych i funkcjonuje w literaturze pod nazwą COGANN⁶⁵. Wyczerpująca klasyfikacja używanych w tej problematyce metod współpracy AG i SN, wraz z ich omówieniem, jest przedstawiona w książce (Rutkowska et al., 1997). Zarówno przesłanki wynikające z dotychczasowych badań nad powyższym zagadnieniem, jak i możliwości efektywnego samodzielnego zastosowania zarówno SN jak i AG w

⁶⁵ Ang.: *Combinations of Genetic Algorithms and Neural Networks*.

procesach wspomagania decyzji inwestycyjnych, skłaniają do podjęcia prób konstrukcji hybrydowych systemów decyzyjnych, wykorzystujących współpracę obu tych narzędzi. W istocie szereg publikacji i raportów zawartych w literaturze dotyczy prowadzonych badań lub zastosowania takich systemów wspomagania decyzji w praktyce - zob. np. (Katz & McCormick, 1994; Wong, 1994).

W systemach hybrydowych, opartych na połączeniu AG i SN, wykorzystuje się często koncepcje wykorzystania algorytmów genetycznych do ogólnie rozumianej konstrukcji sieci neuronowych, które z kolei stanowią zasadnicze narzędzia służące do generowania decyzji inwestycyjnych. Tego typu zastosowania AG mogą polegać m.in. na:

- uczeniu SN przy użyciu AG (ewolucyjny dobór wag połączeń w sieci),
- optymalizacji architektury SN przy wykorzystaniu AG (np. dobór liczby warstw ukrytych, ilości elementów w warstwach ukrytych, topologii połączeń, itp.),
- genetycznym wyborze parametrów uczenia lub reguł uczenia SN,
- wykorzystaniu AG w procesie doboru zestawu zmiennych wejściowych (a niekiedy również zmiennych wyjściowych) dla SN.

Rozważając powyższe możliwości wspomagania procesu konstrukcji i uczenia SN przy zastosowaniu AG, szczególnie obiecująca wydaje się koncepcja genetycznej optymalizacji doboru zmiennych wejściowych dla sieci w celu poprawy efektywności procesu uczenia i redukcji czasu obliczeń. Jak przedstawiono w rozdziale 2, problem doboru wejść dla sieci neuronowej jest szczególnie istotny w przypadku zastosowania SN w zagadnieniach predykcji lub wspomagania decyzji na rynkach finansowych, gdzie liczba potencjalnych zmiennych wejściowych, mogących mieć istotny wpływ na jakość funkcjonowania sieci, jest z reguły bardzo duża (por. podrozdział 2.7.6).

Ponieważ nie istnieją uniwersalne reguły, umożliwiające optymalny dobór wejść dla SN, zaś klasyczne metody statystyczne (np. analiza korelacji między daną zmienną wejściową a wyjściem) często okazują się w tym zagadnieniu mało wiarygodne (zob. np. Thomason, 1996), w związku z tym użycie metod genetycznych, które potwierdziły swoją dużą efektywność w wielu zagadnieniach podobnego typu i o

podobnym stopniu złożoności, może okazać się wysoce celowe (zob. Rutkowska et al., 1997; Guo & Uhrig, 1992). Przydatność koncepcji genetycznego doboru wejść potwierdzają także jej realizacje praktyczne, m.in. w sprzedawanym na rynku oprogramowaniu do wspomagania decyzji inwestycyjnych (np. *Genetica Neuroforecaster*, *NeuroGenetic Optimizer*), a także w oprogramowaniu do celów statystycznych (*Statistica - Neural Networks*).

Uwzględniając przedstawione powyżej przesłanki dla zastosowania AG do wspomagania procesu konstrukcji SN, zdecydowano się wykorzystać tą ideę w zaprezentowanej w pracy metodologii budowy systemów decyzyjnych. W dalszej części pracy zaproponowano zatem pewne własne rozwiązania genetycznego doboru zmiennych wejściowych dla SN oraz parametrów tych zmiennych, a także zaproponowano użycie AG do optymalizacji innych elementów sieci, np. ilości neuronów w warstwie ukrytej, albo typu zmiennej wyjściowej (horyzontu predykcji).

4. Propozycja nowej metodologii konstruowania i realizacji aktywnych strategii inwestycyjnych.

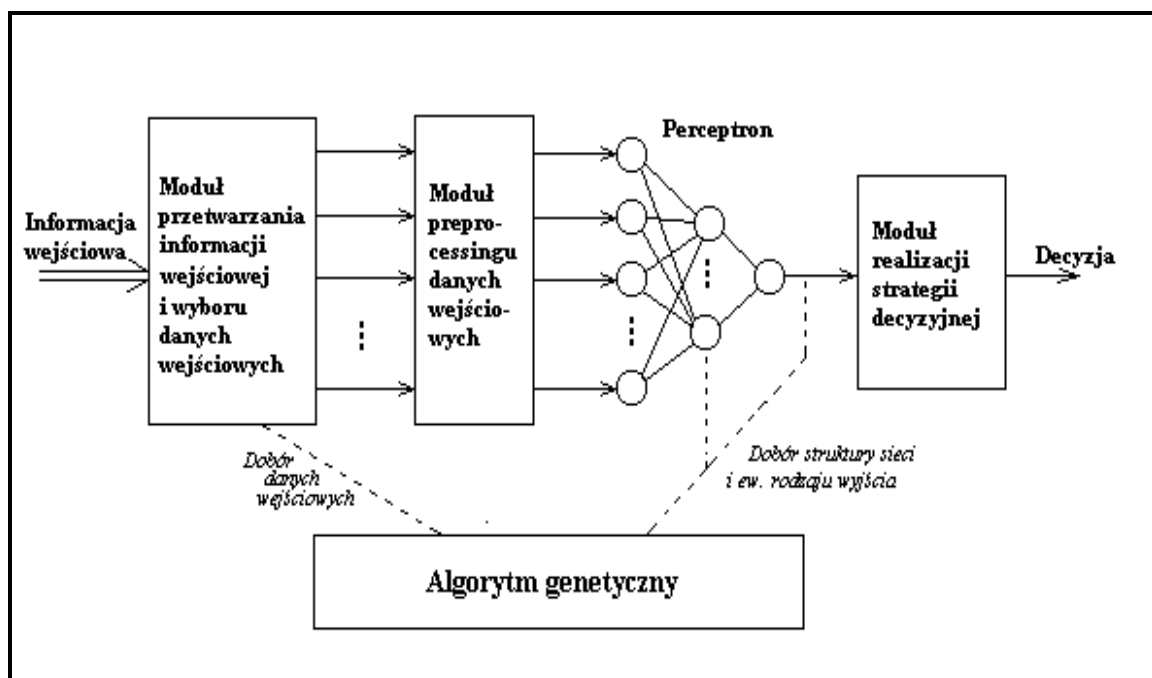
W niniejszym rozdziale przedstawiono propozycję opracowanej na potrzeby tej pracy nowej metodologii, mającej na celu konstruowanie systemów decyzyjnych, umożliwiających realizację aktywnych strategii inwestycyjnych. Strategie te mogą dotyczyć określonego instrumentu finansowego lub obejmować zarządzanie całym portfelem inwestycyjnym, w oparciu o prognozy rynku (*market timing*) lub dobór papierów wartościowych w portfelu (*security selection*). Prezentowaną metodologię oparto na przesłankach uzasadniających celowość jej stosowania (omówionych w rozdziale 1), oraz na wykorzystaniu narzędzi sztucznej inteligencji - sieci neuronowych i algorytmów genetycznych - opisanych w rozdziałach 2 i 3.

W założeniu omawiana w tej pracy metodologia stanowić ma ogólną koncepcję efektywnego wykorzystania SN i AG w postaci pewnego systemu hybrydowego, łączącego oba te narzędzia; jednak należy liczyć się (jak niemal zawsze przy budowie podobnych metodologii) z koniecznością doprecyzowania metody (systemu decyzyjnego) przez badacza lub praktyka do konkretnych warunków jej stosowania. Doprecyzowanie to może być realizowane poprzez dobór pewnej liczby parametrów systemu, mających zasadniczy wpływ na jego charakter i efektywność. W dalszej części pracy przedstawiono wyniki szeregu eksperymentów dotyczących prób zastosowania proponowanej metodologii do wybranych problemów. Zaprezentowane badania albo mogą stanowić podstawę do podjęcia przez przyszłego użytkownika metodologii arbitralnych decyzji przy ustalaniu parametrów, albo też mogą posłużyć jako punkt wyjściowy do przeprowadzenia przez niego własnych poszukiwań w przypadku konkretnego rozpatrywanego problemu. Ponieważ w wybranych przypadkach przedstawiono także efekty zastosowania omawianej metodologii w oparciu o notowania Giełdy Papierów Wartościowych w Warszawie, badania te

mogą też pełnić rolę weryfikacyjną, potwierdzającą przydatność metody w warunkach polskiego rynku finansowego.

4.1. Ogólna struktura systemu decyzyjnego i postulowany sposób jego wykorzystania.

Ogólny schemat systemu decyzyjnego⁶⁶ skonstruowanego w oparciu o prezentowaną metodologię przedstawia rys. 4-1.



Rys. 4-1. Schemat systemu decyzyjnego skonstruowanego w oparciu o zaproponowaną metodę połączenia SN i AG. Algorytm genetyczny oraz proces uczenia sieci neuronowych jest wykorzystywany jedynie w fazie budowy systemu (lub jego dalszego rozwoju). Źródło: opracowanie własne.

⁶⁶ Występujące w schemacie systemu i w dalszym tekście pojęcie „moduł” będzie w niniejszym rozdziale rozumiane w szerszym znaczeniu w stosunku do definicji z rozdziału 1 (zob. przypis 9). Tutaj będzie ono oznaczało dowolny element systemu decyzyjnego realizujący określony sposób przetwarzania informacji, a więc nie tylko element generujący prognozy lub decyzje. Takie rozszerzenie znaczenia tego terminu nie powinno wywołać żadnych nieporozumień.

Rozważany system wykorzystuje na wejściu dostępną informację pochodzącą z rynku finansowego (zarówno aktualną jak i z określonego okresu przeszłości), która jest wykorzystywana w procesie uczenia sieci neuronowych oraz służy do optymalizacji określonych parametrów systemu przy użyciu AG. Następnie ta sama informacja służy do generowania bieżącej decyzji inwestycyjnej w oparciu o gotowy system. Otrzymywana na wyjściu systemu decyzja jest precyzyjną rekomendacją kupna, sprzedaży lub zachowania pozycji (stanu posiadania) dla całego portfela papierów wartościowych, jego określonego składnika, lub dla odrębnego instrumentu finansowego⁶⁷. Decyzja ta może więc zostać łatwo zamieniona na konkretne zlecenie (zlecenia) kupna lub sprzedaży papierów wartościowych, a cały system może posłużyć do praktycznej realizacji i oceny aktywnej strategii inwestycyjnej w procesie zarządzania portfelem papierów wartościowych.

Należy zaznaczyć, że w przypadku praktycznego wykorzystania omawianej metodologii z reguły nie wystarczy skonstruowanie jednego systemu decyzyjnego. W przypadku bowiem stosowania strategii opartej o prognozy całego rynku można rozważyć koncepcję swoistej dywersyfikacji portfela, polegającej na podziale portfela na kilka (lub kilkanaście) części i „powierzenie” zarządzania każdej z tych części innemu systemowi decyzyjnemu, wybranemu ze zbioru systemów najlepiej ocenionych (według przyjętego kryterium). Również w przypadku strategii opartej o proces selekcji papierów wartościowych w portfelu, celowe może się okazać zbudowanie odrębnego systemu decyzyjnego dla każdego papieru mogącego wejść w skład portfela, a ponadto zrealizowanie dywersyfikacji według powyższych zasad.

W dalszej części niniejszego rozdziału omówiono zasady konstrukcji i funkcjonowania poszczególnych elementów wchodzących w skład zaprezentowanego schematu systemu decyzyjnego.

⁶⁷ W zależności od dostępnych w obrocie instrumentów finansowych inwestor może wykorzystywać w takich operacjach także kontrakty terminowe (*futures*) i sprzedaż krótką. Przykładowo zamiast operować całym portfelem akcji opartym o indeks WIG 20, można dokonywać transakcji na silnie skorelowanych z takim portfelem kontraktach terminowych dla WIG 20.

4.2. Moduł przetwarzania informacji wejściowej i wyboru danych wejściowych.

Informację wejściową dla systemu decyzyjnego mogą stanowić wszelkie dane, które mogą mieć choćby niewielki wpływ na zachowanie się rynku w przyszłości i które tym samym powinny stanowić przesłankę dla generowanych przez system decyzji inwestycyjnych. W szczególności mogą to być dane dotyczące takich czynników jak:

- wskaźniki makroekonomiczne (np. stopa inflacji, wysokość stóp procentowych, wybrane wskaźniki charakteryzujące wzrost gospodarczy kraju, itp.),
- sytuacja na rynku pieniężnym (np. wielkości popytu na bony skarbowe, zmiany kursu złotówki w stosunku do innych walut, interwencje NBP, itp.),
- sytuacja na innych giełdach światowych (zmiany głównych indeksów giełdowych, np. DJIA lub NIKKEI),
- dane fundamentalne z poszczególnych spółek giełdowych (np. zmiany wielkości sprzedaży, zysku, wskaźniki pochodzące z analizy finansowej, itp.),
- informacje otrzymane z poszczególnych spółek (np. nowe kontrakty, prognozy zysku, itp.),
- dane o charakterze technicznym, tj. prognozy wynikające z analizy formacji kształtujących się na wykresach, wartości wskaźników technicznych, nachylenia trendów, układy średnich kroczących, oraz inne wartości otrzymane w wyniku analizy zmienności kursów i wolumenu obrotów dla indeksów giełdowych i poszczególnych spółek (uzyskane w wyniku analizy technicznej).

Należy zwrócić uwagę, że przed wprowadzeniem do systemu wszystkie powyższe rodzaje informacji muszą zostać zamienione na postać numeryczną, umożliwiającą analizę ilościową w oparciu o komputerowe przetwarzanie danych.

W badaniach opisywanych w tej pracy zdecydowano się wykorzystać jedynie ostatni z wyżej wymienionych typów informacji (dane o charakterze technicznym)⁶⁸. Tak

⁶⁸ Na taką decyzję miały wpływ głównie następujące czynniki:

- łatwa dostępność odpowiedniej ilości danych umożliwiających efektywne uczenie sieci i wykorzystanie algorytmów genetycznych, a tym samym konstrukcję systemów decyzyjnych i ich wiarygodną ocenę,

więc jako informację wejściową dla systemu potraktowano szeregi czasowe dotyczące wartości notowań, oraz wolumenu obrotów dla określonej spółki, a także wartości indeksów giełdowych oraz wielkości obrotów dla całego rynku. Informacja ta ulega wstępnej transformacji w module realizującym przetwarzanie informacji wejściowej oraz dobór danych wejściowych. Koncepcja takiej transformacji wynika z szeregu przesłanek pochodzących z dotychczasowych badań opisanych w literaturze, wskazujących na fakt, iż sieci neuronowe uczą się i funkcjonują znacznie efektywniej w przypadku zastosowania właściwego przetworzenia „surowej” informacji wejściowej. Idea ta dotyczy również analizy szeregów czasowych, dla których zamiast podawania na wejście SN kolejnych realizacji szeregu zawartych w określonym oknie czasowym, można zastosować alternatywne rozwiązanie polegające na zastosowaniu odpowiednich funkcji przetwarzających te dane i podaniu wynikowych wartości tych funkcji na wejścia sieci neuronowej.

W celu realizacji wstępnego przetwarzania wejściowych informacji opracowano zestaw 110 procedur, których zadaniem jest przekształcenie wymienionych wyżej wejściowych szeregów czasowych na określone wartości przetransformowane, wykorzystywane następnie jako dane wejściowe dla sieci neuronowej. Procedury te stanowią albo proste implementacje wybranych wskaźników analizy technicznej (np. MACD, Stochastic, RSI, Momentum, itd.), albo pewne (zaproponowane przez autora) transformacje tych wskaźników, albo wreszcie pewne funkcje nie wykorzystywane dotychczas w analizie technicznej, jako bezpośrednio nie przydatne do generacji sygnałów kupna/sprzedaży (np. odchylenie standardowe dla jednodniowych względnych zmian kursu, wartość różnicy wybranych średnich kroczących, wartości współczynników kierunkowych prostych reprezentujących n -dniowe trendy), jednak obiecujące jako uzupełniające przesłanki w całości systemu. W większości procedur wykorzystano opisane w podrozdziale 2.7.5 średnie

-
- fakt iż informacja zawarta w danych technicznych odzwierciedla w dużym stopniu inne (wymienione wyżej) rodzaje informacji pochodzących z rynku,
 - numeryczny charakter danych technicznych,
 - efektywność wykorzystania tego typu informacji w początkowym etapie rozwoju omawianej metodologii.

Należy jednak podkreślić, że nie istnieją żadne przeciwwskazania do wykorzystania innych typów informacji (np. danych fundamentalnych) w dalszych badaniach dotyczących przedmiotowej problematyki lub w konkretnych aplikacjach praktycznych.

kroczące, użyte jako narzędzia służące do wstępnej eliminacji szumów („wygładzania” określonych szeregów czasowych); w wielu przypadkach średnie te stanowią zresztą integralny element wybranej metody przetwarzania danych (np. w formule wskaźnika MACD). Dane wejściowe dla powyższych procedur zawarte są w następujących szeregach czasowych⁶⁹:

- notowania kursów danej spółki⁷⁰,
- dzienny wolumen obrotów dla danej spółki wyrażony w sztukach akcji⁷¹,
- wartości Warszawskiego Indeksu Giełdowego (WIG),
- dzienny wolumen obrotów dla całego rynku.

Wykaz wszystkich 110 procedur zawarty jest w zamieszczonej poniżej tabeli 4-1.

Ponadto w przypadku analizy konkretnych spółek giełdowych (nie indeksów), możliwe jest dodatkowo użycie wybranych 81 spośród 110 procedur (zob. tabela 4-1), dla których zamiast wejściowego szeregu czasowego stanowiącego notowania kursów spółki zastosowany jest inny szereg czasowy $d(t)$ określony następująco:

$$d(t) = \ln \left(\frac{kurs(t)}{WIG(t)} \right) \quad (4.1)$$

gdzie: $kurs(t)$ - kurs danej spółki w chwili t

$WIG(t)$ - wartość indeksu WIG w chwili t

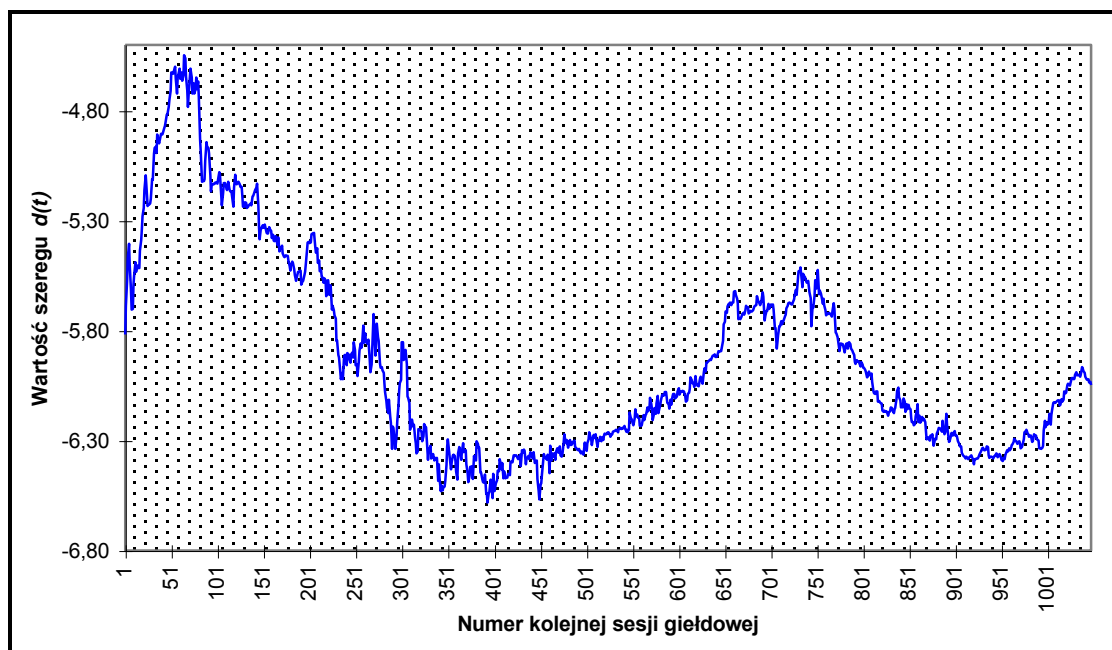
Przebieg tego typu szeregu czasowego dla spółki Exbud S.A. zaprezentowano na rys. 4-2. W przypadku użycia szeregu $d(t)$ zamiast kursów spółki, formuły określające funkcjonowanie niektórych procedur wymagają pewnej modyfikacji⁷², zachowana zostaje jednak istota danej procedury.

⁶⁹ Nie wszystkie szeregi czasowe są wszędzie wykorzystywane; o ile pierwszy z wymienionych szeregów jest używany prawie we wszystkich procedurach, o tyle ostatni z nich jest zastosowany tylko w jednej.

⁷⁰ W przypadku analizy indeksu giełdowego (np. WIG 20) są to wartości tego indeksu.

⁷¹ Dla indeksów giełdowych - w złotych (PLN).

⁷² Jest to związane z logarytmicznym charakterem szeregu $d(t)$. Na przykład jeżeli w standartowej procedurze używającej kursu spółki wymagane jest obliczenie względnej (procentowej) zmiany tego kursu, to w przypadku szeregu $d(t)$ wyznaczamy w to miejsce bezwzględną zmianę wartości: $d(t)-d(t-1)$ (jako różnicę logarytmów). Szczegóły tych modyfikacji zamieszczono w tabeli 4-1.



Rys. 4-2. Wykres przykładowego szeregu czasowego $d(t)$ zdefiniowanego formułą (4.1), określonego dla spółki Exbud S.A w 6-letnim okresie od 16.04.91 (pierwsze notowanie) do 16.04.97. Wyraźnie widoczne trendy średnioterminowe potwierdzają istnienie istotnych składników nielosowych w tym szeregu. Źródło: opracowanie własne.

Szereg czasowy $d(t)$ jest często bardziej użyteczny w badaniach, ponieważ obrazuje nie tyle bezwzględne zmiany notowań danej spółki, ile zmiany tych notowań w stosunku do zmian indeksu całego rynku; odzwierciedla on zatem względną siłę danej spółki na tle rynku. Zwróćmy bowiem uwagę na następującą zależność, pozwalającą na interpretację zmian wartości szeregu $d(t)$:

$$d(t) - d(t-1) = \ln\left(\frac{kurs(t)}{WIG(t)}\right) - \ln\left(\frac{kurs(t-1)}{WIG(t-1)}\right) = \ln\left(\frac{\frac{kurs(t)}{kurs(t-1)}}{\frac{WIG(t)}{WIG(t-1)}}\right) \quad (4.2)$$

Analiza danych zawartych w szeregu $d(t)$ może, zdaniem autora, okazać się szczególnie przydatna dla procesu generowania decyzji inwestycyjnych, toteż za celowe uznano użycie tych właśnie danych (po dokonaniu odpowiedniej

transformacji przez procedury przetwarzania informacji wejściowej) jako wartości wejściowych dla sieci neuronowych w rozważanym systemie decyzyjnym.

Procedury przetwarzania informacji wejściowej mogą posiadać parametry (takie jak np. okres lub okresy użytych średnich kroczących, długość okna czasowego dla pewnych transformacji danych wejściowych, itp.). Zwykle określona procedura posiada od 0 do 3 parametrów (zostały one wyszczególnione w tabeli 4-1). Każdy parametr każdej procedury może przybierać wartości całkowite nieujemne z określonego zakresu, zdefiniowanego osobno dla każdego parametru. Ponadto wewnątrz tych zakresów określono wartości preferowane poszczególnych parametrów, jako przyjęte *a priori* wielkości uznane (arbitralnie lub na podstawie przesłanek pochodzących z publikacji dotyczących analizy technicznej) za prawdopodobnie najlepsze (lub najpowszechniej stosowane) dla danego parametru.

TABELA 4-1

ZESTAWIENIE PROCEDUR WSTĘPNEGO PRZETWARZANIA INFORMACJI WEJŚCIOWEJ
(ZAWARTOŚĆ KOLUMN, OZNACZENIA I ZASTOSOWANE FUNKCJE SPECJALNE OBJAŚNIONO POD TABELĄ)

| Nr | Nazwa procedury (opis) | Formuła obliczania wartości wynikowej dla danej procedury | Para - metry | D | N |
|----|--|---|--------------|---|---|
| 1 | MACD (różnica średnich wykładniczych) | $E_{n,k}(X) - E_{2n+1,k}(X)$ | n | + | 0 |
| 2 | MACD w oparciu o linię sygnałową | $\text{sig}_9(E_{n,k}(X) - E_{2n+1,k}(X))$ | n | + | 0 |
| 3 | MACD: generator impulsów z linią sygnałową | $\text{imp}_{0,5}(\text{sig}_9(E_{n,k}(X) - E_{2n+1,k}(X)))$ | n | + | - |
| 4 | Wskaźnik zmiany ROC (Rate of Change) | $\ln(x_k) - \ln(x_{k-r})$; D: $d_k - d_{k-r}$ | r | + | 1 |
| 5 | ROC z liniami wykupienia i wysprzedania | $\text{ww}_{0,8}^q(\ln(x_k) - \ln(x_{k-r}))$; D: $\text{ww}_{0,8}^q(d_k - d_{k-r})$ | r, q | + | - |
| 6 | ROC: generator impulsów z linią sygnałową | $\text{imp}_{0,5}(\text{sig}_9(\ln(x_k/x_{k-r})))$ D: $\text{imp}_{0,5}(\text{sig}_9(d_k - d_{k-r}))$ | r | + | - |
| 7 | Momentum | $\ln(x_k/S_{n,k}(X))$; D: $d_k - S_{n,k}(D)$ | n | + | 1 |
| 8 | Momentum z liniami wykupienia/wysprzedania | $\text{ww}_{0,8}^q(\ln(x_k/S_{n,k}(X)))$; D: $\text{ww}_{0,8}^q(d_k - S_{n,k}(D))$ | n, q | + | - |

| Nr | Nazwa procedury (opis) | Formuła obliczania wartości wynikowej dla danej procedury | Para - metry | D | N |
|----|--|---|--------------|---|---|
| 9 | Momentum: generator impulsów z linią sygnałową | $\text{imp}_{0,8}(\text{sig}_9(\ln(x_k/S_{n,k}(X))))$ D: $\text{imp}_{0,8}(\text{sig}_9(d_k - S_{n,k}(D)))$ | n | + | - |
| 10 | Wskaźnik wykupienia / wysprzedania (O/O) | $(x_k - \min_{k-r < i \leq k}(x_i)) / (\max_{k-r < i \leq k}(x_i) - \min_{k-r < i \leq k}(x_i)) - 0,5$ | r | + | 1 |
| 11 | Wskaźnik O/O z liniami wykupienia/wysprzedania | $\text{ww}_{0,8}^q \left((x_k - \min_{k-r < i \leq k}(x_i)) / (\max_{k-r < i \leq k}(x_i) - \min_{k-r < i \leq k}(x_i)) - 0,5 \right)$ | r, q | + | - |
| 12 | Wskaźnik względnej siły RSI | $\text{RSI}_{r,k} = \frac{\text{up}_r(g_k)}{\text{up}_r(g_k) + \text{down}_r(g_k)} - \frac{1}{2}$ gdzie: $g_k = (x_k - x_{k-1}) / x_{k-1}$; D: $g_k = d_k - d_{k-1}$ | r | + | 1 |
| 13 | Wskaźnik RSI z liniami wykupienia/wysprzedania | $\text{ww}_{0,8}^q(\text{RSI}_{r,k})$, gdzie $\text{RSI}_{r,k}$ jest obliczane jak wyżej | r, q | + | - |
| 14 | Oscylator cen POS dla średnich zwykłych | $\text{POS}_{n,m} = \left(\epsilon_{n,k}(X) - S_{m,k}(X) \right) S_{m,k}(X)$ D: $\text{POS}_{n,m} = S_{n,k}(D) - S_{m,k}(D)$ | n, m | + | 0 |
| 15 | Oscylator POS impulsowy z linią sygnałową | $\text{imp}_{0,8}(\text{sig}_9(\text{POS}_{n,m}))$, $\text{POS}_{n,m}$ jest obliczane jak wyżej | n, m | + | - |
| 16 | Oscylator cen POE dla średnich wykładniczych | $\text{POE}_{n,m} = \left(\epsilon_{n,k}(X) - E_{m,k}(X) \right) E_{m,k}(X)$ D: $\text{POE}_{n,m} = E_{n,k}(D) - E_{m,k}(D)$ | n, m | + | 0 |
| 17 | Oscylator POE w oparciu o linię sygnałową | $\text{sig}_9(\text{POE}_{n,2n+1})$, gdzie $\text{POE}_{n,2n+1}$ jest obliczane jak wyżej | n | + | 0 |
| 18 | Oscylator POE impulsowy z linią sygnałową | $\text{imp}_{0,5}(\text{sig}_9(\text{POE}_{n,2n+1}))$, POE jest obliczane jak w p.16 | n | + | - |
| 19 | Oscylator wolumenu VOS | $\text{VOS}_{n,m} = \left(\epsilon_{n,k}(\text{VOL}) - S_{m,k}(\text{VOL}) \right) S_{m,k}(\text{VOL})$ | n, m | - | 0 |
| 20 | Oscylator VOS impulsowy z linią sygnałową | $\text{imp}_{0,8}(\text{sig}_9(\text{VOS}_{n,m}))$, $\text{VOS}_{n,m}$ jest obliczane jak wyżej | n, m | - | - |
| 21 | Odchylenie wolumenu | $\left(\text{vol}_k - S_{n,k}(\text{VOL}) \right) S_{n,k}(\text{VOL})$ | n | - | 1 |
| 22 | Wskaźnik różnicy średnich wolumenu | $\ln(S_{n,k}(\text{VOL})) - \ln(S_{m,k}(\text{VOL}))$ | n, m | - | 0 |
| 23 | Odchylenie kursu - wersja 1 | $\ln \left(\epsilon_k / S_{n,k-r}(X) \right)$; D: $d_k - S_{n,k-r}(D)$ | r, n | + | 1 |
| 24 | Odchylenie kursu - wersja 2 | Formuła identyczna jak w procedurze 23, ale zastosowano inne dopuszczalne zakresy zmienności parametrów r, n | r, n | + | 1 |
| 25 | Wskaźnik różnicy średnich kursu | $\ln \left(\epsilon_{n,k}(X) / S_{m,k-r}(X) \right)$; D: $E_{n,k}(D) - S_{m,k-r}(D)$ | r, m, n | + | 0 |

| Nr | Nazwa procedury (opis) | Formuła obliczania wartości wynikowej dla danej procedury | Para - metry | D | N |
|----|--|--|--------------|---|---|
| 26 | Porównanie z WIG - (1) różnica sum stóp zwrotu | $\sum_{i=k-r+1}^k sz(x_i) - \sum_{i=k-r+1}^k sz(wig_i)$ | r | - | 0 |
| 27 | Porównanie z WIG - (2) różnica sum stóp zwrotu | Formuła identyczna jak w procedurze 26, ale zastosowano inny dopuszczalny zakres zmienności parametru r | r | - | 0 |
| 28 | Porównanie z WIG - (3) ilość lepszych stóp zwrotu | Liczba dni dla których $sz(x_i) > sz(wig_i)$, pomniejszona o liczbę dni dla których $sz(x_i) < sz(wig_i)$, gdzie: $i \in [k-r, k]$ | r | - | 0 |
| 29 | Porównanie z WIG - (4) ilość lepszych stóp zwrotu | Formuła identyczna jak w procedurze 28, ale zastosowano inny dopuszczalny zakres zmienności parametru r | r | - | 0 |
| 30 | Porównanie z WIG - (5) dwie kolejne stopy zwrotu | $\begin{cases} +1 & \text{gdy } sz(x_k) - sz(wig_k) > sz(x_{k-1}) - sz(wig_{k-1}) > 0 \\ -1 & \text{gdy } sz(x_k) - sz(wig_k) < sz(x_{k-1}) - sz(wig_{k-1}) < 0 \\ 0 & \text{gdy nie zachodzi żaden z powyższych przypadków} \end{cases}$ | - | - | - |
| 31 | Oscylator skumulowanego wolumenu VAO w oparciu o linię sygnałową | $\text{sig}_p(\text{VAO}_{r,k})$, gdzie $\text{VAO}_{r,k}$ jest określone wzorem: $\text{VAO}_{r,k} = \sum_{j=1}^k \left(\left(x_j - \frac{1}{2} (\max_{j-r < i \leq j} (x_i) + \min_{j-r < i \leq j} (x_i)) \right) \cdot \sum_{i=j-r+1}^j \text{vol}_i \right)$ | r, p | + | 0 |
| 32 | Oscylator Chaikina | $\begin{cases} +1 & \text{gdy } \text{CHK}_r(x_k) > 0 \text{ i } \text{CHK}_r(x_k) > \text{CHK}_r(x_{k-1}) \\ -1 & \text{gdy } \text{CHK}_r(x_k) < 0 \text{ i } \text{CHK}_r(x_k) < \text{CHK}_r(x_{k-1}) \\ 0 & \text{gdy nie zachodzi żaden z powyższych przypadków} \end{cases}$ gdzie: $\text{CHK}_r(x_k) = E_{3,k}(\text{VAO}_{r,k}) - E_{10,k}(\text{VAO}_{r,k})$ | r | + | - |
| 33 | Oscylator Chaikina w oparciu o linię sygnałową | $\text{sig}_p(\text{CHK}_r(x_k))$ | r, p | + | 0 |
| 34 | Dodatni indeks obrotu PVI z linią sygnałową | $\text{sig}_p \left(\sum_{j=2}^k g_j \right)$, gdzie: $g_j = \begin{cases} 0 & \text{gdy } \text{vol}_j \leq \text{vol}_{j-1} \\ \text{sz}(x_j) & \text{gdy } \text{vol}_j > \text{vol}_{j-1} \\ \mathbf{D}: d_j - d_{j-1} \end{cases}$ | p | + | 0 |
| 35 | Ujemny indeks obrotu NVI z linią sygnałową | $\text{sig}_p \left(\sum_{j=2}^k g_j \right)$, gdzie: $g_j = \begin{cases} 0 & \text{gdy } \text{vol}_j > \text{vol}_{j-1} \\ \text{sz}(x_j) & \text{gdy } \text{vol}_j \leq \text{vol}_{j-1} \\ \mathbf{D}: d_j - d_{j-1} \end{cases}$ | p | + | 0 |
| 36 | Oscylator Bollingera BO | $\text{BO}_{r,k} = \left(x_k - \frac{1}{r} \sum_{i=k-r+1}^k x_i \right) / \sigma \llcorner_i: \text{ dla } k-r < i \leq k \rceil$ gdzie σ oznacza odchylenie standardowe | r | + | 1 |
| 37 | Oscylator BO z liniami wykupienia/wysprzedania | $\text{ww}_{0,8}^q(\text{BO}_{r,k})$ | r, q | + | - |
| 38 | Łatwość ruchu (<i>Ease of Movement Value</i>) | $\begin{cases} +1 & \text{gdy } g_{k,r} > 0 \text{ i } g_{k,r} > g_{k-1,r} \\ -1 & \text{gdy } g_{k,r} < 0 \text{ i } g_{k,r} < g_{k-1,r} \\ 0 & \text{gdy nie zachodzą w/w warunki} \end{cases}$ gdzie: | r | + | - |

| Nr | Nazwa procedury (opis) | Formuła obliczania wartości wynikowej dla danej procedury | Para - metry | D | N |
|----|---|--|--------------|---|---|
| | | $g_{i,r} = E_{10,i} \left(\frac{1}{2} h_{i,r} \cdot \left(\frac{\max_{i-r < j \leq i} (x_j) - \min_{i-r < j \leq i} (x_j)}{\sum_{j=i-r+1}^i \text{vol}_j} \right) \right)$ $h_{i,r} = \max_{i-r < j \leq i} (x_j) + \min_{i-r < j \leq i} (x_j) - \max_{i-2r < j \leq i-r} (x_j) - \min_{i-2r < j \leq i-r} (x_j)$ | | | |
| 39 | Wskaźnik bilansu wolumenu OBV | $OBV_{k,r} = \sum_{i=k-r+1}^k \left[x_i - x_{i-1} \right] \cdot \text{vol}_i$ | r | + | 0 |
| 40 | Wskaźnik OBV - wersja zmodyfikowana | $\begin{cases} +1 & \text{gdy } OBV_{k,k-1} > OBV_{k-r,k-r-1} < OBV_{k-2r,k-2r-1} \\ -1 & \text{gdy } OBV_{k,k-1} < OBV_{k-r,k-r-1} > OBV_{k-2r,k-2r-1} \\ 0 & \text{gdy nie są spełnione powyższe warunki} \end{cases}$ | r | + | - |
| 41 | Cenowy trend wolumenu VPT zmodyfikowany średnią adaptacyjną | $\begin{cases} +1 & \text{gdy } 0 < A_{5,k}(OBV_{k,r}) > A_{5,k-1}(OBV_{k-1,r}) \\ -1 & \text{gdy } 0 > A_{5,k}(OBV_{k,r}) < A_{5,k-1}(OBV_{k-1,r}) \\ 0 & \text{gdy nie są spełnione powyższe warunki} \end{cases}$ | r | + | - |
| 42 | Cenowy trend wolumenu VPT - wersja 2 | Formuła identyczna jak w procedurze 41, ale zastosowano inny dopuszczalny zakres zmienności parametru r | r | + | - |
| 43 | Dywergencja pomiędzy VPT a kursem | $\begin{cases} +1 & \text{gdy } A_{10,k}(X) < A_{10,k-1}(X) \text{ i } A_{10,k}(OBV_{k,r}) \geq A_{10,k-1}(OBV_{k-1,r}) \\ -1 & \text{gdy } A_{10,k}(X) > A_{10,k-1}(X) \text{ i } A_{10,k}(OBV_{k,r}) \leq A_{10,k-1}(OBV_{k-1,r}) \\ 0 & \text{gdy nie są spełnione powyższe warunki} \end{cases}$ | r | + | - |
| 44 | Wskaźnik względnej siły RS z linią sygnałową | $\text{sig}_{20} \left(\frac{\sum_{i=k-r+1}^k x_i}{\sum_{i=k-r+1}^k \text{wig}_i} \right)$ | r | - | 0 |
| 45 | Wskaźnik RS dla średniej adaptacyjnej | $\text{sig}_{20} \left(\frac{A_{n,k}(X)}{A_{n,k}(\text{WIG})} \right)$ | n | - | 0 |
| 46 | Wskaźnik kolejnych wzrostów/spadków | $\begin{cases} +1 & \text{gdy } x_k > x_{k-1} > \dots > x_{k-r} \\ -1 & \text{gdy } x_k < x_{k-1} < \dots < x_{k-r} \\ 0 & \text{gdy nie są spełnione w/w warunki} \end{cases}$ | r | + | - |
| 47 | Wskaźnik KAIRI | $\left(\frac{E_k}{E_{n,k}(X)} \right) - 1$ | n | - | 0 |
| 48 | Wskaźnik losowych zmian (<i>Random Walk Indicator</i>) | $\sqrt{r} \cdot \left(\frac{E_k - x_{k-r}}{\sum_{i=k-r+1}^k x_i - x_{i-1} } \right)$ | r | - | 0 |
| 49 | Wskaźnik losowych zmian (wersja 2) | Formuła identyczna jak w procedurze 48, ale zastosowano inny dopuszczalny zakres zmienności parametru r | r | - | 0 |
| 50 | Wskaźnik zmienności VI (<i>Volatility Indicator</i>) | $VI_{k,r} = k \cdot g_{k,r} / \sum_{i=1}^k g_{i,r} \quad \text{gdzie:}$ $g_{i,r} = \left(\frac{\max_{i-r < j \leq i} (x_j) - \min_{i-r < j \leq i} (x_j)}{\sum_{j=i-r+1}^i \text{vol}_j} \right) \cdot (x_i - x_{i-r+1})$ | r | + | 0 |
| 51 | Średnia adaptacyjna dla wskaźnika zmienności | $A_{10,k}(VI_{k,r})$ | r | + | 0 |
| 52 | Zmienność ekstremalna | $\max_{k-r < j \leq k} (x_j) - \min_{k-r < j \leq k} (x_j)$ | r | + | 1 |

| Nr | Nazwa procedury (opis) | Formuła obliczania wartości wynikowej dla danej procedury | Para - metry | D | N |
|----|--|---|--------------|---|---|
| 53 | Odchylenie standardowe dziennych stóp zwrotu | $\sigma (sz(x_i) : \text{dla } k-r < i \leq k)$ D: $\sigma ((x_i - x_{i-1}) : \text{dla } k-r < i \leq k)$ | r | + | 1 |
| 54 | Dzienna zmiana kursu | $(x_k - x_{k-1}) / x_{k-1}$ | - | + | 0 |
| 55 | 2-dniowa zmiana kursu | $(x_k - x_{k-2}) / x_{k-2}$ | - | - | 0 |
| 56 | 3-dniowa zmiana kursu | $(x_k - x_{k-3}) / x_{k-3}$ | - | - | 0 |
| 57 | Odchylenie od średniej wykładniczej (wersja 1) | $x_k - E_{n,k}(X)$ | n | + | 0 |
| 58 | Odchylenie od średniej wykładniczej (wersja 2) | $x_k - E_{n,k}(X)$, dla większych wartości parametru n | n | + | 0 |
| 59 | Odchylenie od średniej zwykłej (wersja 1) | $x_k - S_{n,k}(X)$ | n | + | 0 |
| 60 | Odchylenie od średniej zwykłej (wersja 2) | $x_k - S_{n,k}(X)$, dla większych wartości parametru n | n | + | 0 |
| 61 | Różnica średnich (1) | $A_{10,k}(X) - E_{n,k}(X)$ | n | + | 0 |
| 62 | Różnica średnich (2) | $A_{10,k}(X) - E_{n,k}(X)$, dla większych wartości parametru n | n | + | 0 |
| 63 | Różnica średnich (3) | $A_{10,k}(X) - S_{n,k}(X)$ | n | + | 0 |
| 64 | Różnica średnich (4) | $A_{10,k}(X) - S_{n,k}(X)$, dla większych wartości parametru n | n | + | 0 |
| 65 | Różnica średnich (5) | $E_{n,k}(X) - E_{m,k}(X)$ | n, m | + | 0 |
| 66 | Różnica średnich (6) | $E_{n,k}(X) - S_{r,k}(X)$ | n, r | + | 0 |
| 67 | Różnica średnich (7) | $E_{n,k}(X) - S_{m,k}(X)$ ($m > n$) | n, m | + | 0 |
| 68 | Różnica średnich (8) | $E_{n,k}(X) - S_{r,k}(X)$, dla większych n i r niż w procedurze 66 | n, r | + | 0 |
| 69 | Różnica średnich (9) | $S_{n,k}(X) - S_{m,k}(X)$ | n, m | + | 0 |
| 70 | Generator sygnałów (1) oparty o różnicę średnich | $\begin{cases} +1 & \text{gdy } A_{10,k}(X) > E_{n,k}(X) \geq E_{n,k-1}(X) \text{ i } A_{10,k}(X) > A_{10,k-1}(X) \\ -1 & \text{gdy } A_{10,k}(X) < E_{n,k}(X) \leq E_{n,k-1}(X) \text{ i } A_{10,k}(X) < A_{10,k-1}(X) \\ 0 & \text{gdy nie zachodzi żaden z powyższych warunków} \end{cases}$ | n | + | - |
| 71 | Generator sygnałów (2) oparty o różnicę średnich | Formuła identyczna jak w procedurze 70, ale zastosowano większe dopuszczalne wartości dla parametru n | n | + | - |
| 72 | Generator sygnałów (3) oparty o różnicę średnich | $\begin{cases} +1 & \text{gdy } A_{10,k}(X) > S_{n,k}(X) \geq S_{n,k-1}(X) \text{ i } A_{10,k}(X) > A_{10,k-1}(X) \\ -1 & \text{gdy } A_{10,k}(X) < S_{n,k}(X) \leq S_{n,k-1}(X) \text{ i } A_{10,k}(X) < A_{10,k-1}(X) \\ 0 & \text{gdy nie zachodzi żaden z powyższych warunków} \end{cases}$ | n | + | - |
| 73 | Generator sygnałów (4) oparty o różnicę średnich | $\begin{cases} +1 & \text{gdy } E_{n,k}(X) > E_{m,k}(X) \geq E_{m,k-1}(X) \text{ i } E_{n,k}(X) > E_{n,k-1}(X) \\ -1 & \text{gdy } E_{n,k}(X) < E_{m,k}(X) \leq E_{m,k-1}(X) \text{ i } E_{n,k}(X) < E_{n,k-1}(X) \\ 0 & \text{gdy nie zachodzi żaden z powyższych warunków} \end{cases}$ | n, m | + | - |

| Nr | Nazwa procedury (opis) | Formuła obliczania wartości wynikowej dla danej procedury | Para - metry | D | N |
|----|---|---|--------------|---|---|
| 74 | Generator sygnałów (5) oparty o różnicę średnich | $\begin{cases} +1 & \text{gdy } E_{r,k}(X) > S_{n,k}(X) \geq S_{n,k-1}(X) \text{ i } E_{r,k}(X) > E_{r,k-1}(X) \\ -1 & \text{gdy } E_{r,k}(X) < S_{n,k}(X) \leq S_{n,k-1}(X) \text{ i } E_{r,k}(X) < E_{r,k-1}(X) \\ 0 & \text{gdy nie zachodzi żaden z powyższych warunków} \end{cases}$ | n, r | + | - |
| 75 | Wskaźnik ROC dla wolumenu | $\ln \left(A_{10,k}(\text{VOL}) \right) - \ln \left(A_{10,k-r}(\text{VOL}) \right)$ | r | - | 1 |
| 76 | Zmiana wolumenu | $(\text{vol}_k - \text{vol}_{k-1}) / \text{vol}_{k-1}$ | - | - | 1 |
| 77 | Zmiana średniej adaptacyjnej dla wolumenu | $\left(A_{n,k}(\text{VOL}) - A_{n,k-1}(\text{VOL}) \right) / A_{n,k-1}(\text{VOL})$ | n | - | 1 |
| 78 | Momentum dla wolumenu | $\ln(\text{vol}_k) - \ln \left(A_{n,k}(\text{VOL}) \right)$ | n | - | 1 |
| 79 | Wskaźnik O/O zastosowany dla wolumenu | $(\text{vol}_k - \min_{k-r < i \leq k}(\text{vol}_i)) / (\max_{k-r < i \leq k}(\text{vol}_i) - \min_{k-r < i \leq k}(\text{vol}_i)) - 0,5$ | r | - | 1 |
| 80 | Względne odchylenie wolumenu od średniej | $\left(\text{vol}_k - S_{n,k-r}(\text{VOL}) \right) / S_{n,k-r}(\text{VOL})$ | r, n | - | 1 |
| 81 | Porównanie zmian wolumenu ze zmianą obrotów dla rynku | $\sum_{i=k-r+1}^k \frac{\text{vol}_i - \text{vol}_{i-1}}{\text{vol}_{i-1}} - \sum_{i=k-r+1}^k \frac{\text{volwig}_i - \text{volwig}_{i-1}}{\text{volwig}_{i-1}}$ | r | - | 0 |
| 82 | Wskaźnik KAIRI zastosowany dla wolumenu | $\left(\text{vol}_k / E_{n,k}(\text{VOL}) \right) - 1$ | n | - | 0 |
| 83 | Różnica średnich z wolumenu (1) | $A_{5,k}(\text{VOL}) - S_{n,k}(\text{VOL})$ | n | - | 0 |
| 84 | Różnica średnich z wolumenu (2) | $A_{10,k}(\text{VOL}) - S_{m,k}(\text{VOL})$ | m | - | 0 |
| 85 | Wskaźnik ROC dla średniej adaptacyjnej z kursu | $\ln \left(A_{10,k}(X) \right) - \ln \left(A_{10,k-r}(X) \right)$ D: $A_{10,k}(D) - A_{10,k-r}(D)$ | r | + | 1 |
| 86 | ROC dla średniej adaptacyjnej (wariant 2) | Formuła identyczna jak w procedurze 85, ale zastosowano większe dopuszczalne wartości dla parametru r | r | + | 1 |
| 87 | Wskaźnik ROC dla średniej wykładniczej z kursu | $\ln \left(E_{n,k}(X) \right) - \ln \left(E_{n,k-r}(X) \right)$ D: $E_{n,k}(D) - E_{n,k-r}(D)$ | r, n | + | 1 |
| 88 | ROC dla średniej wykładniczej (wariant 2) | Formuła identyczna jak w procedurze 87, ale zastosowano większe dopuszczalne wartości dla parametru n | r, n | + | 1 |
| 89 | Dzienna zmiana średniej adaptacyjnej | $\left(A_{n,k}(X) / A_{n,k-1}(X) \right) - 1$; D: $A_{n,k}(D) - A_{n,k-1}(D)$ | n | + | 0 |
| 90 | Dzienna zmiana średniej wykładniczej | $\left(E_{n,k}(X) / E_{n,k-1}(X) \right) - 1$; D: $E_{n,k}(D) - E_{n,k-1}(D)$ | n | + | 0 |
| 91 | Zmodyfikowany MACD z linią sygnałową | $\text{sig}_9(g_{n,k})$ gdzie: $g_{n,k} = A_{10,k}(X) - E_{n,k}(X)$ | n | + | 0 |

| Nr | Nazwa procedury (opis) | Formuła obliczania wartości wynikowej dla danej procedury | Para - metry | D | N |
|-----|---|---|--------------|---|---|
| 92 | Zmodyfikowany MACD generujący sygnały w oparciu o linię sygnałową | $\begin{cases} +1 & \text{gdy } \text{sig}_g(g_{n,k}) \geq 0 \text{ i } g_{n,k} > g_{n,k-1} \\ -1 & \text{gdy } \text{sig}_g(g_{n,k}) \leq 0 \text{ i } g_{n,k} < g_{n,k-1} \\ 0 & \text{gdy nie są spełnione powyższe warunki} \end{cases}$ <p style="text-align: center;">gdzie: $g_{n,k} = A_{10,k}(X) - E_{n,k}(X)$</p> | n | + | - |
| 93 | Wskaźnik ROC dla średniej adaptacyjnej z kursu generujący sygnały w oparciu o linię sygnałową | $\begin{cases} +1 & \text{gdy } g_{r,k} \geq 0 \text{ i } g_{r,k} > g_{r,k-1} \\ -1 & \text{gdy } g_{r,k} \leq 0 \text{ i } g_{r,k} < g_{r,k-1} \\ 0 & \text{gdy nie są spełnione w/w warunki} \end{cases}$ <p style="text-align: center;">gdzie:</p> $g_{r,k} = \begin{cases} \text{sig}_g \left[\frac{A_{10,k}(X) - A_{10,k-r}(X)}{A_{10,k}(X) + A_{10,k-r}(X)} \right] - 1 \\ \text{D: } \text{sig}_g \left[\frac{A_{10,k}(D) - A_{10,k-r}(D)}{A_{10,k}(D) + A_{10,k-r}(D)} \right] \end{cases}$ | r | + | - |
| 94 | Momentum dla średniej adaptacyjnej generujący sygnały w oparciu o linię sygnałową | $\begin{cases} +1 & \text{gdy } \text{sig}_g(g_{n,k}) \geq 0 \text{ i } g_{n,k} > g_{n,k-1} \\ -1 & \text{gdy } \text{sig}_g(g_{n,k}) \leq 0 \text{ i } g_{n,k} < g_{n,k-1} \\ 0 & \text{gdy nie są spełnione powyższe warunki} \end{cases}$ <p style="text-align: center;">gdzie:</p> $g_{n,k} = \ln \left(\frac{A_{10,k}(X)}{S_{n,k}(X)} \right); \quad \text{D: } g_{n,k} = A_{10,k}(D) - S_{n,k}(D)$ | n | + | - |
| 95 | Porównanie zmodyfikowanych wskaźników ROC dla spółki i indeksu WIG | $\ln \left(\frac{A_{10,k}(X) \cdot A_{10,k-r}(\text{WIG})}{A_{10,k-r}(X) \cdot A_{10,k}(\text{WIG})} \right)$ | r | - | 0 |
| 96 | Porównanie zmodyfikowanych ROC (wariant 2) | Formuła identyczna jak w procedurze 95, ale zastosowano większe dopuszczalne wartości dla parametru r | r | - | 0 |
| 97 | Wskaźnik trendu DMI | $\text{DMI}_{r,k} = \frac{\text{up}_r(g_k) - \text{down}_r(g_k)}{\text{up}_r(g_k) + \text{down}_r(g_k)}; \quad \text{gdzie: } g_k = x_k - x_{k-1}$ | r | + | 0 |
| 98 | Wskaźnik siły trendu ADX | $\text{ADX}_{r,k} = \begin{cases} +1 & \text{gdy } A_{15,k} \left(\text{DMI}_{r,k} \right) > A_{15,k-1} \left(\text{DMI}_{r,k-1} \right) \\ -1 & \text{gdy } A_{15,k} \left(\text{DMI}_{r,k} \right) < A_{15,k-1} \left(\text{DMI}_{r,k-1} \right) \end{cases}$ | r | + | - |
| 99 | Generator sygnałów w oparciu o DMI i ADX | $\begin{cases} +1 & \text{gdy } \text{ADX}_{r,k} = +1 \text{ i } \text{up}_r(g_k) > \text{down}_r(g_k) \\ -1 & \text{gdy } \text{ADX}_{r,k} = +1 \text{ i } \text{up}_r(g_k) < \text{down}_r(g_k) \\ 0 & \text{gdy } \text{ADX}_{r,k} = -1 \end{cases}$ <p style="text-align: center;">gdzie: $g_k = x_k - x_{k-1}$</p> | r | + | - |
| 100 | Stochastic Slow (1) S1 | $S1_{n,k} = \begin{cases} +1 & \text{gdy } x_k > E_{n,k}(X) > E_{n,k-1}(X) \\ -1 & \text{gdy } x_k < E_{n,k}(X) < E_{n,k-1}(X) \\ S1_{n,k-1} & \text{gdy nie zachodzą w/w warunki} \end{cases}$ | n | + | - |
| 101 | Stochastic Slow (2) S2 | $S2_k = \begin{cases} +1 & \text{gdy } S1_{n,k} = 1 \text{ i } \tilde{E}_k^r < s \\ -1 & \text{gdy } S1_{n,k} = -1 \text{ i } \tilde{E}_k^r > 100 - s \\ 0 & \text{gdy nie zachodzą w/w warunki} \end{cases}$ | n, r, s | + | - |

| Nr | Nazwa procedury (opis) | Formuła obliczania wartości wynikowej dla danej procedury | Para - metry | D | N |
|-----|---|---|--------------|---|---|
| 102 | Stochastic Slow (3) S3 | $S3_k = \begin{cases} +1 & \text{gdy } S1_{n,k}=1 \text{ i } s > \tilde{E}_k^r > \tilde{E}_{k-1}^r < \tilde{E}_{k-2}^r < \tilde{E}_{k-3}^r \\ -1 & \text{gdy } S1_{n,k}=-1 \text{ i } 100-s < \tilde{E}_k^r < \tilde{E}_{k-1}^r > \tilde{E}_{k-2}^r > \tilde{E}_{k-3}^r \\ 0 & \text{gdy } (S1_{n,k}=-1 \text{ i } S3_{k-1}=1) \text{ lub } (S1_{n,k}=1 \text{ i } S3_{k-1}=-1) \\ S3_{k-1} & \text{gdy nie są spełnione powyższe warunki} \end{cases}$ | n, r, s | + | - |
| 103 | Stochastic Slow (4) S4 | $S4_k = \begin{cases} +1 & \text{gdy } S1_{n,k}=1 \text{ i } s > \tilde{E}_k^r > \tilde{E}_{k-1}^r > \tilde{E}_{k-2}^r < \tilde{E}_{k-3}^r < \tilde{E}_{k-4}^r \\ -1 & \text{gdy } S1_{n,k}=-1 \text{ i } 100-s < \tilde{E}_k^r < \tilde{E}_{k-1}^r < \tilde{E}_{k-2}^r > \tilde{E}_{k-3}^r > \tilde{E}_{k-4}^r \\ 0 & \text{gdy } (S1_{n,k}=-1 \text{ i } S4_{k-1}=1) \text{ lub } (S1_{n,k}=1 \text{ i } S4_{k-1}=-1) \\ S4_{k-1} & \text{gdy nie są spełnione powyższe warunki} \end{cases}$ | n, r, s | - | - |
| 104 | Stochastic Slow (5) S5 | $S5_k = \begin{cases} +1 & \text{gdy } \tilde{E}_k^r > s \text{ i } \tilde{E}_{k-1}^r \leq s \\ -1 & \text{gdy } \tilde{E}_k^r < 100-s \text{ i } \tilde{E}_{k-1}^r \geq 100-s \\ 0 & \text{gdy } (S5_{k-1}=1 \text{ i } \tilde{E}_k^r < \tilde{E}_{k-1}^r) \text{ lub } (S5_{k-1}=-1 \text{ i } \tilde{E}_k^r > \tilde{E}_{k-1}^r) \\ S5_{k-1} & \text{gdy nie są spełnione powyższe warunki} \end{cases}$ | r, s | + | - |
| 105 | Stochastic Slow (6) S6 | $S6_k = \text{sig}_{15}(\tilde{E}_k^r)$ | r | + | 0 |
| 106 | Nachylenie trendu (1) | Współczynnik a_h nachylenia prostej regresji liniowej dla szeregu X, obliczony według wzoru (4.12), w którym należy przyjąć: $p(i) = x_{k-h+i}$, gdzie h (parametr) oznacza długość przedziału czasowego dla wyznaczania regresji. | h | + | 0 |
| 107 | Nachylenie trendu (2) | Formuła jak powyżej (<i>przyjęto inny typ normalizacji</i>) | h | + | 1 |
| 108 | Odchylenie kursu od prostej regresji | $x_k - y_{h,k}$ gdzie $y_{h,k}$ jest rzędną punktu prostej regresji, wyznaczonej jak w procedurze 106, dla chwili k i przedziału czasowego h | h | + | 0 |
| 109 | Różnica nachyleń trendów | $a_{h/2} - a_h$ (współczynnik a_i jest zdefiniowany w procedurze 106) | h | + | 0 |
| 110 | Różnica średniej adaptacyjnej i wartości regresji | $A_{h,k}(X) - y_{h,k}$ gdzie $y_{h,k}$ jest zdefiniowane w procedurze 108 | h | + | 1 |

Źródło: opracowanie własne.

Objaśnienia dotyczące tabeli 4-1:

Kolejne kolumny tabeli zawierają:

- numer kolejny procedury,
- przyjętą nazwę procedury (często nazwa ta jest skojarzona z odpowiednim wskaźnikiem analizy technicznej, ale formuła obliczania wartości tego wskaźnika może być tutaj nieco zmodyfikowana),

- formułę, według której wyznaczana jest wartość wynikowa dla danej procedury; niekiedy zamieszczona jest również druga formuła poprzedzona symbolem **D:**, wówczas ma ona zastosowanie dla szeregu czasowego D (czyli $d(t)$) określonego wzorem (4.1),
- wykaz parametrów procedury,
- kolumna D - informację, czy dana procedura może być wykorzystana do przetwarzania szeregu czasowego $d(t)$ określonego wzorem (4.1) (znak $+$), czy też nie jest w tym celu używana (znak $-$); jeżeli szereg $d(t)$ może być wykorzystany, to jest on wtedy podstawiany na miejsce szeregu X w odpowiedniej formule w kolumnie trzeciej, chyba że podano dla niego odrębną formułę poprzedzoną symbolem **D:**.
- kolumna N - typ normalizacji wartości wynikowych dla danej procedury; przyjęto następujące oznaczenia: **0** - normalizacja bez przesunięcia zera (0-offset), **1** - standaryzacja, „-” - brak normalizacji; (zob. podrozdział 4.3.2).

Oznaczenia literowe parametrów:

- **k** - moment czasu (numer sesji giełdowej), w którym wyznaczana jest wartość wynikowa procedury,
- **n, m** - okresy użytych średnich kroczących (jeżeli występują oba symbole, to $m > n$),
- **r** - długość okna czasowego, lub długość kroku wstecz, przy wyznaczaniu wybranych wielkości
- **s** - poziom linii sygnałowych wykupienia / wysprzedania w wybranych oscylatorach
- **h** - długość przedziału czasowego przy wyznaczaniu prostej regresji (trendu) dla szeregu **X**

Oznaczenia literowe szeregów czasowych:

- **D** = $(d_1, d_2, \dots, d_k, \dots)$ - szereg czasowy $d(t)$ określony formułą (4.1)
- **X** = $(x_1, x_2, \dots, x_k, \dots)$ - wejściowy szereg czasowy dla danej procedury, standardowo będący szeregiem kursów danej akcji; jeżeli jednak dozwolone jest zastosowanie szeregu **D**, czyli szeregu $d(t)$ określonego wzorem (4.1) (symbol „+” w kolumnie D) i nie określono dla tej opcji odrębnej formuły (poprzedzonej symbolem **D:**), oznaczenie **X** może się odnosić także do szeregu **D**.
- **VOL** = $(vol_1, vol_2, \dots, vol_k, \dots)$ - szereg reprezentujący wolumen obrotów dla danej akcji w sztukach
- **WIG** = $(wig_1, wig_2, \dots, wig_k, \dots)$ - szereg wartości indeksu giełdowego WIG
- **VOLWIG** = $(volwig_1, volwig_2, \dots, volwig_k, \dots)$ - szereg wielkości obrotu dla rynku podstawowego
- **G** = $(g_1, g_2, \dots, g_k, \dots)$ - szereg pomocniczy, będący określoną transformacją szeregu **D**, **X**, lub **VOL**

Oznaczenia wartości średnich kroczących (zob. podrozdział 2.7.5):

- $S_{n,k}(X)$ - wartość n -dniowej zwykłej średniej kroczącej w chwili k (formuła (2.16)),

- $E_{n,k}(X)$ - wartość n -dniowej wykładniczej średniej kroczącej w chwili k (formuła (2.17)),
- $A_{n,k}(X)$ - wartość n -dniowej adaptacyjnej średniej kroczącej w chwili k (formuła (2.18)),

Definicje funkcji specjalnych:

- $\text{sig}_p(g_k) = g_k - E_{p,k}(G)$

$$\bullet \text{imp}_p(g_k) = \begin{cases} +1 & \text{gdy } g_k > 0 \text{ i } \text{imp}_p(g_{k-1}) \leq 0 \\ -1 & \text{gdy } g_k < 0 \text{ i } \text{imp}_p(g_{k-1}) \geq 0 \\ p \cdot \text{imp}_p(g_{k-1}) & \text{gdy w / w nie są spełnione} \end{cases} \quad \text{przy czym: } \begin{cases} \text{imp}_p(g_1) = 0 \\ 0 < p < 1 \end{cases}$$

$$\bullet \text{ww}_p^q(g_k) = \begin{cases} +1 & \text{gdy } g_k \geq -s_{q,k} \text{ i } g_{k-1} < -s_{q,k-1} \\ -1 & \text{gdy } g_k \leq +s_{q,k} \text{ i } g_{k-1} > +s_{q,k-1} \\ p \cdot \text{ww}_p^q(g_{k-1}) & \text{gdy w / w nie są spełnione} \end{cases} \quad \text{przy czym: } \begin{cases} \text{ww}_p^q(g_1) = 0 \\ 0 < p < 1 \end{cases}$$

gdzie $s_{q,k} = \frac{q}{100} \cdot \sigma(g_i; i \leq k)$ jest wielkością q % odchylenia standardowego szeregu G obliczonego dla wartości tego szeregu nie większych od k .

- $\text{up}_r(g_k) = \sum_{i=k-r+1}^k (g_i; g_i > 0)$ - suma *dodatnich* realizacji szeregu G z okna długości r
- $\text{down}_r(g_k) = - \sum_{i=k-r+1}^k (g_i; g_i < 0)$ - minus suma *ujemnych* realizacji szeregu G z okna dł. r
- $\text{sz}(g_i) = \frac{g_i - g_{i-1}}{g_{i-1}}$ - dzienna stopa zwrotu wyznaczona dla finansowego szeregu G w chwili i

$$\bullet \tilde{E}_i^r = E_{5,i} \left(100 \cdot \frac{x_i - \min_{i-r < j \leq i} (x_j)}{\max_{i-r < j \leq i} (x_j) - \min_{i-r < j \leq i} (x_j)} \right)$$

Możliwość zastosowania różnych zestawów parametrów dla procedur wejściowych pozwala na znaczne zwielokrotnienie liczby możliwych zmiennych wejściowych dla badanych sieci neuronowych, pozwalając przy tym na istotnie większą elastyczność konstrukcji systemu decyzyjnego oraz na zwiększenie potencjału przetwarzania informacji w tym systemie. Z drugiej jednak strony bardziej znaczący staje się dylemat doboru zmiennych wejściowych dla sieci neuronowej, polegający już nie tylko na samym wyborze danych pochodzących z odpowiednich procedur, ale

również na sprecyzowaniu właściwych parametrów dla tych procedur⁷³. Mając na uwadze przeprowadzoną w podrozdziale 2.7.6 dyskusję dotyczącą problematyki optymalnego doboru danych wejściowych dla sieci neuronowej, oraz przedstawione w podrozdziale 3.3 przesłanki zastosowania algorytmów genetycznych w tym procesie, zdecydowano się właśnie na zastosowanie metodologii opartej o AG w rozwiązywaniu tego zagadnienia. W dalszej części pracy zaproponowano zatem algorytm genetyczny dla optymalizacji doboru danych wejściowych dla SN (tj. procedur przetwarzania informacji wejściowej i ich parametrów).

4.3. Moduł preprocessingu danych wejściowych.

Zagadnienie preprocessingu⁷⁴ (wstępnego przetwarzania danych), oraz celowości zastosowania tej operacji dla SN, zostało dość ogólnie przeanalizowane w podrozdziale 2.7. Poniżej przedstawiono przyjęte w rozważanej metodologii operacje transformacji danych (otrzymanych z pierwszego modułu systemu decyzyjnego), realizowane przez drugi z kolei element toru przepływu informacji - moduł preprocessingu danych wejściowych.

4.3.1. Trymowanie (obcinanie) wartości nietypowych.

Wartości nietypowe w danym szeregu czasowym, tj. znacznie odchyłone od średniej, mogą negatywnie wpływać na efektywność procedury normalizacji, ponieważ zwiększają one nadmiernie cały zakres pierwotnych wartości zmiennej wejściowej (zob. podrozdział 2.7.8). W związku z tym celowe jest „przybliżenie” elementów nietypowych do wartości średniej szeregu. Przyjęto następujące reguły trymowania⁷⁵:

$$\begin{array}{l}
 \text{Jeżeli } \mathbf{x} > \mathbf{x}_{\text{sr}} + \gamma_{\text{tr}} \cdot \sigma \quad \text{to} \quad \mathbf{x} \leftarrow \mathbf{x}_{\text{sr}} + \gamma_{\text{tr}} \cdot \sigma \\
 \text{Jeżeli } \mathbf{x} < \mathbf{x}_{\text{sr}} - \gamma_{\text{tr}} \cdot \sigma \quad \text{to} \quad \mathbf{x} \leftarrow \mathbf{x}_{\text{sr}} - \gamma_{\text{tr}} \cdot \sigma
 \end{array} \quad (4.3)$$

⁷³ Zasady budowy systemu dopuszczają możliwość wielokrotnego użycia na wejściach jednej sieci danych otrzymanych z tej samej procedury, ale wywołanej z różnymi parametrami.

⁷⁴ Pojęcie „preprocessing” w terminologii SN jest używane (również w polskojęzycznej literaturze) dla szeroko pojętego wstępnego przetwarzania danych dla sieci; tutaj pojęcie to nieco zawężimy ograniczając je do drugiego etapu przetwarzania danych, w odróżnieniu od transformacji informacji dokonywanej w pierwszym module systemu, omówionym w podrozdziale 4.2.

⁷⁵ Ang. *outliers trimming*.

gdzie: x - analizowana wartość zmiennej wejściowej
 x_{sr} - średnia wszystkich wartości danej zmiennej
 σ - odchylenie standardowe dla danej zmiennej
 γ_{tr} - parametr trymowania.

4.3.2. Normalizacja zmiennych wejściowych.

Uwzględniając uzasadnioną wcześniej potrzebę normalizacji zmiennych wejściowych, dla każdej z nich przyjęto jedną z następujących opcji normalizacji:

- **normalizacja „bez przesunięcia zera” (zero offset)** - polegająca na liniowym przeskalowaniu wartości zmiennej wejściowej w taki sposób, aby zachowany został znak każdej wartości, zaś element o największym module został przeskalowany do wielkości +1 lub -1; operacja taka jest celowa w przypadku, gdy istotny jest znak wartości wejściowej (generowanej przez określoną procedurę wstępnego przetwarzania informacji), jest ona wyrażona przypisaniem:

$$x \leftarrow x / \mathbf{amax} \quad (4.4)$$

gdzie: \mathbf{amax} - największa wartość bezwzględna danej zmiennej wejściowej

- **standaryzacja** - polegająca na odjęciu od każdej wartości średniej arytmetycznej dla danej zmiennej wejściowej, a następnie na podzieleniu wyniku przez dzielnik, stanowiący określoną wielokrotność odchylenia standardowego dla tej zmiennej; nie musi zostać zachowany znak wartości wejściowej, wartości znormalizowane mogą również wykraczać poza przedział $[-1, 1]$; operacja ta opisana jest kolejno realizowanymi przekształceniami (oznaczenia jak wyżej):

$$x \leftarrow x - x_{sr} \quad (4.5)$$

$$x \leftarrow x / (\gamma_{st} \cdot \sigma)$$

gdzie: γ_{st} - parametr przekształcenia

- **normalizacja „bez przesunięcia zera” po uprzednim odjęciu średniej** - jest kombinacją wymienionych wyżej dwóch metod, polega na odjęciu od każdej wartości - średniej arytmetycznej danej zmiennej, obliczeniu dla tak otrzymanych

liczb maksymalnej wartości bezwzględnej - **amax** i następnie na wykonaniu dzielenia przez **amax** ; operacje te można wyrazić następująco:

$$\mathbf{x} \leftarrow \mathbf{x} - \mathbf{x}_{\text{sr}} \quad (4.6)$$

$$\mathbf{x} \leftarrow \mathbf{x} / \mathbf{amax}$$

Ten typ normalizacji jest stosowany m.in. przy realizacji analizy głównych składowych jako metody wstępnej transformacji danych wejściowych. Znormalizowane wielkości posiadają w tym przypadku zerową wartość średnią i zawierają się w przedziale $[-1, 1]$.

- **brak normalizacji** - jest uzasadniony tylko w przypadku, gdy zmienna wejściowa, generowana przez określoną procedurę wstępnego przetwarzania informacji, posiada z natury wartości należące do zbioru odpowiadającego przybliżonemu zakresowi innych zmiennych znormalizowanych; podczas tworzenia opisywanej metodologii za zbiór ten uznano przedział $[-1, 1]$. Na przykład wewnątrz modułu przetwarzania informacji wejściowej zaproponowano szereg opartych na pojedynczych wskaźnikach technicznych procedur, generujących konkretne sygnały kupna/sprzedaży (rekomendacje decyzji inwestycyjnych) wyrażone wartościami: **+1** dla sygnału kupna, **-1** dla sygnału sprzedaży⁷⁶. Tego typu wartości nie wymagają normalizacji przed podaniem ich na wejście SN.

Typ normalizacji określono dla każdej procedury wejściowej (typu zmiennej wejściowej) *a priori* i wyszczególniono w tabeli 4-1.

4.3.3. Skokowe rozdzielenie wartości dodatnich i ujemnych.

Ten typ przetwarzania danych wejściowych ma zastosowanie tylko dla tych zmiennych, dla których zastosowano normalizację „bez przesunięcia zera”, tj. w przypadku, gdy istotną informację dla sieci stanowi znak danej zmiennej. Operacja ta polega na liniowym odwzorowaniu przedziału wartości $(0, 1]$ na przedział $(\gamma_{\text{rw}}, 1]$, oraz przedziału $[-1, 0)$ na przedział $[-1, -\gamma_{\text{rw}})$, gdzie γ_{rw} ($0 < \gamma_{\text{rw}} < 1$) jest określoną

⁷⁶ Możliwe są też wartości: 0 - brak decyzji, a także niekiedy szereg wartości pośrednich z przedziału $(-1, 1)$, oznaczających odpowiednio osłabione sygnały kupna/sprzedaży.

liczbą (parametrem). Przesłanki dla zastosowania tego przekształcenia wynikają z faktu, że sygnały wejściowe bliskie zera są mało znaczące dla SN (mają niewielki wpływ na sygnał wyjściowy). Skoro zatem znak zmiennej stanowić ma informację znaczącą, to zmiana tego znaku powinna być wyraźnie „zauważona” przez sieć; po zastosowaniu wyżej opisanej transformacji zmiana znaku wiąże się ze skokową zmianą danego sygnału wejściowego o przedział nie mniejszy niż $2 \cdot \gamma_{rw}$.

Przekształcenie skokowego rozdzielenia wartości dodatnich i ujemnych opisane jest zależnościami:

$$\begin{aligned} \mathbf{x} &\leftarrow \mathbf{x} \cdot (1 - \gamma_{rw}) + \gamma_{rw} && \text{dla } \mathbf{x} > \mathbf{0} \\ \mathbf{x} &\leftarrow \mathbf{x} \cdot (1 - \gamma_{rw}) - \gamma_{rw} && \text{dla } \mathbf{x} < \mathbf{0} \end{aligned} \quad (4.7)$$

4.3.4. Zastosowanie antysymetrycznej funkcji potęgowej.

W wielu przypadkach można poprawić efektywność funkcjonowania SN poprzez dodatkową transformację wybranych zmiennych wejściowych w oparciu o funkcję potęgową (antysymetryczną). W omawianej metodologii stworzono zatem opcję takiego przekształcenia danych wejściowych, zakładając przy tym możliwość doboru różnych wartości parametru δ - wykładnika potęgi - dla różnych typów normalizacji zmiennych. Transformacja tego typu wyrażona jest formułą:

$$\mathbf{x} \leftarrow \begin{cases} \mathbf{x}^\delta & \text{dla } \mathbf{x} \geq 0 \\ -(-\mathbf{x})^\delta & \text{dla } \mathbf{x} < 0 \end{cases} \quad (4.8)$$

gdzie wykładnik δ ($\delta > 0$) jest parametrem przekształcenia.

4.3.5. Analiza głównych składowych.

Analiza głównych składowych jest stosowana jako metoda wstępnego przetwarzania danych wejściowych (zob. podrozdział 2.7.7) w celu redukcji ilości wejść SN, oraz dekorelacji zmiennych wejściowych (Azoff, 1994). W prezentowanej metodologii zastosowano analizę głównych składowych jako opcjonalne narzędzie służące do dokonania transformacji wektorów wejściowych według przekształcenia liniowego

opisanego zależnością (2.20). W celu wyznaczenia macierzy \mathbf{V} tego przekształcenia wykorzystano metodę opartą o samouczącą się sieć neuronową typu PCA (zob. Haykin, 1994; Osowski, 1996; Grabowski, 1997).

Sieć PCA jest w istocie liniową siecią jednowarstwową o n elementach w warstwie wejściowej i k liniowych neuronach w warstwie wyjściowej, przetwarzającą informację właśnie według formuły (2.20). W tej sieci wektory wag związane z poszczególnymi neuronami wyjściowymi stanowią kolejne wiersze macierzy przekształcenia liniowego $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]^\top$. Wagi te wyznaczone są w iteracyjnym procesie uczenia, w oparciu o pojawiające się na wejściu sieci kolejne elementy ciągu uczącego - wektory $\mathbf{x}(t)$. Do najczęściej stosowanych algorytmów uczenia takiej sieci (czyli wyznaczania współczynników macierzy \mathbf{V}) należy reguła Sangera (zwana również regułą GHA), będąca pewną modyfikacją powszechnie stosowanej w uczeniu nienadzorowanym reguły Hebba. Reguła Sangera wyrażona jest wektorową zależnością⁷⁷ (Osowski, 1996):

$$\mathbf{v}_i(t+1) = \mathbf{v}_i(t) + \eta(t) \cdot y_i(t) \cdot \mathbf{x}'(t) - y_i(t) \cdot \mathbf{v}_i(t) \quad (4.9)$$

gdzie:
$$y_i(t) = \mathbf{v}_i(t) \circ \mathbf{x}(t) = \sum_{j=1}^n v_{ij}(t) x_j(t)$$

$$\mathbf{x}'(t) = \mathbf{x}(t) - \sum_{h=1}^{i-1} \mathbf{v}_h(t) y_h(t)$$

$$i = 1, 2, \dots, k$$

Wykazano (Haykin, 1994), że jeżeli $\mathbf{x}(t)$ jest wektorem losowym o zerowej wartości średniej, to przy założeniach dotyczących współczynnika uczenia $\eta(t)$:

$$\lim_{t \rightarrow \infty} \eta(t) = 0 \quad \text{oraz} \quad \sum_{t=1}^{\infty} \eta(t) = \infty$$

⁷⁷ Wektory oznaczono symbolami pogrubionymi.

wektory wag $\mathbf{v}_1(t), \mathbf{v}_2(t), \dots, \mathbf{v}_k(t)$ zainicjowane losowymi wartościami początkowymi i modyfikowane w trakcie procesu uczenia według iteracyjnej reguły (4.9), będą odpowiednio dążyć przy $t \rightarrow \infty$ do k pierwszych wektorów własnych macierzy kowariancji wektora losowego \mathbf{x} , a zatem wyznaczą k pierwszych głównych składowych tego wektora.

Neuronowa implementacja procedury transformacji wektorów wejściowych opartej o analizę głównych składowych, realizowana w module preprocessingu danych wejściowych, obejmuje:

- normalizację wszystkich zmiennych wejściowych, w taki sposób, aby zapewnić spełnienie warunku zerowej wartości średniej każdej zmiennej - tj. normalizację „bez przesunięcia zera” po uprzednim odjęciu średniej według zależności (4.6); znormalizowane w ten sposób wszystkie dostępne wektory wejściowe (oznaczymy ich ilość symbolem \mathfrak{Q}) stanowiąc będą ciąg uczący dla sieci PCA,
- określenie początkowego współczynnika uczenia $\eta(1)$, zasad dokonywania zmian wartości tego współczynnika w trakcie procesu uczenia, zdefiniowanie liczby epok uczenia E , oraz dokonanie losowej inicjalizacji wag,
- realizację algorytmu uczenia sieci PCA złożonego z następujących kroków:
 - Krok 1.** Podstaw $e = 1, t = 1$;
 - Krok 2.** Dokonaj randomizacji⁷⁸ ciągu uczącego, podstaw $p = 1$;
 - Krok 3.** $\mathbf{x}(t) \leftarrow p$ -ty element ciągu uczącego;
 - Krok 4.** Dokonaj modyfikacji wag według reguły (4.9);
 - Krok 5.** $p \leftarrow p + 1$, $t \leftarrow t + 1$, jeżeli $p \leq \mathfrak{Q}$ to przejdź do **kroku 3**;
 - Krok 6.** $e \leftarrow e + 1$, jeżeli $e \leq E$ to przejdź do **kroku 2**.
- transformację wszystkich \mathfrak{Q} wektorów wejściowych w oparciu o nauczoną sieć PCA, tzn. według zależności (2.20);

⁷⁸ Randomizacja ciągu polega na losowym mieszanii elementów tego ciągu.

- opcjonalną normalizację k -wymiarowych wektorów otrzymanych w wyniku powyższego przekształcenia; aby nie zmienić zerowej korelacji pomiędzy poszczególnymi k głównymi składowymi, stosowana jest przy tej operacji normalizacja „bez przesunięcia zera” (*zero offset*) poszczególnych zmiennych (głównych składowych), która nie zmienia właściwości wzajemnej dekorrelacji tych zmiennych.

Otrzymane w wyniku wyżej opisanej transformacji k -wymiarowe wektory stanowią będą właściwe wzorce wejściowe dla sieci typu *perceptron*, stosowanych do predykcji określonych wskaźników rynkowych.

Podczas prowadzenia badań zgodnie z przyjętym schematem napotkano na kilka problemów, które wymagają rozwiązania. Istotnym zagadnieniem w procesie uczenia sieci PCA okazał się dobór współczynnika uczenia η oraz metody dokonywania jego zmian w czasie. Nie istnieją bowiem uniwersalne reguły określania tego parametru, powszechnie uznawana przez badaczy jest jedynie konieczność stopniowego zmniejszania jego wartości w trakcie procesu uczenia. Ponadto należy zaznaczyć, że dobór zbyt małych wielkości η powoduje znaczne spowolnienie uczenia sieci PCA, z kolei stwierdzono, że zbyt duże wartości tego współczynnika mogą prowadzić do destabilizacji procesu uczenia i do braku jego zbieżności. W proponowanej metodologii, korzystając z metody prób i błędów przyjęto ostatecznie następującą - dającą dobre efekty w sensie zbieżności i szybkości uczenia i spełniającą założenia teoretyczne dla metody Sangera - regułę doboru i zmian parametru η w zależności od kolejnego numeru epoki uczenia e ($e = 1, 2, 3, \dots$) oraz liczby zmiennych wejściowych (neuronów w warstwie wejściowej) n :

$$\eta(e) = \frac{1}{n \cdot \sqrt{e}} \quad (4.10)$$

przy czym $\eta(e)$ oznacza wartość współczynnika uczenia η podczas e -tej epoki.

Wypraktkowano też, że dobre wyniki daje początkowa inicjalizacja wag sieci PCA wartościami liczbowymi z przedziału $[-0,2, 0,2]$, oraz przyjęto, że liczba epok uczenia E jest parametrem systemu.

4.4. Przyjęta architektura sieci neuronowych typu *perceptron*.

W prezentowanym systemie decyzyjnym zastosowano sieci neuronowe typu *perceptron*, używane jako zasadnicze elementy robocze zaangażowane w proces generowania prognoz, będących następnie podstawą dla decyzji kupna lub sprzedaży określonych instrumentów finansowych.

Przyjęto następujące założenia dotyczące architektury wykorzystywanych sieci:

- ilość warstw ukrytych: jedna (z opcją - bez warstw ukrytych),
- liczba neuronów:
 - * w warstwie wejściowej - równa liczbie zmiennych wejściowych, określonej w module przetwarzania informacji wejściowej i doboru danych wejściowych, względnie w module preprocessingu danych wejściowych (w przypadku zastosowania PCA),
 - * w warstwie wyjściowej: jeden,
 - * w warstwie ukrytej: wyznaczana arbitralnie lub genetycznie z przedziału $[2 , 2k+1]$, gdzie: k - liczba wejść sieci.
- charakter połączeń: jednokierunkowy (*feedforward*) według schematu *każdy z każdym* (zob. podrozdział 2.4),
- model neuronu w warstwie ukrytej i wyjściowej: nieliniowy z sigmoidalną funkcją aktywacji, należąca do klasy I (zob. podrozdział 2.3), określoną formułą (2.4).

4.5. Modele i normalizacja zmiennej wyjściowej dla sieci *perceptron*.

Zdefiniowanie typu zmiennej wyjściowej dla sieci neuronowej (oraz związany z tym dobór wartości wzorcowych w ciągu ucącym) jest zagadnieniem o kluczowym znaczeniu (zostało ono wstępnie omówione w podrozdziale 2.7.10). W sytuacji gdy ostatecznym celem systemu jest generowanie decyzji kupna lub sprzedaży określonego instrumentu finansowego, za zmienną wyjściową należy niewątpliwie przyjąć odpowiedni wskaźnik charakteryzujący dynamikę kursu cenowego tego

instrumentu tak, aby prognoza wartości tego wskaźnika mogła stanowić bezpośrednio podstawę do podjęcia odpowiedniej decyzji transakcyjnej. Przy konstrukcji tego typu wskaźnika należy także uwzględnić celowość wykorzystania takich form przetwarzania danych (np. szeregu czasowego kursów danej akcji), jak regresja liniowa oraz średnia krocząca⁷⁹. Operacje te powodują wstępną eliminację szumów i mogą przyczynić się do poprawy efektywności procesu uczenia SN.

W celu zdefiniowania zmiennej wyjściowej dla SN rozważymy dwa szeregi czasowe: kursu danego instrumentu (ceny akcji), oraz wartości indeksu giełdy WIG, oznaczone analogicznie jak w formule (4.1): $kurs(t)$ oraz $WIG(t)$. Oznaczmy przez h - horyzont prognozy dla wybranego wskaźnika, zaś przez $y(t)$ - wartość zmiennej wyjściowej w chwili t . Zaproponowane w omawianym systemie decyzyjnym opcje doboru zmiennej wyjściowej dla SN zaprezentowano kolejno w podrozdziałach od 4.5.1 do 4.5.4. Podrozdział 4.5.5 poświęcono zagadnieniu końcowego przekształcenia (normalizacji) wartości tej zmiennej.

4.5.1. Względna zmiana kursu.

Prostą koncepcję wyznaczenia zmiennej wyjściowej stanowi przyjęcie względnych zmian kursu danego instrumentu finansowego w okresie h według zależności (4.11):

$$y(t) = \frac{kurs(t+h) - kurs(t)}{kurs(t)} \quad (4.11)$$

4.5.2. Trend względnych zmian kursu.

Niech $p(i)$ oznacza względną i -dniową zmianę kursu danego instrumentu, określoną formułą:

$$p(i) = \frac{kurs(t+i) - kurs(t)}{kurs(t)} \quad \text{gdzie} \quad i = 1, 2, \dots, h$$

⁷⁹ W podrozdziale 4.5.3 zaproponowano np. użycie średniej adaptacyjnej, która charakteryzuje się praktycznie brakiem opóźnienia w stosunku do wartości analizowanego szeregu czasowego (zob. podrozdział 2.7.5).

Trend względnych zmian kursu⁸⁰, określony w okresie h , zdefiniujemy jako współczynnik kierunkowy prostej regresji wyznaczonej dla wartości $p(1)$, $p(2)$, ..., $p(h)$ zgodnie ze wzorem (4.12):

$$y(t) = \frac{\left(\sum_{i=1}^h i\right) \cdot \left(\sum_{i=1}^h p(i)\right) - h \cdot \sum_{i=1}^h ip(i)}{\left(\sum_{i=1}^h i\right)^2 - h \cdot \sum_{i=1}^h i^2} \quad (4.12)$$

4.5.3. Najbliższa istotna względna zmiana średniej adaptacyjnej z kursu.

Poniżej zdefiniujemy (najpierw formalnie, a następnie opisowo) sposób wyznaczania wartości zmiennej wyjściowej dla SN, którą będziemy interpretować jako najbliższą istotną względną zmianę średniej adaptacyjnej z kursu. Rozważać będziemy przy tym średnią 5-dniową (por. podrozdział 2.7.5).

Niech $a(i)$ (gdzie $i = 0, 1, 2, \dots, h$) oznacza wartość średniej adaptacyjnej $A_{5,t+i}(kurs)$ wyznaczonej według zależności (2.18) dla kursu cenowego (tzn. \mathbf{x}_k we wzorach (2.17), (2.18) jest wartością $kurs(k)$). Niech $g = a(1) \cdot (1+q/100)$, $d = a(1) \cdot (1-q/100)$ oznaczają poziomy, których przekroczenie uważane jest za istotną zmianę średniej adaptacyjnej z kursu⁸¹, gdzie q jest parametrem reprezentującym wymaganą procentową zmianę średniej. Niech m będzie taką najmniejszą liczbą z zakresu

⁸⁰ W wielu opracowaniach przyjmuje się jako zmienną wyjściową trend samego kursu cenowego, tzn. współczynnik kierunkowy prostej regresji wyznaczonej dla wartości $kurs(t+i)$, $i=1, 2, \dots, h$. Tak zdefiniowana zmienna wyjściowa może być stosowana, jeżeli w okresie obejmującym dane uczące kurs jest w przybliżeniu stabilny. W sytuacji jednak, gdy w rozważanym okresie kurs znacznie się zmienił (np. z początkowej wartości 100 do końcowej 500), wpłynie to zasadniczo na wartości współczynnika kierunkowego prostej regresji i tym samym spowoduje istotnie zmieniający się w czasie zakres wahań zmiennej wyjściowej (wartości wzorcowej) w ciągu uczącym. Spowoduje to zniekształcenie procesu uczenia sieci i zmniejszenie jej przydatności w praktyce, gdyż strategię decyzyjną należy oprzeć raczej o względne a nie bezwzględne zmiany cen. Podobny negatywny efekt osiągniemy także w dopuszczalnym w zaproponowanej metodologii przypadku, polegającym na wykorzystaniu do uczenia określonej sieci *wielu* szeregów czasowych kursów akcji, a więc danych pochodzących z wielu spółek, różniących się niekiedy znacznie bezwzględnymi cenami akcji. Tak więc z przedstawionych wyżej powodów zdecydowano się do wyznaczania trendu przyjąć wartości względnych zmian kursu

⁸¹ Należy zauważyć, że punktem odniesienia dla wartości g i d jest wielkość średniej kursu z dnia następnego, tzn. prawdopodobna cena, po której ewentualnie dokonamy transakcji.

$m = 2, 3, \dots, h$, że $a(m) > g$ lub $a(m) < d$. Wyznaczymy pomocniczą wartość liczbową z następująco:

- jeżeli m nie istnieje, to: $z = a(1)$

- jeżeli $a(m) > g$, to: $z = \max_{h \geq i \geq m} \{ a(i) : \forall_{(j,k): m \leq j < k < i} a(k) \geq a(j) \cdot (1 - q / 100) \}$

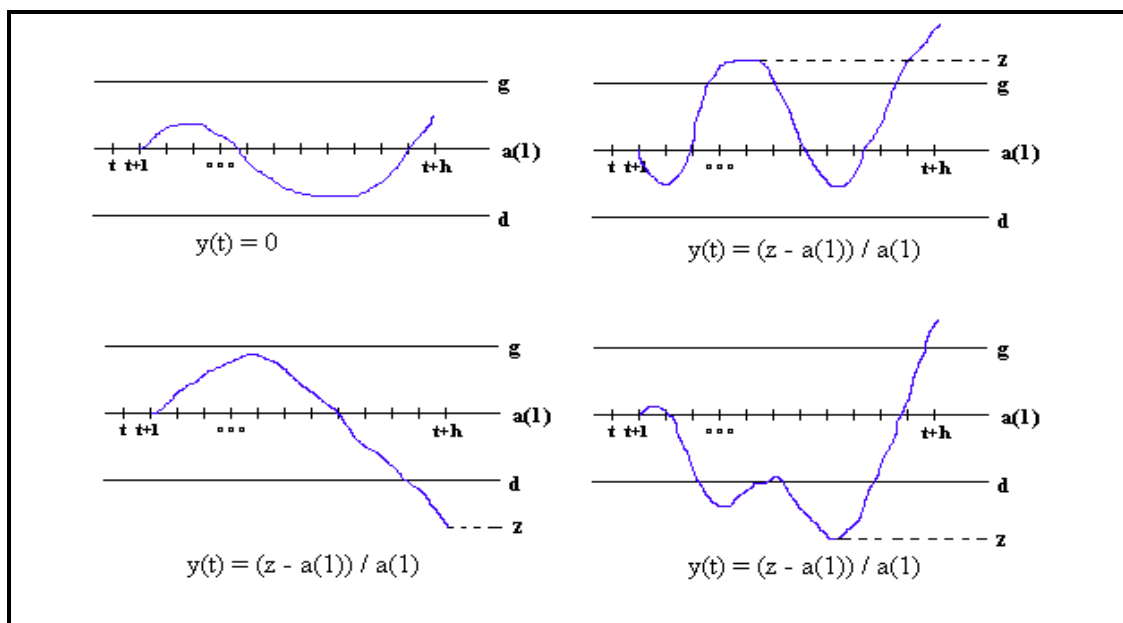
- jeżeli $a(m) < d$, to: $z = \min_{h \geq i \geq m} \{ a(i) : \forall_{(j,k): m \leq j < k < i} a(k) \leq a(j) \cdot (1 + q / 100) \}$

Wartość wyjściową $y(t)$ określimy według formuły:

$$y(t) = \frac{z - a(1)}{a(1)} \quad (4.13)$$

Sformalizowaną powyżej definicję zmiennej wyjściowej $y(t)$ można zinterpretować następująco:

- jeżeli w okresie od $t+1$ do $t+h$ nie nastąpi istotna (q -procentowa) zmiana średniej adaptacyjnej kursu, to $y(t) = 0$;
- jeżeli pierwszą istotną zmianą w tym okresie będzie ponad q -procentowy wzrost średniej w stosunku do jej wartości w momencie $t+1$, to obliczamy w tym okresie maksimum tej średniej nie poprzedzone istotnym spadkiem, po czym za sygnał wyjściowy przyjmujemy względną zmianę średniej od chwili $t+1$ do tego maksimum;
- jeżeli pierwszą istotną zmianą będzie ponad q -procentowy spadek średniej w stosunku do jej wartości w momencie $t+1$, to postępujemy analogicznie obliczając jednak minimum średniej nie poprzedzone istotnym jej wzrostem.



Rys. 4-3. Przykłady obliczania wartości zmiennej wyjściowej $y(t)$ jako najbliższej istotnej względnej zmiany adaptacyjnej średniej kroczącej. Niebieska linia przedstawia przykładowe przebiegi tej średniej. Źródło: opracowanie własne

Tak zdefiniowana zmienna wyjściowa dla SN ma istotną zaletę, polegającą na tym, że w sposób bezpośredni może być zastosowana do generowania sygnałów transakcyjnych, gdyż w istocie samodzielnie określa ona pewną (intuicyjnie oczywistą) strategię inwestycyjną. Jeżeli bowiem wartością wyjściową (prognozowaną) będzie istotny (przy założonym parametrze q) wzrost względny kursu, to wartość ta stanowi swoisty sygnał kupna tym mocniejszy, im większa jest prognozowana zmiana kursu (przed potencjalnym spadkiem). Analogicznie można interpretować ujemną wartość zmiennej wyjściowej jako sygnał sprzedaży. Należy także podkreślić, że podstawowym zadaniem adaptacyjnej średniej kroczącej jest eliminacja zakłóceń występujących w przypadku wyjściowych wartości uczących.

Odpowiednie przykłady graficzne sposobu obliczania wartości pomocniczej z oraz właściwej wielkości $y(t)$ zaprezentowano na rysunku 4-3..

4.5.4. Trend wielkości różnic logarytmów kursu i indeksu.

Opcja ta definiuje zmienną wyjściową jako trend szeregu czasowego d opisanego formułą (4.1), rozumiany jako współczynnik kierunkowy prostej regresji wyznaczonej dla wartości $d(t+1)$, $d(t+2)$, ..., $d(t+h)$:

Wartość wyjściowa $y(t)$ jest wyrażona zależnością (4.12), dla której jednak wielkości $p(i)$ zdefiniujemy następująco:

$$p(i) = d(t+i) = \ln \left(\frac{\text{kurs}(t+i)}{\text{WIG}(t+i)} \right), \quad i = 1, 2, \dots, h \quad (4.14)$$

Ten typ zmiennej wyjściowej stosować można w przypadku generowania prognoz zmian kursu akcji danej spółki w stosunku do zmian indeksu giełdowego⁸². Prognozy tego typu można wykorzystać w procesie aktywnego zarządzania portfelem inwestycyjnym w celu wyboru do portfela tych akcji, dla których przewidywany wzrost kursu w porównaniu do zmiany indeksu jest największy.

4.5.5. Normalizacja wartości wzorcowych zmiennej wyjściowej.

Dysponując danymi historycznymi z pewnego okresu obejmującego punkty czasowe $t = t_1, t_2, \dots, t_b$, (t_b - chwila obecna), można w oparciu o przyjętą formułę wyznaczania wartości zmiennej wyjściowej dla SN zdefiniować wzorcowe wartości wyjściowe $z_i = y(t_i)$ dla $i = 1, 2, \dots, b-h$ (tzn. dla danego okresu skróconego o horyzont predykcji). Tak określone wartości z_i wymagają dalszych transformacji obejmujących:

- opcjonalne trymowanie (obcinanie) wartości nietypowych (ekstremalnych, tzn. znacznie odległych od średniej) według zależności (4.3),
- normalizację, polegającą na takim przekształceniu wartości z_i , aby znalazły się one wewnątrz zbioru wartości funkcji aktywacji neuronu wyjściowego SN; metody normalizacji zmiennej wyjściowej obejmują:

- standardową normalizację wyrażoną przekształceniem (2.21),
- normalizację „bez przesunięcia zera” (wzór (4.4)),
- normalizację „bez przesunięcia zera” po uprzednim odjęciu wartości średniej (formuła (4.6))

Należy zwrócić uwagę na opcjonalną w systemie operację odejmowania średniej wszystkich wyjściowych wartości wzorcowych (trzeci z powyższych typów normalizacji). Stosuje się ją wtedy, gdy istnieje potrzeba wyeliminowania - z wzorcowych danych uczących - składnika odpowiedzialnego za długoterminowy trend w analizowanym okresie. Okazuje się bowiem, że sieć neuronowa w trakcie uczenia „uznaje” ten składnik za informację istotną i uczy się go, a następnie w fazie realizacji generuje sygnały obciążone również tą wielkością. Efekt ten może okazać się niekorzystny z punktu widzenia przydatności sygnałów SN do realizacji określonej strategii inwestycyjnej, toteż w takich przypadkach wskazana jest jego eliminacja.

Zwróćmy także uwagę na fakt, iż znormalizowane wyjściowe wartości wzorcowe zawierają się w przedziale $[-1, 1]$, z kolei przy przyjętej we wzorze (2.4) wartości parametru $a = 1,15$ (zob. podrozdział 4.4), zbiór wartości funkcji aktywacji jest przedziałem $(-1,15 ; 1,15)$. W istocie często zaleca się (Azoff, 1994), aby zakres zmienności wartości wzorcowych był nieco mniejszy od przedziału wartości funkcji aktywacji neuronu wyjściowego, gdyż unika się w ten sposób efektu „nasylenia”, pojawiającego się dla wartości funkcji aktywacji zbliżających się do odpowiednich wartości asymptotycznych tej funkcji.

4.6. Uczenie i testowanie sieci neuronowej.

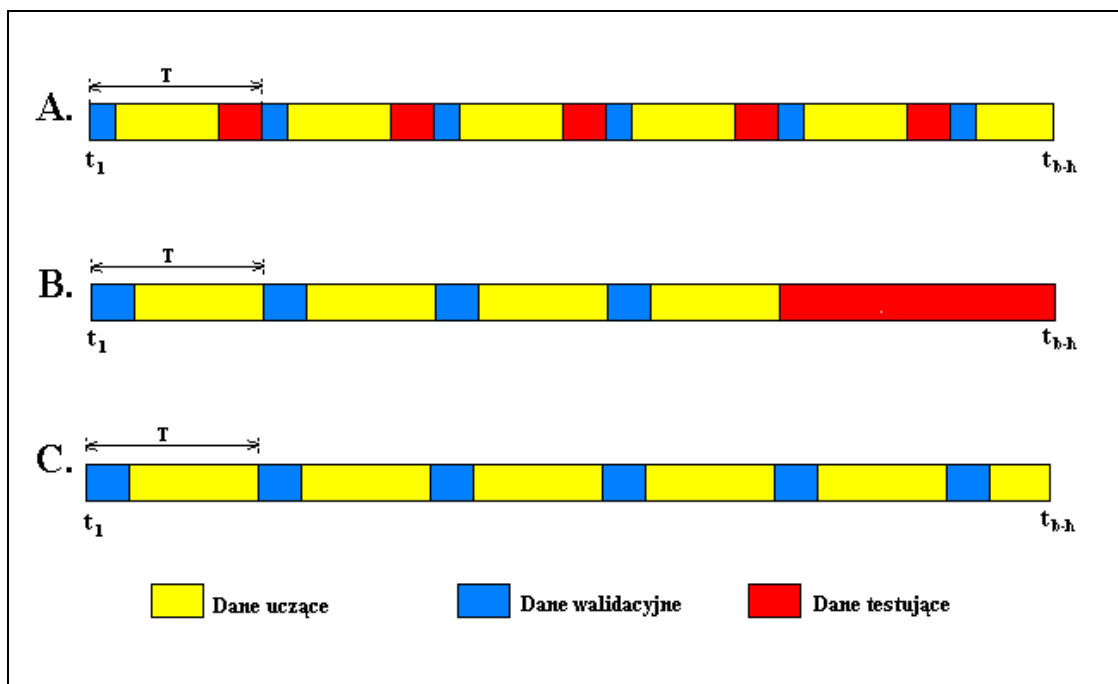
Po zrealizowaniu operacji doboru danych wejściowych i ich ewentualnego preprocessingu, opisanych w podrozdziałach 4.2 i 4.3, oraz po wyznaczeniu i normalizacji wzorcowych wartości wyjściowych (podrozdział 4.5) dysponujemy dla chwil czasowych $t = t_1, t_2, \dots, t_{b-h}$ ciągiem par (v_i, z_i) gdzie v_i oznacza wektor

⁸² por. zależność (4.2)

wejściowy, którego elementy są traktowane jako sygnały poszczególnych neuronów warstwy wejściowej SN, natomiast z_i oznacza znormalizowaną wzorcową wartość wyjściową, będącą poprawną prognozą dla wektora v_i ($i = 1, 2, \dots, b-h$). Ciąg ten należy teraz podzielić na trzy podciągi: uczący, walidacyjny i testujący (zob. podrozdział 2.7.9).

Jeżeli w skład powyższego ciągu wchodzi dane obejmujące pierwsze 50 dni notowań rozważanego instrumentu finansowego, nie zostają one włączone do żadnego z podciągów służących do uczenia i oceny sieci. Obszar tych danych należy bowiem traktować jako swoisty „stan nieustalony”, w którym mogą występować nietypowe zakłócenia związane ze specyfiką początkowych notowań giełdowych danego waloru, oraz w którym nie zostały jeszcze właściwie wyznaczone wartości niektórych średnich kroczących i innych wskaźników, opartych na danych z pewnego okresu przeszłości.

W omawianej tu metodologii dopuszczalne są różne warianty dokonania podziału danych na podciągi uczący, walidacyjny i testujący; zasadnicze metody zaprezentowano schematycznie na rys. 4-4. Wszystkie przedstawione na rys. 4-4 koncepcje podziału danych na poszczególne podciągi opierają się na wstępnym podziale całego okresu t_1, t_2, \dots, t_{b-h} na bloki o długości T (T jest parametrem systemu).



Rys. 4-4. Typowe schematy podziału ciągu danych na podciągi: uczący, walidacyjny i testujący. Źródło: opracowanie własne.

Schemat A polega na podziale każdego bloku na trzy obszary obejmujące dane walidacyjne, uczące i testujące (podział każdego bloku jest identyczny, z wyjątkiem ostatniego, zwykle krótszego bloku), o długościach odpowiednio T_v , T_u , T_t ($T_v + T_u + T_t = T$). Dobór parametrów T_v , T_u , T_t odbywa się w sposób arbitralny i ma istotny wpływ na efektywność funkcjonowania SN, zwróćmy bowiem uwagę, że przyjęcie zbyt małej wielkości T_u/T może spowodować za małą ilość elementów ciągu uczącego, zdefiniowanie zbyt małej proporcji T_v/T może prowadzić do niedostatecznej walidacji sieci i do efektu przeuczenia, z kolei przyjęcie zbyt małego ułamka T_t/T (w przypadku dokonywania końcowej oceny sieci w oparciu o ciąg testujący) prowadzi do niskiej wiarygodności oceny sieci. Duże znaczenie ma też długość bloku T . Zbyt duża wielkość tego parametru (np. określenie tylko jednego lub dwóch bloków) może prowadzić do sytuacji, gdzie na przykład dane w ciągu uczącym oraz dane w ciągu testującym pochodzą (w przeważającej części) z różnych etapów dynamiki rynku (np. sieć będzie się uczyła głównie na danych z okresu hossy, a testowana będzie na danych z okresu bessy). Z kolei zbyt małe T spowoduje niekorzystny efekt wynikający ze znaczącej autokorelacji kolejnych

elementów ciągu (w konsekwencji np. dane walidacyjne będą istotnie skorelowane z następującymi po nich danymi uczącymi, co pogorszy procedurę niezależnej walidacji i wywoła efekt przeuczenia).

W schemacie **B** przyjęto koncepcję umieszczenia ciągu testującego na samym końcu dostępnych danych w myśl idei testowania sieci na najbardziej aktualnych danych. Znaczącą rolę dla oceny i późniejszej efektywności sieci odgrywa przyjęta długość ciągu testującego. Z kolei schemat **C** dotyczy przypadku, gdy poprzestajemy na ocenie sieci dokonanej na podstawie ciągu walidacyjnego (i ewentualnie uczącego). W obu tych rozwiązaniach również istotny jest dobór parametrów T_v i T_u (aktualne są tu uwagi przedstawione w związku z zasadami ich doboru dla schematu **A**).

W proponowanej metodologii przyjmuje się także możliwość uczenia sieci na danych dotyczących kilku różnych instrumentów finansowych o podobnych charakterystykach (np. akcji spółek o podobnej kapitalizacji lub z tego samego sektora gospodarki). Koncepcja taka opiera się z jednej strony na idei zapewnienia odpowiednio licznego ciągu uczącego (Azoff, 1994; podrozdział 2.7.9), z drugiej strony na założeniu, że sieć posiadająca dobrą zdolność generalizacji będzie efektywnie funkcjonować w predykcji cen wielu akcji o podobnym charakterze dynamiki (toteż uczenie sieci można realizować przy wykorzystaniu danych z kilku spółek). Aby jednak zapewnić wiarygodność oceny sieci, nie należy doprowadzić do nakładania się na siebie obszarów danych uczących, walidacyjnych i testujących pochodzących z różnych spółek (inaczej mówiąc: obszary poszczególnych typów danych powinny być przyporządkowane tym samym chronologicznym okresom⁸³).

Uczenie sieci odbywa się według klasycznego algorytmu wstecznej propagacji błędów (zob. podrozdział 2.5), z randomizacją ciągu uczącego podczas każdej epoki uczenia. Z kolei wstępna ocena sieci (zob. podrozdział 2.7.11) jest dokonywana na bieżąco w oparciu o średnie wartości błędu walidacji i błędu uczenia, przypadające

⁸³ Kursy spółek giełdowych są silnie skorelowane, toteż jeżeli np. dane pochodzące z określonego okresu dla spółki X wykorzystamy do uczenia sieci, a dane z tego samego okresu, ale ze spółki Y - do testowania lub walidacji, to wzajemne korelacje między tymi danymi spowodują sztuczne zawyżenie (zbyt optymistyczne oszacowanie) bieżącej lub końcowej oceny sieci, co może skutkować przeuczeniem lub błędną (nadmiernie optymistyczną) oceną końcową sieci.

na jeden element ciągu walidacyjnego lub uczącego. W ogólnym przypadku bieżącą wstępną ocenę sieci Θ można wyrazić zależnością:

$$\Theta = \frac{p \cdot \frac{E_u}{i_u} + q \cdot \frac{E_v}{i_v}}{p + q}, \quad p \leq q \quad (4.15)$$

gdzie: E_u, E_v - aktualne wartości błędów odpowiednio uczenia i walidacji,
 i_u, i_v - ilości elementów w ciągu odpowiednio uczącym i walidacyjnym,
 p, q - przyjęte współczynniki, $p \leq q$, często przyjmuje się $p = 0$.

Bieżąca wstępna ocena sieci jest dokonywana tylko wtedy, gdy w danej epoce błąd uczenia E_u osiągnął swoje minimum. Z kolei warunkiem koniecznym przeprowadzenia korekty zapamiętanego najlepszego zestawu wag jest osiągnięcie w tej epoce minimalnej (najlepszej) oceny sieci Θ .

Podczas realizacji uczenia sieci bardzo istotne jest zagadnienie właściwego początkowego doboru oraz bieżącej modyfikacji współczynnika uczenia η oraz współczynnika bezwładności (*momentum*) μ . Jak wiadomo z literatury problem ten nie doczekał się uniwersalnych rozwiązań i wydaje się, że jego rozstrzygnięcie zależy w dużej mierze od specyfiki analizowanego zagadnienia. Poniżej zaprezentowano własną propozycję⁸⁴ zestawu reguł (R0, ..., R3) dokonujących zmian współczynników η oraz μ (*momentum*)⁸⁵ w oparciu o bieżącą analizę spadku błędu uczenia w kolejnych epokach uczenia SN.

Nazwijmy *sukcesem* zdarzenie polegające na poprawie (spadku) błędu uczenia w bieżącej epoce w stosunku do wartości tego błędu w poprzedniej epoce uczenia; natomiast *porażką* nazwijmy zdarzenie przeciwne, polegające na braku poprawy

⁸⁴ Prezentowane propozycje otrzymano stosując metodę „prób i błędów”, z uwzględnieniem licznych przesłanek wynikających z dotychczas przeprowadzonych własnych badań autora, dotyczących doboru tych współczynników, oraz z przedstawionych w literaturze wyników innych badaczy. Jak wypraktykowano, przedstawiona tutaj metoda daje dobre rezultaty w procesie uczenia SN w analizie finansowych szeregów czasowych.

⁸⁵ Współczynnik μ odnosi się do całej epoki uczenia, a nie do pojedynczej iteracji dla jednego elementu ciągu uczącego, tzn. określa długość składowej wektora korekty wag z całej epoki, równoległej do analogicznego wektora korekty wag z poprzedniej epoki.

błędu uczenia. Niech l_s oznacza liczbę *kolejnych* (następujących po sobie nieprzerwanie w kolejnych epokach) sukcesów, zaś l_p - liczbę *kolejnych* porażek (oczywiście zawsze jedna z tych dwóch wartości jest zerem). Przyjmijmy następujące, sprawdzane i realizowane po każdej epoce, iteracyjne reguły modyfikacji współczynników η i μ :

R0 : Początkowe wartości w pierwszej epoce: $\eta = 0,05 \quad \mu = 0,5$

R1 : Jeżeli $l_s \geq 2$ to $\mu \leftarrow \mu + \alpha$; ale jeżeli $\mu > 0,9$ to $\mu \leftarrow 0,9$

R2 : Jeżeli $l_s \geq 5$ to $\eta \leftarrow \eta + \alpha$

R3 : Jeżeli $l_p \geq 1$ to $\mu \leftarrow \mu \cdot \beta$ oraz $\eta \leftarrow \eta \cdot \beta$

Stałe α oraz β (spełniające warunki: $\alpha > 0$, $0 < \beta < 1$), określające wielkość modyfikacji współczynników η i μ , stanowią parametry systemu (algorytmu uczenia)⁸⁶.

Zwróćmy uwagę, że reguły R1 i R2 decydują o zwiększeniu obu współczynników w sytuacji, gdy następuje systematyczny spadek błędu uczenia i zachodzi podejrzenie, że proces uczenia mógłby być szybszy (bo aktualnie jest realizowany zbyt małymi krokami). Z kolei w przypadku kolejnych *porażek* zachodzi niebezpieczeństwo wystąpienia oscylacji wywołujących ryzyko utraty zbieżności procesu uczenia, toteż w tym przypadku reguła R3 powoduje odpowiednie zmniejszanie obu współczynników, co w konsekwencji powoduje skrócenie wektorów korekty wag w kolejnych iteracjach.

Ważnym parametrem procesu uczenia SN jest dobierana arbitralnie maksymalna ilość epok uczenia M . Jednak przed zrealizowaniem M epok uczenie może zostać przerwane w dwóch następujących przypadkach:

- nastąpi znaczne pogorszenie bieżącej wartości oceny sieci Θ (zob. wzór (4.15)) w stosunku do najlepszej dotychczas uzyskanej oceny Θ' ; dla tego przypadku zaproponowano następującą regułę, w której e oznacza numer bieżącej epoki:

jeżeli ($\Theta > 1,3 \cdot \Theta'$ oraz $e > 0,03 \cdot M$) to przerwij uczenie.

⁸⁶ Wypraktykowano, iż dobre efekty daje przyjęcie stosunkowo niewielkiej wartości α (rzędu 0,01 - 0,05), oraz wartości β niewiele mniejszej od 1 (np. 0,95).

- wystąpi brak poprawy najlepszej dotychczasowej oceny sieci Θ' w pewnym dość długim okresie uczenia, co można wyrazić następującą zaproponowaną tutaj regułą (ε oznacza numer epoki, w której ostatni raz wystąpiła poprawa oceny):

jeżeli ($e > 1,5 \cdot \varepsilon$ oraz $e > \varepsilon + 0,03 \cdot M$) to przerwij uczenie.

W celu unikania zatrzymania się procesu uczenia sieci w niekorzystnym minimum lokalnym funkcji błędu uczenia lub funkcji oceny sieci, uczenie danej sieci jest powtarzane wielokrotnie (liczba powtórzeń P jest przyjętym arbitralnie⁸⁷ parametrem), przy czym za każdym powtórzeniem proces rozpoczyna się z różnych, losowo wygenerowanych, punktów przestrzeni wag sieci. Współrzędne tych punktów startowych (wartości początkowe wag sieci) są generowane losowo⁸⁸ (ze stałą funkcją gęstości prawdopodobieństwa) z przedziału $[-0,2, 0,2]$. Jako ostateczny, wynikowy zbiór wag dla nauczonej sieci przyjmowany jest ten spośród P końcowych zestawów wag otrzymanych w trakcie P powtórzeń procesu uczenia, dla którego sieć uzyskała najlepszą ocenę wstępną.

Sposób końcowej oceny sieci w proponowanej metodologii zależy od przyjętego arbitralnie kryterium (będzie ono określane w konkretnych, opisanych dalej badaniach), w szczególności może nim być:

- najlepsza wstępna ocena uzyskana w trakcie P powtórzeń procesu uczenia (tzn. wstępna ocena najlepszej sieci zostaje uznana za jej ocenę końcową),
- ocena testowa - dokonana w oparciu o ciąg testujący już po zakończeniu operacji uczenia (wszystkich P powtórzeń) - dla sieci posiadającej najlepszą ocenę wstępną; możliwe do zastosowania są oba warianty wymienione w podrozdziale 2.7.11, tzn. wykorzystanie funkcji błędu dla ciągu testującego, lub ocena efektywności strategii inwestycyjnej dla danych z ciągu testującego (lub walidacyjnego),

⁸⁷ Zbyt duże P znacznie wydłuża czas obliczeń podczas konstrukcji modelu, z kolei zbyt małe P powoduje ryzyko zakończenia procesu konstrukcji (doboru wag) danej sieci w zbyt wysoko położonym minimum lokalnym funkcji błędu.

⁸⁸ W praktyce wykorzystano generator liczb pseudolosowych.

- metoda mieszana, polegająca na hybrydowym połączeniu obu powyższych sposobów (wielkość zawierająca jako składniki zarówno ocenę wstępną, jak i ocenę testową uzyskaną dla ciągu testującego).

4.7. Koncepcja algorytmu genetycznego.

Uwzględniając przedstawione w podrozdziale 3.3 przesłanki wykorzystania algorytmów genetycznych jako narzędzi wspomagających konstrukcję modeli opartych na sieciach neuronowych, w ramach prezentowanej metodologii konstrukcji systemów decyzyjnych zastosowano AG do:

- doboru zestawu zmiennych wejściowych, rozumianych tutaj jako zbiór procedur przetwarzania informacji wejściowej oraz ich parametrów (zob. podrozdział 4.2 - tabela 4-1),
- określenia wymiaru warstwy ukrytej SN,
- ewentualnego⁸⁹ doboru horyzontu prognozy (zob. podrozdział 4.5).

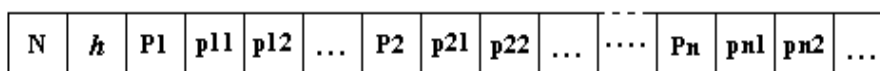
Poniżej zaproponowano oryginalną modyfikację klasycznego algorytmu genetycznego (podrozdział 3.1), przystosowaną przez autora do realizacji powyższych celów.

4.7.1. Zastosowany sposób genetycznego kodowania rozwiązań.

Zastosowano chromosomy o zmiennej długości oraz metodę kodowania poszczególnych genów (elementów chromosomu) za pomocą liczb całkowitych⁹⁰. Długość chromosomu zależy przede wszystkim od ilości reprezentowanych przez niego zmiennych wejściowych, a także od liczby parametrów procedur definiujących te zmienne. Schemat pojedynczego chromosomu prezentuje rys. 4-5.

⁸⁹ Często w badaniach wielkość tą dobierano arbitralnie.

⁹⁰ Takie rozwiązanie różni się od zdecydowanej większości realizacji AG, w których wykorzystuje się chromosomy o stałej długości oraz alfabet binarny.



Rys. 4-5. Schemat chromosomu w zaproponowanym algorytmie genetycznym. Źródło: opracowanie własne.

Na schemacie przedstawionym na rys. 4-5 przyjęto następujące oznaczenia genów:

N - liczba całkowita reprezentująca wielkość warstwy ukrytej w stosunku do warstwy wejściowej, należąca do przedziału⁹¹ $[0, 20000]$; jeżeli przez **m** oznaczymy liczbę neuronów warstwy ukrytej, zaś przez **k** - ilość neuronów warstwy wejściowej, to $m = [k \cdot N / 10000] + 2$, gdzie $[x]$ oznacza cechę⁹² liczby **x**. Tak więc zgodnie z założeniem przyjętym w podrozdziale 4.4: $m \in [2, 2k + 1]$.

h - horyzont prognozy (zob. podrozdział 4.5), $h \in [2, 20]$.

P1, P2, ..., Pn - numery procedur wstępnego przetwarzania informacji wejściowej, zastosowanych w module przetwarzania informacji wejściowej⁹³; jeżeli nie stosowano redukcji wejść SN (analizy głównych składowych), to $n = k$, zaś kolejne procedury o numerach **P1, P2, ..., Pn** generują dane podawane na kolejne wejścia SN, w przypadku wykorzystania PCA do redukcji liczby wejść sieci mamy $n \geq k$.

pji - oznacza wartość **i**-tego parametru **j**-tej procedury przetwarzania danych ($j = 1, 2, \dots, n$); parametry te zawsze przyjmują wartości całkowite z określonego dla każdego z nich osobno zakresu, parametry dla poszczególnych procedur wyszczególniono w tabelicy 4-1.

⁹¹ Długość tego przedziału wynika z potrzeby zachowania stosunkowo dużej precyzji określania proporcji wielkości warstwy ukrytej do warstwy wejściowej w sieci, przy jednoczesnym spełnieniu wymogu, aby liczba **N** była całkowita. W celu wyznaczenia tej proporcji (podlegającej później niewielkim modyfikacjom) wartość **N** pochodząca z tego przedziału należy podzielić przez 10000.

⁹² Największa liczba całkowita nie większa od danej liczby.

⁹³ Numery procedur zasadniczych (od 1 do 110) wyszczególniono w pierwszej kolumnie tabeli 4-1, w przypadku zastosowania dodatkowych 81 procedur dla szeregu czasowego $d(t)$ (zob. podrozdział 4.2), numery tych procedur określone są kolejno jako liczby całkowite z przedziału od 111 do 191.

4.7.2. Dobór funkcji celu i funkcji przystosowania.

Niech $\Theta'(c)$ oznacza końcową ocenę sieci o architekturze jednoznacznie określonej przez chromosom c , wyznaczoną w oparciu o metodologię opisaną w podrozdziałach 2.7.11 oraz 4.6. Funkcję celu f , określoną na zbiorze wszystkich możliwych chromosomów, zdefiniujemy bezpośrednio zależnością:

$$f(c) = \Theta'(c) \quad (4.16)$$

Zadaniem algorytmu będzie zatem znalezienie takiego chromosomu (chromosomów), dla którego funkcja celu f osiąga minimum, czyli znalezienie takiego zestawu określanych przez chromosom parametrów, dla których sieć neuronowa uzyskała najlepszą ocenę (najmniejszą liczbową wartość oceny).

Z kolei zdefiniujemy funkcję przystosowania $g(c)$ - jako pewną transformację funkcji $f(c)$ - w celu wyznaczania wartości przystosowania poszczególnych chromosomów przed zastosowaniem operatorów genetycznych (zob. podrozdział 3.1.1). Opierając się na licznych przesłankach opisanych w literaturze dotyczącej AG (podrozdział 3.1.3), a także na badaniach własnych (Morajda, 1997b), funkcję tą określimy następująco:

$$g(c) = \begin{cases} \rho \cdot (f_{sr} - f(c)) + 1 & \text{dla } f(c) \leq f_{sr} \\ e^{\rho \cdot (f_{sr} - f(c))} & \text{dla } f(c) > f_{sr} \end{cases} \quad (4.17)$$

gdzie: $\rho = \frac{A}{f_{sr} - f_{min}}$

f_{sr} - średnia arytmetyczna z wartości $f(c)$ liczona dla wszystkich chromosomów z całej populacji

f_{min} - minimum z wartości $f(c)$ obliczane dla wszystkich chromosomów z całej populacji

A - współczynnik równowagi, regulujący proporcję: szybkość zbieżności / różnicowanie populacji ($A > 0$).

Tak zdefiniowana funkcja przystosowania preferuje chromosomy, dla których określone przez nie sieci zostały ocenione lepiej od średniej, dając przy tym szansę

(choć niewielkie) chromosomom, które nie spełniają tego warunku. Należy zauważyć, że wartości funkcji g są zawsze dodatnie, oraz rosną gdy wartości funkcji celu maleją, osiągając maksimum w punkcie, w którym funkcja f osiąga minimum. Jak wynika z obserwacji autora, funkcja taka dość dobrze nadaje się do oceny chromosomów będącej z kolei podstawą operacji selekcji, oraz zapewnia wysoką efektywność algorytmu.

Występujący w zależności (4.17) współczynnik A decyduje o równowadze pomiędzy szybkością zbieżności algorytmu, a zróżnicowaniem populacji. Przyjęcie większych wartości A powoduje szybszą zbieżność, ale odbywa się to kosztem mniejszego różnicowania chromosomów, co może przyczynić się do utraty możliwości poszukiwań lepszych osobników w innych obszarach przestrzeni rozwiązań. Dobór mniejszej wartości tego współczynnika spowoduje, że w operacji selekcji uwzględnione będzie większe spektrum osobników, co zwiększy szanse na znalezienie rozwiązania optymalnego, ale spowolni zbieżność algorytmu. Właściwą wartość A najlepiej dobrać na drodze eksperymentalnej, w opisanych dalej badaniach przyjęto

$A = 50$.

4.7.3. Propozycje zmodyfikowanych operatorów genetycznych i innych parametrów algorytmu.

Początkowa populacja chromosomów jest tworzona losowo lub częściowo losowo z uwzględnieniem dozwolonych zakresów wartości poszczególnych genów. Każdy chromosom populacji początkowej określa taką samą początkową liczbę wejść SN , tzn. genów definiujących procedury wejściowe P_j (wartość n dla wszystkich osobników populacji początkowej jest jednakowa - stanowi ona parametr algorytmu). Podczas generowania wartości genów dla chromosomów początkowej populacji możliwe są dwa warianty, analogiczne do omówionych w dalszej części typów mutacji:

- wariant pierwszy zakłada, iż wartości genów generuje się z całego dozwolonego zakresu zmienności danego genu losowo, przy założeniu równomiernego rozkładu gęstości prawdopodobieństwa;

- w drugim wariacie postępuje się tak jak wyżej jedynie w przypadku genów reprezentujących numery procedur wejściowych, pozostałym genom są przypisywane pewne dobrane *a priori* wartości preferowane (uznane na podstawie posiadanej wiedzy za prawdopodobnie najlepsze wartości dla określanych przez geny parametrów), po czym wartości te są poddawane mutacji typu 2 (zob. niżej).

Operacja utworzenia kolejnej populacji wymaga zdefiniowania trzech stałych parametrów: I_B , I_C , I_O , gdzie $I_B + I_C + I_O = L$ (L oznacza liczebność całej populacji), przy czym:

- I_B najlepiej przystosowanych osobników z danej populacji przechodzi bez zmian do następnej populacji (nie podlegają oni też mutacji),
- I_C osobników kolejnej populacji jest tworzonych w wyniku operacji krzyżowania i mutacji, na podstawie wybranych chromosomów z populacji poprzedniej,
- I_O osobników jest kopiowanych z innych pokrewnych populacji (jeżeli takie istnieją) - mogą to być np. końcowe populacje otrzymane w wyniku realizacji AG dla tego samego zadania (instrumentu finansowego), ale przy innych wartościach startowych (np. innej populacji początkowej), lub też populacje otrzymane dla innych instrumentów finansowych o podobnym charakterze (np. akcje innej spółki); kopiowaniu podlegają chromosomy najlepiej przystosowane, albo wybrane losowo z prawdopodobieństwem proporcjonalnym do wartości funkcji przystosowania, skopiowane chromosomy nie podlegają mutacji. Tego typu operacja ma na celu zwiększenie zróżnicowania populacji, unikanie przedwczesnej zbieżności do lokalnego minimum funkcji celu i wyznaczenie nowych kierunków poszukiwań.

O ile podczas pierwszej i trzeciej z powyższych operacji chromosom nie ulega zmianie, o tyle sprecyzowania wymaga druga operacja obejmująca kolejno: wybór pary chromosomów do krzyżowania, przeprowadzenie krzyżowania, oraz mutację wybranych genów otrzymanej pary osobników.

Wybór pary chromosomów do krzyżowania odbywa się losowo, z prawdopodobieństwem proporcjonalnym do wartości funkcji przystosowania $g(c)$. Geny wchodzące w skład tej pary tworzą w wyniku krzyżowania (wymiany) dwa

nowe chromosomy c_A i c_B , przy czym operacja ta odbywa się w dwóch określonych następująco etapach:

1. każdy z genów określających wartości N oraz h , oraz numery procedur P_j jest losowo przyporządkowywany do chromosomu c_A lub c_B z prawdopodobieństwem 0,5, przy czym w każdym z chromosomów c_A i c_B musi znaleźć się jeden gen N i jeden gen h , oraz co najmniej dwa geny definiujące procedury P_j (sieć musi posiadać minimum dwa wejścia); przyporządkowanie określonego genu P_j do danego chromosomu c_A lub c_B powoduje jednoczesne przyporządkowanie do tego samego chromosomu wszystkich następujących po nim genów p_{ji} definiujących parametry tej procedury,
2. jeżeli w chromosomie c_A istnieją takie same (określone przez numery P_j) procedury jak w c_B (w obu nowopowstałych chromosomach istnieją geny P_j o tej samej wartości, jeden w c_A a drugi w c_B), to dla takich par następuje krzyżowanie (częściowa wymiana) parametrów p_{ji} tych procedur w taki sposób, że każda odpowiadająca sobie para parametrów podlega wymianie między chromosomami z prawdopodobieństwem 0,5.

Należy zwrócić uwagę na zastosowane w opracowanej metodzie niekonwencjonalne rozwiązanie operacji krzyżowania, uniezależniające wynik tej operacji od wzajemnego położenia (sąsiedztwa) genów definiujących procedury wejściowe w chromosomach tworzących parę „rodzicielską”⁹⁴. Koncepcja ta wynika z faktu, iż wzajemne położenie określonych wejść w sieci neuronowej nie ma znaczenia (wejścia mogą być położone w dowolnej kolejności). Przyjęto zatem założenie, że w wyniku operacji krzyżowania każda para genów definiujących wejścia sieci (procedury wejściowe), niezależnie od wzajemnego położenia tych genów, z równym prawdopodobieństwem znajdzie się w tym samym chromosomie jak i zostanie rozdzielona.

⁹⁴ W klasycznym algorytmie genetycznym opartym na krzyżowaniu jednopunktowym lub dwupunktowym, im geny są położone bliżej siebie, tym większe prawdopodobieństwo zdarzenia, że po krzyżowaniu znajdą się one w tym samym chromosomie.

Operacja krzyżowania jest powtarzana wielokrotnie tak, aby uzyskać I_C nowych osobników. Nie wyklucza się wielokrotnego udziału tych samych chromosomów jako „rodziców” w procesie krzyżowania.

Dla powstałych w wyniku krzyżowania osobników realizowana jest z kolei operacja mutacji, polegająca na zmianie wartości niektórych genów, wybranych losowo z określonym *a priori* prawdopodobieństwem. W przypadku jeżeli mutacji podlega gen określający numer procedury wejściowej P_j , nowa wartość genu jest dobierana losowo z zakresu określonego przez liczbę wszystkich możliwych procedur, a następnie adekwatnie do wyznaczonej nowej procedury wejściowej również losowo dobierane są jej parametry w taki sposób, jak podczas tworzenia populacji początkowej. W przypadku mutacji genu N , h , lub parametrów procedur p_{ji} , omawiana metodologia obejmuje dwa warianty:

- **Typ mutacji 1** - polega na wyznaczeniu nowej wartości podlegającego mutacji genu w sposób losowy z określonego dla danego genu zakresu zmienności, przy założeniu równomiernego rozkładu prawdopodobieństwa dla wszystkich wartości z danego zakresu;
- **Typ mutacji 2** - polega na skokowej zmianie (zwiększeniu lub zmniejszeniu) wartości genu podlegającego mutacji o określoną (niżej zdefiniowaną) wielkość s (w przypadku generowania populacji początkowej według wariantu drugiego, opisanego na początku tego podrozdziału, ta skokowa zmiana dla wartości preferowanej kodowanego przez gen parametru może nastąpić - z równym prawdopodobieństwem - o wielkość 0 , $\pm s$, $\pm 2s$, lub $\pm 3s$). Wielkość s jest *a priori* zadana bezpośrednio (dla wartości N), lub wyznaczona jako r procent z wartości genu przed mutacją (dla parametrów h oraz p_{ji}), gdzie r jest zadaną dla każdego parametru osobno liczbą z przedziału (0% , 50%); podczas takiej skokowej zmiany wartości genu kontrolowane jest przekroczenie granic przedziału zmienności danego parametru.

Prawdopodobieństwo mutacji stanowi parametr systemu. W zaprezentowanych dalej badaniach wartości prawdopodobieństwa mutacji⁹⁵ przyjmowano z zakresu od 0,005 do 0,01.

Każdorazowo przy realizacji badań należy zdefiniować ilość populacji (iteracji), po których następuje zakończenie funkcjonowania algorytmu genetycznego. Za rozwiązania końcowe przyjmuje się pewną przyjętą liczbę nauczonych sieci neuronowych, które uzyskały najlepszą ocenę końcową w ostatnim pokoleniu (określone są one przez najlepiej przystosowane chromosomy ostatniej populacji).

4.8. Moduł realizacji strategii decyzyjnej.

Moduł ten generuje decyzje transakcyjne (kupna lub sprzedaży) w oparciu sygnał najlepszej (najlepiej ocenionej) sieci neuronowej otrzymanej podczas etapu konstrukcji systemu. Bardziej złożona strategia może uwzględniać sygnały wielu najlepszych sieci, i generować decyzje na podstawie określonej reguły logicznej przetwarzającej te sygnały. Należy jednak zwrócić uwagę, iż w przypadku praktycznego wykorzystania SN w systemie wspomagania decyzji inwestycyjnych ocena końcowa sieci nie musi być jedynym i ostatecznym kryterium zakwalifikowania danej sieci jako elementu strategii inwestycyjnej - wskazana jest stała bieżąca kontrola funkcjonowania danej sieci lub dodatkowa jej ocena na danych z ostatniego okresu przeszłości nie uwzględnionych przy budowie modelu.

W celu otrzymania aktualnego sygnału danej SN, należy na wejście systemu podać aktualne dane pochodzące z rynku i dokonać identycznego wstępnego przetwarzania tych danych jak na etapie konstrukcji systemu i uczenia sieci, z zachowaniem tych samych reguł i parametrów przetwarzania danych wejściowych. Sygnał otrzymany na wyjściu sieci poddawany jest następnie operacji tzw. postprocessingu (jest to transformacja wielkości wyjściowej odwrotna do normalizacji zmiennej wyjściowej). Otrzymana w rezultacie wartość może być interpretowana jako określona przez typ

⁹⁵ W literaturze zaleca się przyjmowanie stosunkowo małych wartości prawdopodobieństwa mutacji w AG.

zastosowanej zmiennej wyjściowej prognoza zachowania się kursu danego instrumentu w przyszłości.

Należy zwrócić uwagę, że postprocessing wartości wyjściowej, mimo że umożliwia precyzyjną interpretację tej wartości, nie zawsze jest operacją konieczną w przypadku zastosowania sygnału wyjściowego sieci jako danej wejściowej dla reguły decyzyjnej. Zwróćmy bowiem uwagę, że dla każdego typu zmiennej wyjściowej SN dodatnie wartości wyjściowe związane są z prognozą wzrostu kursu danego instrumentu, zaś wartości ujemne - z prognozą spadku kursu.

Najprostsza strategia decyzyjna wykorzystująca sygnały jednej sieci opiera się na następujących regułach⁹⁶, w których y oznacza sygnał wyjściowy sieci, zaś λ jest odpowiednio dobranym parametrem strategii, reprezentującym szerokość „strefy nieczułości” metody:

- jeżeli $y > \lambda$ to kupuj,
- jeżeli $y < -\lambda$ to sprzedaj.

W proponowanej metodologii dobór reguły określającej strategię inwestycyjną dokonywany jest arbitralnie. W dalszych badaniach można jednak próbować wykorzystać w tym celu określone metody takie, jak np. algorytmy genetyczne.

⁹⁶ Nie wymieniamy tu na razie oczywistych warunków dodatkowych, np. tego, że aby sprzedać akcje trzeba je posiadać, aby zaś zrealizować decyzje kupna należy dysponować w danej chwili gotówką.

5. Analiza rezultatów zastosowania przedstawionej metodologii w strategii opartej o prognozy indeksu WIG 20.

Poniżej zaprezentowano rezultaty wykorzystania zaproponowanej metodologii w zagadnieniu prognozowania zmian indeksu giełdowego WIG 20 oraz wyniki zastosowania tych prognoz bezpośrednio do realizacji transakcji kupna lub sprzedaży tego indeksu. Do konstrukcji systemu wykorzystano dane dotyczące indeksów WIG 20 oraz WIG z okresu od 14 kwietnia 1994 (pierwsze notowanie WIG 20)⁹⁷ do 30 września 1997.

5.1. Dobór parametrów systemu.

Przyjęto następujące parametry konstrukcji systemu według zaproponowanej w rozdziale 4 metodologii:

Parametry modułu przetwarzania informacji wejściowej

- Zastosowano 110 procedur wstępnego przetwarzania informacji wejściowej, wyszczególnionych w tabeli 4-1. Nie zdecydowano się na wykorzystanie dodatkowych procedur opartych o szereg czasowy $d(t)$ (opisany formułą 4.1) z uwagi na silną korelację indeksów WIG 20 i WIG.

⁹⁷ Ponieważ pierwsze 50 notowań indeksu WIG 20 posłużyło jedynie w celu zdeterminowania odpowiednich średnich kroczących i innych wskaźników wykorzystujących dane z pewnego okresu przeszłości (stosowanych w procedurach wstępnego przetwarzania danych), właściwy początek danych użytych do uczenia i walidacji sieci odpowiada dacie 2 sierpnia 1994 (wartość WIG 20 w tym dniu wynosiła 1082,4).

Parametry modułu preprocessingu danych wejściowych

- parametr trymownia wartości nietypowych (formuła (4.3)) $\gamma_{tr} = 4$
- w przypadku zmiennych wejściowych podlegających standaryzacji (wzór (4.5)) zastosowano parametr $\gamma_{st} = 1,5$
- zastosowano skokowe rozdzielenie wartości dodatnich i ujemnych dla zmiennych normalizowanych „bez przesunięcia zera” (formuła (4.7)) z parametrem $\gamma_{rw} = 0,25$
- nie wykorzystano analizy głównych składowych do wstępnego przetwarzania danych

Parametry architektury i uczenia sieci typu *perceptron*

- przyjęto strukturę sieci zdefiniowaną w podrozdziale 4.4, zawierającą jedną warstwę ukrytą z genetycznie wyznaczaną liczbą neuronów
- jako zmienną wyjściową przyjęto wielkość najbliższej istotnej względnej zmiany 5-dniowej adaptacyjnej średniej kroczącej z kursu (zob. podrozdział 4.5.3 - formuła (4.13)), przy założeniu horyzontu $h = 5$ oraz parametru $q = 2,5$; należy zwrócić uwagę, że horyzont h został tu ustalony arbitralnie i nie podlega genetycznej modyfikacji
- zrezygnowano z obcinania wartości nietypowych dla zmiennej wyjściowej i przyjęto dla niej standardową formułę normalizacji wyrażoną wzorem (2.21)
- z ciągu danych wyeliminowano te wzorce, dla których zmienna wyjściowa przyjmuje wartość zero, po czym pozostałe elementy ciągu danych służących do konstrukcji modelu podzielono na ciąg uczący i ciąg walidacyjny według schematu **C** (rys. 4-4, podrozdział 4.6) przyjmując $T_u = 50$ oraz $T_v = 10$ (zrezygnowano z wykorzystania ciągu testującego do oceny sieci, przyjmując wstępną ocenę sieci jako końcową)
- ocenę sieci realizowano w oparciu o kryterium (4.15) przyjmując $p = 1$, oraz zmienny parametr q według reguły: $q = (E_v i_u) / (E_u i_v)$
- jako wartość oceny sieci (podawaną m.in. przy prezentacji wyników badań)

przyjęto wielkość $4 \cdot \Theta$, gdzie Θ wynika z zależności (4.15)⁹⁸

- przyjęto następujące wartości stałych decydujących o modyfikacji współczynników uczenia i bezwładności w trakcie procesu uczenia: $\alpha = 0,05$, $\beta = 0,95$ (zob. podrozdział 4.6)
- przyjęto wartości parametrów funkcji aktywacji (2.4): $a = 1,15$ oraz $b = 2$
- ustalono maksymalną liczbę epok uczenia sieci $M = 5000$, oraz liczbę powtórzeń procesu uczenia każdej sieci $P = 8$ (w niektórych badaniach opisanych w podrozdziale 5.4 przyjmowano $P = 10$, lub więcej)
- uczenie realizowano metodą wstecznej propagacji błędów (zob. podrozdział 2.5) przyjmując kwadratową funkcję błędu określoną wzorem (2.9).

Parametry algorytmu genetycznego

- przyjęto wielkość populacji $L = 100$ oraz parametry $I_B = 30$, $I_C = 70$, $I_O = 0$
- mutację realizowano w oparciu o typ mutacji 2 (zob. podrozdział 4.7.3)
- przyjęto wartości prawdopodobieństwa mutacji: dla genu N równe $0,01$, natomiast dla pozostałych genów równe $0,005$.
- założono liczbę iteracji (populacji) równą 30
- określono początkową liczbę wejść SN (w populacji początkowej) równą 12

Parametry strategii decyzyjnej

Przyjęto możliwość dokonywania codziennych transakcji polegających na zakupie indeksu WIG 20 za całą posiadaną w danej chwili gotówkę, albo całkowitej sprzedaży indeksu; w danej chwili możemy zatem posiadać portfel złożony w 100% z indeksu, albo w 100% z gotówki. Strategia decyzyjna jest oparta o sygnały jednej, wybranej sieci neuronowej i składa się z następujących reguł:

- jeżeli ($y(t) > 0,1$ i posiadamy gotówkę), to na sesji $t+1$ kup indeks

⁹⁸ Skalowanie liniowe z przyjętym arbitralnie czynnikiem 4 nie ma żadnego wpływu na istotę dokonanych analiz (zastosowano je jednak z uwagi na sposób funkcjonowania wykorzystanego do obliczeń programu komputerowego - dla zachowania spójności danych wyjściowych programu z prezentowanymi dalej wynikami).

- jeżeli ($y(t) < -0,1$ i posiadamy indeks) , to na sesji $t+1$ sprzedaj indeks
- jeżeli nie zachodzi żaden z w/w warunków to zachowaj aktualny stan portfela.

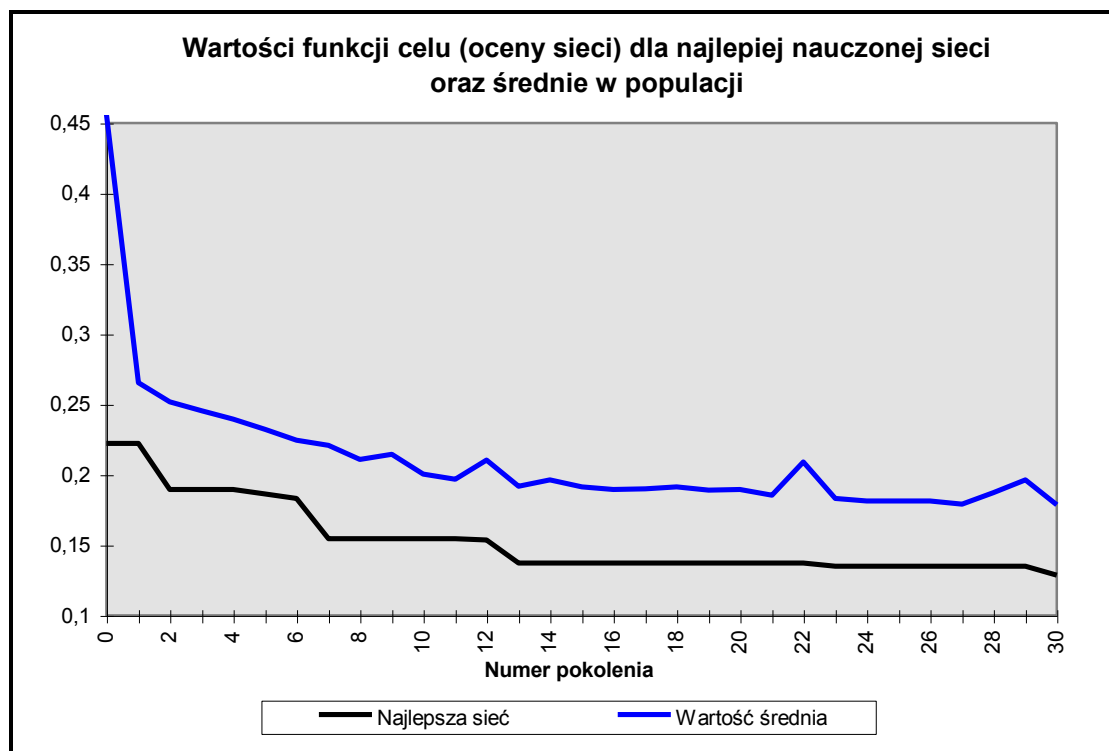
przy czym $y(t)$ jest sygnałem sieci neuronowej bez postprocessingu, otrzymanym po uwzględnieniu wyników notowań na sesji t ⁹⁹.

Aby urealnić ocenę strategii decyzyjnej opartej o SN, przyjęto dodatkowo możliwość uwzględnienia kosztów transakcyjnych (prowizji maklerskich).

5.2. Analiza otrzymanych wyników oraz efektywności metody.

Realizację algorytmu genetycznego przeprowadzono na komputerze PC Pentium 200. Poniżej na rys. 5-1 zaprezentowano wykres obrazujący spadek funkcji celu (oceny sieci) dla najlepszego osobnika z bieżącej populacji oraz przebieg średniej wartości oceny sieci w tej populacji, w zależności od numeru iteracji (pokolenia) AG.

⁹⁹ Należy zwrócić uwagę na przyjęte tutaj, istniejące w warunkach realnych opóźnienie: decyzję wygenerowaną na podstawie wyników sesji t realizujemy dopiero na sesji $t+1$.



Rys. 5-1. Zmiany wartości funkcji celu (oceny sieci) w trakcie realizacji algorytmu genetycznego. Źródło: opracowanie własne.

Dla 100 sieci otrzymanych w wyniku realizacji algorytmu w ostatnim pokoleniu zbadano wyniki realizacji założonej strategii decyzyjnej dla danych użytych do uczenia i walidacji sieci, tzn. pochodzących z okresu od 2.08.1994 (WIG20 = 1082,4) do 30.09.1997 (WIG20 = 1733,2). W dalszej części pracy te właśnie dane wykorzystane w procesie konstrukcji modelu, tzn. podczas uczenia, walidacji i ewentualnie końcowego testowania sieci, oznaczymy symbolem **IN-S** (ang. *in-sample*), w przeciwieństwie do danych nie biorących udziału w procesie uczenia i oceny sieci, które oznaczymy symbolem **OUT-S** (ang. *out-of-sample*). Jako zasadniczy efekt końcowy zastosowania strategii aktywnej, opartej o sygnały danej sieci, przyjęto całkowity zysk w rozważanym okresie czasu, uzyskany w wyniku realizacji tej strategii, czyli różnicę pomiędzy końcową a początkową wartością portfela. Dla ułatwienia dokonania porównań ze strategią pasywną przyjęto początkową wartość aktywów portfela równą wartości indeksu w pierwszym dniu

analizowanego okresu. Ponadto założono koszty transakcyjne (prowizję) w wysokości 0,5% od kwoty transakcji¹⁰⁰.

W następnym etapie zbadano wyniki realizacji wypracowanej strategii decyzyjnej dla danych **OUT-S**, czyli dla danych nie biorących udziału w procesie konstrukcji systemu, pochodzących z okresu ponad czterech i pół miesięcy¹⁰¹, tzn. od 1.10.1997 (WIG20 = 1740,5) do 19.02.1998 (WIG20 = 1759,8). Analizy tej dokonano dla 30 sieci, dla których oceny uzyskane podczas konstrukcji systemu były najlepsze (odpowiednie wartości funkcji celu były najmniejsze). Rozważono przy tym dwa przypadki dotyczące kosztów transakcyjnych: prowizja zerowa, oraz prowizja w wysokości 0,5%.

Oprócz kryterium całkowitego zysku w danym okresie, przy analizie wyników uwzględniono także inne kryterium oceny strategii - największą stratę (*ang. maximum drawdown*), którą oznaczono symbolem **dd**. Wielkość tą definiuje się jako największy względny (wyrażony w procentach) spadek wartości aktywów portfela liczony od maksymalnej poprzedzającej do minimalnej następującej po niej wielkości, w obrębie rozważanego okresu czasu. Formalnie wartość **dd** określono zależnością:

$$dd = \max_{(i,j): b \leq i < j \leq g} \left\{ \frac{s(i) - s(j)}{s(i)} \cdot 100 \right\} \quad (5.1)$$

gdzie: $s(k)$ - stan aktywów portfela w chwili k

b, g - odpowiednio początek i koniec analizowanego okresu czasu

Wartość ta jest również, obok zysku, często podawanym w literaturze wskaźnikiem charakteryzującym efektywność danej strategii (zob. np. Refenes, 1995, str.72). Stanowi ona swoistą miarę ryzyka, określającą potencjalną stratę inwestora stosującego daną strategię inwestycyjną, stąd też niekiedy właśnie ona jest uznawana

¹⁰⁰ Przyjęto tzw. prowizję liniową (stały ułamek wartości transakcji, niezależny od tej wartości).

¹⁰¹ W tym okresie nastąpił początkowo znaczny spadek, a następnie wzrost wartości indeksu WIG 20; datę końcową okresu dobrano tak, aby wartość indeksu w dniu końcowym była w przybliżeniu równa wartości indeksu w dniu początkowym analizowanego okresu, co zapewni rzetelność porównań ze strategią pasywną.

w praktyce za ważniejszy wskaźnik, niż całkowity zysk. Niska wartość **dd** świadczy o niewielkim ryzyku i o dość systematycznym charakterze kumulowania zysku.

Wyniki wyżej opisanych badań efektywności aktywnej strategii inwestycyjnej dla 30-tu sieci¹⁰² posiadających najlepsze oceny (tzn. najmniejsze wartości liczbowe oceny uzyskane na etapie uczenia sieci) zestawiono w tabeli 5-1 (sieci uporządkowano według wartości tych ocen - od najmniejszej, tzn. najlepszej, do największej). W kolejnych kolumnach tabeli przedstawiono:

- numer sieci (sieci uporządkowano rosnąco według wartości ocen)
- wartość oceny (funkcji celu) otrzymaną na etapie konstrukcji systemu
- zysk całkowity dla danych IN-S przy zastosowaniu kosztów transakcyjnych 0,5%
- wartość maksymalnej straty **dd** dla danych IN-S
- zysk całkowity dla danych OUT-S przy zastosowaniu kosztów transakcyjnych 0%
- wartość maksymalnej straty **dd** dla danych OUT-S (koszty 0%)
- zysk całkowity dla danych OUT-S przy zastosowaniu kosztów transakcyjnych 0,5%
- wartość maksymalnej straty **dd** dla danych OUT-S (koszty 0,5%).

W przedostatnim wierszu zestawiono odpowiednie wartości średnie z poszczególnych kolumn, obliczone dla wszystkich 30-tu sieci. W wierszu ostatnim podano wartości zysku całkowitego oraz maksymalnej straty **dd**, otrzymane w wyniku zastosowania w danym okresie strategii pasywnej, polegającej na zakupie indeksu WIG 20 w pierwszym dniu i konsekwentnym utrzymaniu tego stanu portfela przez cały okres. Minimalne i maksymalne wartości z poszczególnych kolumn wyłuszczone.

¹⁰² Rezultaty obrazujące zysk w okresie IN-S dla całej końcowej populacji obejmującej 100 sieci przedstawiono graficznie na rys.5-5.

TABELA 5-1.

| LP | OCENA | IN-S prow. 0,5% | | OUT-S prow. 0 % | | OUT-S prow. 0,5% | |
|--------------------------|----------|-----------------|-------------|-----------------|-------------|------------------|-------------|
| | | ZYSK | dd % | ZYSK | dd % | ZYSK | dd % |
| 0 | 0,128748 | 3370,0 | 21,0 | 578,9 | 6,7 | 400,6 | 8,6 |
| 1 | 0,134784 | 5430,3 | 23,2 | 607,3 | 6,5 | 362,7 | 8,3 |
| 2 | 0,137198 | 3161,5 | 20,5 | 12,2 | 17,6 | -122,5 | 19,3 |
| 3 | 0,138813 | 4893,0 | 17,1 | -142,5 | 17,7 | -228,1 | 20,2 |
| 4 | 0,139809 | 7126,7 | 14,9 | 157,4 | 12,3 | 11,5 | 14,1 |
| 5 | 0,143805 | 5872,5 | 23,5 | 201,2 | 11,7 | 97,2 | 12,5 |
| 6 | 0,144080 | 10195,2 | 13,0 | 734,6 | 11,1 | 590,5 | 11,1 |
| 7 | 0,144428 | 3694,9 | 17,5 | 23,9 | 19,6 | -144,0 | 25,1 |
| 8 | 0,147373 | 6149,1 | 17,0 | -19,7 | 18,8 | -152,0 | 24,3 |
| 9 | 0,148079 | 6259,9 | 10,2 | 90,5 | 14,2 | -100,2 | 19,2 |
| 10 | 0,148566 | 1463,8 | 21,8 | 166,1 | 9,5 | 37,2 | 11,8 |
| 11 | 0,148914 | 6279,6 | 11,3 | 239,4 | 17,0 | 105,5 | 21,8 |
| 12 | 0,149469 | 1897,9 | 25,2 | 61,1 | 15,2 | -94,0 | 19,4 |
| 13 | 0,149633 | 3383,2 | 19,2 | 111,9 | 14,4 | 4,0 | 16,1 |
| 14 | 0,150403 | 5853,8 | 23,1 | 167,4 | 13,3 | 20,7 | 15,0 |
| 15 | 0,150435 | 2614,4 | 24,5 | 94,7 | 16,8 | -63,3 | 19,7 |
| 16 | 0,150892 | 7744,7 | 12,7 | 135,5 | 12,7 | -51,5 | 17,8 |
| 17 | 0,150978 | 3434,2 | 18,6 | 13,4 | 12,6 | -97,0 | 13,5 |
| 18 | 0,151941 | 2142,1 | 16,9 | -154,5 | 20,0 | -261,7 | 21,6 |
| 19 | 0,152075 | 1200,2 | 24,9 | 267,6 | 6,2 | 188,9 | 8,1 |
| 20 | 0,152255 | 4144,3 | 26,7 | 55,9 | 16,0 | -65,5 | 17,6 |
| 21 | 0,152504 | 2971,5 | 18,3 | 350,4 | 11,5 | 189,6 | 13,1 |
| 22 | 0,152551 | 3380,5 | 24,7 | 133,0 | 9,0 | 6,3 | 11,8 |
| 23 | 0,152716 | 5811,5 | 17,7 | -27,8 | 12,9 | -167,4 | 17,0 |
| 24 | 0,153463 | 4365,8 | 22,6 | -126,6 | 13,3 | -213,0 | 15,9 |
| 25 | 0,153596 | 3199,1 | 14,2 | 102,4 | 7,4 | -4,9 | 9,2 |
| 26 | 0,154053 | 5988,8 | 14,6 | -14,5 | 12,6 | -163,1 | 18,8 |
| 27 | 0,154461 | 4932,1 | 24,7 | 219,1 | 17,1 | 105,0 | 19,6 |
| 28 | 0,154491 | 8440,1 | 16,4 | -214,4 | 22,7 | -387,0 | 28,6 |
| 29 | 0,154639 | 4486,5 | 17,4 | 275,5 | 9,9 | 102,0 | 11,7 |
| Wart. średnie: | | 4662,9 | 19,1 | 136,6 | 13,5 | -3,1 | 16,4 |
| Strateg. pasywna: | | 650,8 | 53,0 | 19,3 | 26,6 | 19,3 | 26,6 |

Źródło: opracowanie własne.

Z analizy danych przedstawionych w tabeli 5-1 wynikają następujące wnioski:

- Dla danych IN-S całkowity zysk otrzymany w wyniku zastosowania strategii aktywnej opartej o SN jest dla wszystkich sieci znacznie wyższy niż w przypadku strategii pasywnej (pomimo uwzględnienia 0,5% prowizji od transakcji). Średnia (dla 30 sieci) stopa zwrotu w ponadtrzyletnim okresie IN-S, tj. od 2.08.1994 do 30.09.1997 wyniosła **430,8%**, maksymalna stopa zwrotu (dla sieci nr 6) osiągnęła

941,9%, podczas gdy wielkość ta dla strategii pasywnej wyniosła **60,1%**. Ponadto należy zwrócić uwagę na wysokość maksymalnej straty *dd*, która w przypadku strategii aktywnej wyniosła średnio **19,1%**, podczas gdy dla strategii pasywnej osiągnęła aż **53,0%**; wskaźnik ten jest zatem zdecydowanie lepszy dla aktywnych strategii opartych o SN. Dobre rezultaty decyzyjne uzyskane dla danych użytych do konstrukcji i walidacji modelu są co prawda zjawiskiem normalnym w uczących się systemach decyzyjnych, jednak w rozważanym tutaj zagadnieniu wyniki te niewątpliwie świadczą o wysokiej efektywności sieci neuronowych jako nieliniowych modeli decyzyjnych służących do wspomaganie decyzji inwestycyjnych.

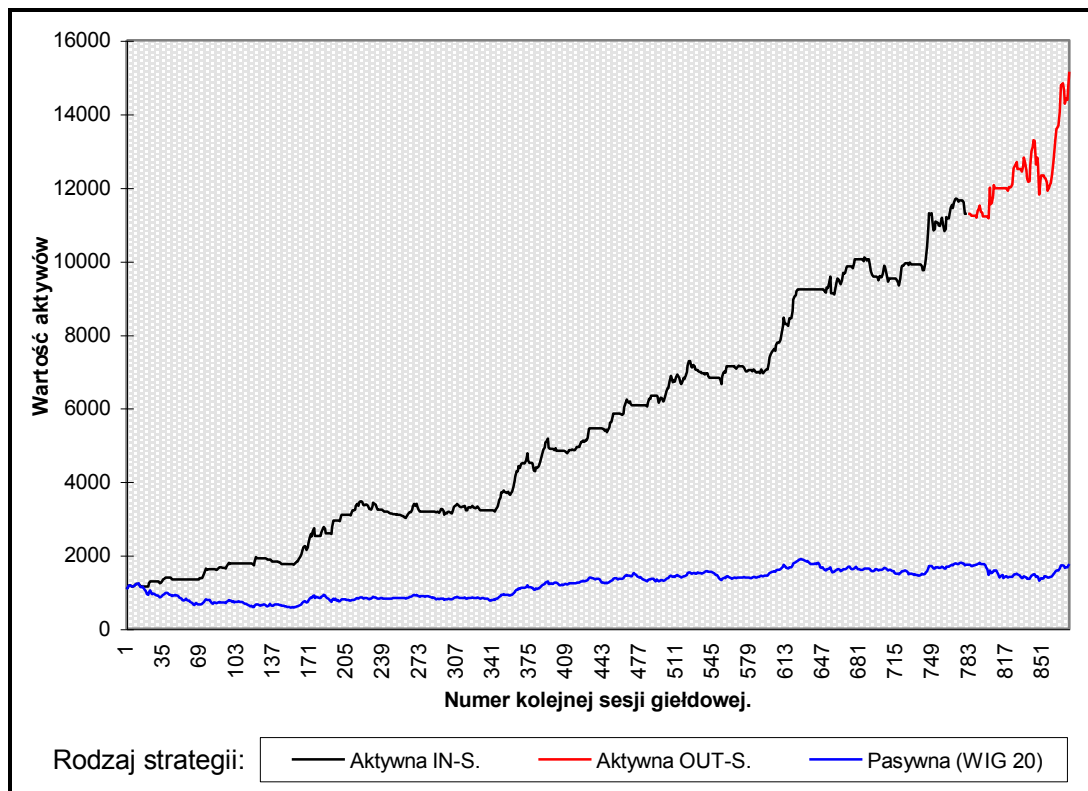
- Dla okresu OUT-S przy założeniu braku kosztów transakcyjnych zyski osiągane w wyniku zastosowania strategii opartych o sygnały SN były przeciętnie lepsze, choć w zasadzie porównywalne z rezultatami strategii pasywnej: dla 21 sieci były one lepsze, a dla 9 niższe od zysku z zakupu i zachowania indeksu WIG 20. Średnia stopa zwrotu w ok. 4,5-miesięcznym okresie OUT-S (1.10.1997 - 19.02.1998) wyniosła dla strategii aktywnej **7,85%**, podczas gdy w przypadku indeksu WIG 20 była ona równa **1,11%**; maksymalna stopa zwrotu w tym okresie (sieć nr 6) osiągnęła **42,21%**. Wysokość maksymalnej straty *dd* dla strategii aktywnej wyniosła średnio **13,5%**, zaś w przypadku inwestycji pasywnej **26,6%**; wskaźnik ten okazał się zatem tutaj dla sieci prawie dwukrotnie mniejszy, a więc istotnie lepszy, co może stanowić argument dla stosowania aktywnej metody inwestowania. Zatem nawet jeżeli stwierdzimy brak istotnych różnic pomiędzy zyskiem całkowitym uzyskanym w wyniku zastosowania tych dwóch rodzajów strategii, to niewątpliwie strategię aktywne oparte na SN należy uznać za obarczone przeciętnie mniejszym ryzykiem.
- Dla okresu OUT-S przy założeniu kosztów transakcyjnych 0,5% strategia pasywna przyniosła nieco większy zysk niż przeciętna strategia aktywna: średnia stopa zwrotu dla strategii opartej na SN wyniosła **-0,18%** (dla strategii pasywnej: **+1,11%**). Maksymalna stopa zwrotu w tym okresie (sieć nr 6) osiągnęła **33,93%**. Należy zwrócić uwagę na wielkość maksymalnej straty *dd*, która średnio dla

strategii aktywnej (**16,4%**) jest nadal znacznie niższa niż w przypadku pasywnej inwestycji w indeks WIG 20 (**26,6%**).

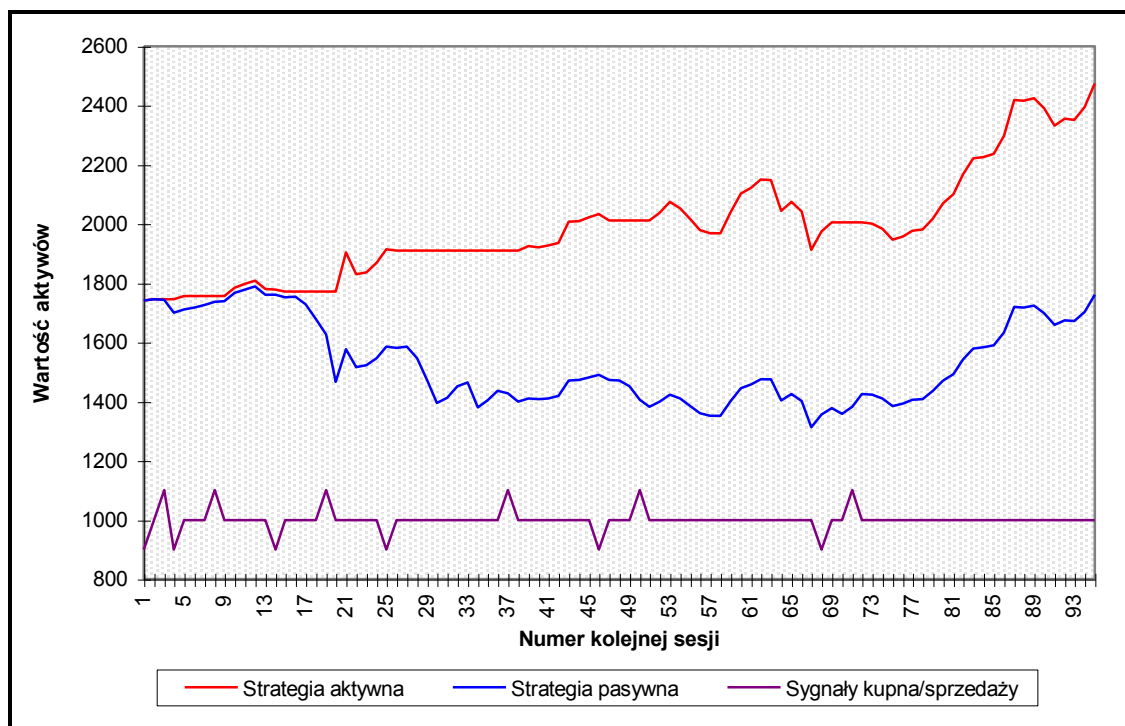
Analizując otrzymane wyniki można generalnie stwierdzić, że efektywność aktywnych strategii inwestycyjnych, bardzo wysoka dla danych IN-S, spada znacznie przy zastosowaniu danych nie użytych do budowy systemu (OUT-S) (szczególnie w przypadku uwzględnienia istniejących w rzeczywistości kosztów transakcyjnych), pozostając jednak nadal na obiecującym poziomie. Godnym podkreślenia faktem jest przy tym przeciętnie mniejsze ryzyko rozważanych tutaj aktywnych strategii opartych na SN, wyrażone wielkością maksymalnej straty *dd*, w stosunku do analogicznej wartości wyznaczonej dla strategii pasywnej. Generalnie można zatem uznać otrzymane rezultaty za dość obiecującą inspirację do dalszych badań i testowania modeli opartych o SN.

Należy także zauważyć, że wśród analizowanego zestawu SN istnieją sieci (np. sieć nr 6), które wykazują stałą, ponadprzeciętną efektywność (również dla danych OUT-S). Być może poszukiwanie tego typu efektywnych sieci powinno stanowić klucz do konstrukcji dobrych aktywnych strategii decyzyjnych opartych o prognozy zachowania się rynku.

Rezultaty otrzymane w wyniku realizacji aktywnej strategii wykorzystującej sygnały generowane przez sieć nr 6 zaprezentowano na rysunkach 5-2 oraz 5-3. Rys. 5-2 przedstawia przyrost aktywów portfela w czasie (mierzonym kolejnymi numerami sesji giełdowych) dla danych IN-S oraz OUT-S, w porównaniu ze strategią pasywną (indeksem WIG 20), przy założeniu kosztów transakcyjnych w wysokości 0,5%. Z kolei rys. 5-3 prezentuje analogiczne porównanie tylko dla okresu OUT-S, przy zerowej prowizji od transakcji; na rysunku tym przedstawiono dodatkowo sygnały kupna („iglice” skierowane w górę), oraz sprzedaży (skierowane w dół), generowane w oparciu o wartości wyjściowe sieci (transakcje odpowiadające tym sygnałom są realizowane na następnej sesji po ich wygenerowaniu). Analizę rezultatów zastosowania tej sieci w dłuższym okresie OUT-S, wraz z dyskusją zagadnień douczania lub ponownego uczenia sieci, przeprowadzono w podrozdziałach 5.4.7 oraz 5.4.8.



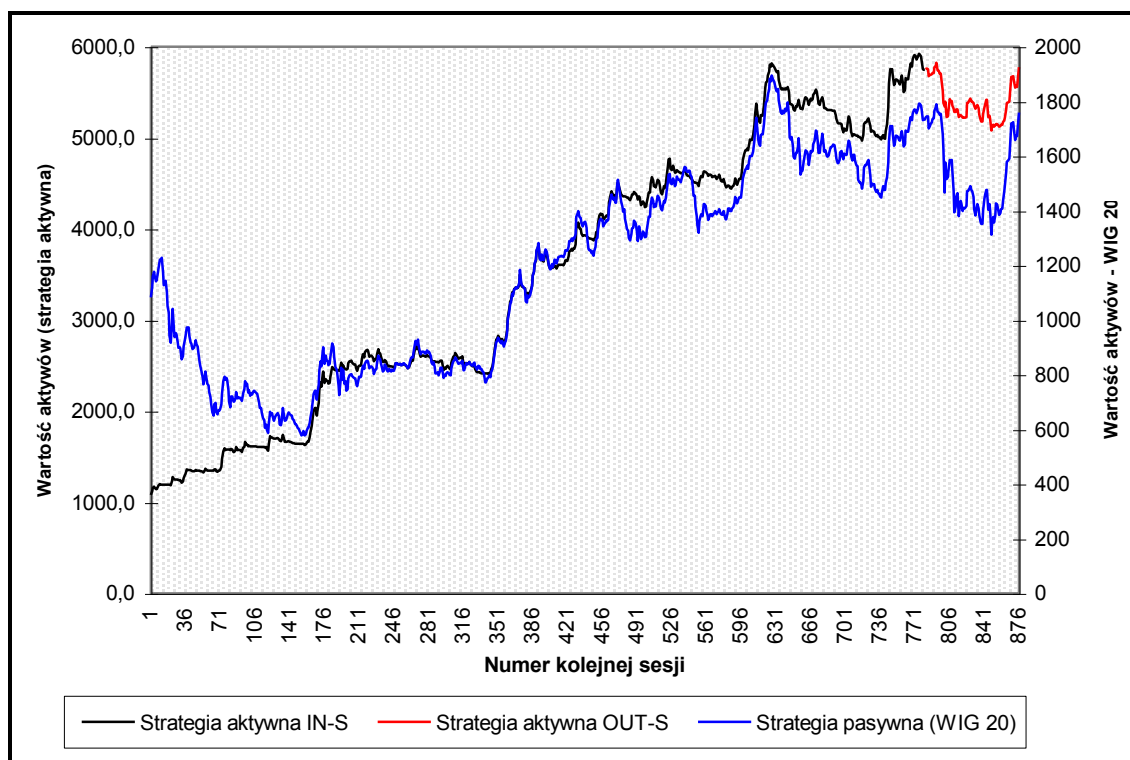
Rys. 5-2. Przyrost aktywów portfela w czasie dla aktywnej strategii inwestycyjnej opartej o sygnały generowane przez sieć nr 6, dla okresu IN-S oraz OUT-S przy założeniu kosztów transakcyjnych w wysokości 0,5%, w porównaniu z pasywną inwestycją w indeks WIG20. Inwestycja początkowa: 1082,4. Źródło: opracowanie własne.



Rys. 5-3. Przyrost aktywów portfela przy wykorzystaniu aktywnej strategii inwestycyjnej opartej o sygnały generowane przez sieć nr 6, w porównaniu ze strategią pasywną. Okres OUT-S, zerowe koszt transakcyjne, inwestycja początkowa: 1740,5. Przedstawiono sygnały kupna (skierowane w górę), oraz sprzedaży (skierowane w dół), podlegające realizacji na następnej sesji. Źródło: opracowanie własne.

Dokonajmy jeszcze analizy graficznej (opartej o wykresy zmienności aktywów) „przeciętnej” dla 30 sieci aktywnej strategii inwestycyjnej. Strategię taką można analizować rozważając w każdym momencie czasu średnią arytmetyczną aktywów z 30-tu początkowo równych portfeli, z których każdy wykorzystuje sygnały transakcyjne innej sieci. Tego typu „przeciętną” strategię można praktycznie zrealizować dokonując odpowiedniej dywersyfikacji aktywów początkowych - w omawianym tu przypadku dzieląc aktywa początkowe na 30 równych części i „powierzać” zarządzanie każdej z tych części innej sieci neuronowej. Wyniki realizacji tej strategii dla danych IN-S oraz OUT-S, w porównaniu ze strategią pasywną, prezentuje rys. 5-4. Na rysunku tym - w celu umożliwienia dokonania lepszych porównań - wartość aktywów dla strategii aktywnej przedstawiono na lewej osi pionowej, natomiast odpowiednie wartości dla strategii pasywnej (wartości indeksu WIG 20) - stosownie powiększone na osi prawej. Początkowa wartość aktywów jest w obu przypadkach ta sama i wynosi 1082,4. W badaniach

uwzględniono koszty transakcyjne w wysokości 0,5%. Należy zwrócić uwagę na widoczne na wykresie mniejsze spadki wartości portfela zarządzanego przy wykorzystaniu strategii aktywnej w okresach osłabienia koniunktury giełdowej (z faktem tym wiąże się też niższa średnia wartość wskaźnika *dd* dla strategii aktywnych).

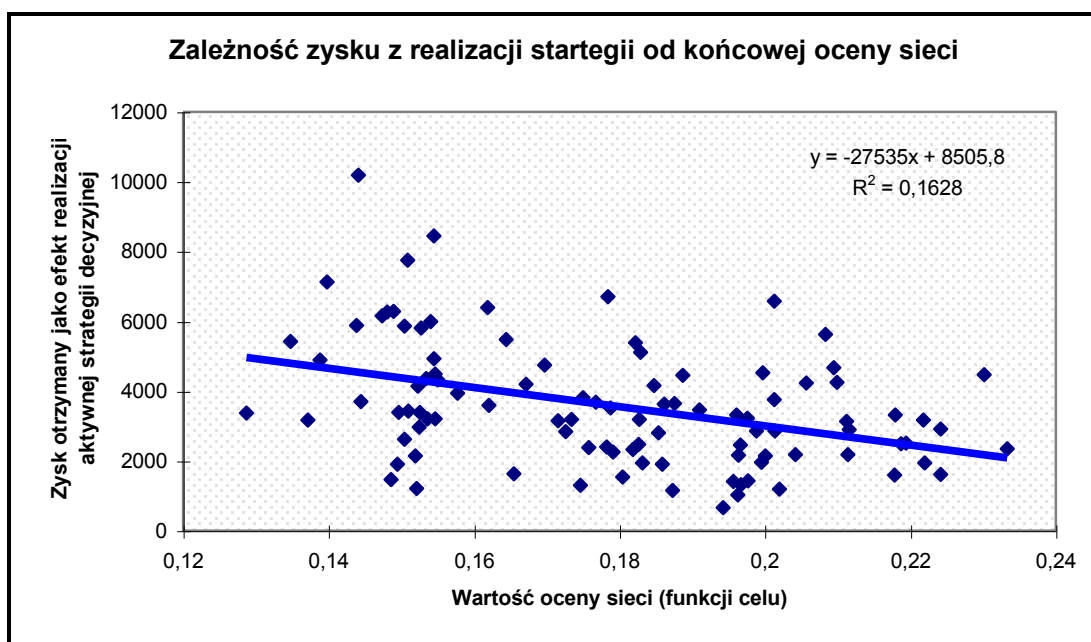


Rys. 5-4. Przyrost aktywów portfela w czasie - dla „przeciętnej” aktywnej strategii inwestycyjnej uśrednionej dla 30-tu sieci - porównanie ze strategią pasywną (indeksem WIG 20). Przedstawiono dane dla okresu IN-S oraz OUT-S przy założeniu kosztów transakcyjnych w wysokości 0,5%. Inwestycja początkowa: 1082,4. Lewa oś pionowa odnosi się do strategii aktywnej, natomiast prawa oś (posiadająca inną skalę) - do strategii pasywnej. Źródło: opracowanie własne.

5.3. Analiza wzajemnych relacji pomiędzy ilościowymi miarami efektywności sieci.

Postaramy się teraz ustalić, jakie czynniki wyznaczają sieć odznaczającą się dobrymi właściwościami w praktycznej realizacji zadań inwestycyjnych. Dokonajmy najpierw

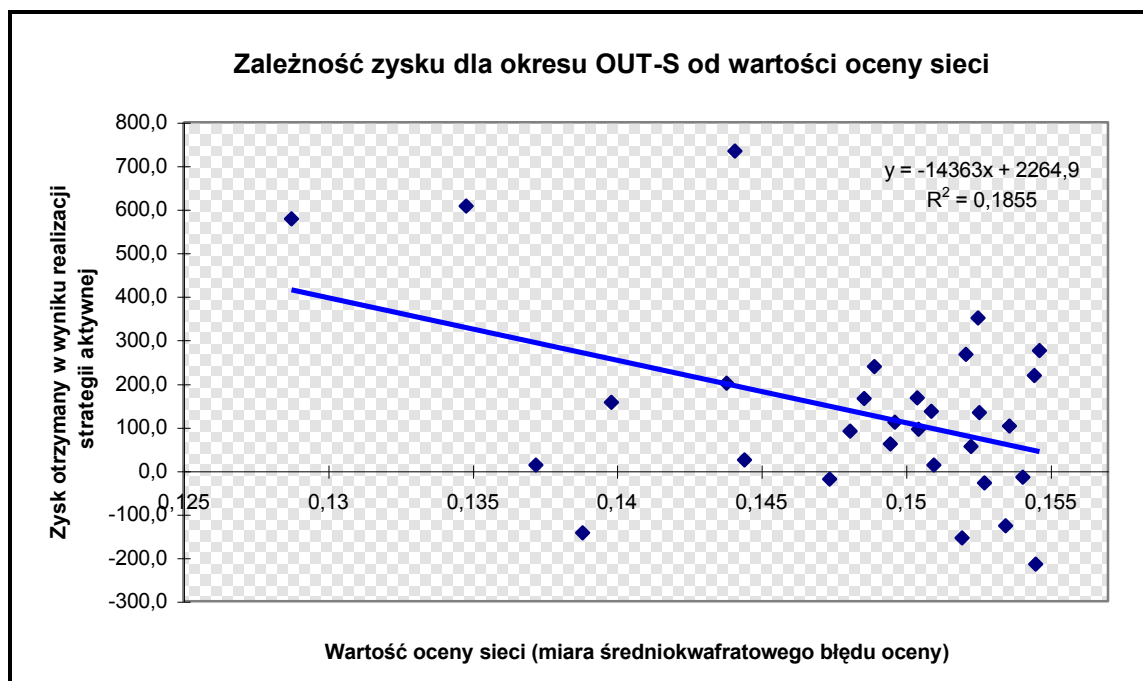
analizy wielkości zysku otrzymanego w wyniku realizacji strategii aktywnej opartej o daną sieć w okresie IN-S (2.08.94 - 30.09.97) w zależności od końcowej oceny tej sieci (wartości funkcji celu uzyskanej na etapie konstrukcji sieci podczas realizacji algorytmu genetycznego). Graficzną reprezentację tej zależności dla 100 sieci wchodzących w skład ostatniej populacji AG, wraz z analizą regresji liniowej, przedstawia rys. 5-5. Przyjęto istnienie kosztów transakcyjnych w wysokości 0,5%.



Rys. 5-5. Zależność zysku pochodzącego z zastosowania opartej na SN aktywnej strategii decyzyjnej od końcowej oceny tej SN uzyskanej na etapie konstrukcji modelu (realizacji algorytmu genetycznego) dla danych IN-S z okresu 2.08.94 - 30.09.97. Przedstawiono wyniki dla 100 sieci wchodzących w skład ostatniej populacji, oraz analizę regresji liniowej obrazującej tę zależność. Źródło: opracowanie własne.

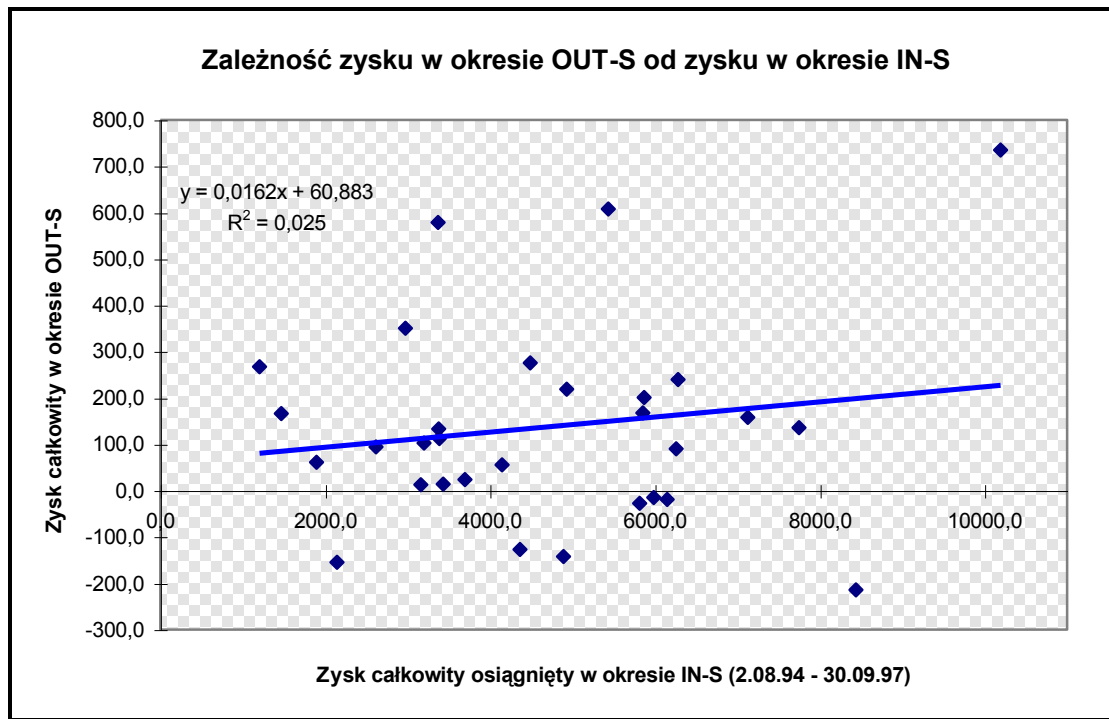
Jak wynika z powyższego wykresu, lepsza (tzn. mniejsza) ocena sieci implikuje przeciętnie lepsze efekty strategii inwestycyjnej wykorzystującej tą sieć (większe wartości zysku), co jest zgodne z oczekiwaniami. Fakt ten potwierdza kierunek nachylenia prostej regresji liniowej dla wyznaczonych danych. Jednak niewielka wartość współczynnika determinacji $R^2 = 0,1628$ wskazuje na stosunkowo słabą zależność pomiędzy tymi wielkościami (jedynie w około 16% wielkość osiąganego zysku wyjaśniana jest przez ocenę sieci), co jest także łatwo zauważalne na wykresie.

Z kolei rozważmy zyski w okresie OUT-S (1.10.97 - 19.02.98) zrealizowane przez 30 sieci posiadających najlepszą ocenę (zob. tabela 5-1) przy zerowych kosztach transakcyjnych. Na rys. 5-6 przedstawiono zależność tego zysku od wartości oceny sieci, natomiast na kolejnym rys. 5-7 zaprezentowano zależność zysku osiągniętego w rozpatrywanym okresie OUT-S od zysku wygenerowanego przez daną sieć w okresie IN-S. Próba dokonania analizy regresji liniowej ujawnia stosunkowo słabe związki pomiędzy rozważanymi wielkościami. Można stwierdzić, że zależność zysku OUT-S od wartości oceny, zobrazowana na rys. 5-6, posiada ten sam charakter co odpowiednia zależność zysku IN-S od oceny (rys. 5-5): analogiczne nachylenie prostej regresji, podobna jest też wartość współczynnika determinacji $R^2 = 0,1855$. Z kolei analiza regresji przedstawiona na rys. 5-7 pokazuje praktyczny brak związku pomiędzy zyskiem osiągniętym w okresie OUT-S, a zyskiem w okresie IN-S. Interesująco jest tu jednak uwidoczniony (znajdujący się w prawym górnym narożniku wykresu) element nietypowy (reprezentujący analizowaną poprzednio sieć nr 6), który zdecydowanie odstaje (na korzyść) od pozostałych elementów (sieci) pod względem obu przedstawionych na wykresie wartości.



Rys. 5-6. Zależność zysku pochodzącego z zastosowania aktywnej strategii decyzyjnej dla danych OUT-S (z okresu 1.10.97 - 19.02.98) od końcowej oceny danej sieci neuronowej uzyskanej na etapie konstrukcji modelu. Przedstawiono wyniki dla 30

sieci posiadających najlepszą (najmniejszą w sensie wartości) ocenę, oraz analizę regresji liniowej obrazującej tę zależność. Źródło: opracowanie własne.



Rys. 5-7. Zależność zysku otrzymanego w wyniku zastosowania danej sieci do realizacji aktywnej strategii decyzyjnej w okresie OUT-S (1.10.97 - 19.02.98) od wielkości analogicznego zysku z okresu IN-S (2.08.94 - 30.09.97). Przedstawiono wyniki dla 30 sieci, oraz analizę regresji liniowej obrazującej tę zależność. Źródło: opracowanie własne.

Brak mocnego związku pomiędzy wartością oceny sieci uzyskaną podczas jej uczenia, a efektem realizacji aktywnej strategii decyzyjnej opartej o sygnały generowane przez tę sieć może wskazywać na celowość poszukiwania i zastosowania alternatywnych metod oceny sieci podczas fazy uczenia SN oraz realizacji algorytmu genetycznego. Tradycyjne, powszechnie wykorzystywane w większości zastosowań SN reguły walidacji i testowania sieci oparte o kwadratową funkcję błędu, jakkolwiek również tutaj spełniają swoje zadanie, jednak tylko w pewnym stopniu wyjaśniają efektywność SN stanowiących elementy finansowych systemów decyzyjnych. Szczególnie słaba jest zależność zysku w okresie OUT-S od oceny sieci, oraz od zysku w okresie IN-S.

Alternatywne metody oceny sieci mogą być związane np. z bezpośrednią analizą stóp zwrotu uzyskiwanych w zadanych okresach czasu w wyniku realizacji strategii opartej o daną sieć neuronową. Pewne propozycje oraz próby wykorzystania tego typu koncepcji przedstawiono w dalszej części pracy.

5.4. Badania doboru wybranych parametrów systemu i ich wpływu na efektywność metody.

Poniżej zaprezentowano przeprowadzone przez autora badania efektywności analizowanej metodologii w zależności od doboru wybranych parametrów systemu decyzyjnego (uznanych *a priori* za potencjalnie istotne), oraz przedstawiono wyniki wybranych analiz funkcjonowania i efektywności skonstruowanych w ramach systemu sieci neuronowych. Badania te odnoszą się do rozważanego przykładu dotyczącego generowania sygnałów transakcyjnych dla indeksu WIG 20 i obejmują kolejno zagadnienia: wpływu wartości wag początkowych SN, analizy przebiegu procesu uczenia sieci, optymalnej wielkości warstwy ukrytej, celowości zastosowania analizy głównych składowych do wstępnego przetwarzania danych, doboru funkcji błędu, doboru parametrów funkcji aktywacji, oraz analizy celowości douczania sieci.

5.4.1. Zależność efektywności sieci od wartości wag początkowych.

W przedstawionych poniżej badaniach analizie poddano efekty uczenia i realizacji strategii inwestycyjnej dla pojedynczej sieci neuronowej o określonej strukturze, danych wejściowych oraz zmiennej wyjściowej, ale przy różnych zbiorach wartości wag początkowych, generowanych losowo przed rozpoczęciem procesu uczenia. Do badań przyjęto uzyskaną w wyniku realizacji algorytmu genetycznego sieć nr 6, której funkcjonowanie zostało zaprezentowane w podrozdziale 5.2 i uznane za wysoce efektywne. Sieć ta posiada 20 elementów w warstwie wejściowej (20 danych wejściowych), 7 neuronów w warstwie ukrytej, oraz 155 połączeń wagowych, których wartości są modyfikowane w procesie uczenia.

Dla tej sieci proces uczenia przeprowadzono 200 razy, rozpoczynając go za każdym razem od różnych, losowo wygenerowanych¹⁰³ wartości wag początkowych. Do uczenia i walidacji zastosowano dokładnie te same parametry i dane, które zostały przyjęte w rozważanym tu zagadnieniu generowania decyzji transakcyjnych dla indeksu WIG 20 (zob. podrozdział 5.1). Rezultaty poszczególnych 200-tu operacji uczenia i realizacji SN, uporządkowanych według rosnących (pogarszających się) wartości końcowych ocen sieci, zestawiono w tabeli 5-2. W kolejnych kolumnach tabeli wyszczególniono:

- numer kolejny procesu (od 0 do 199),
- osiągniętą wartość końcowej oceny sieci,
- zysk osiągnięty w wyniku realizacji aktywnej strategii inwestycyjnej opartej o sygnały danej sieci (zob. podrozdział 5.1) w okresie IN-S (od 2.08.1994 do 30.09.1997) przy założeniu kosztów transakcyjnych w wysokości 0,5%,
- wielkość maksymalnej straty **dd** w okresie IN-S,
- analogiczny zysk w okresie OUT-S (1.10.97 - 19.02.98) przy zerowych kosztach transakcyjnych,
- wielkość maksymalnej straty **dd** w okresie OUT-S.

W celu ułatwienia analizy danych zawartych w tabeli 5-2, oraz dokonania odpowiednich porównań z efektami strategii pasywnej, przytoczymy tu jeszcze raz następujące dane:

- wartość początkowa aktywów portfela (odpowiadająca wartości indeksu WIG 20) dla okresu IN-S: **1082,4** , dla okresu OUT-S: **1740,5**.
- zysk osiągnięty w wyniku realizacji strategii pasywnej (różnica pomiędzy końcową i początkową wartością WIG 20 w danym okresie): IN-S **650,8** , OUT-S **19,3**.
- maksymalna strata **dd** dla strategii pasywnej: IN-S **53,0** % , OUT-S **26,6** %.

¹⁰³ Wagi początkowe stanowiły liczby losowe generowane z przedziału $[-0,2 ; 0,2]$ według rozkładu jednostajnego przy użyciu generatora liczb pseudolosowych. Funkcji randomizacji używano przed każdym procesem uczenia.

TABELA 5-2

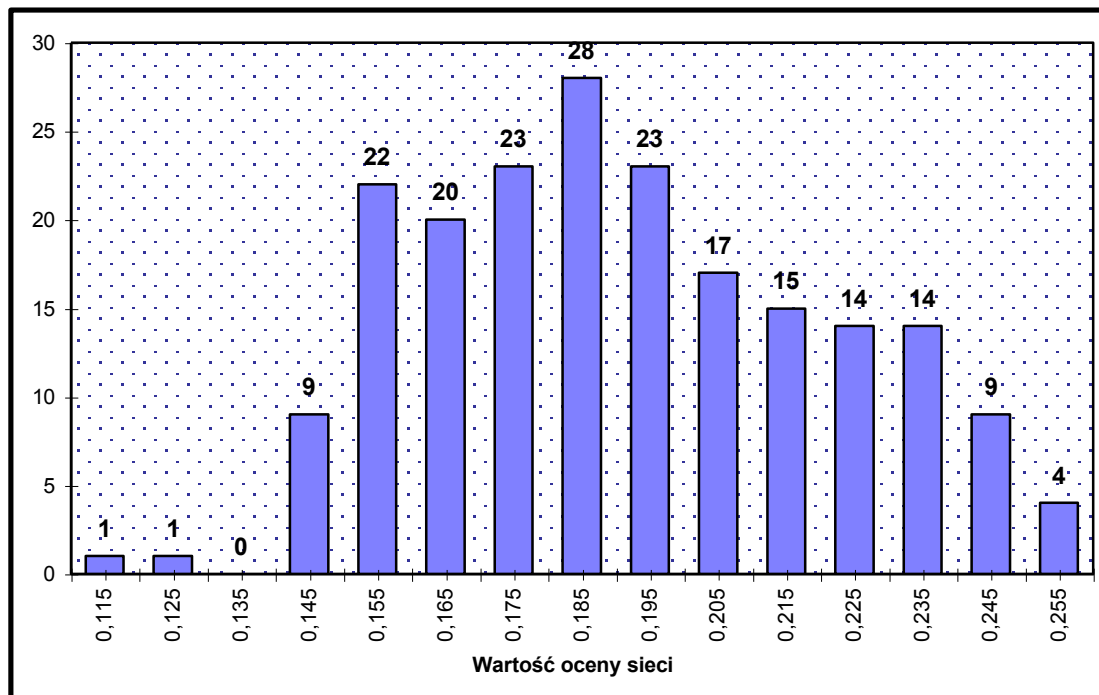
| Nr | Ocena | IN-S prow. 0,5% | | OUT-S prow. 0% | | Nr | Ocena | IN-S prow. 0,5% | | OUT-S prow. 0% | |
|----|---------|-----------------|------|----------------|------|-----|---------|-----------------|------|----------------|------|
| | | Zysk | dd % | Zysk | dd % | | | Zysk | dd % | Zysk | dd % |
| 0 | 0,11378 | 13296,1 | 10,3 | 129,3 | 13,3 | 100 | 0,18876 | 4123,4 | 18,5 | 470,9 | 7,1 |
| 1 | 0,12683 | 12660,2 | 11,6 | 549,0 | 7,4 | 101 | 0,18881 | 10290,4 | 12,8 | 577,6 | 11,1 |
| 2 | 0,14008 | 8671,2 | 16,5 | 654,9 | 11,0 | 102 | 0,18944 | 3708,5 | 24,4 | 329,6 | 9,7 |
| 3 | 0,14246 | 7868,4 | 15,1 | 561,0 | 11,0 | 103 | 0,18968 | 6428,5 | 21,8 | 205,3 | 7,3 |
| 4 | 0,14321 | 9850,1 | 17,2 | 514,3 | 12,6 | 104 | 0,19021 | 9090,8 | 11,7 | 200,0 | 12,6 |
| 5 | 0,14365 | 10971,9 | 11,4 | 524,1 | 12,3 | 105 | 0,19054 | 5287,3 | 19,6 | 56,0 | 13,1 |
| 6 | 0,14588 | 7875,4 | 16,3 | 677,4 | 11,0 | 106 | 0,19104 | 4915,7 | 14,9 | 277,6 | 7,7 |
| 7 | 0,14628 | 8396,0 | 11,2 | 415,3 | 7,4 | 107 | 0,19177 | 5174,9 | 20,5 | 637,0 | 11,1 |
| 8 | 0,14797 | 8945,1 | 13,7 | 86,9 | 10,1 | 108 | 0,19334 | 4423,0 | 20,8 | 217,7 | 8,6 |
| 9 | 0,14830 | 7497,7 | 16,9 | 568,1 | 5,7 | 109 | 0,19347 | 7326,5 | 12,7 | 416,2 | 12,6 |
| 10 | 0,14958 | 8114,0 | 17,5 | 651,1 | 11,1 | 110 | 0,19391 | 10540,4 | 13,2 | 712,7 | 5,4 |
| 11 | 0,15197 | 8735,1 | 18,6 | 560,8 | 12,3 | 111 | 0,19402 | 4239,1 | 17,6 | 465,0 | 9,6 |
| 12 | 0,15367 | 9110,0 | 12,8 | 146,7 | 12,6 | 112 | 0,19504 | 7737,7 | 13,7 | 406,0 | 12,7 |
| 13 | 0,15412 | 9900,9 | 14,1 | 303,4 | 11,0 | 113 | 0,19532 | 4376,9 | 16,4 | 146,1 | 10,1 |
| 14 | 0,15492 | 7004,3 | 16,6 | 436,1 | 11,0 | 114 | 0,19559 | 5352,8 | 11,6 | 506,6 | 9,6 |
| 15 | 0,15499 | 9800,4 | 12,1 | 185,9 | 12,3 | 115 | 0,19560 | 9843,4 | 10,8 | -83,1 | 13,0 |
| 16 | 0,15531 | 11935,3 | 11,6 | 516,9 | 11,1 | 116 | 0,19586 | 5557,7 | 13,3 | -24,7 | 15,6 |
| 17 | 0,15536 | 7947,5 | 10,1 | 173,6 | 9,8 | 117 | 0,19588 | 6303,6 | 16,9 | 273,7 | 12,6 |
| 18 | 0,15589 | 12196,2 | 15,2 | 322,8 | 12,3 | 118 | 0,19622 | 5679,9 | 18,9 | 96,8 | 11,0 |
| 19 | 0,15602 | 11571,4 | 13,2 | 453,8 | 12,3 | 119 | 0,19668 | 7921,4 | 18,7 | 121,5 | 9,7 |
| 20 | 0,15626 | 10763,1 | 10,9 | 432,1 | 7,8 | 120 | 0,19822 | 9115,2 | 18,9 | 35,6 | 11,0 |
| 21 | 0,15629 | 7508,3 | 15,8 | 298,8 | 7,7 | 121 | 0,19825 | 4790,5 | 15,6 | 297,3 | 13,1 |
| 22 | 0,15643 | 7635,8 | 19,4 | 622,7 | 11,0 | 122 | 0,19851 | 3177,8 | 27,3 | 121,3 | 12,6 |
| 23 | 0,15784 | 6819,8 | 13,9 | 72,9 | 12,6 | 123 | 0,19879 | 4763,6 | 21,8 | 235,8 | 9,7 |
| 24 | 0,15826 | 9287,7 | 10,3 | 193,9 | 11,0 | 124 | 0,19885 | 6741,1 | 17,3 | 253,7 | 11,1 |
| 25 | 0,15871 | 9540,1 | 12,8 | 469,8 | 12,3 | 125 | 0,19958 | 4731,9 | 22,9 | 337,8 | 12,6 |
| 26 | 0,15899 | 9704,9 | 11,1 | 58,6 | 15,6 | 126 | 0,19997 | 6890,6 | 14,6 | 301,1 | 12,1 |
| 27 | 0,15905 | 8676,7 | 18,9 | 621,5 | 8,4 | 127 | 0,20052 | 12837,4 | 10,7 | 398,3 | 11,0 |
| 28 | 0,15924 | 5307,8 | 14,7 | 371,6 | 12,4 | 128 | 0,20114 | 4735,3 | 19,3 | 116,4 | 11,0 |
| 29 | 0,15934 | 7738,4 | 18,0 | 760,9 | 5,7 | 129 | 0,20238 | 3435,2 | 21,6 | 177,9 | 8,4 |
| 30 | 0,15939 | 4768,9 | 21,5 | 300,1 | 7,9 | 130 | 0,20270 | 4301,2 | 19,6 | 200,5 | 12,6 |
| 31 | 0,15942 | 8856,3 | 14,1 | 111,1 | 12,8 | 131 | 0,20276 | 4743,2 | 24,4 | 198,4 | 15,0 |
| 32 | 0,15977 | 8465,9 | 14,3 | 632,6 | 12,3 | 132 | 0,20302 | 2219,3 | 26,8 | 224,1 | 22,8 |
| 33 | 0,16006 | 8010,3 | 13,8 | 266,0 | 7,4 | 133 | 0,20413 | 7378,3 | 12,0 | 217,9 | 9,1 |
| 34 | 0,16054 | 8578,7 | 11,4 | 43,3 | 12,9 | 134 | 0,20437 | 13589,7 | 10,8 | 160,6 | 18,6 |
| 35 | 0,16141 | 9667,8 | 11,6 | 431,0 | 7,4 | 135 | 0,20521 | 3515,1 | 22,1 | 179,9 | 18,0 |
| 36 | 0,16152 | 8691,8 | 11,6 | 551,9 | 9,6 | 136 | 0,20521 | 5218,9 | 15,5 | 185,7 | 11,1 |
| 37 | 0,16173 | 8869,5 | 10,1 | 441,1 | 5,7 | 137 | 0,20542 | 4466,0 | 19,4 | 156,9 | 13,8 |
| 38 | 0,16179 | 10012,3 | 11,6 | 552,9 | 11,0 | 138 | 0,20558 | 3843,7 | 17,3 | 226,9 | 9,1 |
| 39 | 0,16201 | 9552,1 | 11,6 | 130,4 | 11,1 | 139 | 0,20573 | 7204,0 | 10,2 | 291,7 | 11,1 |
| 40 | 0,16219 | 8833,0 | 15,0 | 576,5 | 11,0 | 140 | 0,20671 | 5739,8 | 12,1 | 428,6 | 6,0 |
| 41 | 0,16247 | 5987,6 | 17,9 | 316,6 | 11,2 | 141 | 0,20682 | 4085,1 | 18,9 | 226,2 | 6,3 |
| 42 | 0,16497 | 6681,8 | 17,2 | 412,8 | 14,2 | 142 | 0,20738 | 5233,1 | 14,6 | 363,8 | 9,7 |
| 43 | 0,16551 | 8989,4 | 9,8 | 178,6 | 9,7 | 143 | 0,20962 | 2986,8 | 20,4 | 434,0 | 11,1 |
| 44 | 0,16570 | 7196,1 | 13,3 | 312,8 | 12,6 | 144 | 0,21067 | 2978,9 | 29,2 | 185,3 | 15,5 |
| 45 | 0,16674 | 8009,8 | 10,6 | 233,0 | 13,5 | 145 | 0,21381 | 6776,0 | 17,9 | 196,1 | 12,6 |
| 46 | 0,16744 | 5841,8 | 13,1 | 318,5 | 7,7 | 146 | 0,21385 | 6535,0 | 19,4 | 392,9 | 11,1 |
| 47 | 0,16779 | 8574,4 | 9,9 | 456,5 | 11,0 | 147 | 0,21404 | 8871,8 | 11,8 | 207,5 | 12,3 |
| 48 | 0,16797 | 11249,7 | 10,6 | 272,5 | 13,7 | 148 | 0,21419 | 3541,7 | 21,8 | 289,6 | 7,7 |
| 49 | 0,16808 | 8126,2 | 17,7 | 511,5 | 5,7 | 149 | 0,21423 | 3625,6 | 23,1 | 275,8 | 12,6 |
| 50 | 0,16872 | 9287,2 | 17,5 | 253,5 | 12,6 | 150 | 0,21605 | 4660,5 | 18,9 | 85,1 | 9,7 |
| 51 | 0,16899 | 8951,5 | 18,2 | 494,3 | 11,1 | 151 | 0,21731 | 4797,8 | 16,9 | 23,6 | 13,3 |
| 52 | 0,16986 | 8811,2 | 18,9 | 134,9 | 12,6 | 152 | 0,21752 | 3912,8 | 21,6 | 239,3 | 9,1 |
| 53 | 0,17004 | 6440,8 | 18,9 | 95,1 | 12,6 | 153 | 0,21818 | 4029,7 | 19,1 | 375,4 | 5,8 |
| 54 | 0,17091 | 9169,7 | 9,9 | 263,5 | 5,7 | 154 | 0,21891 | 1009,1 | 48,8 | 19,3 | 26,6 |
| 55 | 0,17107 | 3940,8 | 15,1 | 423,2 | 18,0 | 155 | 0,21919 | 2978,5 | 33,7 | 213,7 | 13,6 |
| 56 | 0,17349 | 9616,9 | 11,1 | 191,2 | 12,6 | 156 | 0,21948 | 2166,2 | 20,6 | -208,3 | 23,5 |
| 57 | 0,17359 | 9361,1 | 11,0 | 250,9 | 14,6 | 157 | 0,21958 | 4585,6 | 17,2 | 436,0 | 11,1 |
| 58 | 0,17425 | 9095,1 | 15,3 | 160,4 | 9,7 | 158 | 0,21974 | 3152,6 | 17,0 | 241,2 | 11,9 |
| 59 | 0,17476 | 9282,4 | 13,4 | 452,3 | 12,6 | 159 | 0,22121 | 2654,8 | 28,4 | -73,9 | 15,3 |
| 60 | 0,17491 | 6894,6 | 15,6 | 671,9 | 7,8 | 160 | 0,22137 | 6141,0 | 15,7 | 26,4 | 20,3 |
| 61 | 0,17504 | 6733,4 | 11,8 | 42,3 | 12,6 | 161 | 0,22158 | 4277,6 | 15,0 | 291,2 | 11,5 |
| 62 | 0,17522 | 8747,0 | 15,4 | 594,8 | 12,3 | 162 | 0,22159 | 3893,3 | 22,7 | 306,3 | 5,7 |
| 63 | 0,17600 | 10038,4 | 18,9 | 96,1 | 11,0 | 163 | 0,22192 | 6255,8 | 18,8 | -35,0 | 20,3 |
| 64 | 0,17709 | 5432,1 | 19,6 | 145,9 | 7,3 | 164 | 0,22214 | 3593,9 | 22,1 | 222,4 | 11,3 |
| 65 | 0,17796 | 7082,6 | 15,9 | 423,3 | 12,3 | 165 | 0,22294 | 3922,6 | 19,4 | 505,3 | 5,9 |
| 66 | 0,17881 | 6309,1 | 18,9 | 194,6 | 9,7 | 166 | 0,22371 | 10413,7 | 10,2 | 11,2 | 15,9 |

| Nr | Ocena | IN-S prow. 0,5% | | OUT-S prow. 0% | | Nr | Ocena | IN-S prow. 0,5% | | OUT-S prow. 0% | |
|----|---------|-----------------|------|----------------|------|-----|---------|-----------------|------|----------------|------|
| | | Zysk | dd % | Zysk | dd % | | | Zysk | dd % | Zysk | dd % |
| 67 | 0,17894 | 8487,6 | 13,2 | 65,7 | 12,1 | 167 | 0,22422 | 2257,6 | 31,0 | 359,8 | 13,6 |
| 68 | 0,17895 | 3625,6 | 19,1 | 334,2 | 18,0 | 168 | 0,22490 | 1926,2 | 25,8 | 82,7 | 13,6 |
| 69 | 0,17928 | 8394,2 | 15,4 | 577,2 | 11,9 | 169 | 0,22859 | 4308,9 | 21,8 | 420,3 | 5,7 |
| 70 | 0,17930 | 9016,2 | 9,9 | 222,0 | 5,7 | 170 | 0,22896 | 1335,7 | 42,7 | -38,9 | 26,6 |
| 71 | 0,17951 | 9090,2 | 12,2 | 100,9 | 9,8 | 171 | 0,22928 | 948,4 | 52,8 | 19,3 | 26,6 |
| 72 | 0,17988 | 8473,0 | 11,2 | 100,1 | 20,3 | 172 | 0,22933 | 5676,3 | 15,3 | 124,2 | 13,7 |
| 73 | 0,17992 | 11215,8 | 9,9 | 162,2 | 12,3 | 173 | 0,23030 | 4505,0 | 16,6 | 309,0 | 5,6 |
| 74 | 0,17994 | 8492,4 | 13,9 | -10,7 | 13,5 | 174 | 0,23177 | 5376,6 | 18,9 | 206,8 | 12,6 |
| 75 | 0,17999 | 5807,0 | 18,7 | 356,6 | 9,6 | 175 | 0,23356 | 2803,0 | 24,7 | -16,8 | 29,7 |
| 76 | 0,18065 | 5669,7 | 20,2 | 491,6 | 12,6 | 176 | 0,23402 | 1678,7 | 29,3 | 106,5 | 19,8 |
| 77 | 0,18118 | 9210,3 | 10,3 | 523,2 | 8,4 | 177 | 0,23468 | 2091,3 | 32,9 | 206,6 | 5,8 |
| 78 | 0,18119 | 8100,3 | 13,1 | 330,0 | 7,7 | 178 | 0,23495 | 2793,0 | 32,2 | -1,1 | 11,1 |
| 79 | 0,18171 | 9648,3 | 12,1 | 443,6 | 14,9 | 179 | 0,23560 | 4700,6 | 19,4 | 309,0 | 9,1 |
| 80 | 0,18191 | 6604,3 | 18,9 | 226,1 | 12,6 | 180 | 0,23665 | 1700,5 | 27,3 | 31,1 | 9,4 |
| 81 | 0,18308 | 6773,9 | 10,3 | 331,6 | 9,1 | 181 | 0,23709 | 2202,6 | 29,0 | -218,2 | 18,5 |
| 82 | 0,18318 | 5887,7 | 17,1 | 189,3 | 12,6 | 182 | 0,23747 | 5600,6 | 15,5 | -17,2 | 13,3 |
| 83 | 0,18327 | 7678,3 | 20,2 | 191,8 | 17,2 | 183 | 0,23762 | 3407,8 | 24,8 | 79,2 | 14,3 |
| 84 | 0,18346 | 5232,9 | 19,8 | 327,5 | 8,6 | 184 | 0,23795 | 2286,6 | 31,6 | -3,1 | 9,5 |
| 85 | 0,18394 | 5965,8 | 16,8 | 202,6 | 9,7 | 185 | 0,23902 | 1631,8 | 24,2 | 240,1 | 9,9 |
| 86 | 0,18401 | 4508,5 | 18,9 | 475,2 | 9,7 | 186 | 0,23972 | 5832,6 | 12,1 | 198,7 | 7,7 |
| 87 | 0,18403 | 6777,4 | 12,5 | 336,0 | 7,7 | 187 | 0,24029 | 924,9 | 38,6 | -135,3 | 18,0 |
| 88 | 0,18457 | 7508,0 | 22,1 | 276,0 | 11,1 | 188 | 0,24121 | 2156,5 | 22,4 | -285,0 | 21,8 |
| 89 | 0,18549 | 8341,8 | 10,8 | 113,6 | 11,0 | 189 | 0,24151 | 3344,6 | 15,1 | -122,1 | 19,1 |
| 90 | 0,18606 | 6433,5 | 19,6 | 143,1 | 9,7 | 190 | 0,24168 | 1639,2 | 28,0 | -9,6 | 27,8 |
| 91 | 0,18629 | 6602,1 | 19,6 | 45,7 | 15,9 | 191 | 0,24515 | 2459,4 | 22,1 | 107,5 | 6,9 |
| 92 | 0,18647 | 7728,7 | 18,7 | 273,8 | 12,6 | 192 | 0,24543 | 773,5 | 45,5 | 235,1 | 18,0 |
| 93 | 0,18699 | 6095,8 | 13,0 | 150,4 | 8,4 | 193 | 0,24662 | 1600,0 | 33,4 | 19,3 | 26,6 |
| 94 | 0,18741 | 3940,1 | 19,7 | 537,5 | 11,1 | 194 | 0,24681 | 6208,5 | 18,7 | 296,0 | 12,6 |
| 95 | 0,18742 | 9829,0 | 13,6 | -6,1 | 13,2 | 195 | 0,24688 | 1578,6 | 34,5 | -168,5 | 19,6 |
| 96 | 0,18749 | 4427,9 | 22,2 | 323,0 | 7,4 | 196 | 0,25145 | 5633,4 | 22,8 | 507,2 | 9,3 |
| 97 | 0,18750 | 7454,1 | 22,4 | 558,2 | 11,0 | 197 | 0,25457 | 1837,5 | 26,1 | 159,2 | 18,2 |
| 98 | 0,18770 | 6223,5 | 19,6 | 376,0 | 13,5 | 198 | 0,25522 | 2149,5 | 28,6 | 202,4 | 9,5 |
| 99 | 0,18782 | 6351,9 | 15,9 | 379,3 | 12,6 | 199 | 0,25589 | 2602,6 | 28,1 | 138,8 | 11,1 |

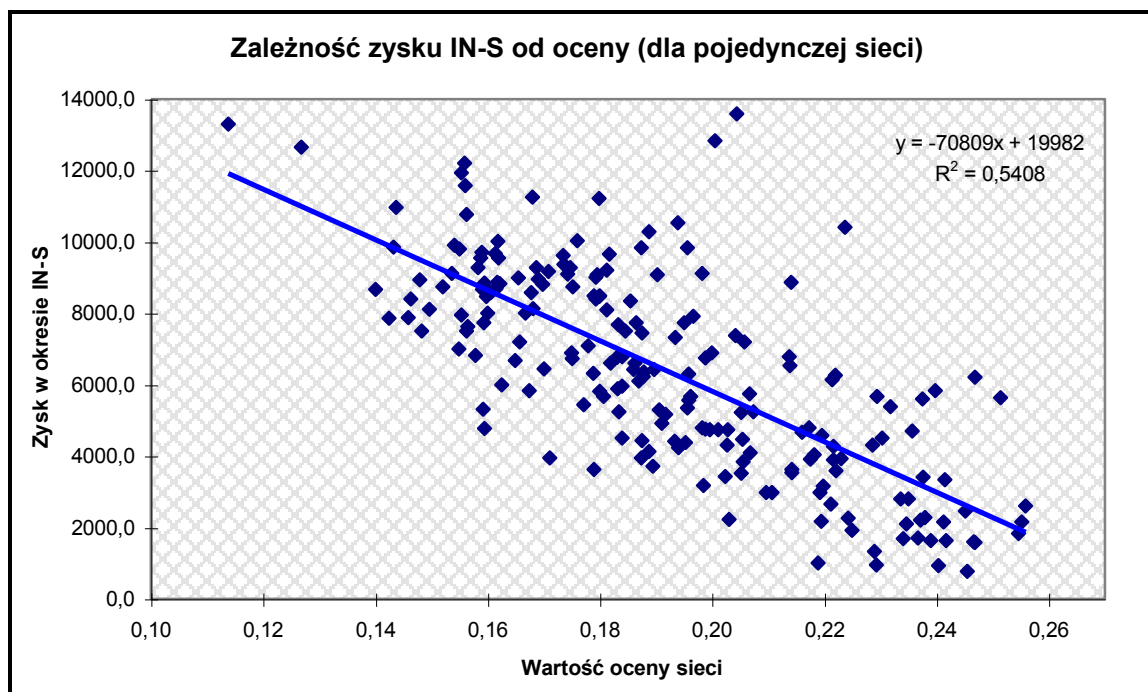
Źródło: opracowanie własne.

Z danych zawartych w tabeli 5-2 wynika, że efekty uczenia SN, oraz późniejszej realizacji strategii opartej o sygnały nauczonej sieci, w sposób wyraźny zależą od wartości początkowych wag tej sieci. Wynika to z faktu, że proces uczenia w każdym przypadku zostaje zakończony (przy założonych kryteriach zatrzymania tego procesu - zob. podrozdział 4.6) w pobliżu innego lokalnego minimum funkcji oceny, dając w efekcie inną końcową wartość oceny sieci. Wartości te wahają się tutaj w dość szerokim zakresie od **0,11378** do **0,25589** (średnia wartość oceny: **0,19210**). Na rys. 5-8 zaprezentowano histogram obrazujący rozkład wielkości końcowych ocen sieci dla przeprowadzonych 200-tu operacji uczenia SN.

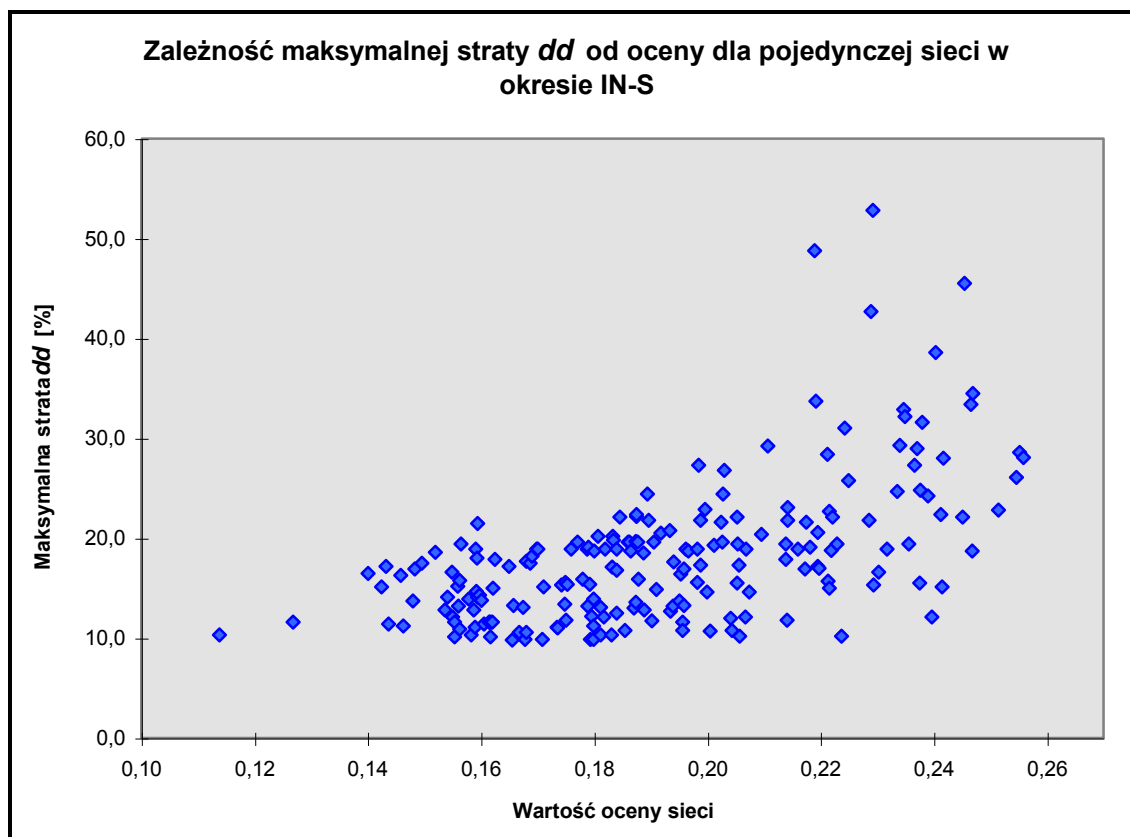
Następnie zbadano zależność zysku osiągniętego w okresie IN-S w wyniku realizacji aktywnej strategii opartej o sygnały nauczonej sieci, od oceny tej sieci w każdym z 200-tu przypadków, dołączając odpowiednią analizę regresji liniowej (rys. 5-9). Dokonano również graficznej analizy zależności maksymalnej straty **dd** w okresie IN-S od oceny sieci (rys. 5-10).



Rys. 5-8. Histogram prezentujący rozkład wartości ocen sieci, otrzymany w przypadku 200 niezależnych operacji uczenia tej samej sieci neuronowej przy różnych wartościach wag początkowych. Źródło: opracowanie własne.



Rys. 5-9. Zależność zysku osiągniętego w okresie IN-S w wyniku zastosowania strategii aktywnej, od oceny sieci generującej sygnały transakcyjne dla tej strategii. Na wykresie przedstawiono też prostą i równanie regresji liniowej dla tej zależności. Źródło: opracowanie własne.



Rys. 5-10. Zależność maksymalnej straty *dd* w okresie IN-S od wartości oceny sieci.

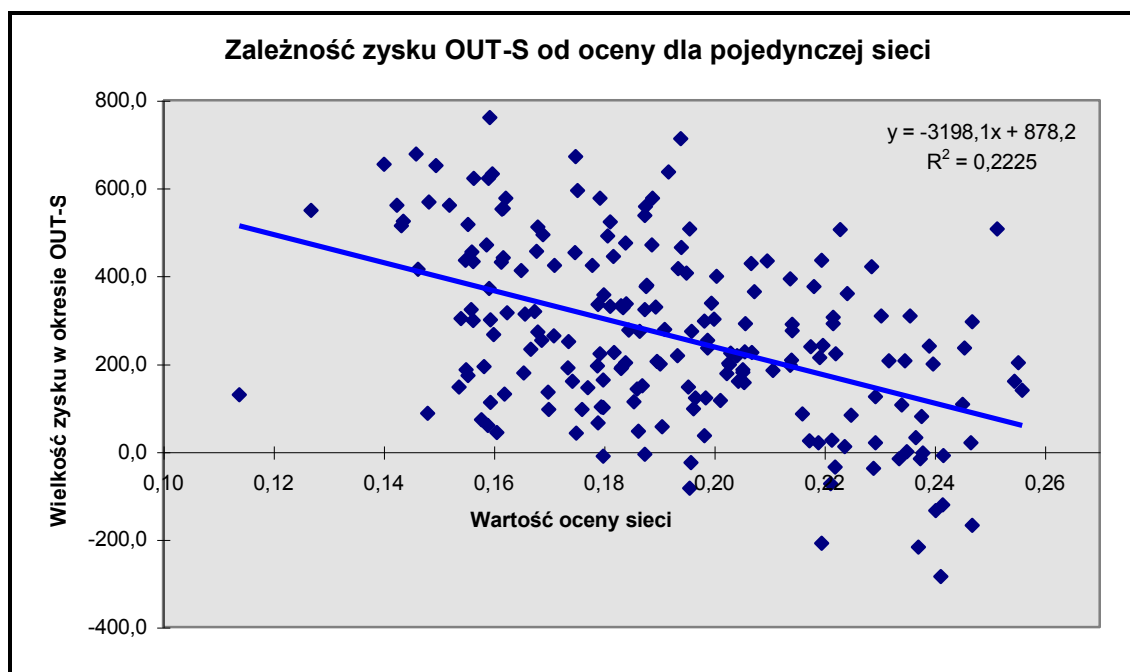
Źródło: opracowanie własne.

Z analizy rys. 5-9 wynika dość wyraźna zależność pomiędzy oceną sieci, uzyskaną w efekcie uczenia, a zyskiem w okresie IN-S osiągniętym dzięki realizacji decyzji transakcyjnych wygenerowanych przez daną sieć. Zależność tą można zobrazować prostą regresji, dla której uzyskano współczynnik determinacji $R^2 \approx 0,54$.

Interesujące wnioski można też wyciągnąć z analizy rys. 5-10. Widoczne jest występowanie pewnego poziomu (o wartości około 10%), poniżej którego nie otrzymano żadnej wartości maksymalnej straty *dd*. Poziom ten można zinterpretować jako minimalne ryzyko (wielkość straty), z jakim zawsze musi liczyć się inwestor stosując określoną strategię opartą o nawet najlepiej nauczoną daną sieć neuronową. Wyraźnie przeciętnie lepsze (mniejsze) wartości *dd* zanotowano dla przypadków, w których uzyskano lepsze (mniejsze) wartości oceny sieci. Z kolei w sytuacjach gorszych ocen końcowych sieci (prawa część wykresu) odpowiednie wielkości

maksymalnej straty *dd* są nie tylko przeciętnie wyższe, ale ich rozrzut jest znacznie większy (wskazuje to na znaczny wzrost ryzyka związanego ze stosowaniem „źle” nauczonej sieci).

Zbadajmy jeszcze zależność zysku osiągniętego w okresie OUT-S przy zastosowaniu aktywnej strategii (zerowe koszty transakcyjne) od wartości oceny sieci. Zależność tą (wraz z analizą regresji liniowej) zaprezentowano na rys. 5-11. Generalnie posiada ona podobny charakter, jak w przypadku analogicznego zysku otrzymanego w okresie IN-S (rys. 5-9): poprawa, tzn. spadek wartości oceny sieci implikuje przeciętnie wyższe wartości zysku, fakt ten obrazuje nachylenie prostej regresji. Jednak w tym przypadku związek zysku w okresie OUT-S z oceną sieci jest słabszy (współczynnik determinacji $R^2 \approx 0,22$).



Rys. 5-11. Zależność zysku osiągniętego w okresie OUT-S w wyniku zastosowania strategii aktywnej, od oceny danej sieci, wraz z analizą regresji liniowej. *Źródło: opracowanie własne.*

Interpretacja wyżej zaprezentowanych rezultatów badań prowadzi do następujących wniosków:

- Aktywna strategia inwestycyjna oparta o sygnały wyjściowe uzyskane z rozważanej sieci neuronowej przyniosła przeciętnie stosunkowo dobre wyniki

zarówno w okresie IN-S (średni zysk **6379,47** , średnia stopa zwrotu dla całego okresu **589,4%**, co daje średnią miesięczną stopę zwrotu¹⁰⁴ w wysokości **4,76%**), jak i w okresie OUT-S (średni zysk **263,84** , średnia stopa zwrotu dla całego okresu **15,2%**, co przynosi średnią miesięczną stopę zwrotu w wysokości **3,07%**). Należy przy tym pamiętać, że architektura tej sieci oraz zestaw danych wejściowych zostały otrzymane w wyniku realizacji algorytmu genetycznego.

- Zysk w okresie IN-S, zysk w okresie OUT-S, oraz maksymalna strata *dd* dla generowanej przez sieć aktywnej strategii inwestycyjnej wykazują zależność od uzyskanej oceny sieci (wartości te są przeciętnie tym lepsze, im lepsza jest ocena sieci), z kolei wartość końcowej oceny zależy od wag początkowych.
- Zależność zysku w okresie OUT-S od oceny sieci jest stosunkowo słaba (fakt ten potwierdza m.in. wnioski z badań opisanych w podrozdziale 5.3).
- W celu uzyskania jak najlepszej oceny sieci (i tym samym możliwie najlepszych efektów realizacji sieci) należy w trakcie konstrukcji (uczenia) sieci wielokrotnie powtarzać proces uczenia rozpoczynając go od różnych zestawów wartości początkowych wag. Ilość tych powtórzeń, wykonywanych na etapie realizacji algorytmu genetycznego, jest definiowana przez parametr *P* (zob. podrozdział 4.6), który okazuje się być ważną wielkością decydującą o efektywności całego procesu konstrukcji systemu decyzyjnego. Oczywiście należy dążyć do określenia możliwie dużej wartości parametru *P*, jednak ograniczeniem jest tu czas realizacji powtarzanych operacji uczenia dla tej samej sieci i tym samym łączny czas konstrukcji systemu (wykonywania algorytmu genetycznego), który jest w przybliżeniu proporcjonalny do wielkości *P* (w analizowanym tutaj przypadku przyjęto *P* = 8).
- Po zakończeniu algorytmu genetycznego na ogół celowe jest przeprowadzenie dla otrzymanych wynikowych sieci neuronowych możliwie dużej liczby dodatkowych operacji uczenia tych sieci (przy ustalonej architekturze sieci, danych wejściowych i parametrach uczenia, ale dla różnych zbiorów wag początkowych). Pozwoli to

¹⁰⁴ Wielkość ta została tu wyznaczona (w celach porównawczych) tak jak miesięczna stopa oprocentowania, przy której po upływie całego analizowanego okresu, dla danej początkowej wartości aktywów, osiągniemy określony średni zysk (dla IN-S: 6379,47; dla OUT-S: 263,84).

na ewentualną poprawę efektywności funkcjonowania tych sieci. Jednak realizacja tej operacji dla już dobrze przetestowanych sieci wykazujących wysoką efektywność w okresie OUT-S, może okazać się niecelowa i przyczynić się do pogorszenia właściwości sieci (zob. podrozdział 5.4.8)!

- Interesującym i ważnym tematem badawczym może być zagadnienie doboru właściwego (nielosowego) zestawu wag początkowych dla sieci neuronowych generujących prognozy lub sygnały transakcyjne dla finansowych szeregów czasowych. Zagadnienie to nie będzie jednak rozważane w tej pracy.

5.4.2. Analiza przykładowego procesu uczenia sieci neuronowej.

Na rys. 5-12 zaprezentowano przebieg przykładowego procesu uczenia sieci neuronowej rozważanej w podrozdziale 5.4.1, z zastosowaniem tych samych danych uczących i parametrów uczenia. Zaprezentowano zmiany jednostkowego¹⁰⁵ błędu uczenia, jednostkowego błędu walidacji, oraz bieżącej wartości oceny sieci w kolejnych epokach uczenia. Wartość oceny sieci jest wyrażona zależnością $4 \cdot \Theta$, gdzie wielkość Θ jest określona wzorem (4.15) przy parametrach p i q wyznaczanych według reguły zaproponowanej w podrozdziale 5.1. Formułę obliczania oceny sieci można zatem przedstawić następująco:

$$\text{Ocena} = 4\Theta = 4 \cdot \frac{p \cdot \frac{E_u}{i_u} + q \cdot \frac{E_v}{i_v}}{p + q}, \quad \text{gdzie: } p = 1, \quad q = \frac{E_v i_u}{E_u i_v}$$

gdzie: E_u, E_v - bieżące wartości całkowitych błędów uczenia i walidacji,

i_u, i_v - ilości elementów w ciągu odpowiednio uczącym i walidacyjnym,

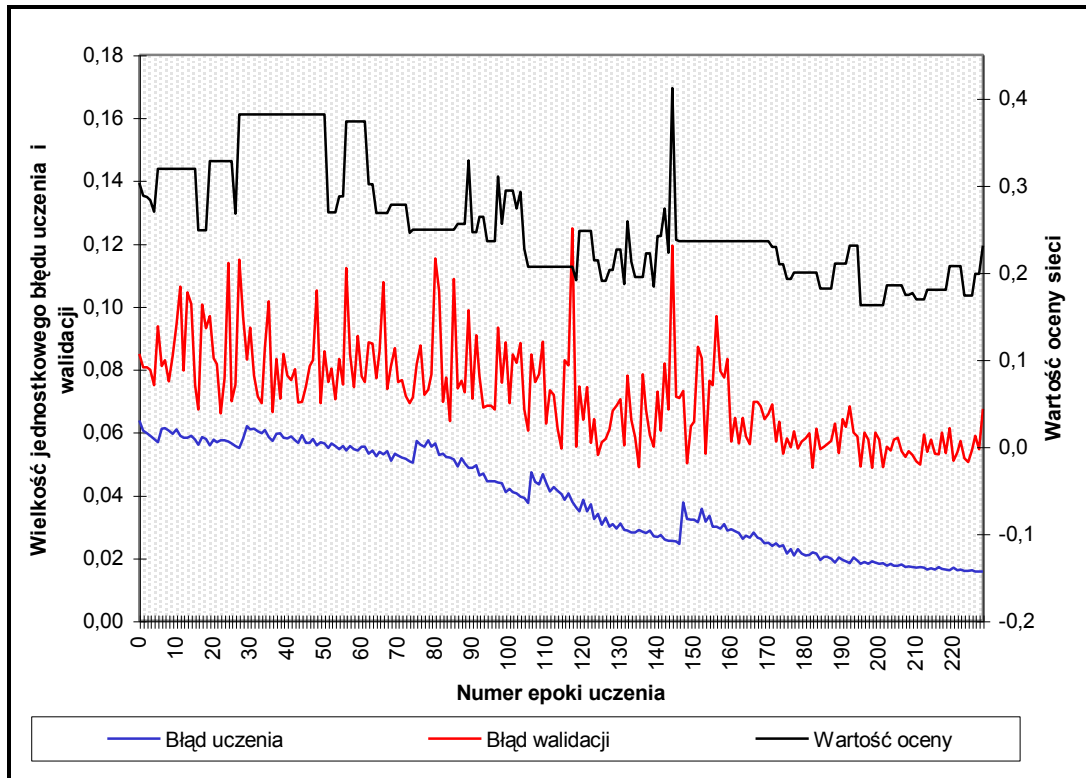
Jeżeli jednostkowe błędy uczenia i walidacji oznaczmy odpowiednio:

$$e_u = \frac{E_u}{i_u} \quad e_v = \frac{E_v}{i_v}$$

¹⁰⁵ Przez błąd jednostkowy rozumiemy tutaj błąd przypadający na jeden element danego ciągu, np. jednostkowy błąd uczenia otrzymujemy przez podzielenie całkowitego błędu uczenia przez ilość elementów ciągu uczącego.

to formułę służącą do wyznaczania wartości oceny sieci można wyrazić też wzorem¹⁰⁶:

$$Ocena = 4\Theta = 4 \frac{e_u^2 + e_v^2}{e_u + e_v}$$



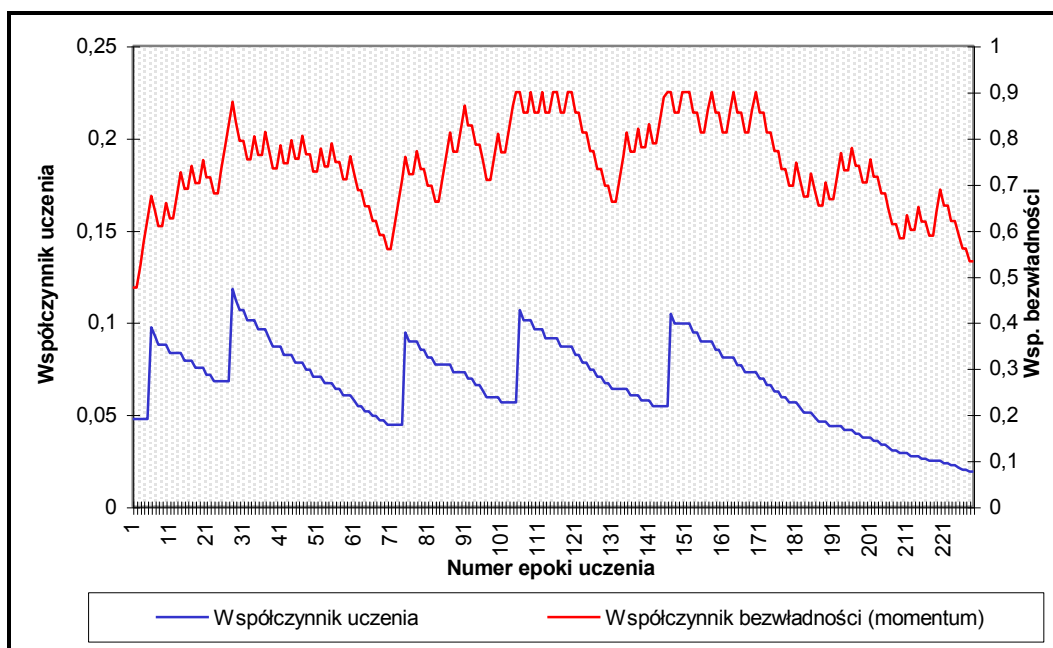
Rys. 5-12. Przebieg przykładowego procesu uczenia sieci neuronowej w rozważanym systemie decyzyjnym dla indeksu WIG 20. Przedstawiono zmiany jednostkowego (przypadającego na jeden element ciągu) błędu uczenia i błędu walidacji (lewa oś pionowa), oraz wartości oceny sieci (prawa oś). Skokowe zmiany błędu uczenia występują w momentach skokowej modyfikacji (wzrostu) współczynnika uczenia η .
Źródło: opracowanie własne.

Wartość oceny jest obliczana i modyfikowana tylko wtedy, gdy błąd uczenia osiągnie swoje bieżące minimum, tzn. najniższą wartość od początku procesu uczenia. Z kolei osiągnięcie bieżącego minimum przez wielkość oceny sieci powoduje zapamiętanie

¹⁰⁶ Warto zauważyć, że spośród wszystkich możliwych par wartości błędów jednostkowych (e_u, e_v) takich, że ich suma $e_u + e_v$ jest stała, ocena osiąga wartość minimalną (najlepszą) dla przypadku $e_u = e_v$. Właściwość ta preferuje sytuacje, gdy jednostkowe błędy uczenia i walidacji są podobne (różnią się możliwie mało), co z kolei zapobiega w pewien sposób występowaniu efektu przeuczenia sieci (charakteryzującego się znaczącym

w danej chwili aktualnego zestawu wag jako najlepszego dotychczas uzyskanego zbioru wag sieci. Ostatni zapamiętany podczas procesu uczenia najlepszy zestaw wag determinuje ostateczne wartości połączeń wagowych w nauczonej sieci.

Na kolejnym rysunku 5-13 przedstawiono zmiany współczynnika uczenia η oraz współczynnika bezwładności μ podczas prezentowanego przykładowego procesu uczenia SN. Zmiany te odbywają się według reguł podanych w podrozdziale 4.6.



Rys. 5-13. Przebiegi zmian współczynnika uczenia η oraz współczynnika bezwładności μ (momentum) w trakcie analizowanego procesu uczenia. Źródło: opracowanie własne.

W przedstawionym tutaj przykładzie zrealizowanych zostało 228 epok uczenia sieci. Najlepszą wartość oceny sieci, wynoszącą **0,162620**, osiągnięto w 195-tej epoce. Przyczyną zatrzymania procesu uczenia było tutaj pogorszenie (znaczący wzrost) wartości oceny sieci (po przekroczeniu limitu 150 epok), spełniające pierwsze kryterium przerwania uczenia (zob. podrozdział 4.6).

spadkiem błędu uczenia przy jednoczesnym wzroście błędu walidacji, a więc nadmiernym zwiększeniem się różnicy tych błędów).

Analizując przebieg zademonstrowanego procesu uczenia można zauważyć następujące fakty:

- błąd walidacji charakteryzuje się dużą zmiennością (wariancją); zmienność ta zmniejsza się wyraźnie w końcowym etapie uczenia,
- błąd uczenia spada znacznie szybciej niż błąd walidacji; może to świadczyć o tendencji do występowania efektu przeuczenia sieci (środkiem zapobiegawczym użytym w relacjonowanych badaniach jest zaproponowana reguła wyznaczania oceny sieci),
- widoczne w pewnych momentach skokowe wzrosty błędu uczenia mogą być tu spowodowane zbyt dużymi modyfikacjami (wzrostami) współczynnika uczenia η w tych właśnie chwilach (według reguł określonych w podrozdziale 4.6); zjawisko to nie wydaje się mieć znaczącego wpływu na końcowy efekt uczenia (wartości błędów uczenia i walidacji stabilizują się), jednak może niepotrzebnie go wydłużyć.

Powyższa analiza może prowadzić do następujących wniosków:

- Podczas konstrukcji finansowych systemów predykcyjnych lub decyzyjnych opartych o sieci neuronowe, celowe jest przeprowadzenie graficznej analizy procesu uczenia sieci - operacja ta ułatwić może optymalizację ważnych parametrów uczenia (takich jak np. współczynniki uczenia η i bezwładności μ) oraz reguł ich modyfikacji (te wielkości i reguły mogą w dużym stopniu zależeć od specyfiki rozpatrywanego problemu).
- Ważnym elementem systemu jest dobór właściwych kryteriów zatrzymania procesu uczenia. Jak próbowano wykazać wyżej, zagadnienie to również może być rozwiązywane na podstawie graficznej analizy przebiegu uczenia. Należy jednak zachować ostrożność przy wyborze kryteriów zakończenia uczenia opartych o wzrost błędu walidacji, błąd ten może bowiem wykazywać dużą amplitudę wahań - po gwałtownym wzroście jego wartości często następuje równie gwałtowny spadek. Dobrym kryterium wydaje się być brak poprawy bieżącej oceny sieci w trakcie określonej liczby epok. Kryterium to ma proste uzasadnienie: ze względu na czas obliczeń niecelowa jest realizacja zbyt dużej

liczby epok przy braku poprawy oceny sieci. Jako interesujące, ale godne dalszej analizy, kryterium przerwania uczenia można też zaproponować spadek zmienności (wariancji) błędu walidacji poniżej określonej wartości.

- Regułę wyznaczania bieżącej oceny sieci należy dobrać w taki sposób, aby wyeliminować tendencję do występowania efektu przeuczenia sieci, ale jednocześnie zmniejszyć jej wrażliwość na chwilowe oscylacje błędu.

5.4.3. Dylemat doboru optymalnej wielkości warstwy ukrytej.

W niniejszym podrozdziale zaprezentowano wyniki badań wpływu wielkości warstwy ukrytej SN na jej efektywność, wyrażoną wartością oceny danej sieci. W badaniach wykorzystano 30 sieci, które w rozważanym tu zagadnieniu decyzyjnym, po zakończeniu realizacji algorytmu genetycznego, uzyskały najlepsze oceny. Przez pojęcie „sieć” należy tu rozumieć właściwie zestaw określonych przez AG zmiennych wejściowych, gdyż struktura sieci (wymiar warstwy ukrytej) podlegała właśnie zmianom w ramach referowanych badań. Pozostałe parametry sieci, a także zmienną wyjściową oraz parametry procesu uczenia, pozostawiono podczas badań bez zmian, żeby nie maskować poszukiwanego efektu wpływu rozmiaru warstwy ukrytej na wyniki.

W trakcie badań dokonano ponownego uczenia każdej sieci, rozpoczynając te procesy od losowo generowanych wag początkowych (dla danej sieci proces uczenia realizowano dziesięciokrotnie - tzn. przyjęto parametr $P = 10$). Operacje uczenia wszystkich sieci powtarzano dla różnych, arbitralnie określanych, rozmiarów warstwy ukrytej, przy czym rozmiar ten definiowano nie w formie bezwzględnej liczby neuronów, a w postaci stosunku liczby elementów warstwy ukrytej do liczby elementów warstwy wejściowej (oznaczmy ten stosunek symbolem ω)¹⁰⁷. Badania zrealizowano dla wybranych wartości ω z zakresu od zera (brak warstwy ukrytej) do 2 (por. dyskusja w podrozdziałach 2.4, 2.7.3). Szczegółowe rezultaty badań

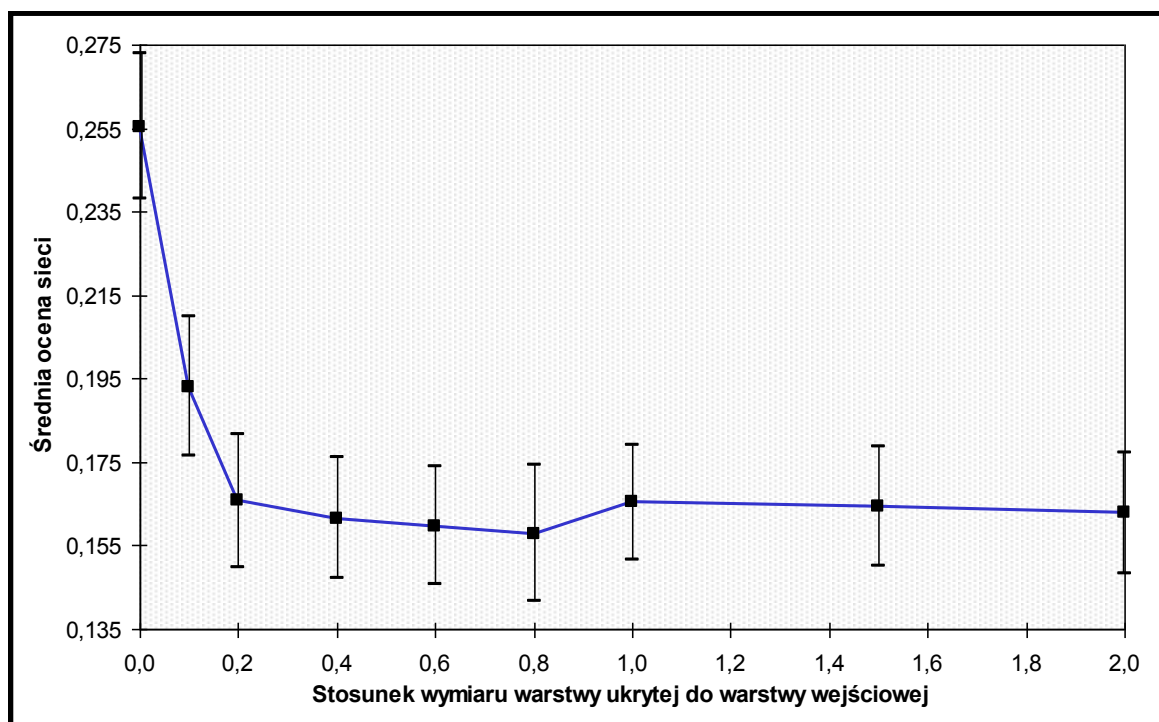
¹⁰⁷ W praktyce dla zadanej wartości ω wyznaczano odpowiednią liczbę neuronów warstwy ukrytej, zaokrąglając tę liczbę do najbliższej wielkości całkowitej.

przedstawiono w tabeli 5-3, a następnie zobrazowano je w postaci graficznej na rys. 5-14 w formie zależności średniej oceny dla 30-tu badanych sieci od wielkości ω .

TABELA 5-3

| Numer sieci | Wartości oceny sieci dla poszczególnych rozmiarów warstwy ukrytej | | | | | | | | |
|--------------------------|---|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | 0,0 | 0,1 | 0,2 | 0,4 | 0,6 | 0,8 | 1,0 | 1,5 | 2,0 |
| 1 | 0,27116 | 0,19524 | 0,16946 | 0,16978 | 0,14995 | 0,16053 | 0,15644 | 0,15227 | 0,16867 |
| 2 | 0,27439 | 0,20815 | 0,15777 | 0,15976 | 0,18825 | 0,15508 | 0,15605 | 0,15490 | 0,16352 |
| 3 | 0,25021 | 0,20525 | 0,18004 | 0,19008 | 0,14621 | 0,17732 | 0,15195 | 0,15543 | 0,19481 |
| 4 | 0,26774 | 0,19732 | 0,15994 | 0,13650 | 0,15168 | 0,16085 | 0,18006 | 0,18129 | 0,17100 |
| 5 | 0,24637 | 0,17352 | 0,13504 | 0,14292 | 0,12868 | 0,12346 | 0,14969 | 0,14495 | 0,13545 |
| 6 | 0,22784 | 0,17345 | 0,17205 | 0,16948 | 0,16789 | 0,15533 | 0,18287 | 0,15962 | 0,17644 |
| 7 | 0,25022 | 0,22750 | 0,16236 | 0,15774 | 0,15654 | 0,12646 | 0,15276 | 0,15048 | 0,14374 |
| 8 | 0,27073 | 0,21125 | 0,17704 | 0,15183 | 0,14660 | 0,15962 | 0,16345 | 0,16318 | 0,15115 |
| 9 | 0,25444 | 0,19993 | 0,14190 | 0,15631 | 0,14626 | 0,13860 | 0,14334 | 0,17400 | 0,14927 |
| 10 | 0,24231 | 0,17838 | 0,15374 | 0,14066 | 0,15233 | 0,13857 | 0,14061 | 0,14679 | 0,14266 |
| 11 | 0,25648 | 0,19630 | 0,17930 | 0,18149 | 0,15367 | 0,13722 | 0,18020 | 0,15888 | 0,16792 |
| 12 | 0,25539 | 0,17282 | 0,15683 | 0,13710 | 0,15812 | 0,15615 | 0,15851 | 0,15836 | 0,16346 |
| 13 | 0,21008 | 0,18248 | 0,17603 | 0,16352 | 0,19505 | 0,14784 | 0,17231 | 0,17468 | 0,16637 |
| 14 | 0,27484 | 0,22123 | 0,18809 | 0,16727 | 0,15416 | 0,16153 | 0,18613 | 0,17203 | 0,17702 |
| 15 | 0,22932 | 0,16765 | 0,16825 | 0,16239 | 0,18057 | 0,15932 | 0,15646 | 0,16676 | 0,16391 |
| 16 | 0,26456 | 0,21502 | 0,15118 | 0,16408 | 0,15593 | 0,16016 | 0,16396 | 0,16134 | 0,15751 |
| 17 | 0,24422 | 0,19683 | 0,16851 | 0,17681 | 0,17613 | 0,19331 | 0,18037 | 0,19172 | 0,18278 |
| 18 | 0,26762 | 0,19928 | 0,14808 | 0,14714 | 0,14660 | 0,17468 | 0,15480 | 0,18053 | 0,15620 |
| 19 | 0,27181 | 0,18937 | 0,15530 | 0,17104 | 0,14890 | 0,16200 | 0,16555 | 0,16294 | 0,15944 |
| 20 | 0,24794 | 0,15879 | 0,14821 | 0,15445 | 0,16293 | 0,16211 | 0,14810 | 0,14205 | 0,16504 |
| 21 | 0,26120 | 0,19325 | 0,17260 | 0,14417 | 0,15654 | 0,14918 | 0,16071 | 0,14856 | 0,15156 |
| 22 | 0,23841 | 0,21260 | 0,19920 | 0,14498 | 0,14616 | 0,15002 | 0,15549 | 0,20271 | 0,15308 |
| 23 | 0,25540 | 0,18275 | 0,16077 | 0,15813 | 0,15331 | 0,16247 | 0,17407 | 0,17593 | 0,16418 |
| 24 | 0,23621 | 0,18308 | 0,14836 | 0,15919 | 0,17252 | 0,16073 | 0,17982 | 0,16596 | 0,14822 |
| 25 | 0,27382 | 0,21178 | 0,16134 | 0,15894 | 0,15377 | 0,15256 | 0,15247 | 0,15548 | 0,16941 |
| 26 | 0,24896 | 0,18527 | 0,17409 | 0,17184 | 0,16286 | 0,16045 | 0,17570 | 0,17919 | 0,18518 |
| 27 | 0,28343 | 0,17733 | 0,15421 | 0,16389 | 0,16413 | 0,16787 | 0,18637 | 0,16731 | 0,14737 |
| 28 | 0,28286 | 0,17834 | 0,18656 | 0,17764 | 0,16695 | 0,20212 | 0,18447 | 0,16248 | 0,18914 |
| 29 | 0,24010 | 0,19713 | 0,16347 | 0,17704 | 0,17001 | 0,16052 | 0,17160 | 0,14405 | 0,16962 |
| 30 | 0,26276 | 0,20094 | 0,19982 | 0,18906 | 0,17553 | 0,15912 | 0,17697 | 0,17562 | 0,14708 |
| Średnia | 0,25536 | 0,19307 | 0,16565 | 0,16151 | 0,15961 | 0,15784 | 0,16538 | 0,16432 | 0,16271 |
| Odchyl. standard. | 0,01738 | 0,01680 | 0,01585 | 0,01448 | 0,01416 | 0,01631 | 0,01376 | 0,01446 | 0,01448 |

Źródło: opracowanie własne.



Rys. 5-14. Zależność średniej oceny dla 30-tu sieci od wzajemnej proporcji wielkości warstwy ukrytej do warstwy wejściowej. Graficznie przedstawiono odchylenia standardowe od wartości średniej dla poszczególnych punktów wykresu. Źródło: opracowanie własne.

Z analizy wykresu zaprezentowanego na rys. 5-14 wypływają następujące wnioski:

- Architektury sieci nie posiadające warstwy ukrytej ($\omega = 0$) okazują się praktycznie nieefektywne w rozważanym zagadnieniu (i prawdopodobnie w większości prób wykorzystania SN w celu prognozowania lub generowania decyzji inwestycyjnych na rynkach finansowych) - świadczy o tym bardzo duża (zła) wartość oceny sieci.
- Średnia ocena sieci początkowo szybko spada (poprawia się) wraz ze wzrostem ω , po czym stabilizuje się po przekroczeniu wartości $\omega \approx 0,2$. Można zatem jako optymalną przyjąć (dla tego problemu) wielkość warstwy ukrytej wynoszącą 20% - 40% wymiaru warstwy wejściowej (dalsze zwiększanie warstwy ukrytej praktycznie nie przynosi poprawy efektywności sieci).

Należy zwrócić uwagę, iż stosowanie zbyt dużych wartości ω (tutaj np. $\omega > 0,8$) nie tylko nie poprawia efektów funkcjonowania sieci, ale przyczynia się do wzrostu czasu (kosztów) uczenia sieci oraz pogłębia wrażliwość sieci na efekt przeuczenia. Podkreślmy także, że w innych typach zagadnień, oraz przy innej specyfice danych służących do konstrukcji sieci (np. przy dłuższych ciągach uczących lub innej wzajemnej korelacji zmiennych wejściowych) optymalny zakres wartości ω może być inny - celowe jest zatem przeprowadzenie badań doboru wielkości warstwy ukrytej odrębnie dla każdego problemu.

Zwróćmy uwagę na fakt, iż w rozważanym zagadnieniu, w trakcie realizacji algorytmu genetycznego, stosunek wielkości warstwy ukrytej do warstwy zerowej był optymalizowany genetycznie (gen reprezentujący wielkość N - zob. podrozdział 4.7). Średnia, uzyskana po zakończeniu AG, wartość ω dla 30-tu rozpatrywanych tutaj sieci wyniosła¹⁰⁸ **0,94** (przy odchyleniu standardowym $\pm 0,26$). Dla analizowanej sieci nr 6 (zob. podrozdział 5.2), uznanej za wyjątkowo efektywną w sensie zysku generowanego przez opartą o nią strategię, wartość ω wynosi **0,35** (sieć ta posiada **7** neuronów w warstwie ukrytej).

Należy zauważyć, że w omówionym w niniejszej pracy algorytmie genetycznym optymalizowaną funkcją celu była funkcja oceny sieci. Z rezultatów badań zawartych w niniejszym podrozdziale wynika z kolei, że przebieg tej funkcji w zależności od wartości argumentu ω ma wyraźnie „płaski” charakter w zakresie ω od 0,2 do 2; minimum tej funkcji jest więc trudne do znalezienia, a nawet odległe od tego minimum wartości ω mogą zostać uznane przez algorytm za wystarczająco dobre. Jednakże, jak wynika z przeprowadzonej dyskusji, zdecydowanie należy preferować możliwie najmniejsze (z ujawnionego optymalnego zakresu) wielkości ω . Celowe może okazać się zatem uwzględnienie w minimalizowanej przez algorytm genetyczny funkcji celu dodatkowego składnika, reprezentującego „karę” za

¹⁰⁸ Przykładowo dla sieci o 20-tu wejściach wartość ω może się wahać od 0,1 do 2,05 (por. przyjęte ograniczenia dla liczby neuronów w warstwie ukrytej - podrozdział 4.4). Średnia liczba wejść dla analizowanych 30-tu sieci wynosi **24,33** (przy odchyleniu standardowym $\pm 3,49$).

nadmierną wielkość (złożoność) struktury sieci (np. całkowitej liczby wag sieci, lub właśnie wartości ω)¹⁰⁹.

5.4.4. Celowość zastosowania analizy głównych składowych jako elementu wstępnego przetwarzania danych.

Metodę analizy głównych składowych (PCA), jako formy wstępnego przetwarzania danych dla sieci typu *perceptron*, przedstawiono ogólnie w podrozdziale 2.7.7. Następnie w podrozdziale 4.3.5 zaprezentowano (wykorzystywaną w zaproponowanej metodologii) regułę Sangera oraz algorytm uczenia sieci neuronowej PCA (dokonującej właściwej transformacji PCA), a także przeprowadzono dyskusję zastosowania tej metody w procesie konstrukcji systemu decyzyjnego.

W przypadku zastosowania dla danych wejściowych SN analizy PCA, istotną sprawą jest optymalne określenie następujących wielkości:

- parametru k , tzn. liczby uwzględnionych głównych składowych¹¹⁰,
- parametru E , tzn. liczby epok uczenia sieci PCA.

Poniżej przedstawiono wyniki badań zmierzających do właściwego oszacowania tych wartości w rozważanym zagadnieniu, rozważono także problem redukcji warstwy ukrytej SN, oraz zaprezentowano szereg wniosków i uwag wynikających z przeprowadzonych badań.

5.4.4.1. Badanie wpływu liczby zastosowanych głównych składowych na efektywność sieci.

Analizując kwestię zdefiniowania liczby głównych składowych stwierdzono, że wyniki badań będą miały walor ogólności jeśli problem będzie sformułowany nie tyle w kategorii bezpośredniego określenia parametru k , co w sensie wyznaczenia procentowego stosunku ilości głównych składowych do całkowitej ilości

¹⁰⁹ Koncepcja tego typu dodatkowego składnika, wyrażającego stopień złożoności sieci, została zaproponowana przez niektórych badaczy w dyskusjach problemu optymalizacji topologii sieci - zob. np. (Azoff, 1994).

¹¹⁰ Zakładamy wykorzystanie przez sieć k pierwszych, tzn. najbardziej informatywnych głównych składowych.

pierwotnych danych wejściowych, tzn. wartości $\xi = (k/n) \cdot 100\%$. Jest to podyktowane występowaniem różnych wielkości n dla poszczególnych SN otrzymanych w wyniku realizacji algorytmu genetycznego. Do badań użyto 30-tu sieci neuronowych posiadających najlepsze wartości oceny po zakończeniu AG (zob. podrozdział 5.2, tabela 5-1). Dla każdej z tych sieci zrealizowano (przy niezmiennych dla danej SN pierwotnych danych wejściowych, parametrach uczenia, wielkości ω , oraz zmiennej wyjściowej) szereg niezależnych procesów uczenia, stosując analizę głównych składowych jako dodatkowe narzędzie wstępnego przetwarzania danych wejściowych i przyjmując kolejno wartości $\xi = 10\%, 20\%, 30\%, \dots, 100\%$.

Zwróćmy uwagę na fakt, iż przy zastosowaniu transformacji danych wejściowych z wykorzystaniem metody PCA, struktura sieci typu *perceptron* ulega istotnej redukcji: wymiar warstwy wejściowej zmniejsza się z n do k elementów (gdzie po zaokrągleniu $k \approx \xi \cdot n / 100\%$), a także odpowiednio maleje wielkość warstwy ukrytej z $n \cdot \omega$ do około¹¹¹ $k \cdot \omega$. Przy tak dokonywanej redukcji SN, dla sieci o dużych rozmiarach, liczba podlegających modyfikacji podczas uczenia wag spada w przybliżeniu proporcjonalnie do **kwadratu** wielkości stopnia redukcji warstwy wejściowej (wymiar tej warstwy jest określony przez ilość wykorzystanych głównych składowych).

Dokonując uczenia sieci PCA (przy wykorzystaniu algorytmu omówionego w podrozdziale 4.3.5) przyjęto¹¹² liczbę epok uczenia $E = 500$. Ponadto dla każdej wartości ξ dokonano analizy dwóch przypadków:

- wykonano dodatkową normalizację k -wymiarowych wektorów otrzymanych w wyniku transformacji PCA przy zastosowaniu metody „bez przesunięcia zera” (*zero offset*) (metoda ta nie zmienia właściwości wzajemnej dekorelacji poszczególnych k zmiennych), (tabela 5-5);

¹¹¹ Liczbę neuronów w warstwie ukrytej wyznaczamy na podstawie określonej przez chromosom wartości N (zob. podrozdział 4.7). Zmniejszenie liczby wejść sieci powoduje (dla ustalonego N) odpowiednie zmniejszenie rozmiaru warstwy ukrytej, następuje też zaokrąglenie jej wielkości do liczby całkowitej.

¹¹² Próby zastosowania większej liczby epok uczenia sieci PCA, np. $E = 1500$, prowadziły praktycznie do takich samych wyników badań; zob. też dyskusja zagadnienia doboru właściwej wartości E , zamieszczona poniżej.

- zrezygnowano z realizacji powyższej dodatkowej normalizacji (tabela 5-4).

Ponadto, w celu dokonania obiektywnego porównania wyników uczenia przeprowadzono dla każdej sieci neuronowej ponowny, niezależny proces uczenia bez wykorzystania analizy głównych składowych dla transformacji danych wejściowych. W każdym z analizowanych przypadków uczenia, dla każdej z rozpatrywanych sieci, proces uczenia powtarzano dziesięciokrotnie, rozpoczynając go od różnych, losowo wygenerowanych, wag początkowych (parametr $P = 10$).

TABELA 5-4

| Wartości ocen poszczególnych sieci w przypadku braku dodatkowej normalizacji wektorów otrzymanych w wyniku PCA | | | | | | | | | | | |
|--|--|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Numer sieci | Procent wykorzystanych głównych składowych | | | | | | | | | | Bez PCA |
| | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% | |
| 1 | 0,27266 | 0,23964 | 0,22146 | 0,17297 | 0,12552 | 0,16250 | 0,12379 | 0,12711 | 0,16034 | 0,14708 | 0,14566 |
| 2 | 0,27668 | 0,22667 | 0,20370 | 0,19214 | 0,13106 | 0,17773 | 0,16238 | 0,15688 | 0,14054 | 0,13059 | 0,15736 |
| 3 | 0,27222 | 0,26248 | 0,16958 | 0,20151 | 0,15808 | 0,16418 | 0,16545 | 0,15791 | 0,15868 | 0,16185 | 0,16302 |
| 4 | 0,27652 | 0,26109 | 0,24194 | 0,24212 | 0,23139 | 0,16619 | 0,17193 | 0,18102 | 0,17307 | 0,16410 | 0,18837 |
| 5 | 0,27058 | 0,21391 | 0,20789 | 0,18857 | 0,17285 | 0,16709 | 0,14951 | 0,14577 | 0,14951 | 0,15000 | 0,13923 |
| 6 | 0,26973 | 0,24433 | 0,24391 | 0,24372 | 0,19008 | 0,17881 | 0,17386 | 0,19188 | 0,16821 | 0,17677 | 0,16103 |
| 7 | 0,27691 | 0,24161 | 0,21580 | 0,20449 | 0,17946 | 0,17738 | 0,13784 | 0,15156 | 0,16796 | 0,14111 | 0,14106 |
| 8 | 0,27656 | 0,27416 | 0,22951 | 0,20600 | 0,17001 | 0,15919 | 0,15817 | 0,15866 | 0,16672 | 0,15930 | 0,16144 |
| 9 | 0,25886 | 0,23450 | 0,24512 | 0,22839 | 0,18871 | 0,16281 | 0,15318 | 0,16082 | 0,16413 | 0,15614 | 0,14173 |
| 10 | 0,24855 | 0,24211 | 0,19587 | 0,17280 | 0,16172 | 0,15054 | 0,13261 | 0,16657 | 0,14570 | 0,15314 | 0,16189 |
| 11 | 0,25252 | 0,24679 | 0,22015 | 0,19018 | 0,20005 | 0,19119 | 0,17173 | 0,16111 | 0,14073 | 0,18010 | 0,17522 |
| 12 | 0,26005 | 0,22701 | 0,22349 | 0,17081 | 0,16031 | 0,16229 | 0,12145 | 0,15472 | 0,13825 | 0,14544 | 0,14779 |
| 13 | 0,25040 | 0,22598 | 0,20369 | 0,16822 | 0,15551 | 0,16195 | 0,17394 | 0,17265 | 0,16218 | 0,18418 | 0,16671 |
| 14 | 0,27506 | 0,25415 | 0,21853 | 0,20031 | 0,15300 | 0,15990 | 0,14655 | 0,15447 | 0,15919 | 0,15446 | 0,17417 |
| 15 | 0,25091 | 0,23840 | 0,20904 | 0,19274 | 0,17381 | 0,16619 | 0,15741 | 0,17560 | 0,15414 | 0,16338 | 0,14836 |
| 16 | 0,26743 | 0,26385 | 0,22372 | 0,18534 | 0,15456 | 0,14831 | 0,15774 | 0,15296 | 0,16467 | 0,12924 | 0,15143 |
| 17 | 0,25665 | 0,22244 | 0,22122 | 0,17858 | 0,15977 | 0,17635 | 0,14866 | 0,16485 | 0,14778 | 0,16590 | 0,17027 |
| 18 | 0,27840 | 0,26634 | 0,20668 | 0,18570 | 0,17513 | 0,16547 | 0,15336 | 0,15418 | 0,14371 | 0,15868 | 0,13518 |
| 19 | 0,24577 | 0,24840 | 0,18341 | 0,18738 | 0,15710 | 0,14937 | 0,17397 | 0,17113 | 0,16431 | 0,16145 | 0,16002 |
| 20 | 0,26677 | 0,25210 | 0,20196 | 0,20798 | 0,20484 | 0,19375 | 0,14349 | 0,16220 | 0,17691 | 0,16534 | 0,17291 |
| 21 | 0,27522 | 0,22849 | 0,21988 | 0,21124 | 0,16603 | 0,14480 | 0,16321 | 0,14752 | 0,15543 | 0,15858 | 0,14901 |
| 22 | 0,25693 | 0,25436 | 0,23684 | 0,19893 | 0,17989 | 0,19064 | 0,17812 | 0,19768 | 0,18045 | 0,19995 | 0,17492 |
| 23 | 0,26885 | 0,24991 | 0,25401 | 0,18277 | 0,15421 | 0,16228 | 0,13304 | 0,15902 | 0,16947 | 0,16304 | 0,16602 |
| 24 | 0,24500 | 0,23999 | 0,23317 | 0,20881 | 0,17094 | 0,14411 | 0,14239 | 0,14313 | 0,15594 | 0,15817 | 0,16528 |
| 25 | 0,28394 | 0,24509 | 0,23130 | 0,18749 | 0,14112 | 0,14266 | 0,14718 | 0,16130 | 0,14757 | 0,16647 | 0,16153 |
| 26 | 0,27499 | 0,25122 | 0,23985 | 0,23944 | 0,18719 | 0,17231 | 0,16722 | 0,16506 | 0,16768 | 0,16415 | 0,17593 |
| 27 | 0,27562 | 0,24404 | 0,21323 | 0,17936 | 0,17335 | 0,16000 | 0,16351 | 0,16173 | 0,16004 | 0,16043 | 0,17075 |
| 28 | 0,26107 | 0,20700 | 0,21093 | 0,17317 | 0,17105 | 0,17285 | 0,15990 | 0,17681 | 0,17441 | 0,15792 | 0,18030 |
| 29 | 0,24605 | 0,23829 | 0,22757 | 0,18454 | 0,16678 | 0,15936 | 0,16930 | 0,14199 | 0,15648 | 0,15093 | 0,16152 |
| 30 | 0,27140 | 0,26016 | 0,24107 | 0,22319 | 0,20708 | 0,18697 | 0,18937 | 0,20058 | 0,14965 | 0,17970 | 0,18415 |
| Średnia | 0,26541 | 0,24348 | 0,21982 | 0,19696 | 0,17069 | 0,16591 | 0,15634 | 0,16256 | 0,15879 | 0,16025 | 0,16174 |
| Odchyl. standard. | 0,01160 | 0,01573 | 0,01889 | 0,02132 | 0,02233 | 0,01391 | 0,01649 | 0,01610 | 0,01134 | 0,01468 | 0,01386 |

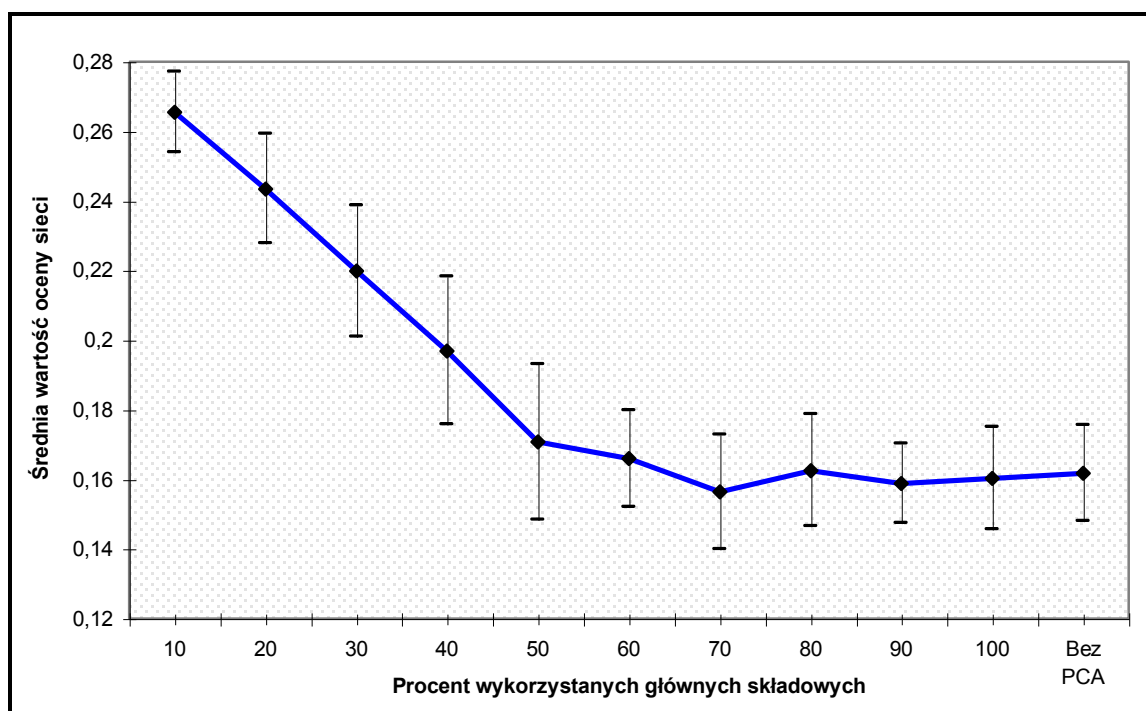
Źródło: opracowanie własne.

TABELA 5-5

| Wartości ocen poszczególnych sieci w przypadku zastosowania dodatkowej normalizacji wektorów otrzymanych w wyniku PCA | | | | | | | | | | | |
|---|--|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Numer sieci | Procent wykorzystanych głównych składowych | | | | | | | | | | Bez PCA |
| | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% | |
| 1 | 0,27365 | 0,23296 | 0,22063 | 0,16554 | 0,16462 | 0,14717 | 0,14872 | 0,17990 | 0,15407 | 0,18266 | 0,14566 |
| 2 | 0,27503 | 0,23157 | 0,20998 | 0,20314 | 0,15007 | 0,16959 | 0,15873 | 0,18720 | 0,16499 | 0,17339 | 0,15736 |
| 3 | 0,27341 | 0,27269 | 0,22743 | 0,22861 | 0,21155 | 0,18961 | 0,17850 | 0,18342 | 0,15732 | 0,17235 | 0,16302 |
| 4 | 0,27521 | 0,27027 | 0,24627 | 0,24641 | 0,23284 | 0,17318 | 0,17292 | 0,18364 | 0,16595 | 0,16478 | 0,18837 |
| 5 | 0,26367 | 0,25585 | 0,21276 | 0,18459 | 0,16922 | 0,15595 | 0,14650 | 0,14917 | 0,14770 | 0,15622 | 0,13923 |
| 6 | 0,26763 | 0,24451 | 0,24878 | 0,24348 | 0,18367 | 0,19123 | 0,16941 | 0,19013 | 0,17263 | 0,17965 | 0,16103 |
| 7 | 0,27692 | 0,25419 | 0,20452 | 0,20587 | 0,17023 | 0,15998 | 0,17442 | 0,19102 | 0,16679 | 0,19794 | 0,14106 |
| 8 | 0,27409 | 0,27152 | 0,22585 | 0,21960 | 0,19338 | 0,17678 | 0,17085 | 0,15806 | 0,16663 | 0,16808 | 0,16144 |
| 9 | 0,26034 | 0,25269 | 0,25606 | 0,24439 | 0,20550 | 0,19865 | 0,18577 | 0,18840 | 0,19771 | 0,18720 | 0,14173 |
| 10 | 0,26271 | 0,26109 | 0,22074 | 0,16953 | 0,19001 | 0,15966 | 0,17147 | 0,15871 | 0,15016 | 0,16188 | 0,16189 |
| 11 | 0,26916 | 0,27078 | 0,22600 | 0,21504 | 0,15622 | 0,17107 | 0,17259 | 0,17527 | 0,17774 | 0,15513 | 0,17522 |
| 12 | 0,26775 | 0,23083 | 0,23148 | 0,16919 | 0,15019 | 0,14623 | 0,17594 | 0,16557 | 0,14869 | 0,15030 | 0,14779 |
| 13 | 0,26055 | 0,23982 | 0,20143 | 0,18894 | 0,18816 | 0,19010 | 0,17730 | 0,16104 | 0,16766 | 0,16484 | 0,16671 |
| 14 | 0,27344 | 0,26217 | 0,26257 | 0,20604 | 0,18106 | 0,14719 | 0,17184 | 0,16775 | 0,17175 | 0,17706 | 0,17417 |
| 15 | 0,26212 | 0,25051 | 0,22723 | 0,18727 | 0,16043 | 0,17959 | 0,19398 | 0,19364 | 0,19625 | 0,19590 | 0,14836 |
| 16 | 0,26574 | 0,26509 | 0,23467 | 0,19490 | 0,20066 | 0,15025 | 0,17646 | 0,18621 | 0,16773 | 0,18942 | 0,15143 |
| 17 | 0,26223 | 0,25268 | 0,23907 | 0,22772 | 0,19956 | 0,18071 | 0,18037 | 0,18916 | 0,17390 | 0,18425 | 0,17027 |
| 18 | 0,27722 | 0,26811 | 0,20419 | 0,19712 | 0,18263 | 0,17111 | 0,15343 | 0,15647 | 0,14453 | 0,15566 | 0,13518 |
| 19 | 0,26172 | 0,24828 | 0,21775 | 0,18809 | 0,16308 | 0,18497 | 0,18077 | 0,17617 | 0,18331 | 0,17150 | 0,16002 |
| 20 | 0,27247 | 0,26507 | 0,22208 | 0,19906 | 0,16956 | 0,17493 | 0,16722 | 0,16601 | 0,18092 | 0,16457 | 0,17291 |
| 21 | 0,27637 | 0,24838 | 0,20798 | 0,20903 | 0,15997 | 0,14309 | 0,17039 | 0,15998 | 0,15486 | 0,16641 | 0,14901 |
| 22 | 0,25598 | 0,25533 | 0,25332 | 0,21693 | 0,20939 | 0,20555 | 0,19571 | 0,18994 | 0,19642 | 0,18860 | 0,17492 |
| 23 | 0,26997 | 0,26718 | 0,26131 | 0,18499 | 0,17393 | 0,15853 | 0,17918 | 0,16880 | 0,18333 | 0,17427 | 0,16602 |
| 24 | 0,25079 | 0,25387 | 0,22393 | 0,20653 | 0,18810 | 0,15607 | 0,16506 | 0,16629 | 0,18565 | 0,18597 | 0,16528 |
| 25 | 0,27873 | 0,24426 | 0,21033 | 0,18383 | 0,14949 | 0,15464 | 0,15536 | 0,15926 | 0,13762 | 0,14080 | 0,16153 |
| 26 | 0,27290 | 0,26089 | 0,25927 | 0,24765 | 0,19376 | 0,18411 | 0,19456 | 0,18719 | 0,19931 | 0,19552 | 0,17593 |
| 27 | 0,27860 | 0,24690 | 0,25710 | 0,18623 | 0,18555 | 0,18884 | 0,18402 | 0,16885 | 0,17632 | 0,14671 | 0,17075 |
| 28 | 0,27020 | 0,25768 | 0,24047 | 0,20656 | 0,16623 | 0,17949 | 0,18748 | 0,19369 | 0,18682 | 0,16508 | 0,18030 |
| 29 | 0,25698 | 0,24231 | 0,23273 | 0,19014 | 0,19523 | 0,18378 | 0,16175 | 0,17884 | 0,16527 | 0,19358 | 0,16152 |
| 30 | 0,26675 | 0,26828 | 0,27412 | 0,22992 | 0,21024 | 0,18212 | 0,17713 | 0,18143 | 0,20017 | 0,19457 | 0,18415 |
| Średnia | 0,26841 | 0,25486 | 0,23200 | 0,20488 | 0,18182 | 0,17181 | 0,17326 | 0,17537 | 0,17141 | 0,17348 | 0,16174 |
| Odchyl. standard | 0,00737 | 0,01221 | 0,02008 | 0,02329 | 0,02113 | 0,01703 | 0,01259 | 0,01318 | 0,01739 | 0,01587 | 0,01386 |

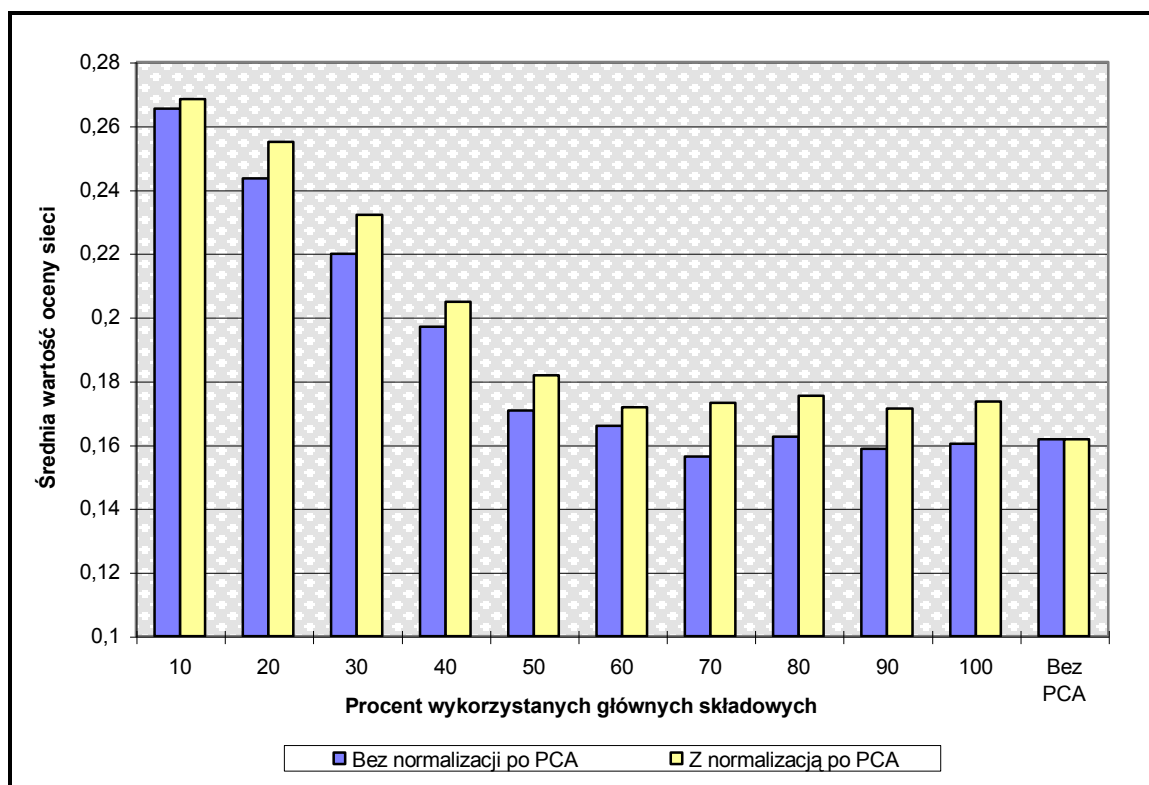
Źródło: opracowanie własne

Wyniki liczbowe zrealizowanych badań dla obu opisanych wyżej przypadków zamieszczono w tabelach 5-4 oraz 5-5. Rezultaty badań zaprezentowano także w formie graficznej na rys. 5-15 oraz 5-16. Tabele 5-4 oraz 5-5 zawierają w kolejnych kolumnach wartości oceny poszczególnych sieci dla parametru ξ zmieniającego się od 10% do 100%, oraz dla przypadku bez zastosowania transformacji PCA. Na rys. 5-15 zobrazowano zależność średniej oceny sieci dla przypadku rezygnacji z wykorzystania dodatkowej normalizacji wektorów po przekształceniu PCA, od procentowej wielkości ξ , przy czym dla poszczególnych punktów wykresu naniesiono także w postaci słupków błędu odchylenia standardowe ocen 30-tu analizowanych sieci. Rys. 5-16 prezentuje w postaci wykresu słupkowego średnie wartości oceny sieci dla 30-tu sieci,



Rys. 5-15. Zależność średniej oceny sieci od wartości parametru ξ , wyrażającego proporcję liczby użytych głównych składowych do ilości pierwotnych danych wejściowych. Zamieszczone rezultaty dotyczą przypadku braku zastosowania dodatkowej normalizacji danych po transformacji PCA. Słupki błędów przedstawiają odchylenia standardowe od wartości średniej dla poszczególnych punktów wykresu. Należy zwrócić uwagę na istotne minimum średniej oceny sieci (punkt, dla którego ta ocena jest najlepsza) dla $\xi = 70\%$. Źródło: opracowanie własne.

w zależności od wielkości ξ , dla obu badanych przypadków (bez dodatkowej normalizacji wektorów po zrealizowaniu transformacji PCA, oraz przy wykorzystaniu takiej normalizacji), umożliwiając porównanie tych wartości. Na obu wykresach zamieszczono także otrzymane rezultaty dla przypadku braku wykorzystania analizy PCA do transformacji danych wejściowych dla SN.



Rys. 5-16. Zależność średniej oceny sieci od wartości parametru ξ (procentu wykorzystanych głównych składowych). Przedstawiono porównawczo średnie wartości dla 30-tu sieci dla obu badanych przypadków, obejmujących odpowiednio zastosowanie dodatkowej normalizacji danych po transformacji PCA (słupki żółte) i brak takiej normalizacji (słupki niebieskie). Źródło: opracowanie własne.

Graficzna analiza rezultatów zaprezentowanych na rysunkach 5-15 oraz 5-16 prowadzi do stwierdzenia następujących faktów:

- i) Efektywność sieci neuronowych, wyrażona liczbowo poprzez ocenę sieci, początkowo wyraźnie wzrasta wraz ze wzrostem parametru ξ (wartość oceny spada), osiągając jednak stały poziom dla ξ wynoszącego ok. 50 - 60%. Przy

dalszym wzroście liczby głównych składowych efektywność sieci pozostaje stała.

- ii) Zastosowanie dodatkowej normalizacji zmiennych wejściowych SN, otrzymanych po transformacji PCA, przyniosło dla każdej wartości ξ gorsze efekty (większe średnie wartości oceny), niż wykorzystanie jako wejść SN głównych składowych bez normalizacji. Świadczy to, iż tego typu normalizacja, proponowana m.in. przez Azoffa (Azoff, 1994), okazuje się (przynajmniej w rozważanym tu zagadnieniu) niepożądana.
- iii) Transformacja danych wejściowych przy wykorzystaniu analizy głównych składowych (bez dodatkowej normalizacji) bezpośrednio nie poprawia efektywności funkcjonowania sieci neuronowej (można to wywnioskować z porównania wartości oceny sieci dla ξ z zakresu od 50% do 100% z przypadkiem uczenia SN bez zastosowania analizy PCA), nie przynosi też widocznych efektów sama dekorelacja zmiennych ($\xi = 100\%$). Bardzo istotną korzyścią użycia przekształcenia PCA może być natomiast redukcja ilości zmiennych wejściowych SN oraz powiązana z tym redukcja rozmiarów całej sieci, co w dalszej konsekwencji prowadzi do znacznego przyspieszenia procesu uczenia SN oraz poprawy efektywności konstrukcji systemu decyzyjnego.

Przedstawione wyżej fakty, zwłaszcza i) oraz ii), potwierdzają również w rozważanym tu zagadnieniu powszechnie akceptowaną tezę, iż kolejne główne składowe, posiadające stopniowo malejącą wariancję, dostarczają odpowiednio coraz mniejszej ilości istotnej informacji. Z wniosku i) wynika, iż końcowe ok. 50% (w tym przypadku) składowych n -wymiarowego wektora wejściowego uzyskanego po transformacji PCA, nie wnosi praktycznie żadnej potencjalnie użytecznej informacji dla sieci neuronowej.

Interesująca jest również interpretacja faktu ii). Jak bowiem wiadomo, w przypadku sieci neuronowej typu *perceptron*, zmienne wejściowe dostarczające relatywnie dużych (w sensie wartości bezwzględnej) oraz podlegających znacznym wahaniom (o stosunkowo dużej wariancji) sygnałów wejściowych, posiadają istotnie większy

wpływ na wartość wyjściową sieci niż pozostałe zmienne¹¹³. Zastosowanie dodatkowej normalizacji głównych składowych (po transformacji PCA), jakkolwiek nie zmienia właściwości dekorelacji tych składowych, wprowadza jednak dla nich „równouprawnienie” w sensie wyrównania zakresów ich zmienności i utraty informacji o ich wariancji. Stwierdzono jednak, że operacja taka prowadzi do pogorszenia efektywności sieci. Można postawić hipotezę, że źródłem tego paradoksu jest fakt, że w sieci pożądane jest, aby bardziej informatywne (początkowe) główne składowe, zachowując większą wariancję, posiadały również większe znaczenie jako sygnały wejściowe SN. Jest to swoisty sposób dodatkowego „ważenia” (wartościowania) zmiennych wejściowych pochodzących z transformacji opartej o metodę analizy głównych składowych.

5.4.4.2. Celowość proporcjonalnej redukcji warstwy ukrytej sieci.

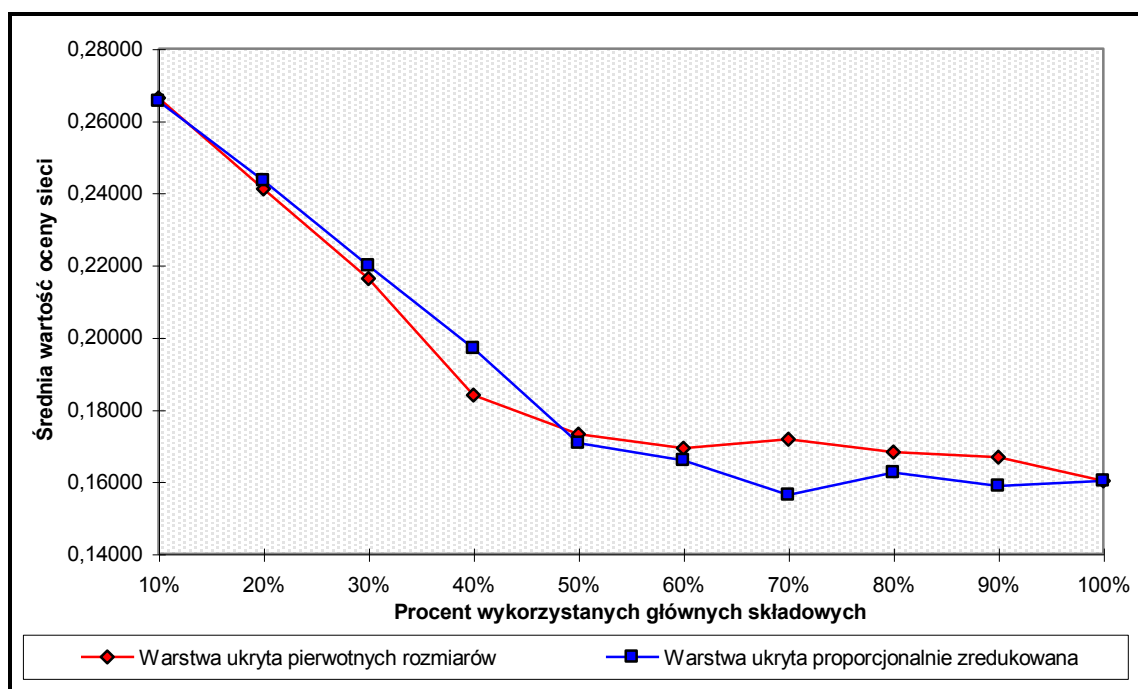
W dotychczasowych badaniach przyjmowano założenie, iż równocześnie z redukcją wymiaru warstwy wejściowej sieci typu *perceptron* (w następstwie ograniczenia liczby zmiennych wejściowych związanego z zastosowaniem transformacji PCA), proporcjonalnej redukcji ulega również wymiar warstwy ukrytej tej sieci. Redukcja ta odbywa się w taki sposób, aby określony przez algorytm genetyczny stosunek wielkości warstwy ukrytej do warstwy wejściowej ω został (w przybliżeniu) zachowany. W niniejszej pracy zbadano jednak dodatkowo także przypadek zakładający pozostawienie pierwotnego rozmiaru warstwy ukrytej sieci, pomimo zmniejszenia liczby wejść sieci wskutek przekształcenia PCA. Wyniki tych badań w formie otrzymanych wartości średnich ocen dla 30-tu sieci, dla poszczególnych wielkości ξ , zamieszczono w tabeli 5-6 oraz zaprezentowano graficznie na rys. 5-17, porównując je z analogicznymi rezultatami uzyskanymi dla przypadku proporcjonalnej redukcji warstwy ukrytej (bez dodatkowej normalizacji wektorów po PCA).

¹¹³ Fakt ten uzasadnia celowość normalizacji zmiennych wejściowych.

TABELA 5-6

| Warstwa ukryta | Średnie wartości oceny sieci dla różnych procentowych ilości wykorzystanych głównych składowych | | | | | | | | | |
|----------------|---|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| Niezmieniona | 0,26634 | 0,24114 | 0,21629 | 0,18392 | 0,17311 | 0,16924 | 0,17173 | 0,16822 | 0,16672 | 0,16025 |
| Zredukowana | 0,26541 | 0,24348 | 0,21982 | 0,19696 | 0,17069 | 0,16591 | 0,15634 | 0,16256 | 0,15879 | 0,16025 |

Źródło: opracowanie własne.



Rys. 5-17. Zależność średniej oceny sieci od wartości ξ (procentu wykorzystanych głównych składowych) dla przypadku pozostawienia pierwotnego rozmiaru warstwy ukrytej, w porównaniu z wariantem zakładającym proporcjonalną redukcję tej warstwy. Źródło: opracowanie własne.

Jak wynika z analizy otrzymanych rezultatów, zachowanie pierwotnego rozmiaru warstwy ukrytej podczas redukcji ilości wejść sieci przy wykorzystaniu PCA z reguły **nie poprawia** wartości oceny sieci (uzyskane wielkości liczbowe średnich ocen nie różnią się istotnie od analogicznych wartości otrzymanych dla przypadku proporcjonalnej redukcji wymiaru warstwy ukrytej). Dla zakresu $\xi > 50\%$ średnie oceny sieci posiadających nie zmniejszony rozmiar tej warstwy są nawet nieznacznie gorsze. Fakty te potwierdzają wnioski zawarte w podrozdziale 5.4.3 dotyczące

celowości maksymalnej możliwej redukcji liczby neuronów w warstwie ukrytej SN. Należy także pamiętać, że operacja ta przyczynia się dodatkowo do skrócenia czasu uczenia sieci.

5.4.4.3. Dobór liczby epok uczenia sieci PCA.

Analizując generalnie celowość zastosowania transformacji danych wejściowych przy wykorzystaniu metody PCA, należy zasadniczo rozważyć kompromis pomiędzy:

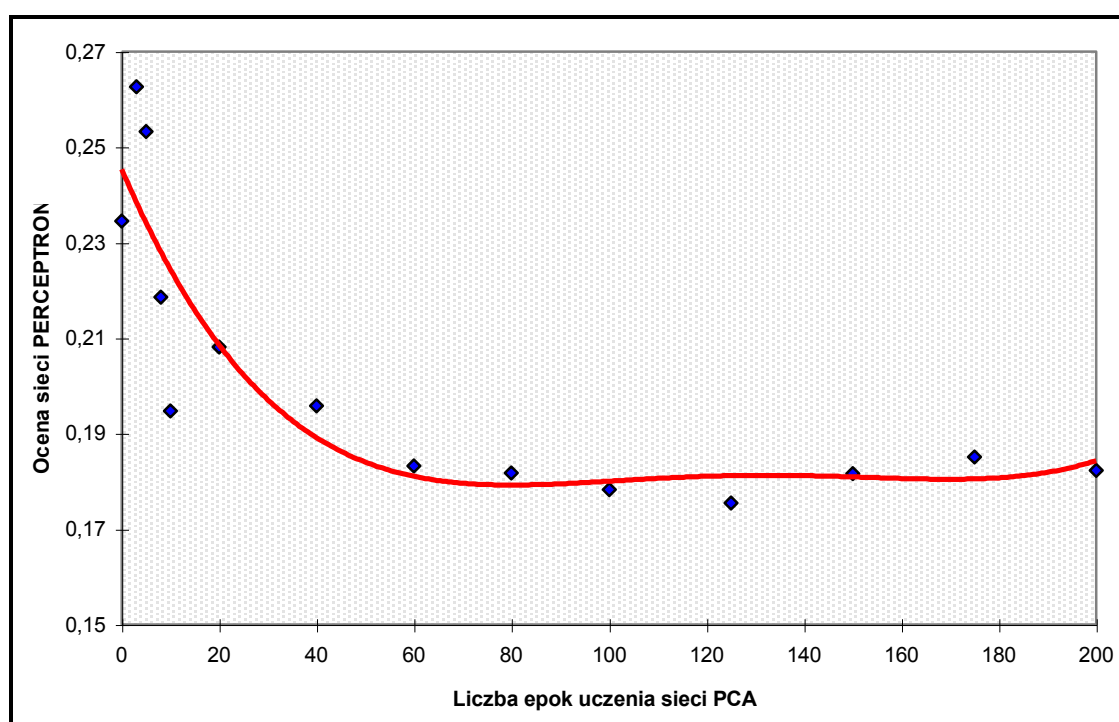
- generalną oszczędnością wysiłku związanego z budową systemu, wynikającą ze skrócenia czasu i potencjalnej poprawy efektywności uczenia SN typu *perceptron* wskutek redukcji struktury sieci,
- wzrostem czasu przeznaczanego na konstrukcję wzorców wejściowych dla SN wskutek dodatkowych nakładów czasu komputerowego poświęconych na transformację PCA (głównie uczenie sieci PCA).

Drugi z powyższych czynników jest zależny w głównej mierze od parametru E , definiującego liczbę epok uczenia sieci PCA. Optymalny dobór parametru E okazuje się bardzo istotny, gdyż jego zbyt mała wartość prowadzi do niewystarczająco dokładnego wyznaczenia głównych składowych, z kolei zbyt duża liczba epok E prowadzi do nadmiernego, w praktyce niepotrzebnego wydłużenia czasu uczenia sieci PCA. Za celowe należy więc uznać przeprowadzenie dla danego zagadnienia badań, mających na celu oszacowanie właściwej (optymalnej) wartości E .

Przykładowe rezultaty tego typu badań w rozpatrywanym tu problemie zobrazowano w formie graficznej na rys. 5-18, gdzie przedstawiono zależność końcowej oceny zasadniczej SN typu *perceptron* od liczby epok uczenia sieci PCA, służącej do wstępnego przetwarzania danych wejściowych dla tej sieci. Badanie przeprowadzono dla zestawu pierwotnych danych wejściowych przy wybraniu jako obiektu badań analizowanej uprzednio sieci nr 6, oraz dla parametru $\xi = 50\%$ (spowodowało to redukcję struktury sieci z 20 do 10 neuronów w warstwie wejściowej, oraz z 7 do 4 elementów w warstwie ukrytej). Proces uczenia tej sieci rozpoczynano każdorazowo od nowa od tych samych warunków początkowych, jednak trening przeprowadzano

przy pomocy danych przetransformowanych do postaci PCA przy użyciu sieci PCA uczonej przy zastosowaniu różnych wartości E .

Jak wynika z rys. 5-18, efektywność w ten sposób trenowanej sieci *perceptron* początkowo wzrasta (wartość oceny spada) wraz ze wzrostem liczby epok E , po czym po przekroczeniu pewnej wartości stabilizuje się na określonym poziomie. Pozwala to w tym przypadku na oszacowanie pożądanej wielkości E równej około **80**.



Rys. 5-18. Przykładowa zależność efektywności SN typu *perceptron*, wyrażona wartością oceny tej sieci, od liczby epok uczenia E sieci PCA służącej do wstępnej transformacji danych. W tym przykładzie przyjęto $\xi = 50\%$ (sieć *perceptron* posiada 10 elementów w warstwie wejściowej związanych z pierwszymi 10-ma głównymi składowymi, oraz 4 elementy w warstwie ukrytej). Uczenie sieci *perceptron* rozpoczynano zawsze od tych samych warunków początkowych. Regresji dokonano wielomianem 4-go stopnia. Źródło: opracowanie własne.

Przy doborze parametru E należy dodatkowo brać pod uwagę, iż wielkość ta zależy m.in. od następujących czynników:

- Ilości wzorców wchodzących w skład ciągu uczącego sieci PCA (tzn. wielkości \mathfrak{S} , zob. podrozdział 4.3.5) - zwykle jest to łączna ilość wzorców ciągu uczącego, walidacyjnego i testującego wybranego dla sieci *perceptron*. Generalnie wypraktykowano, że E można zmniejszyć gdy wartość \mathfrak{S} wzrasta.
- Liczby wyznaczanych głównych składowych - im jest ona mniejsza tym można przyjąć mniejszą wartość E . Zasada ta wynika ze sposobu funkcjonowania zastosowanego algorytmu uczenia sieci PCA, powodującego najszybszy spadek błędu wyznaczania początkowych głównych składowych, przy coraz wolniejszym spadku tego błędu dla kolejnych dalszych składowych.

5.4.4.4. Metoda neuronowa a metoda klasyczna przekształcenia PCA.

Decydując się na zastosowanie transformacji PCA dla danych wejściowych można również rozważyć wykorzystanie dla ich wyznaczania metody klasycznej, czyli algorytmu numerycznego opartego o wyznaczanie wektorów własnych macierzy kowariancji poszczególnych zmiennych wejściowych. W pewnych przypadkach takie podejście klasyczne może okazać się szybsze niż uczenie sieci PCA, pozwala ono też na dokładniejsze wyznaczenie głównych składowych. Jednak w wielu zagadnieniach, zwłaszcza wtedy, gdy celowa jest redukcja stosunkowo dużej liczby zmiennych wejściowych SN do niewielkiej ich ilości, a wysoka precyzja wyznaczania głównych składowych nie jest konieczna (a z takimi właśnie sytuacjami spotykamy się na ogół w problemach wstępnego przetwarzania danych finansowych dla SN) podejście neuronowe do wyznaczania PCA, oparte o regułę Sangera, może okazać się bardziej efektywne¹¹⁴. Efektywność neuronowych metod transformacji PCA wzrośnie też znacznie przy zastosowaniu do obliczeń coraz powszechniej dostępnych komputerowych architektur równoległych. Fakty te przyczyniły się m.in. do wykorzystania właśnie sieci neuronowej PCA (a nie klasycznego algorytmu) w zaproponowanej w niniejszej pracy metodologii¹¹⁵.

¹¹⁴ Interesujące badania właściwości sieci neuronowych PCA uczonych w oparciu o regułę Sangera zawarto w pracy (Grabowski, 1997).

¹¹⁵ Dodatkowym motywem zastosowania neuronowego podejścia do analizy głównych składowych był charakter badań zawartych w niniejszej pracy, ukierunkowany na wykorzystanie sieci neuronowych jako metod sztucznej inteligencji.

5.4.5. Dobór funkcji błędu w procesie uczenia sieci.

W niniejszym podrozdziale zamieszczono wyniki badań zależności efektywności SN w analizowanym zagadnieniu decyzyjnym dla indeksu WIG 20, od postaci funkcji błędu, użytej na etapie uczenia SN (zob. podrozdział 2.5). W badaniach wykorzystano 30 sieci, pochodzących z optymalizacji algorytmem genetycznym, o najlepszych ocenach końcowych. Był to ten sam zestaw SN, który został użyty m.in. w badaniach opisanych w podrozdziałach 5.2, 5.3, 5.4.3 i 5.4.4.

Każda sieć została poddana szeregowi niezależnych procesów uczenia, z zastosowaniem tych samych, opisanych w podrozdziale 5.1 parametrów uczenia, lecz przy wykorzystaniu każdorazowo innej funkcji błędu F . Dla każdego typu funkcji F proces uczenia powtarzano 10 razy ($P = 10$). Analizie poddano funkcję określoną ogólną formułą (2.11) przy zmieniającym się wykładniku u (zauważmy, że dla $u = 1$ funkcja (2.11) jest tożsama z funkcją (2.10), zaś dla $u = 2$ jest ona tożsama z funkcją kwadratową (2.9)), oraz funkcję CR-EN określoną zależnością (2.12) (ang. *cross-entropy*)¹¹⁶. Jako miarę efektywności SN przyjęto tu średni (dla 30-tu sieci) zysk osiągnięty w wyniku zastosowania aktywnej strategii w okresie IN-S (bez uwzględniania kosztów transakcyjnych). Wybrano takie właśnie kryterium, ponieważ bezpośrednie porównanie wartości oceny sieci nie ma sensu, gdyż zależą one istotnie od typu funkcji F . Rezultaty badań zamieszczono w formie liczbowej w tabeli 5-7 (gdzie podano wielkości zysku dla poszczególnych sieci przy zastosowaniu poszczególnych typów funkcji błędu podczas ich uczenia) oraz zobrazowano graficznie na rys. 5-19.

Analiza wykresu zaprezentowanego na rys. 5-19 nie pozwala na jednoznaczne wskazanie najefektywniejszego typu funkcji błędu F . Dodatkowa trudność jednoznacznej interpretacji otrzymanych rezultatów wynika ze stosunkowo dużych wartości odchyień standardowych wielkości zysku otrzymanego dla poszczególnych

¹¹⁶ Aby zastosowanie tej funkcji miało sens, należało zmienić parametr a sigmoidalnej funkcji aktywacji f nadając mu wartość $a = 1$.

sieci. Rozważając jednak funkcję (2.11) dla różnych wartości parametru¹¹⁷ u można przyjąć, iż spełnia ona dość efektywnie swoje zadanie dla u większych od 1 oraz nie większych od 2. Najlepszy rezultat uzyskano dla $u = 1,2$, ale użycie powszechnie stosowanej kwadratowej funkcji błędu ($u = 2$) również przyniosło dość dobry, porównywalny efekt. Zastosowanie zbyt dużych wartości u (tzn. przyjęcie zbyt dużej wrażliwości na wielkość błędów) powoduje wyraźne pogorszenie efektywności funkcjonowania sieci. Warto także zauważyć, że zastosowanie znacznie bardziej kosztownej (w sensie złożoności obliczeniowej) funkcji typu CR-EN nie przynosi zauważalnych korzyści, zatem jej stosowanie nie wydaje się celowe.

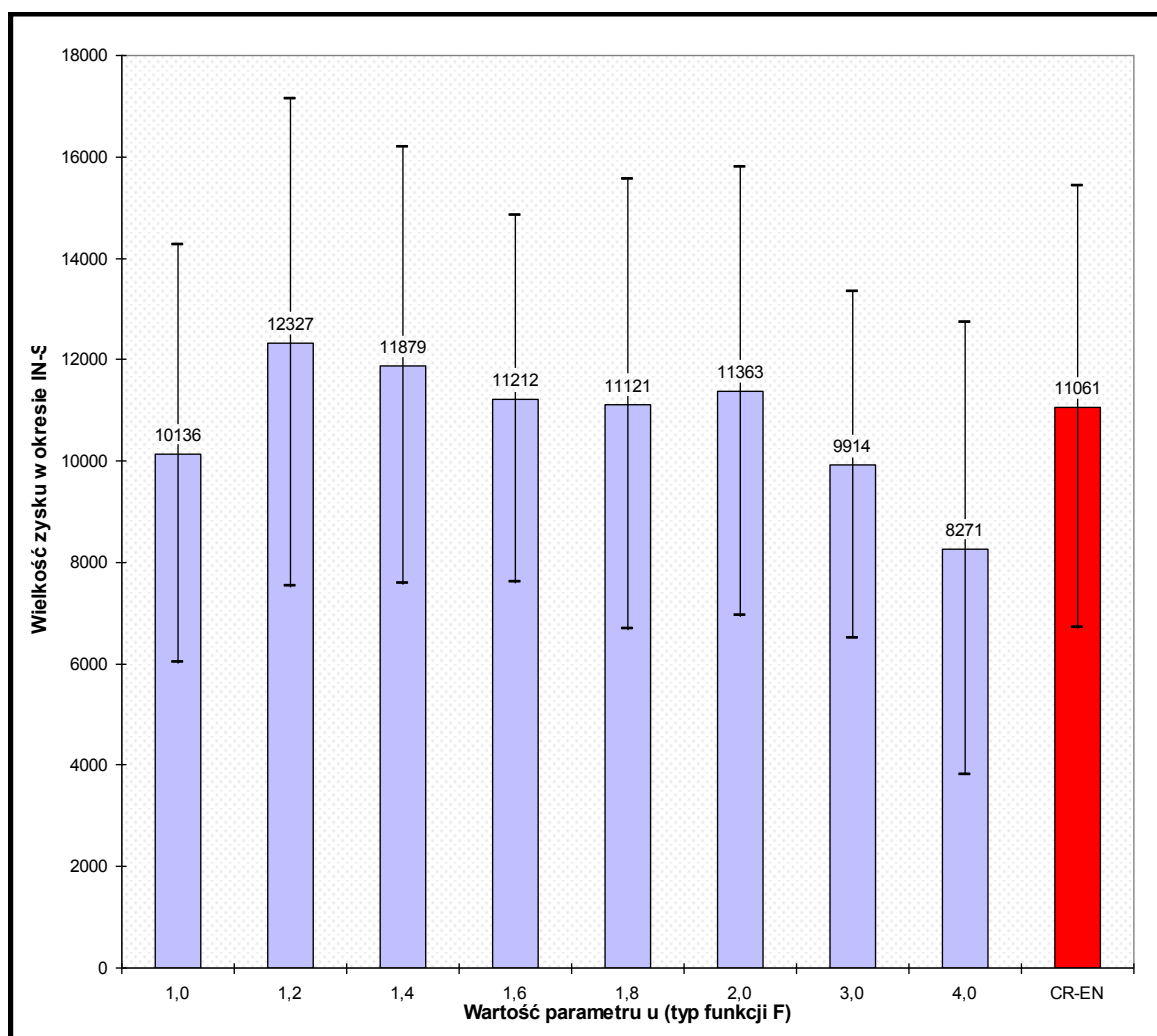
TABELA 5-7

| Numer sieci | Wartości zysku w okresie IN-S dla poszczególnych sieci | | | | | | | | |
|-------------|--|-------|-------|-------|-------|-------|-------|-------|---------------|
| | Wartość parametru u dla funkcji (2.11) | | | | | | | | Funkcja CR-EN |
| | 1,0 | 1,2 | 1,4 | 1,6 | 1,8 | 2,0 | 3,0 | 4,0 | |
| 1 | 6112 | 7886 | 11145 | 9396 | 10601 | 6942 | 11003 | 13135 | 17165 |
| 2 | 10694 | 13070 | 19234 | 12954 | 9891 | 12868 | 9059 | 6165 | 11154 |
| 3 | 7038 | 19825 | 7262 | 10612 | 8999 | 13562 | 6502 | 8001 | 5123 |
| 4 | 17565 | 12410 | 12039 | 6680 | 9820 | 13013 | 12485 | 5418 | 6433 |
| 5 | 5784 | 14686 | 15669 | 15523 | 11952 | 11416 | 8791 | 4604 | 22591 |
| 6 | 13123 | 9255 | 14156 | 11347 | 7418 | 14119 | 10212 | 9571 | 10764 |
| 7 | 19406 | 29064 | 14449 | 12318 | 14441 | 17190 | 10884 | 4385 | 16479 |
| 8 | 7653 | 10712 | 6545 | 15107 | 6763 | 9757 | 10525 | 5272 | 7381 |
| 9 | 11338 | 7273 | 9091 | 9267 | 15146 | 10876 | 8421 | 4624 | 9170 |
| 10 | 8801 | 11852 | 9275 | 12764 | 9332 | 12953 | 7027 | 14661 | 10566 |
| 11 | 5868 | 12452 | 9911 | 14265 | 15838 | 9739 | 7576 | 4863 | 10760 |
| 12 | 9508 | 12130 | 7658 | 6210 | 7989 | 13793 | 3394 | 6716 | 10170 |
| 13 | 12081 | 8556 | 7333 | 8361 | 11651 | 4944 | 6239 | 6815 | 9140 |
| 14 | 13295 | 8499 | 15143 | 6535 | 14681 | 9953 | 6038 | 8576 | 19950 |
| 15 | 12453 | 6432 | 22244 | 6901 | 7233 | 6948 | 8489 | 11755 | 7228 |
| 16 | 6320 | 10133 | 12518 | 11250 | 7703 | 12583 | 8481 | 3301 | 6172 |
| 17 | 7400 | 14544 | 19993 | 13097 | 21070 | 13709 | 13820 | 5206 | 13925 |
| 18 | 19953 | 20183 | 11030 | 18016 | 15921 | 10187 | 11304 | 10778 | 10471 |
| 19 | 10930 | 11950 | 8228 | 8839 | 9225 | 4753 | 15960 | 6006 | 6392 |
| 20 | 5223 | 7170 | 12062 | 12035 | 10621 | 14965 | 11905 | 18613 | 15692 |
| 21 | 5929 | 14078 | 13121 | 13037 | 11082 | 11361 | 12748 | 9928 | 12706 |
| 22 | 7150 | 7931 | 6030 | 10009 | 7994 | 12132 | 9358 | 10362 | 9661 |
| 23 | 8397 | 7223 | 11379 | 8318 | 8317 | 7340 | 9797 | 10058 | 13939 |

¹¹⁷ Parametr u reprezentuje wrażliwość funkcji błędu F na szczególnie duże wartości różnic pomiędzy sygnałem sieci y a wartością pożądaną z . Dla $u = 1$ funkcja jest najmniej wrażliwa na wielkość tych różnic, w miarę wzrostu u wrażliwość ta rośnie.

| | | | | | | | | | |
|-----------------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|-------------|--------------|
| 24 | 10486 | 10181 | 6498 | 9862 | 8293 | 7208 | 11535 | 3569 | 6881 |
| 25 | 8721 | 13588 | 12222 | 6695 | 11047 | 10770 | 8649 | 7400 | 12410 |
| 26 | 4367 | 12038 | 9213 | 11807 | 6577 | 5286 | 9613 | 6066 | 8380 |
| 27 | 9607 | 12952 | 9712 | 11139 | 12456 | 13346 | 21206 | 3742 | 10361 |
| 28 | 9684 | 16524 | 10545 | 10766 | 10190 | 9759 | 8695 | 12207 | 10998 |
| 29 | 15252 | 9742 | 12252 | 10475 | 5447 | 11486 | 5471 | 4684 | 4356 |
| 30 | 13933 | 17463 | 20413 | 22768 | 25933 | 27938 | 12245 | 21638 | 15420 |
| Średnia | 10136 | 12327 | 11879 | 11212 | 11121 | 11363 | 9914 | 8271 | 11061 |
| Odchyl. stand. | 4117 | 4798 | 4305 | 3608 | 4434 | 4414 | 3417 | 4457 | 4361 |

Źródło: opracowanie własne.



Rys. 5-19. Efektywność SN wyrażona średnim zyskiem w okresie IN-S, w zależności od parametru u funkcji F określonej wzorem (2.11) (kolor niebieski), oraz dla funkcji F typu *cross-entropy* określonej formułą (2.12) (kolor czerwony). Słupki błęd reprezentują odchylenia standardowe wielkości zysku dla poszczególnych punktów.

Źródło: opracowanie własne.

5.4.6. Dobór parametrów funkcji aktywacji neuronu.

Poniżej zaprezentowano wyniki badań zależności efektywności sieci od doboru wartości parametrów a oraz b funkcji aktywacji neuronu. Funkcja ta jest funkcją sigmoidalną określoną następującą formułą (wzór (2.4)):

$$f(x) = \frac{2a}{1 + e^{-bx}} - a$$

W badaniach tych zastosowano tą samą metodologię i ten sam zbiór sieci neuronowych oraz parametrów uczenia, co w poprzednim podrozdziale 5.4.5. Jako miary efektywności sieci przyjęto zarówno wartość końcowej oceny sieci, jak i zysk osiągnięty dzięki aktywnej strategii generowanej przez sieć w okresie IN-S. Rezultaty badań dla zmieniającego się w zakresie od 1 do 2 parametru a zaprezentowano w tabeli 5-8 (wartości ocen), w tabeli 5-9 (wartości zysku) oraz na rys. 5-20.

TABELA 5-8

| Numer sieci | Wartości ocen poszczególnych sieci dla kolejnych wielkości parametru a | | | | | | | | | |
|-------------|--|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | $a =$ | 1,0 | 1,075 | 1,15 | 1,225 | 1,30 | 1,45 | 1,60 | 1,80 | 2,00 |
| 1 | | 0,14730 | 0,14921 | 0,17658 | 0,15578 | 0,17418 | 0,17123 | 0,17816 | 0,18452 | 0,24959 |
| 2 | | 0,19550 | 0,15856 | 0,15098 | 0,15714 | 0,16537 | 0,16564 | 0,17163 | 0,19635 | 0,19250 |
| 3 | | 0,16084 | 0,18318 | 0,15738 | 0,17654 | 0,18547 | 0,16251 | 0,16525 | 0,20682 | 0,23748 |
| 4 | | 0,18865 | 0,18141 | 0,17946 | 0,16283 | 0,17158 | 0,17438 | 0,17502 | 0,16815 | 0,20257 |
| 5 | | 0,14605 | 0,12949 | 0,14470 | 0,14784 | 0,14578 | 0,15903 | 0,16596 | 0,17477 | 0,23104 |
| 6 | | 0,17881 | 0,17141 | 0,15225 | 0,15885 | 0,15700 | 0,18320 | 0,17342 | 0,18179 | 0,20727 |
| 7 | | 0,15598 | 0,15235 | 0,14317 | 0,14979 | 0,13952 | 0,18310 | 0,20045 | 0,23438 | 0,24163 |
| 8 | | 0,15731 | 0,16756 | 0,16631 | 0,15716 | 0,16726 | 0,17007 | 0,18321 | 0,19630 | 0,20442 |
| 9 | | 0,14572 | 0,14778 | 0,15436 | 0,16842 | 0,16036 | 0,18472 | 0,16247 | 0,22166 | 0,21186 |
| 10 | | 0,14752 | 0,14460 | 0,13804 | 0,14178 | 0,15295 | 0,14158 | 0,17159 | 0,20038 | 0,16688 |
| 11 | | 0,15609 | 0,18995 | 0,19737 | 0,19688 | 0,19929 | 0,18745 | 0,20263 | 0,20513 | 0,20933 |
| 12 | | 0,17382 | 0,14644 | 0,15352 | 0,15149 | 0,15402 | 0,16875 | 0,18634 | 0,17279 | 0,20876 |
| 13 | | 0,17928 | 0,17503 | 0,16997 | 0,15462 | 0,16709 | 0,16205 | 0,16307 | 0,19186 | 0,19732 |
| 14 | | 0,17075 | 0,16405 | 0,17378 | 0,17083 | 0,16953 | 0,18980 | 0,19648 | 0,18357 | 0,24061 |
| 15 | | 0,17811 | 0,17000 | 0,17573 | 0,15582 | 0,18156 | 0,17730 | 0,16447 | 0,19998 | 0,20074 |
| 16 | | 0,17979 | 0,17753 | 0,17601 | 0,16903 | 0,16579 | 0,18274 | 0,20044 | 0,20528 | 0,22271 |
| 17 | | 0,18646 | 0,19655 | 0,16341 | 0,15216 | 0,18170 | 0,17070 | 0,19988 | 0,19285 | 0,21512 |
| 18 | | 0,16709 | 0,17218 | 0,18406 | 0,15035 | 0,14614 | 0,19583 | 0,15910 | 0,19205 | 0,20487 |
| 19 | | 0,16262 | 0,16269 | 0,18903 | 0,18226 | 0,14650 | 0,18458 | 0,18198 | 0,18524 | 0,18596 |
| 20 | | 0,14847 | 0,15249 | 0,16802 | 0,15499 | 0,14720 | 0,17777 | 0,17300 | 0,19394 | 0,20263 |
| 21 | | 0,15022 | 0,15957 | 0,16333 | 0,15070 | 0,14596 | 0,17720 | 0,19137 | 0,18486 | 0,23102 |
| 22 | | 0,15456 | 0,16666 | 0,17125 | 0,18307 | 0,19065 | 0,18105 | 0,20441 | 0,20996 | 0,23587 |
| 23 | | 0,16135 | 0,16212 | 0,15221 | 0,16554 | 0,13673 | 0,17177 | 0,17071 | 0,17876 | 0,23270 |

| | | | | | | | | | |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 24 | 0,15986 | 0,15693 | 0,17770 | 0,17670 | 0,17093 | 0,16319 | 0,17124 | 0,19318 | 0,19231 |
| 25 | 0,16795 | 0,16183 | 0,15662 | 0,16872 | 0,16594 | 0,17459 | 0,16900 | 0,19612 | 0,20275 |
| 26 | 0,17651 | 0,17338 | 0,18332 | 0,19720 | 0,18556 | 0,19501 | 0,19312 | 0,23394 | 0,23756 |
| 27 | 0,18509 | 0,18272 | 0,16806 | 0,18020 | 0,18639 | 0,18858 | 0,19722 | 0,20588 | 0,19968 |
| 28 | 0,15641 | 0,16659 | 0,16476 | 0,16948 | 0,16253 | 0,18694 | 0,18615 | 0,18203 | 0,21085 |
| 29 | 0,16695 | 0,18915 | 0,17202 | 0,16706 | 0,17386 | 0,15378 | 0,18755 | 0,21112 | 0,23074 |
| 30 | 0,16144 | 0,16790 | 0,17509 | 0,16487 | 0,15308 | 0,17431 | 0,14357 | 0,18609 | 0,21130 |
| Średnia | 0,16555 | 0,16598 | 0,16662 | 0,16460 | 0,16500 | 0,17529 | 0,17963 | 0,19566 | 0,21393 |
| Odchyl. stand. | 0,01403 | 0,01508 | 0,01416 | 0,01396 | 0,01625 | 0,01238 | 0,01544 | 0,01613 | 0,01937 |

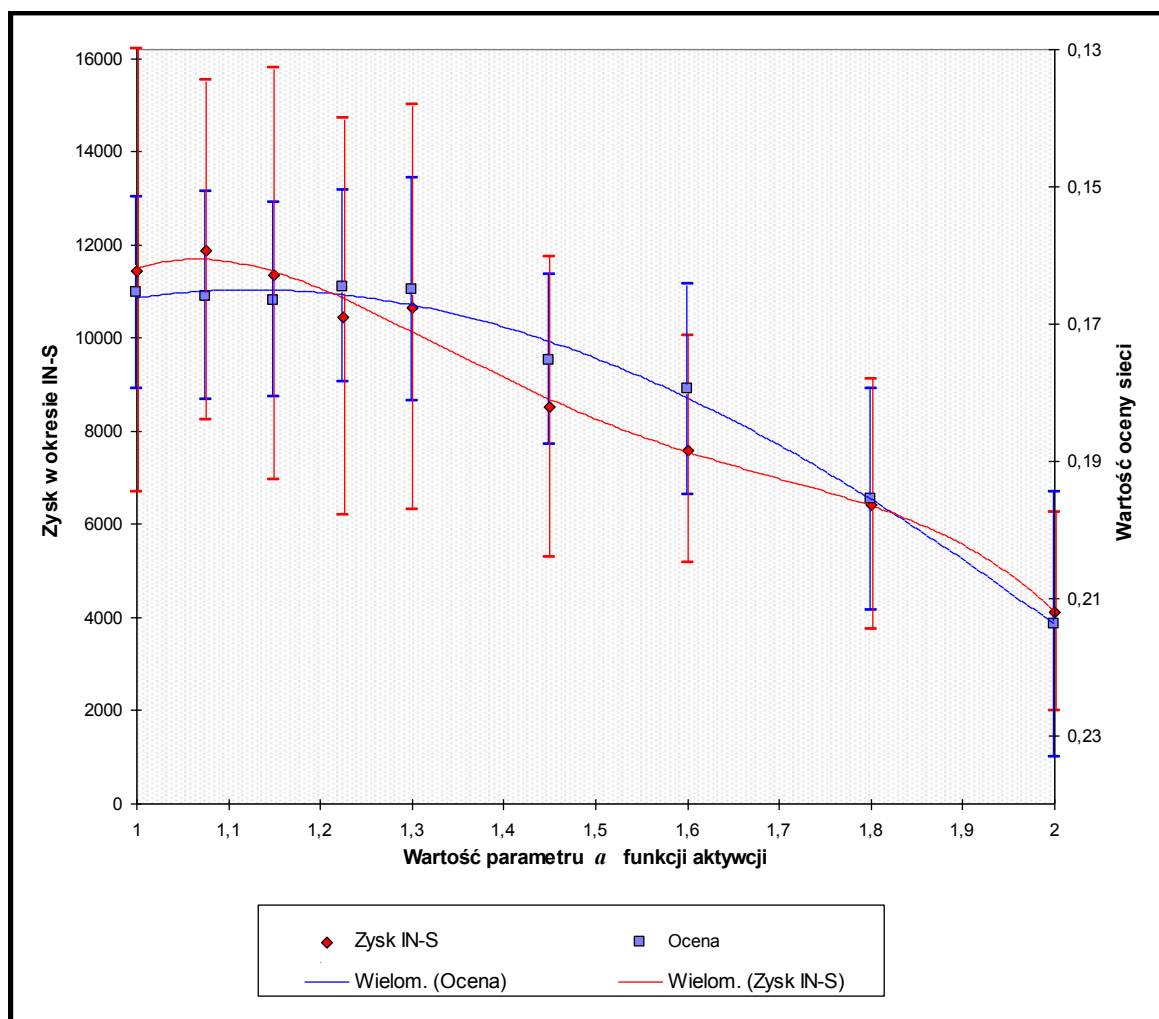
Źródło: opracowanie własne.

TABELA 5-9

| Numer sieci | Wartości zysku wygenerowanego w okresie IN-S przez poszczególne sieci dla kolejnych wielkości parametru α | | | | | | | | |
|----------------|--|----------------|----------------|----------------|----------------|---------------|---------------|---------------|---------------|
| | $\alpha =$ | 1,0 | 1,075 | 1,15 | 1,225 | 1,30 | 1,45 | 1,60 | 1,80 |
| 1 | 16800,2 | 20087,2 | 6942,1 | 13448,2 | 6757,6 | 7954,2 | 5814,7 | 10587,2 | 1212,4 |
| 2 | 7673,1 | 7258,3 | 12867,8 | 12396,5 | 14324,0 | 9772,0 | 11504,5 | 5352,3 | 7706,0 |
| 3 | 10848,7 | 16272,5 | 13561,7 | 11945,8 | 7631,6 | 12496,2 | 11861,3 | 5777,9 | 1533,2 |
| 4 | 11023,0 | 6664,0 | 13012,9 | 4785,7 | 8991,0 | 4035,1 | 11859,6 | 8685,2 | 4523,8 |
| 5 | 16537,3 | 15065,2 | 11415,5 | 15341,2 | 12063,9 | 14341,8 | 8266,3 | 7500,1 | 2236,4 |
| 6 | 9760,7 | 12585,9 | 14119,3 | 13528,2 | 14426,1 | 2786,8 | 9074,9 | 4689,1 | 3607,7 |
| 7 | 15505,0 | 13713,1 | 17190,1 | 13853,6 | 18440,1 | 17427,3 | 6785,5 | 1093,4 | 1202,4 |
| 8 | 10277,6 | 8012,0 | 9756,5 | 9657,0 | 7511,7 | 7737,0 | 5438,7 | 5204,3 | 4727,0 |
| 9 | 12312,5 | 12639,4 | 10876,3 | 8731,0 | 12985,0 | 8073,4 | 9079,0 | 5072,9 | 7127,2 |
| 10 | 14793,1 | 18914,6 | 12953,4 | 10141,8 | 6204,4 | 11520,6 | 8887,4 | 4829,7 | 5666,6 |
| 11 | 21994,5 | 8340,0 | 9739,2 | 8085,9 | 7323,3 | 6469,8 | 6656,6 | 4773,4 | 3422,9 |
| 12 | 8768,6 | 15371,6 | 13793,0 | 9299,9 | 12162,1 | 6826,1 | 10083,2 | 12408,2 | 3737,2 |
| 13 | 7138,3 | 8414,2 | 4944,1 | 10860,9 | 6768,7 | 11075,4 | 10141,3 | 6708,7 | 2127,6 |
| 14 | 10299,2 | 9387,2 | 9952,6 | 11482,1 | 8751,2 | 7964,6 | 8350,8 | 8238,5 | 4027,1 |
| 15 | 8140,5 | 10217,2 | 6948,1 | 11915,1 | 5421,2 | 5325,5 | 7070,9 | 9085,7 | 3671,0 |
| 16 | 6639,1 | 15829,3 | 12582,6 | 7160,7 | 5678,1 | 10015,2 | 4112,8 | 5180,8 | 3360,6 |
| 17 | 11838,7 | 12993,8 | 13708,7 | 26043,5 | 16946,6 | 6384,1 | 3768,3 | 5962,4 | 3071,4 |
| 18 | 9102,0 | 12894,0 | 10186,9 | 11957,2 | 15701,2 | 8042,9 | 9554,9 | 5803,1 | 6588,8 |
| 19 | 10503,5 | 7889,1 | 4753,1 | 7398,3 | 12217,1 | 7436,9 | 7744,3 | 6432,2 | 9892,2 |
| 20 | 11099,5 | 9244,8 | 14965,3 | 8796,8 | 15220,3 | 10055,2 | 7670,3 | 7652,9 | 4867,9 |
| 21 | 13108,7 | 11118,5 | 11361,2 | 6162,9 | 6954,0 | 5832,8 | 4860,6 | 6351,4 | 2128,7 |
| 22 | 10419,0 | 10523,7 | 12132,3 | 8997,5 | 7962,0 | 7471,0 | 5556,9 | 3279,1 | 1375,2 |
| 23 | 8610,5 | 13495,4 | 7339,7 | 9062,1 | 18003,0 | 8551,7 | 5154,3 | 6933,4 | 3301,3 |
| 24 | 5410,5 | 13029,5 | 7208,0 | 7508,5 | 6425,4 | 12523,4 | 7174,9 | 3588,7 | 4584,1 |
| 25 | 6878,3 | 9996,0 | 10770,3 | 8336,7 | 13478,4 | 6455,9 | 6782,0 | 6555,8 | 3597,2 |
| 26 | 8641,2 | 10656,5 | 5285,9 | 3180,5 | 7643,4 | 7270,7 | 5049,4 | 1835,9 | 2572,9 |
| 27 | 10219,2 | 5158,9 | 13346,2 | 6218,9 | 6169,9 | 4062,5 | 4316,0 | 4457,0 | 7625,2 |
| 28 | 15827,8 | 12339,8 | 9759,4 | 14316,9 | 17259,2 | 12745,6 | 9877,8 | 13335,3 | 5648,9 |
| 29 | 5739,5 | 11061,2 | 11485,8 | 7524,7 | 5110,2 | 6302,2 | 4485,6 | 6902,8 | 3164,6 |
| 30 | 27111,4 | 17276,8 | 27938,1 | 15120,1 | 14918,2 | 8514,4 | 10785,3 | 8154,7 | 5134,4 |
| Średnia | 11434,0 | 11881,7 | 11363,2 | 10441,9 | 10648,3 | 8515,7 | 7592,3 | 6414,4 | 4114,7 |
| Odchyl. stand. | 4758,6 | 3654,4 | 4414,4 | 4258,7 | 4345,5 | 3222,9 | 2438,3 | 2686,0 | 2120,3 |

Źródło: opracowanie własne.

Należy zwrócić uwagę, iż w zasadzie istotne znaczenie ma nie tyle bezpośrednia wartość parametru a , ile stosunek wielkości a do granicy przedziału zmienności znormalizowanych wyjściowych wartości pożądanych z_i w ciągu uczącym. W omawianym systemie granica tego przedziału wynosi zawsze 1 (wartości z_i są normalizowane według zależności (2.21) do przedziału $[-1, +1]$).



Rys. 5-20. Zależność średniego zysku w okresie IN-S (lewa oś pionowa) oraz średniej oceny sieci (prawa, odwrócona oś pionowa) - jako miar efektywności sieci - od wartości parametru a sigmoidalnej funkcji aktywacji neuronu danej wzorem (2.4). Liniami ciągłymi przedstawiono odpowiednie krzywe regresji, zrealizowanej wielomianami 4-go stopnia. Słupki pionowe reprezentują odchylenia standardowe dla poszczególnych punktów. Źródło: opracowanie własne.

Przedstawione wyżej rezultaty prowadzą do spostrzeżenia, iż parametr a funkcji aktywacji powinien być równy lub (lepiej) nieznacznie (o około 10% - 15%) większy w stosunku do górnej granicy zakresu zmienności znormalizowanych uczących wartości pożądanych z_i (tzn. do przyjętej tutaj wartości 1). Nadmierne zwiększenie parametru a prowadzi do znaczącego pogorszenia efektywności sieci (spadku zysku IN-S oraz pogorszenia oceny sieci). Zaobserwowano także w tym przypadku wydłużenie czasu uczenia sieci. Wynika to niewątpliwie z faktu, iż w przypadku stosowania zbyt dużych wartości parametru a neurony sieci wykorzystują do transformacji sygnałów jedynie ograniczony, środkowy, a więc praktycznie prostoliniowy fragment funkcji aktywacji, co z kolei utrudnia lub wręcz uniemożliwia modelowanie zjawisk nieliniowych. Płynące z tego faktu wnioski potwierdzają również pośrednio przydatność zastosowania w problematyce finansowej nieliniowych modeli neuronów i jednocześnie świadczą o niskiej efektywności liniowych SN w tego typu zagadnieniach.

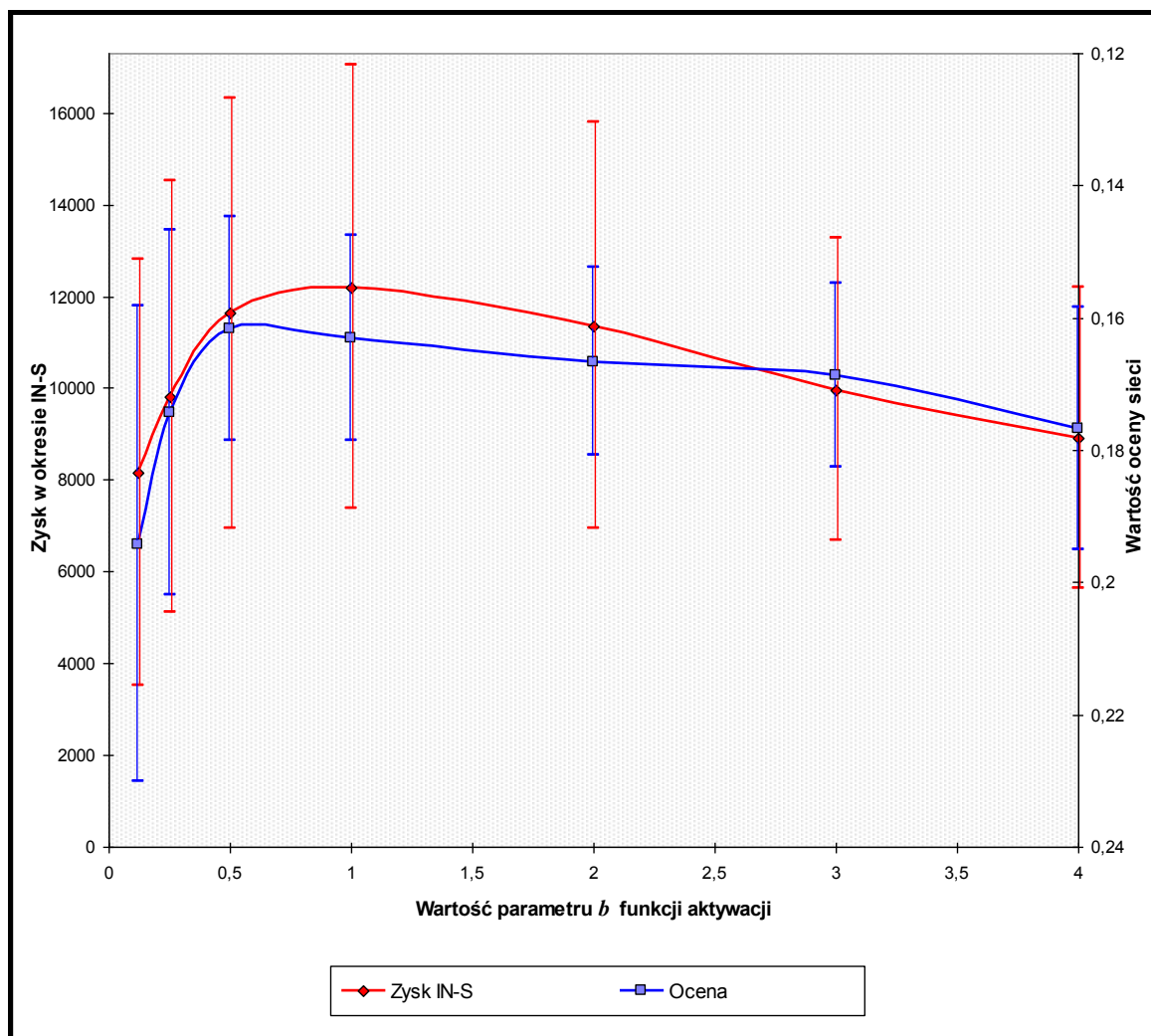
TABELA 5-10

| $b =$ | | Wartości oceny oraz zysku wygenerowanego w okresie IN-S przez poszczególne sieci dla kolejnych wielkości parametru b | | | | | | | | | | | | | |
|-------|--|--|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | | 0,12 | | 0,25 | | 0,50 | | 1,00 | | 2,00 | | 3,00 | | 4,00 | |
| Sieć | | Ocena | Zysk | Ocena | Zysk | Ocena | Zysk | Ocena | Zysk | Ocena | Zysk | Ocena | Zysk | Ocena | Zysk |
| 1 | | 0,18006 | 10869,4 | 0,15562 | 13536,1 | 0,13804 | 13350,8 | 0,14496 | 12390,6 | 0,17658 | 6942,1 | 0,17266 | 8365,5 | 0,17425 | 6394,0 |
| 2 | | 0,17194 | 13898,8 | 0,14242 | 15472,4 | 0,17003 | 8620,6 | 0,15415 | 15435,8 | 0,15098 | 12867,8 | 0,17101 | 8657,6 | 0,17297 | 8172,8 |
| 3 | | 0,25017 | 2861,0 | 0,17107 | 12467,1 | 0,15265 | 8656,7 | 0,18981 | 6103,8 | 0,15738 | 13561,7 | 0,15059 | 14131,0 | 0,15699 | 13122,8 |
| 4 | | 0,25800 | 1798,2 | 0,15926 | 9957,6 | 0,15151 | 23927,3 | 0,18086 | 5349,7 | 0,17946 | 13012,9 | 0,17249 | 11930,4 | 0,17397 | 9446,3 |
| 5 | | 0,15489 | 12220,7 | 0,15761 | 13376,4 | 0,14471 | 15093,7 | 0,14077 | 18130,8 | 0,14470 | 11415,5 | 0,15779 | 11078,5 | 0,17151 | 6009,0 |
| 6 | | 0,18508 | 8360,2 | 0,17947 | 8126,0 | 0,17743 | 5636,0 | 0,17628 | 15703,0 | 0,15225 | 14119,3 | 0,17271 | 6546,7 | 0,17335 | 8451,1 |
| 7 | | 0,16678 | 14778,1 | 0,16070 | 22352,9 | 0,17164 | 16365,4 | 0,14951 | 24961,5 | 0,14317 | 17190,1 | 0,15582 | 19418,8 | 0,17696 | 9073,7 |
| 8 | | 0,19957 | 7394,4 | 0,18562 | 9327,6 | 0,16425 | 12551,6 | 0,16472 | 13056,1 | 0,16631 | 9756,5 | 0,16389 | 5893,1 | 0,16502 | 10287,6 |
| 9 | | 0,25943 | 2363,7 | 0,16613 | 8823,3 | 0,19269 | 7351,3 | 0,17443 | 9588,0 | 0,15436 | 10876,3 | 0,14198 | 13477,9 | 0,14602 | 12170,5 |
| 10 | | 0,12364 | 17562,7 | 0,13164 | 20427,0 | 0,13303 | 12474,3 | 0,13916 | 22579,2 | 0,13804 | 12953,4 | 0,14319 | 10458,8 | 0,18873 | 8961,6 |
| 11 | | 0,21826 | 5057,7 | 0,21627 | 6830,9 | 0,18439 | 19456,5 | 0,17912 | 11746,0 | 0,19737 | 9739,2 | 0,18790 | 4436,4 | 0,19359 | 12134,1 |
| 12 | | 0,24632 | 2626,2 | 0,16088 | 10295,5 | 0,15449 | 13268,5 | 0,15831 | 8235,7 | 0,15352 | 13793,0 | 0,16179 | 8996,2 | 0,18329 | 9447,3 |
| 13 | | 0,16783 | 9319,1 | 0,16664 | 8461,8 | 0,16057 | 9215,2 | 0,17376 | 4659,1 | 0,16997 | 4944,1 | 0,17836 | 7158,1 | 0,17388 | 8490,5 |
| 14 | | 0,17537 | 12257,6 | 0,17483 | 11612,5 | 0,16840 | 12900,5 | 0,15940 | 14244,1 | 0,17378 | 9952,6 | 0,16330 | 9720,2 | 0,18138 | 7316,5 |
| 15 | | 0,23125 | 2720,3 | 0,17479 | 8306,7 | 0,16080 | 6804,0 | 0,15717 | 10613,6 | 0,17573 | 6948,1 | 0,17513 | 5719,1 | 0,18395 | 6142,0 |
| 16 | | 0,18885 | 6652,6 | 0,17247 | 13815,8 | 0,16445 | 10911,7 | 0,16275 | 11461,6 | 0,17601 | 12582,6 | 0,16467 | 9160,8 | 0,18049 | 6458,4 |
| 17 | | 0,19037 | 6589,0 | 0,18643 | 8135,7 | 0,18164 | 22226,8 | 0,17641 | 19150,2 | 0,16341 | 13708,7 | 0,19267 | 9249,9 | 0,16799 | 15479,3 |
| 18 | | 0,17593 | 9447,2 | 0,16493 | 9938,6 | 0,14230 | 13414,8 | 0,15098 | 9311,2 | 0,18406 | 10186,9 | 0,15780 | 12032,6 | 0,16058 | 7987,3 |
| 19 | | 0,18414 | 7209,7 | 0,16506 | 6925,0 | 0,15601 | 9722,0 | 0,17287 | 7794,9 | 0,18903 | 4753,1 | 0,17170 | 13195,5 | 0,16711 | 7719,6 |
| 20 | | 0,14622 | 13927,8 | 0,14328 | 9745,2 | 0,12270 | 15525,8 | 0,14755 | 11952,8 | 0,16802 | 14965,3 | 0,17193 | 10095,6 | 0,15499 | 9025,1 |
| 21 | | 0,16589 | 10219,9 | 0,16908 | 6468,0 | 0,16377 | 7364,0 | 0,15544 | 14114,4 | 0,16333 | 11361,2 | 0,14996 | 11092,3 | 0,19189 | 4246,1 |
| 22 | | 0,24632 | 1863,2 | 0,24075 | 1817,9 | 0,16542 | 10764,0 | 0,17367 | 8198,4 | 0,17125 | 12132,3 | 0,20158 | 10368,9 | 0,22904 | 1326,2 |

| | | | | | | | | | | | | | | |
|---------------------|----------------|---------------|----------------|---------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|---------------|----------------|---------------|
| 23 | 0,17728 | 8641,1 | 0,19104 | 5914,5 | 0,14883 | 10188,9 | 0,12775 | 12640,0 | 0,15221 | 7339,7 | 0,16838 | 5841,0 | 0,17582 | 4849,4 |
| 24 | 0,19980 | 7142,0 | 0,17916 | 5249,4 | 0,17622 | 6080,3 | 0,16285 | 8159,0 | 0,17770 | 7208,0 | 0,16061 | 7916,5 | 0,16343 | 6980,1 |
| 25 | 0,18450 | 5428,7 | 0,15751 | 8075,2 | 0,15586 | 7456,5 | 0,16672 | 11515,3 | 0,15662 | 10770,3 | 0,16589 | 8493,1 | 0,21763 | 7559,1 |
| 26 | 0,23921 | 2844,7 | 0,23503 | 2182,0 | 0,18484 | 6034,4 | 0,18878 | 7458,1 | 0,18332 | 5285,9 | 0,18610 | 9959,4 | 0,18272 | 12377,1 |
| 27 | 0,16853 | 15877,3 | 0,20645 | 11256,6 | 0,18793 | 9386,9 | 0,18274 | 8793,4 | 0,16806 | 13346,2 | 0,18880 | 5375,6 | 0,16839 | 8870,2 |
| 28 | 0,16176 | 12871,7 | 0,14725 | 11052,9 | 0,16121 | 7959,6 | 0,15054 | 11118,5 | 0,16476 | 9759,4 | 0,17407 | 14064,1 | 0,18262 | 8665,8 |
| 29 | 0,23228 | 1905,9 | 0,23009 | 1414,6 | 0,17161 | 9925,0 | 0,17453 | 14366,8 | 0,17202 | 11485,8 | 0,17409 | 13770,4 | 0,20098 | 15807,4 |
| 30 | 0,17755 | 9875,7 | 0,13936 | 13169,2 | 0,14242 | 16343,4 | 0,15524 | 17115,5 | 0,17509 | 27938,1 | 0,16471 | 12498,5 | 0,14279 | 14445,6 |
| Średnia | 0,19424 | 8152,8 | 0,17436 | 9817,8 | 0,16166 | 11632,6 | 0,16304 | 12198,2 | 0,16662 | 11363,2 | 0,16872 | 9970,1 | 0,17674 | 8913,9 |
| Odch. stand. | 0,03593 | 4642,3 | 0,02767 | 4701,7 | 0,01693 | 4682,9 | 0,01546 | 4832,0 | 0,01416 | 4414,4 | 0,01396 | 3289,9 | 0,01829 | 3270,6 |

Źródło: opracowanie własne.

Z kolei badając analogiczną zależność efektywności sieci (wyrażonej miarami wielkości zysku w okresie IN-S dla strategii aktywnej oraz wartości oceny sieci) od parametru b funkcji aktywacji (2.4), odpowiedzialnego za nachylenie tej funkcji (por. rys. 2-5), uzyskano wyniki badań, które zaprezentowano w tabeli 5-10 oraz na rys. 5-21.



Rys. 5-21. Zależność średniego zysku w okresie IN-S (lewa oś pionowa) oraz średniej oceny sieci (prawa, odwrócona oś pionowa) - jako miar efektywności sieci - od wartości parametru b funkcji aktywacji neuronu (2.4). Słupki pionowe reprezentują odchylenia standardowe dla poszczególnych punktów. Źródło: opracowanie własne.

Analiza zamieszczonych wyżej rezultatów potwierdza duże znaczenie właściwego doboru parametru b funkcji aktywacji. W rozpatrywanym problemie optymalna wartość b zawiera się w przybliżonym zakresie od 0,5 do 1 (przedział ten może mieć charakter uniwersalny dla szerszej klasy zagadnień finansowych, jednak jego granice mogą także zależeć od szeregu innych parametrów, np. od zakresu generowanych losowo wag początkowych lub od współczynnika uczenia sieci). Przyjęcie wielkości b spoza optymalnego przedziału prowadzi do istotnego pogorszenia efektywności sieci; zaobserwowano też, iż dla zbyt dużej wartości b (większej od 3) w pewnych przypadkach następuje utrata zbieżności procesu uczenia.

5.4.7. Analiza czasowych zmian efektywności aktywnej strategii opartej o wybraną sieć.

Poniżej zaprezentowano analizę skuteczności aktywnej strategii inwestycyjnej opartej o sieć nr 6 (uznanej na podstawie przesłanek wynikających z podrozdziałów 5.2 i 5.3 za najbardziej efektywną) w porównaniu ze strategią pasywną. Badania przeprowadzono dla rocznego okresu OUT-S (tzn. dla danych nie użytych do konstrukcji sieci), obejmującego 248 sesji giełdowych - od 1.10.1997 do 30.09.1998 (dane uczące i walidacyjne zawierały się w okresie od 14.04.1994 do 30.09.1997). Analiza ta stanowi kontynuację i uzupełnienie badań przedstawionych w podrozdziale 5.2 i obejmuje znacznie szerszy okres OUT-S¹¹⁸. Metodologia badań oraz przyjęte parametry są takie same jak w podrozdziale 5.1 (dla strategii aktywnej założono zerowe koszty transakcyjne).

Rezultaty analizy przedstawiono na rys. 5-22 oraz w tabelach 5-11 i 5-12. Na rys. 5-22 zademonstrowano zmiany wartości aktywów dla strategii aktywnej oraz pasywnej w czasie (w rozważanym rocznym okresie OUT-S), przy założeniu jednakowej początkowej wartości portfeli równej¹¹⁹ 1740,5. W tabeli 5-11 zamieszczono wartości miesięcznych stóp zwrotu w rozważanym okresie (liczonych od ostatniej sesji miesiąca poprzedzającego¹²⁰ do ostatniej sesji danego miesiąca) dla strategii aktywnej oraz pasywnej (indeksu WIG 20). Z kolei kolumny tabeli 5-12 zawierają wybrane wskaźniki efektywności strategii inwestycyjnej, umożliwiające dokonanie analizy porównawczej obu metod inwestowania.

¹¹⁸ Chronologicznie analizę tę przeprowadzono znacznie później niż badania zawarte w podrozdziale 5.2.

¹¹⁹ Wartość indeksu WIG 20 w dniu 1 października 1997.

¹²⁰ Przy obliczaniu miesięcznych stóp zwrotu za wielkość początkową aktywów przyjęto zatem wartość WIG 20 na dzień 30.09.1997, tzn. 1733,2).



Rys. 5-22. Porównanie zmian wartości aktywów w czasie w przypadku wykorzystania strategii aktywnej opartej o sieć nr 6 (przy zerowych kosztach transakcyjnych) ze strategią pasywną (inwestycja w indeks WIG 20) dla rocznego okresu OUT-S od 1.10.1997 do 30.09.1998. Źródło: opracowanie własne.

TABELA 5-11

| Strategia | Wartości miesięcznych stóp zwrotu w rocznym okresie OUT-S (1.10.97 - 30.09.98) [w procentach] | | | | | | | | | | | |
|---------------------|---|----------|---------------|---------|-------|--------|---------------|--------|---------------|--------|----------|---------------|
| | paździer- -nik | listopad | grudzie- ń | styczeń | luty | marzec | kwiecie- ń | maj | czerwie- c | lipiec | sierpień | wrzesie- ń |
| Aktywna | 5,99 | 4,85 | 10,21 | -1,06 | 19,01 | -3,12 | -1,53 | 1,08 | 0,54 | 4,03 | -5,20 | 12,68 |
| Pasywna (WIG 20) | -12,19 | -7,28 | 3,31 | 2,44 | 21,70 | -5,60 | 7,41 | -11,05 | -1,18 | 5,89 | -31,82 | 1,62 |

Źródło: opracowanie własne.

TABELA 5-12

| Strategia | Zysk w całym okresie OUT-S | Maksymalna strata <i>dd</i> | Średnia ^{*)} dzienna stopa zwrotu | Odchylenie standardowe dziennych stóp zwrotu | Średnia ^{*)} miesięczna stopa zwrotu | Odchylenie standardowe miesięcznych stóp zwrotu |
|------------------|----------------------------|-----------------------------|--|--|---|---|
| Aktywna | 952,0 | 17,4 % | 0,19 % | 1,84 % | 3,96 % | 7,11 % |
| Pasywna (WIG 20) | - 552,2 | 43,4 % | - 0,12 % | 2,55 % | - 2,23 % | 13,11 % |

^{*)} Średnia arytmetyczna stóp zwrotu dla wszystkich okresów: dziennych (między kolejnymi sesjami) lub miesięcznych.

Źródło: opracowanie własne.

Przeprowadzając (w celach porównawczych) analogiczne badania efektywności aktywnej strategii inwestycyjnej dla wszystkich 30-tu sieci neuronowych otrzymanych w wyniku realizacji algorytmu genetycznego i analizowanych w poprzednich podrozdziałach, otrzymano dla rozważanego okresu OUT-S:

- średni zysk (strata) dla 30-tu sieci: – **135,1** (WIG 20 stracił w tym czasie – 552,2);
- średnia dla 30-tu sieci wartość maksymalnej straty *dd*: **28,3 %** .

Analizując wyżej przedstawione rezultaty należy stwierdzić, iż strategia aktywna, oparta o sygnały analizowanej sieci neuronowej nr 6, okazała się w rozpatrywanym okresie OUT-S zdecydowanie bardziej efektywna od strategii pasywnej (typu „kup i trzymaj”) polegającej na inwestycji w indeks WIG 20. Świadczą o tym:

- różnica między wartością końcową portfela dla strategii aktywnej i pasywnej, wynosząca ponad 1500; różnica ta powiększała się dość systematycznie, w efekcie w danym okresie strategia aktywna przyniosła zysk wynoszący około 55% wartości początkowej portfela podczas gdy strategia pasywna przyniosła ponad 30-procentową stratę;
- średnie wartości dziennych i miesięcznych stóp zwrotu,
- wartości maksymalnej straty *dd* jako miary ryzyka inwestycyjnego - dla strategii aktywnej wskaźnik ten jest ok. 2,5-krotnie mniejszy (lepszy),

- wartości odchylenia standardowego dla dziennych i miesięcznych stóp zwrotu (stanowiące również często przyjmowane miary ryzyka) - znacznie niższe dla strategii aktywnej.

Analizując miesięczne stopy zwrotu w obu przypadkach należy jednak zauważyć, że w czterech miesiącach (styczeń, luty, kwiecień i lipiec) strategia aktywna przyniosła gorsze wyniki (zysk) niż pasywna. Można stąd wywnioskować, iż efektywna w długim okresie strategia inwestycyjna oparta o SN może w krótszych przedziałach czasowych generować straty, a fakt ten nie powinien stanowić podstawy do jej negatywnej oceny.

Zauważmy także, iż rozważana tu sieć neuronowa jest zdecydowanie bardziej efektywna od pozostałych 30-tu sieci. Pomimo iż przeciętne rezultaty strategii aktywnych dla 30-tu sieci w sensie wartości końcowego zysku i maksymalnej straty *dd* są lepsze od wyników zastosowania strategii pasywnej, to jednak są one wyraźnie gorsze od efektów wykorzystania wyróżnionej sieci nr 6.

5.4.8. Badanie przydatności douczania i ponownego uczenia wybranej efektywnej sieci.

Poniżej zamieszczono wyniki badań celowości douczania analizowanej sieci nr 6 (wyróżniającej się wysoką efektywnością w świetle badań omówionych w podrozdziałach 5.2, 5.3 i 5.4.7) przy wykorzystaniu pojawiających się stale nowych wzorców uczących w trakcie realizacji strategii. Zbadano również efekty prób dodatkowego wielokrotnego ponawiania procesu uczenia sieci nr 6 - przy niezmienniej architekturze i zestawie danych wejściowych - w celu ewentualnego osiągnięcia lepszej oceny sieci. Badania przeprowadzono przy przyjęciu tej samej metodologii oraz takich samych parametrów jak poprzednio (zob. podrozdział 5.1), dla rocznego okresu OUT-S od 1.10.97 do 30.09.98 (identycznego jak w podrozdziale 5.4.7).

Rozważmy trzy sieci neuronowe, oznaczone odpowiednio 6A, 6B i 6C:

- **Sieć 6A** jest siecią nr 6 w niezmiennionej postaci (bez dokonanej jakiegokolwiek modyfikacji wag).

- **Sieć 6B** początkowo posiada wartości wag jak w sieci nr 6, jednak wagi te są modyfikowane na końcu każdego kolejnego miesiąca w procesie douczania. Proces ten polega na dodaniu do ciągu uczącego i/lub walidacyjnego nowych wzorców uczących z ostatniego miesiąca, a następnie na przeprowadzeniu operacji dodatkowego uczenia sieci przy rozpoczęciu realizacji algorytmu uczenia od dotychczasowych wartości wag. Tak douczona sieć jest wykorzystywana do generowania sygnałów transakcyjnych przez następny miesiąc, po czym dokonywane jest kolejne douczanie sieci, itd.
- **Sieć 6C** jest siecią opartą na architekturze i zestawie zmiennych wejściowych sieci nr 6, przeprowadzono jednak dla niej ponowny, znacznie szerszy proces uczenia, w trakcie którego wykonano $P = 50$ (a więc wielokrotnie więcej) powtórzeń uczenia sieci (rozpoczynając je od losowych wartości początkowych wag), oraz przyjęto 3-krotnie większą (tj. $M = 15000$) maksymalną ilość epok uczenia. Zestaw danych uczących oraz pozostałe parametry pozostawiono bez zmian.

Dla tak skonstruowanych sieci zrealizowano aktywne strategie inwestycyjne w rozpatrywanym okresie OUT-S (1.10.1997 - 30.09.1998). Rezultaty przedstawiono w tabeli 5-13 oraz na rys. 5-23.



Rys. 5-23. Porównanie zmian wartości aktywów w czasie dla strategii aktywnych opartych o sieci 6A, 6B i 6C (przy zerowych kosztach transakcyjnych) oraz dla strategii pasywnej (inwestycja w indeks WIG 20) dla rocznego okresu OUT-S od 1.10.1997 do 30.09.1998. Źródło: opracowanie własne.

TABELA 5-13

| Baza strategii | Wartość oceny sieci | Zysk w całym okresie OUT-S | Maksymalna strata <i>dd</i> | Średnia dzienna stopa zwrotu | Odchylenie standardowe dziennych stóp zwrotu |
|----------------|------------------------|----------------------------|-----------------------------|------------------------------|--|
| Sieć 6A | 0,144080 | 952,0 | 17,4 % | 0,19 % | 1,84 % |
| Sieć 6B | 0,167324 ^{*)} | 149,4 | 34,3 % | 0,05 % | 2,08 % |
| Sieć 6C | 0,114775 | - 313,3 | 44,6 % | - 0,06 % | 1,79 % |
| Indeks WIG 20 | — | - 552,2 | 43,4 % | - 0,12 % | 2,55 % |

^{*)} Dla sieci 6B jest to wartość oceny na końcu okresu OUT-S (po wszystkich operacjach douczania wykonanych do 31.08.98)

Źródło: opracowanie własne.

Analiza zaprezentowanych wyżej wyników badań pozwala zauważyć, iż sieci 6B i 6C wykazują istotnie mniejszą skuteczność generowania prawidłowych sygnałów

transakcyjnych w rozważanym okresie OUT-S, niż sieć 6A (choć oparte o sieci 6B i 6C strategie aktywne są nadal bardziej efektywne od strategii pasywnej - inwestycji w indeks WIG 20). Na tej podstawie można wyprowadzić następujące wnioski:

- Douczanie sieci w oparciu o nowo pojawiające się obserwacje nie musi prowadzić do poprawy efektywności modelu decyzyjnego opartego o daną sieć neuronową (wbrew poglądom wielu badaczy).
- Próby dokładniejszego nauczenia sieci (poszukiwania kolejnych minimów funkcji oceny sieci) również nie muszą powodować poprawy efektywności strategii decyzyjnej opartej o tę sieć, mogą natomiast prowadzić do pogorszenia skuteczności modelu decyzyjnego. Choć (jak wynika z badań opisanych w podrozdziale 5.4.1) operacja taka prowadzi na ogół do poprawy funkcjonowania sieci w okresie IN-S, to jednak może okazać się ryzykowna dla sieci już dobrze przetestowanych w dość długim okresie OUT-S (pewnym potwierdzeniem tej tezy wydaje się też być fakt słabej zależności zysku w okresie OUT-S od wartości oceny sieci - por. podrozdz. 5.3, 5.4.1). Potwierdza to rozpowszechnioną opinię, że długotrwałe uczenie pogarsza zdolności sieci do generalizacji wyników.

6. Koncepcja aktywnej strategii zarządzania portfelem akcji w oparciu o zaproponowaną metodologię oraz dyskusja uzyskanych efektów.

Zaprezentowaną w rozdziale 4 metodologię można stosować nie tylko w celu generowania decyzji kupna lub sprzedaży całego portfela akcji, opartych o prognozy zachowania się danego rynku, tak jak to przedstawiono w rozdziale 5. Roboczą tezę sformułowaną na tym etapie pracy jest przypuszczenie, że wykorzystując w systemie decyzyjnym informację wejściową dotyczącą dynamiki kursów poszczególnych składników portfela (akcji), można na wyjściu uzyskać sygnały, które następnie można interpretować jako decyzje kupna lub sprzedaży danej pojedynczej akcji albo też jako rekomendacje dotyczące zwiększenia lub zmniejszenia udziału danej akcji w portfelu¹²¹. System generujący te sygnały może mieć przy tym charakter uniwersalny (ten sam dla wszystkich akcji), albo można skonstruować zestaw systemów - dla każdej akcji osobny. Poniżej zaproponowano koncepcję aktywnej strategii zarządzania portfelem inwestycyjnym (konstrukcji i bieżącej modyfikacji składu portfela) w oparciu o sygnały wyjściowe zestawu sieci neuronowych (lub pojedynczej sieci) skonstruowanych na podstawie przedstawionej metodologii.

¹²¹ W celu generowania decyzji kupna/sprzedaży akcji na podstawie sygnałów pochodzących z sieci neuronowej konieczne jest dodatkowo zdefiniowanie strategii decyzyjnej (ostatniego modułu systemu decyzyjnego); natomiast metoda wyznaczania udziałów poszczególnych akcji w całości portfela może być oparta bezpośrednio na sygnałach wyjściowych sieci neuronowych, co też zostało zaproponowane w podrozdziale 6.1.

6.1. Metodologia aktywnego zarządzania portfelem inwestycyjnym.

Rozważmy zagadnienie konstrukcji portfela inwestycyjnego, w skład którego mogą wchodzić akcje N spółek (oznaczonych numerami $1, 2, \dots, N$), które *a priori* traktujemy jednakowo (nie stosując żadnych preferencji). Oczywiście zakładamy, że w skład aktywów portfela może wchodzić również gotówka, pomijamy natomiast możliwość uwzględniania w portfelu instrumentów wolnych od ryzyka (np. obligacji lub bonów skarbowych), a także pomijamy możliwość stosowania krótkiej sprzedaży akcji¹²².

Dla każdej z rozważanych N spółek zbudujemy według opisanej w rozdziale 4 metodologii oddzielną sieć neuronową¹²³, generującą na wyjściu sygnały, których wartość określa prognozowaną dynamikę kursu akcji danej spółki w przyjętym horyzoncie czasowym. Analizując charakterystykę zaproponowanych w podrozdziale 4.5 typów zmiennych wyjściowych można przyjąć ogólną zasadę interpretacji sygnałów wyjściowych SN, polegającą na tym, że sygnały dodatnie stanowią rekomendacje kupna (posiadania) akcji danej spółki (przy czym im większa wartość bezwzględna sygnału wyjściowego tym rekomendacja ta jest silniejsza), natomiast sygnały ujemne (z tą samą uwagą dotyczącą ich wielkości) stanowią odpowiednie rekomendacje sprzedaży akcji. Poprzestając tutaj jedynie na powyższym stwierdzeniu i na podkreśleniu faktu, iż dobór typu zmiennej wyjściowej SN (a także innych parametrów systemu) w bardzo istotnym stopniu determinuje charakter i efektywność strategii zarządzania portfelem, pozostawimy propozycje określenia tych parametrów wraz z dyskusją do zaprezentowania w dalszej części pracy, poświęconej omówieniu wyników.

¹²² W chwili powstawania niniejszej pracy na Giełdzie Papierów Wartościowych w Warszawie nie istniała możliwość stosowania krótkiej sprzedaży, nietrudno jednak uogólnić dalej zaprezentowaną metodologię w taki sposób, aby możliwość tą uwzględnić.

¹²³ Ogólnie można przyjąć, że jest to sieć, która w procesie konstrukcji systemu (realizacji algorytmu genetycznego) uzyskała najlepszą ocenę końcową. Można jednak także analizować pewien zbiór najlepszych sieci uwzględniając wypadkowy (średni arytmetyczny lub średni ważony) sygnał wszystkich tych sieci. Inne podejście jest związane z zastosowaniem tej samej (uniwersalnej) sieci (zbioru sieci) dla wszystkich spółek lub dla określonej klasy spółek.

Niech $x(t)$ oznacza wartość aktywów portfela w chwili t (wyznaczoną na podstawie cen poszczególnych akcji w chwili t), natomiast przez $p_{got}(t), p_1(t), p_2(t), \dots, p_N(t)$ oznaczymy ułamkowe udziały odpowiednio: gotówki, wartości posiadanych akcji spółki 1, wartości posiadanych akcji spółki 2, ..., wartości posiadanych akcji spółki N - w całości aktywów portfela $x(t)$ w chwili t tak, aby spełniony był poniższy warunek¹²⁴:

$$p_{got}(t) + p_1(t) + p_2(t) + \dots + p_N(t) = 1 \quad \text{przy czym} \quad p_j(t) \geq 0, \quad i = 1, 2, \dots, N$$

Niech $s_1(t), s_2(t), \dots, s_N(t)$ oznaczają sygnały wyjściowe sieci neuronowych (interpretowane zgodnie z wyżej przedstawioną generalną zasadą), wygenerowane w chwili¹²⁵ t . Opierając się na wartościach tych sygnałów oraz na aktualnym składzie portfela określimy zamierzone transakcje dla chwili (dla sesji giełdowej) $t + 1$ według niżej zaprezentowanego algorytmu modyfikacji składu portfela. Generalną koncepcją wykorzystaną w tym algorytmie jest dążenie do sytuacji, w której ułamkowe udziały wartości akcji poszczególnych spółek w chwili (po sesji giełdowej) $t + 1$: $p_1(t+1), p_2(t+1), \dots, p_N(t+1)$ są proporcjonalne do odpowiednich sygnałów $s_1(t), s_2(t), \dots, s_N(t)$ (po wyeliminowaniu sygnałów niedodatnich oraz przy pewnych dodatkowych założeniach i ograniczeniach).

W zaproponowanym niżej algorytmie uwzględniono następujące opcje¹²⁶:

1. zastosowanie parametru s_{min} stanowiącego minimalną wartość sygnału sieci $s_j(t)$ niezbędnego do włączenia do portfela akcji spółki nie będącej aktualnie w składzie portfela¹²⁷ (jeżeli akcje danej spółki znajdują się w portfelu, parametr ten nie ma znaczenia dla transakcji tymi akcjami) - zob. krok 1 algorytmu,
2. możliwość określenia maksymalnej liczby spółek mogących wejść w skład portfela - parametr H ,

¹²⁴ Warunek ten reprezentuje przedstawione wyżej założenia dotyczące potencjalnego składu portfela oraz braku krótkiej sprzedaży; dopuszcza się przy tym możliwość pojawienia się niewielkiego ujemnego udziału gotówki (kredyt), przy czym w prezentowanej metodologii przyjęto zasadę dążenia udziału gotówki do zera.

¹²⁵ Przy znanych już cenach akcji i składzie portfela w momencie t .

¹²⁶ Parametry poszczególnych opcji mogą być dobierane przez inwestora (menedżera).

¹²⁷ Parametr ten decyduje o ignorowaniu znaczenia zbyt słabych sygnałów sieci przy dołączaniu nowych spółek do portfela.

3. uwzględnienie kosztów transakcyjnych (prowizji maklerskiej¹²⁸), określonych przez parametr r , wyrażający prowizję w formie ułamkowej (np. 0,005) od kwoty transakcji,
4. możliwość ograniczenia ułamkowego udziału $p_j(t)$ akcji danej spółki w całości aktywów portfela - do wartości parametru p_{max} ,
5. wprowadzenie parametru Δ definiującego minimalną dozwoloną wielkość transakcji akcjami danej spółki jako wartość $\Delta \cdot x(t)$ ¹²⁹.

Można przypuszczać, że do ewentualnych zastosowań praktycznych mogłaby wystarczyć znacznie bardziej uproszczona wersja algorytmu; jednak wprowadzenie pewnych niuansów i opcji w zaproponowanym niżej rozwiązaniu może być uzasadnione potrzebą wykorzystania algorytmu również do celów badawczych, mających na celu komputerowe testowanie aktywnych strategii zarządzania portfelem w zadanym okresie czasu na podstawie historycznych danych.

Na podstawie przedstawionych wyżej założeń skonstruowano **algorytm modyfikacji składu portfela** o następującej postaci:

Krok 1. Dla akcji spółek 1, 2, ..., N wyznacz kolejno wartości q_1, q_2, \dots, q_N

$$q_j = \begin{cases} 0 & \text{jeżeli } s_j(t) \leq 0 \text{ lub } (p_j(t) = 0 \text{ i } s_j(t) < s_{min}) \\ s_j(t) & \text{w przeciwnym przypadku} \end{cases}$$

W trakcie realizacji algorytmu wartości q_j powinny dążyć do zamierzonych nowych udziałów ułamkowych poszczególnych spółek w portfelu. W początkowym kroku 1 eliminujemy te akcje, dla których sygnał $s_j(t)$ jest niedodatni, a także nie włączamy do portfela nowych spółek, dla których sygnał, chociaż jest dodatni, nie przekroczył minimalnego progu (zob. opcja 1). Dla pozostałych akcji wartości początkowe udziałów w portfelu q_j są (zgodnie z przyjętą regułą) proporcjonalne do

¹²⁸ Przyjmujemy tutaj tzw. prowizję liniową (niezależną od wysokości transakcji).

¹²⁹ Jeżeli wartość transakcji ma być mniejsza od $\Delta \cdot x(t)$, nie dokonujemy zmiany stanu posiadania akcji danej spółki; warunek ten pozwoli uniknąć zbyt małych i częstych transakcji (ma on istotne znaczenie przy stosunkowo wysokich kosztach transakcyjnych). Parametr Δ można też interpretować jako minimalną wymaganą zmianę ułamkowego udziału akcji określonej spółki w portfelu.

odpowiednich sygnałów $s_j(t)$ (wartości te nie są jeszcze tutaj odpowiednio skalowane). Zawsze przy tym spełniony jest warunek $q_j \geq 0$ ($j=1,2,\dots,N$).

Krok 2. Oblicz wartość K równą ilości spółek, dla których $q_j > 0$ ($j=1,2,\dots,N$).

Krok 3. Jeżeli $K > H$, to znajdź $K-H$ najmniejszych dodatnich wartości q_j ($j=1,2,\dots,N$), po czym za te wartości podstaw 0.

Ograniczamy (jeżeli zachodzi konieczność) liczbę spółek w portfelu do H , eliminując akcje spółek, dla których sygnał SN jest najslabszy.

Krok 4. Oblicz współczynnik korekty φ związany z częścią kosztów transakcyjnych

$$\varphi = 1 + \frac{\delta}{1 - p_{got}(t)} \quad \text{gdzie:} \quad \delta = \sum_{j:q_j=0} r \cdot p_j(t)$$

Zadaniem współczynnika φ jest korekta wielkości zamierzonych transakcji w taki sposób, aby uwzględnić wartość kosztów transakcyjnych wynikających ze sprzedaży tych akcji, które zamierzamy sprzedać w całości ($q_j = 0$); całkowita wysokość tych kosztów wynosi $\delta x(t)$.

Krok 5. Oblicz sumę S wszystkich wartości q_j :
$$S = \sum_{j=1}^N q_j$$

Krok 6. Jeżeli $S = 0$ (tzn. $q_j = 0$ dla $j=1,2,\dots,N$), to przejdź do kroku 14; w przeciwnym przypadku podstaw $j \leftarrow 1$.

Krok 7. Jeżeli $q_j = 0$, to przejdź do kroku 12

Krok 8. $q_j \leftarrow q_j / S$ (skalowanie wartości q_j)

Krok 9. Jeżeli $q_j > p_{max}$, to $q_j \leftarrow p_{max}$

Ograniczenie ułamkowego udziału spółki j w portfelu - zob. opcja 4.

Krok 10. Jeżeli $|q_j - p_j(t) \cdot \varphi| < \Delta$, to $q_j \leftarrow p_j(t) \cdot \varphi$

Jeżeli zamierzona zmiana ułamkowego udziału spółki j w portfelu jest mniejsza od

minimalnej wartości Δ , korygujemy wielkość q_j w taki sposób, aby nie realizować w ogóle transakcji (według reguły przedstawionej w kroku 14) - zob. opcja 5.

Krok 11. Jeżeli $q_j > p_j(t)$, to $q_j \leftarrow \frac{q_j + p_j(t) \cdot \varphi \cdot r}{1 + r}$

Jeżeli $q_j < p_j(t)$, to $q_j \leftarrow \frac{q_j - p_j(t) \cdot \varphi \cdot r}{1 - r}$

Modyfikujemy wartość q_j w taki sposób, aby uwzględnić koszty transakcyjne związane z operacją kupna lub sprzedaży akcji spółki j - zob. opcja 3.

Krok 12. Jeżeli $j < N$, to ($j \leftarrow j + 1$; przejdź do kroku 7)

Krok 13. Jeżeli krok 5 (pętla krok 5 - krok 13) był wykonywany mniej niż 5 razy to przejdź do kroku 5.

Operacje określone w krokach 5, 6, ..., 11 należy powtarzać wykonując kilka iteracji w celu otrzymania wystarczająco dokładnych wartości q_j (nie można tego wykonać w pojedynczej iteracji); tutaj przyjęto liczbę iteracji równą 5.

Krok 14. Dla akcji spółek 1, 2, ..., N zrealizuj transakcje według reguł:

- Jeżeli $q_j = 0$, to sprzedaj (jeśli posiadasz) wszystkie akcje spółki j
- Jeżeli $q_j > 0$, to dokonaj transakcji akcjami spółki j za kwotę

$$\mathit{trans}(j) = x(t) \cdot (q_j - p_j(t) \cdot \varphi)$$

($\mathit{trans}(j) > 0$ oznacza kupno, $\mathit{trans}(j) < 0$ oznacza sprzedaż).

Kwota transakcji wynika z pomnożenia aktualnej wartości portfela przez różnicę ułamkowych udziałów akcji danej spółki w portfelu (z uwzględnieniem współczynnika korekcyjnego φ).

Zaprezentowany wyżej algorytm modyfikacji składu portfela realizuje aktywną strategię zarządzania portfelem akcji w oparciu o sygnały odpowiednio skonstruowanych i testowanych (według opracowanej tutaj metodologii) sieci neuronowych, oraz w oparciu o przedstawioną wyżej regułę proporcjonalnego

udziału akcji spółki w portfelu do wielkości odpowiedniego sygnału. Uwzględnienie kosztów transakcyjnych pozwala nie tylko na urealnienie procesu zarządzania portfelem, ale także na zbadanie efektywności strategii aktywnej w zależności od wielkości tych kosztów (wysokości prowizji maklerskiej). W pewnych przypadkach celowe może okazać się ograniczenie tych kosztów, toteż dla zaproponowanego algorytmu przewidziano dodatkowo następujące opcje:

- zastąpienie bieżących wartości sygnałów sieci $s_1(t)$, $s_2(t)$, ..., $s_M(t)$ odpowiednimi średnimi adaptacyjnymi dla poszczególnych sygnałów; operacja ta pozwala na wyeliminowanie występujących na wyjściu sieci szumów o wysokiej częstotliwości i może być przydatna zwłaszcza przy zastosowaniu dłuższych horyzontów prognozy; w trakcie badań zaobserwowano, że dobre rezultaty przynosi zastosowanie okresu średniej adaptacyjnej równej wielkości przyjętego (dla SN) horyzontu prognozy h ;
- dokonywanie modyfikacji składu portfela w okresach większych niż jednodniowe (np. nie częściej niż co tydzień);

W dalszej części pracy zaproponowano algorytm symulacji procesu zarządzania portfelem inwestycyjnym w określonym przedziale czasu na podstawie omówionej wyżej metodologii, a następnie przeprowadzono badania efektywności tej metody.

6.2. Algorytm symulacji procesu aktywnego zarządzania portfelem.

W celu przeprowadzenia badań efektywności wyżej zaproponowanej metody wspomagania zarządzania portfelem inwestycyjnym, w oparciu o odpowiednio skonstruowane i nauczone sieci neuronowe, zastosowano technikę komputerowej symulacji procesu zarządzania, opartą o sygnały SN oraz wykorzystującą rzeczywiste dane pochodzące z rynku akcji w rozważanym okresie czasu. Przyjęto przy tym założenie, że wygenerowane (po ogłoszeniu wyników bieżącej sesji giełdowej) przez algorytm modyfikacji składu portfela zlecenia dotyczące transakcji są realizowane zawsze na kolejnej sesji, z uwzględnieniem kursów obowiązujących na fixingu tej sesji (nie są tu uwzględniane notowania ciągłe). Założono ponadto, że zlecenia

(transakcje) te są realizowane w całości i dokładnie według wielkości kwot (a nie liczby sztuk akcji) określonych przez ten algorytm. Procedurę symulacji procesu zarządzania portfelem przedstawiono poniżej, zachowując oznaczenia przyjęte w podrozdziale 6.1.

Algorytm symulacji procesu zarządzania portfelem akcji

Krok 1. Podstaw $t \leftarrow 0$, przyjmij wartości początkowe:

$$s_j(0) = 0 \text{ oraz } p_j(0) = 0 \quad (j=1, 2, \dots, N); \quad p_{got}(0) = 1$$

$$x(0) = \mathbf{wig}(0) \quad (\text{oznaczamy przez } \mathbf{wig}(t) \text{ wartość indeksu WIG w chwili } t)$$

Początkowo aktywa portfela zawierają jedynie gotówkę, natomiast sygnały SN są dla wszystkich akcji zerowe. Przyjęcie początkowej wartości aktywów równej wielkości indeksu WIG ma na celu ułatwienie porównania strategii aktywnej z pasywną.

Krok 2. Podstaw $t \leftarrow t + 1$.

t oznacza numer kolejnej sesji giełdowej w analizowanym okresie czasu ($t = 1, 2, \dots, T$).

Krok 3. Zrealizuj procedurę modyfikacji składu portfela (w oparciu o algorytm opisany w podrozdziale 6.1), która na podstawie wartości $s_j(t-1)$, $p_j(t-1)$ ($j=1, 2, \dots, N$), $p_{got}(t-1)$, oraz $x(t-1)$ wyznaczy zlecenia transakcji w kwotach $\mathbf{trans}(j)$ ($j=1, 2, \dots, N$), oraz zamierzone nowe proporcje udziałów akcji poszczególnych spółek w portfelu q_j ($j=1, 2, \dots, N$), (w przybliżeniu wartości q_j będą odpowiadać wielkościom rzeczywistym $p_j(t)$).

Transakcje (dla których zlecenia zostały wygenerowane w oparciu o dane z chwili $t-1$) zostaną zrealizowane na bieżącej sesji t .

Krok 4. Oblicz całkowitą wysokość kosztów transakcyjnych (provizji maklerskiej):

$$\mathbf{prow} = r \cdot \sum_{j=1}^N |\mathbf{trans}(j)|$$

Krok 5. Na podstawie wyników sesji t wyznacz aktualną wartość aktywów portfela $x(t)$ według formuły:

$$x(t) = x(t-1) \cdot \left(p_{got}(t-1) + \sum_{j=1}^N \frac{p_j(t-1) \cdot k_j(t)}{k_j(t-1)} \right) - \text{prow}$$

gdzie przez $k_j(t)$ oznaczamy kurs akcji spółki j na sesji t .

Transakcje $\text{trans}(j)$ zawarte na sesji t nie wpływają bezpośrednio na zmianę wartości aktywów portfela na tej sesji (wpływają jedynie na wysokość całkowitych kosztów transakcyjnych prow). Wpływ na zmianę wartości portfela mają zmiany kursów poszczególnych jego składników (akcji), w powiązaniu z udziałami tych akcji w całości aktywów.

Krok 6. Wyznacz aktualne wielkości udziałów poszczególnych akcji w portfelu:

$p_j(t)$ ($j=1, 2, \dots, N$) według zależności:

$$p_j(t) = \begin{cases} \frac{p_j(t-1) \cdot k_j(t) \cdot x_j(t-1)}{k_j(t-1) \cdot x(t)} + \frac{\text{trans}(j)}{x(t)} - \frac{r \cdot |\text{trans}(j)|}{x(t)} & \text{jeżeli } q_j > 0 \\ 0 & \text{jeżeli } q_j = 0 \end{cases}$$

przy czym jeżeli okaże się, że $p_j(t) < 0$, to przyjmij $p_j(t) = 0$.

W górnym wierszu powyższej zależności pierwszy składnik reprezentuje zmianę wartości akcji spółki j w portfelu, spowodowaną zmianą kursu tych akcji, oraz zmianą wartości całego portfela (np. jeżeli kurs akcji j wzrósł, a wartość portfela spadła, to udział wartościowy tych akcji w całym portfelu zwiększył się); element ten nie był możliwy do uwzględnienia podczas wyznaczania wartości q_j , gdyż nieznane były wtedy wartości $k_j(t)$ i $x_j(t)$. Drugi składnik wyrażenia reprezentuje zmianę udziału akcji j w portfelu w wyniku dokonanej transakcji, zaś trzeci składnik reprezentuje koszty tej transakcji. Dolny warunek przedstawionej zależności wyraża zasadę sprzedaży wszystkich akcji spółki j , jeżeli otrzymamy $q_j = 0$.

Krok 7. Oblicz udział gotówki w portfelu:

$$p_{got}(t) = 1 - \sum_{j=1}^N p_j(t)$$

Krok 8. W oparciu o bieżące dane oraz skonstruowane sieci neuronowe wyznacz aktualne sygnały wyjściowe SN: $s_j(t)$, ($j=1, 2, \dots, N$).

Krok 9. Jeżeli $t < T$ to przejdź do kroku 2.

Krok 10. Dokonaj końcowej oceny efektów aktywnej strategii zarządzania portfelem.

6.3. Badania efektywności metody dla wybranego zestawu akcji.

Poniżej zaproponowano parametry konstrukcji systemu decyzyjnego oraz zaprezentowano rezultaty wykorzystania omawianej metody aktywnego zarządzania portfelem inwestycyjnym przy założeniu, że w skład portfela mogą wchodzić akcje 12-tu spółek o najmniejszej kapitalizacji, notowane na rynku podstawowym Giełdy Papierów Wartościowych w Warszawie. Jako kryterium wyboru przyjęto wartość rynkową spółki (kapitalizację), określoną w dniu 16.04.1997¹³⁰ - wyselekcjonowano te spółki z rynku podstawowego, dla których wartość ta nie przekraczała 50 mln zł¹³¹. Są to następujące firmy: Tonsil, Próchnik, Swarzędz, Wólczanka, Espebepe, Remak, Bytom, BWR, KPBP Bick, Mostostal Gdańsk, Pekpol, Oława.

Do uczenia i walidacji sieci (konstrukcji systemu) użyto danych z notowań wszystkich 12-tu wyżej wymienionych spółek, pochodzących z 33-miesięcznego okresu od 31.03.1994 do 31.12.1996. Następnie dokonano oceny efektywności skonstruowanego systemu w oparciu o dane OUT-S z rocznego okresu od 1.01.1997 do 1.01.1998¹³².

¹³⁰ Źródło: *Parkiet* z dnia 17.04.1997.

¹³¹ Kryterium to przyjęto arbitralnie.

¹³² Należy zwrócić uwagę, iż oba wymienione przedziały czasowe charakteryzują się występowaniem zarówno rosnących, jak i malejących trendów rynkowych, zaś wartości indeksu WIG na początku i na końcu danego okresu różnią się nieznacznie; fakt ten stanowił jedno z kryteriów doboru właśnie takich okresów czasowych do konstrukcji, a następnie oceny efektywności systemu.

6.3.1. Zaproponowane parametry systemu decyzyjnego.

W rozpatrywanym zagadnieniu założono, iż system decyzyjny będzie oparty na pojedynczej sieci neuronowej, wspólnej dla wszystkich 12-tu wyżej wymienionych akcji. Przyjęto zatem, iż cały zbiór danych wykorzystany do uczenia, walidacji i końcowej oceny tej sieci będzie obejmował dane (z określonego przedziału czasu) pochodzące ze wszystkich 12-tu spółek. Poniżej zaproponowano parametry dla tak zdefiniowanego systemu decyzyjnego, realizującego aktywną strategię zarządzania portfelem inwestycyjnym.

Parametry modułu przetwarzania informacji wejściowej

- Wykorzystano jedynie 81 dodatkowych procedur wstępnego przetwarzania informacji wejściowej dla szeregu czasowego $d(t)$ opisanego formułą 4.1 (zob. tabela 4-1) i zastosowano je właśnie do tego szeregu; podejście to oparte zostało na przesłance, iż w rozważanym problemie bardziej istotny od dynamiki samego kursu określonej akcji jest charakter jego zmian w stosunku do indeksu giełdy.

Parametry modułu preprocessingu danych wejściowych

- parametr trymownia wartości nietypowych (formuła (4.3)) $\gamma_{tr} = 4$
- w przypadku zmiennych wejściowych podlegających standaryzacji (wzór (4.5)) zastosowano parametr $\gamma_{st} = 1,5$
- zastosowano skokowe rozdzielenie wartości dodatnich i ujemnych dla zmiennych normalizowanych „bez przesunięcia zera” (formuła (4.7)) z parametrem $\gamma_{rw} = 0,05$
- nie wykorzystano analizy głównych składowych do wstępnego przetwarzania danych

Parametry architektury i uczenia sieci typu *perceptron*

- przyjęto strukturę sieci zdefiniowaną w podrozdziale 4.4, zawierającą jedną warstwę ukrytą z genetycznie wyznaczoną liczbą neuronów
- jako zmienną wyjściową przyjęto współczynnik nachylenia prostej regresji (nachylenie trendu) dla szeregu czasowego $d(t)$ (według opisu w podrozdziale

4.5.4); prosta regresji jest wyznaczana dla horyzontu czasowego h , który podlega genetycznej adaptacji w trakcie realizacji algorytmu genetycznego ($h \in [2, 20]$)

- zrezygnowano z obcinania wartości nietypowych dla zmiennej wyjściowej
- z ciągu danych wyeliminowano te wzorce, dla których zmienna wyjściowa (ucząca) przyjmuje wartości oddalone od jej średniej arytmetycznej o mniej niż 25% wielkości odchylenia standardowego¹³³
- dla wielkości wzorcowych zmiennej wyjściowej zastosowano operację odejmowania ich wartości średniej, po czym przeprowadzono dla nich normalizację według standardowej formuły wyrażonej wzorem (2.21)
- elementy ciągu danych służących do konstrukcji modelu podzielono na ciąg uczący i ciąg walidacyjny według schematu **C** (rys. 4-4, podrozdział 4.6) przyjmując $T_u = 50$ oraz $T_v = 50$
- bieżącą (wstępną) ocenę sieci realizowano wyłącznie w oparciu o błąd średniokwadratowy dla ciągu walidacyjnego (dla kryterium (4.15) przyjęto $p = 0$)
- przyjęto następujące wartości stałych decydujących o modyfikacji współczynników uczenia i bezwładności w trakcie procesu uczenia: $\alpha = 0,05$, $\beta = 0,95$ (zob. podrozdział 4.6)
- przyjęto wartości parametrów funkcji aktywacji (2.4): $a = 1,05$ oraz $b = 1,00$
- ustalono maksymalną liczbę epok uczenia sieci $M = 100$, oraz liczbę powtórzeń procesu uczenia każdej sieci $P = 1$ (tak dobrane wartości wynikają ze stosunkowo dużej wielkości ciągu uczącego)
- uczenie realizowano metodą wstecznej propagacji błędów (zob. podrozdział 2.5) przyjmując kwadratową funkcję błędu określoną wzorem (2.9).

Kryterium końcowej oceny sieci neuronowej

Końcowej oceny sieci dokonywano na podstawie ciągu walidacyjnego (podobnie jak w przypadku oceny wstępnej), ale według innej reguły - opartej o ocenę

¹³³ Decydując się na eliminację tych elementów ciągu danych oparto się na empirycznie uzasadnionym założeniu, iż sieć neuronową należy uczyć w oparciu o wzorce prezentujące zdecydowaną, istotną zmianę kursu akcji; wzorce „przeciętne” wnoszą stosunkowo niewiele użytecznej dla celów decyzyjnych informacji, wydłużając przy tym znacznie długość ciągu uczącego.

efektywności sygnałów wyjściowych nauczonej sieci w przyjętej strategii zarządzania portfelem. W tym celu dla każdego osobnego przedziału czasu o długości $T_v = 50$, obejmującego spójny zbiór danych wchodzących w skład ciągu walidacyjnego, realizowano procedurę zarządzania hipotetycznym portfelem złożonym z indeksu WIG oraz akcji tej spółki, której dotyczą dane z analizowanego przedziału czasu. Udział akcji tej spółki w portfelu w danej chwili był proporcjonalny do sygnału wyjściowego SN, reprezentującego prognozowaną zmianę jej kursu (założono przy tym także możliwość krótkiej sprzedaży akcji w przypadku ujemnego sygnału sieci). Jako końcową ocenę sieci przyjęto określoną w procentach wielkość, wyrażającą stosunek liczby tych okresów o długości T_v z ciągu walidacyjnego, dla których zysk z takiego portfela był wyższy od zysku z pasywnej inwestycji w indeks WIG, do całkowitej ilości analizowanych okresów.

Należy zauważyć, że tak zdefiniowane kryterium oceny sieci uwzględnia bardzo istotny czynnik, jakim jest stabilność efektywności SN w czasie (systematyczność generowania zysków). Kryterium to preferuje bowiem te sieci, które stale generują ponadprzeciętne zyski w określonych przedziałach czasu (choćby niewiele większe od zysków z pasywnej inwestycji w indeks), niż sieci, które jednorazowo mogą wygenerować znaczący zysk, ale w pozostałym okresie nie przynoszą właściwych efektów. Dodatkowo kryterium to gwarantuje uniwersalność sieci, tzn. możliwość zastosowania jej do prognozowania akcji wielu spółek, wchodzących w skład określonej klasy przedsiębiorstw.

Parametry algorytmu genetycznego

- przyjęto wielkość populacji $L = 150$ oraz parametry $I_B = 42$, $I_C = 99$, $I_O = 9$
- mutację realizowano w oparciu o typ mutacji 1 (zob. podrozdział 4.7.3)
- ustalono prawdopodobieństwo mutacji dla wszystkich genów równe 0,01
- założono liczbę iteracji (populacji) równą 60
- określono początkową liczbę wejść SN (w populacji początkowej) równą 10

Parametry procedury zarządzania portfelem

Proces aktywnego zarządzania portfelem inwestycyjnym dla akcji 12-tu wymienionych powyżej spółek o niskiej kapitalizacji realizowano w oparciu o zaprezentowany w podrozdziale 6.1 algorytm modyfikacji składu portfela, przy założeniu możliwości dokonywania codziennych korekt tego składu. Korekty te opierały się bezpośrednio na wartościach sygnałów wyjściowych $s_j(t)$ sieci neuronowej otrzymanej w wyniku konstrukcji systemu, przy czym sygnały te, przed wykorzystaniem ich w procedurze modyfikacji portfela, poddawane były operacji postprocessingu (por. podrozdział 4.8).

W zastosowanym algorytmie nie wprowadzano opcjonalnych ograniczeń (zob. podrozdział 6.1) i przyjęto następujące wartości parametrów: $s_{min} = 0$, $H = 12$, $p_{max} = 1$, $\Delta = 0$. Ponadto przyjęto zerowe koszty transakcyjne ($r = 0$).

6.3.2. Wyniki badań efektywności zaproponowanej metody zarządzania portfelem.

W pierwszym etapie badań dokonano - w oparciu o zaprezentowaną w niniejszej pracy metodologię - konstrukcji systemu decyzyjnego dla rozpatrywanego zagadnienia. Efektem końcowym tego etapu (otrzymanym w wyniku realizacji algorytmu genetycznego) było skonstruowanie odpowiedniej sieci neuronowej, generującej sygnały dla algorytmu modyfikacji składu portfela. Sieć ta posiada 10 wejść oraz 18 neuronów w warstwie ukrytej (wartości te osiągnięte zostały na drodze genetycznej optymalizacji). Ponadto uzyskany również w efekcie genetycznej adaptacji horyzont prognozy dla zmiennej wyjściowej wynosi $h = 2$ (oznacza to, że w praktyce sieć prognozuje zmianę szeregu czasowego $d(t)$ od następnej sesji do sesji bezpośrednio po niej następującej).

Wykaz zmiennych wejściowych dla otrzymanej sieci neuronowej zamieszczono w tabeli 6-1. W kolejnych kolumnach tabeli wyszczególniono: numer procedury wstępnego przetwarzania informacji wejściowej pochodzący z tabeli 4-1, nazwę danej procedury, oraz jej parametry. Wymienione procedury są stosowane do

przetwarzania danych pochodzących z szeregu czasowego $d(t)$, określonego formułą 4.1.

Wartość końcowej oceny dla tej sieci według kryterium przyjętego w podrozdziale 6.3.1 wyniosła **93,6 %**.

TABELA 6-1

| Nr procedury w tabeli 4-1 | Nazwa zmiennej (procedury) wejściowej | Parametry procedury |
|---------------------------|--|---------------------|
| 1 | Wskaźnik MACD (różnica średnich wykładniczych) | n = 18 |
| 9 | Wskaźnik momentum generujący impulsy w oparciu o linię sygnałową | n = 28 |
| 31 | Oscylator skumulowanego wolumenu VAO z linią sygnałową | r = 8 p = 10 |
| 50 | Wskaźnik zmienności (<i>volatility</i>) VI | r = 27 |
| 61 | Różnica średnich kroczących (wersja 1) | n = 18 |
| 63 | Różnica średnich kroczących (wersja 3) | n = 11 |
| 87 | Wskaźnik zmiany ROC dla średniej wykładniczej | r = 25 n = 7 |
| 98 | Wskaźnik siły trendu ADX | r = 12 |
| 104 | Wskaźnik <i>Stochastic Slow S5</i> | r = 25 s = 22 |
| 106 | Nachylenie trendu (wariant 1) | h = 19 |

Źródło: opracowanie własne.

Z kolei dokonano oceny efektywności systemu decyzyjnego opartego o otrzymaną sieć w rocznym okresie OUT-S (dane z tego okresu nie zostały użyte do uczenia, walidacji i oceny sieci) od 1 stycznia 1997 do 1 stycznia 1998. Zasadniczym elementem tej oceny była analiza przyrostu aktywów portfela (zysku z portfela), dla którego realizowano w rozpatrywanym okresie aktywną strategię zarządzania opartą o sygnały rozważanej sieci neuronowej. Dokonano również porównania zmian wartości portfela z analogicznymi zmianami indeksu WIG (reprezentującymi zyski z zastosowania strategii pasywnej polegającej na inwestycji w ten indeks), oraz ze zmianami odpowiednio skonstruowanego nieważonego indeksu dla analizowanych 12-tu spółek (indeks ten reprezentuje stan hipotetycznego portfela złożonego ze

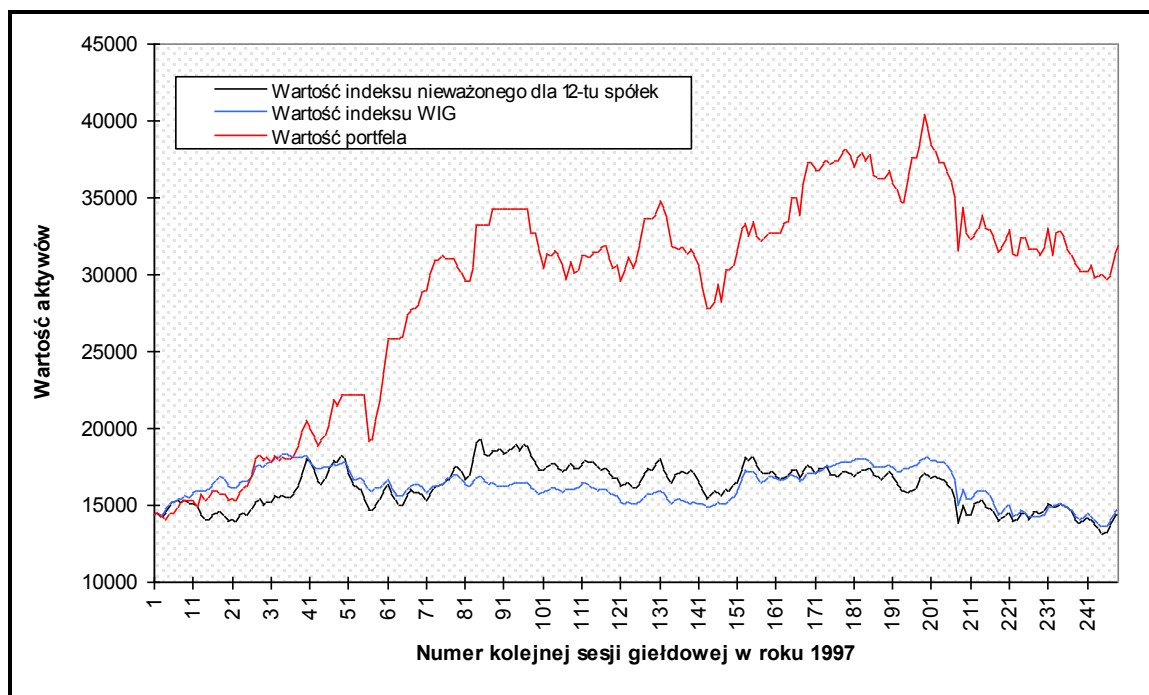
wszystkich 12-tu akcji przy założeniu, że udziały poszczególnych akcji w portfelu są jednakowe i nie ulegają zmianom).

Badania przeprowadzono przy wykorzystaniu algorytmu symulacji procesu aktywnego zarządzania portfelem, opisanego w podrozdziale 6.2 (algorytm zaimplementowano w języku programowania C). Początkowy stan aktywów portfela przyjęto w wysokości 14484,0 - jest to wielkość równa wartości indeksu WIG w drugim dniu analizowanego okresu - 3.01.1997 (informacja z pierwszego dnia - 2.01.1997 - została wykorzystana jedynie do wygenerowania sygnałów sieci dla transakcji, koniecznych do zrealizowania w dniu następnym). Rezultaty badań w postaci sumarycznych wartości zysku z poszczególnych metod inwestowania dla całego okresu zawarto w tabeli 6-2.

TABELA 6-2

| Wartość portfela i osiągnięty zysk | Aktywna strategia zarządzanie portfelem | Strategia pasywna oparta o indeks WIG | Strategia pasywna oparta o indeks nieważony dla 12-tu analizowanych spółek |
|------------------------------------|---|---------------------------------------|--|
| Stan początkowy portfela | +14484,0 | +14484,0 | +14484,0 |
| Stan końcowy portfela | +31597,2 | +14668,0 | +14389,1 |
| Osiągnięty zysk całkowity | +17113,2 | +184,0 | -94,9 |

Źródło: opracowanie własne.

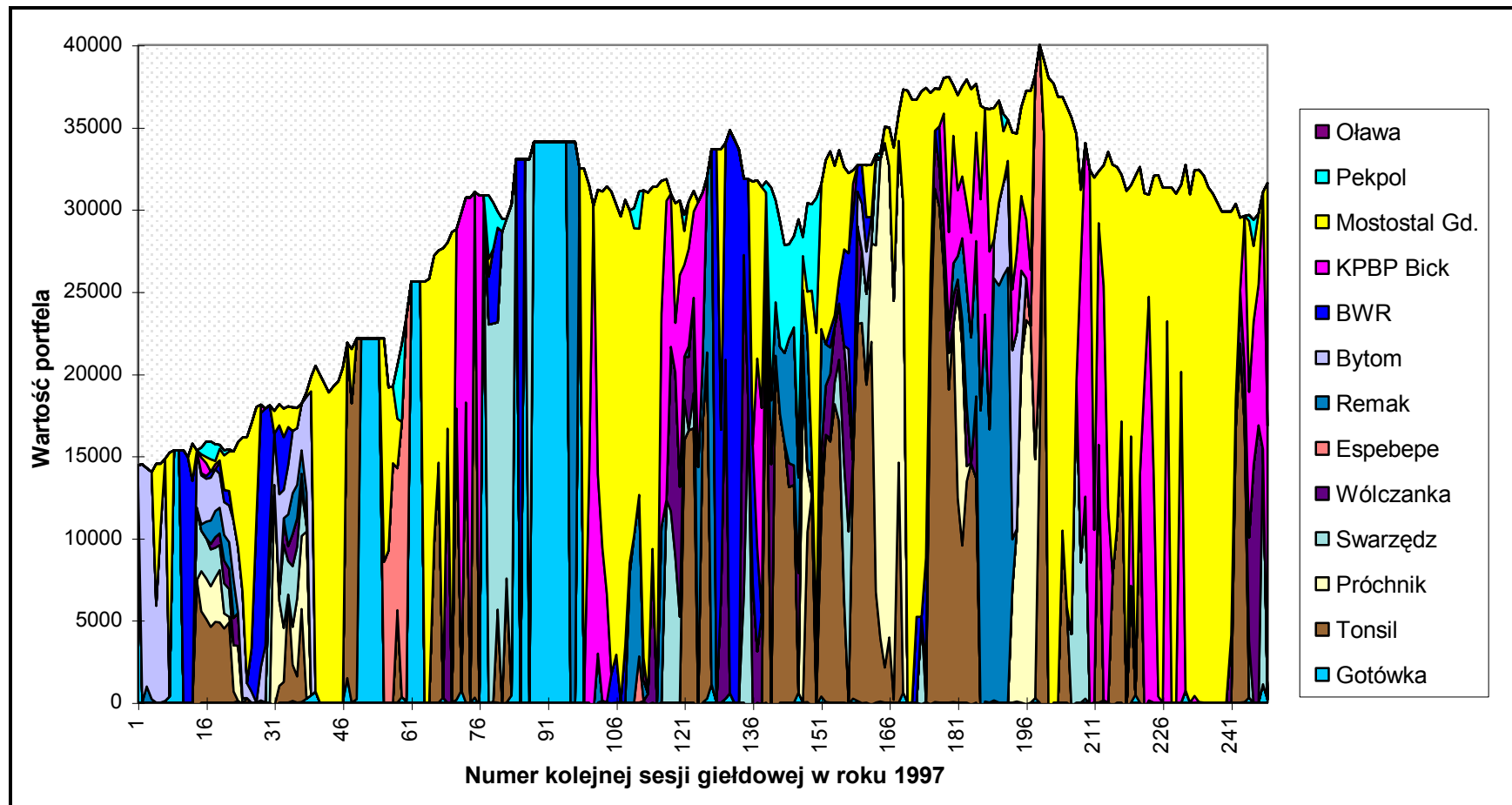


Rys. 6-1. Zmiany wartości aktywów w badanym okresie obejmującym rok 1997 dla portfela zarządzanego przy wykorzystaniu sygnałów sieci neuronowej, inwestycji pasywnej w indeks WIG, oraz inwestycji pasywnej w nieważony indeks 12-tu rozważanych spółek (portfel o stałym składzie, zawierającym równe udziały wszystkich akcji). Źródło: opracowanie własne.

Na rys. 6-1 zaprezentowano przebiegi zmian wartości: aktywów portfela zarządzanego przy wykorzystaniu SN, indeksu WIG, oraz indeksu nieważonego dla 12-tu analizowanych spółek w rozpatrywanym okresie czasu (dwa ostatnie przebiegi reprezentują odpowiednie strategie pasywne).

Z kolei na rys. 6-2 zaprezentowano w formie graficznej wielkości oraz przebieg zmian udziałów akcji poszczególnych spółek w całości aktywów portfela, w funkcji czasu (numeru sesji giełdowych w analizowanym okresie czasu, obejmującym rok 1997). Poszczególne akcje oznaczono różnymi kolorami, wysokość pola o określonym kolorze na pionowym przekroju wykresu dla danej chwili czasu reprezentuje wartość udziału oznaczonej tym kolorem spółki w portfelu. Osobnym kolorem oznaczono też udział gotówki w portfelu (zdarzały się sytuacje, kiedy sygnały wyjściowe sieci dla wszystkich spółek były ujemne - wówczas po sprzedaży wszystkich akcji portfel zawierał wyłącznie gotówkę).

Dokonując analizy otrzymanych wyników należy stwierdzić, że w rozpatrywanym przypadku uzyskano obiecujące rezultaty dla procesu aktywnego zarządzania portfelem inwestycyjnym w oparciu o sygnały odpowiednio skonstruowanej i nauczonej sieci neuronowej. Zastosowana metoda pozwoliła na osiągnięcie w okresie jednego roku zysku w wysokości około 118% początkowej inwestycji, podczas gdy strategię pasywne, nie zakładające korekt składu portfela w oparciu o prognozy dynamiki kursu dla poszczególnych jego składników, nie przyniosły praktycznie żadnych efektów. Rezultaty te stanowią obiecującą przesłankę do dalszych badań nad wykorzystaniem zaproponowanej metodologii w procesie aktywnego zarządzania portfelem inwestycyjnym.



Rys. 6-2. Graficzny schemat udziałów akcji poszczególnych spółek w całości aktywów portfela oraz zmiany tych udziałów w analizowanym okresie czasu, obejmującym rok 1997. Źródło: opracowanie własne.

Należy jednocześnie zwrócić uwagę, że ewentualne praktyczne aplikacje zaprezentowanej metody muszą zostać poprzedzone dużą ilością badań i testów w celu sprecyzowania optymalnych parametrów dla omawianej metodologii, a także dokonania ewentualnych jej modyfikacji. Trzeba bowiem podkreślić, iż w trakcie badań nie zawsze udawało się osiągnąć wystarczająco obiecujące rezultaty (np. znacząco gorsze efekty dla okresów testowych OUT-S osiągnano przy innych zestawach parametrów, nie udało się także uzyskać zadowalających rezultatów zarządzania portfelem - w okresach OUT-S - dla innych klas akcji obejmujących spółki o większej kapitalizacji). Oddzielnym problemem wymagającym zbadania jest zminimalizowanie wrażliwości metody na występujące w rzeczywistości koszty transakcyjne.

Rozwiązanie powyższych zagadnień wymaga wielu badań o charakterze empirycznym. Na obecnym etapie można postawić hipotezę, iż zaproponowana metodologia może okazać się bardzo efektywna w praktycznych zastosowaniach, jednak efektywność ta jest uzależniona od odpowiednio i starannie dobranego zestawu jej parametrów.

7. Podsumowanie otrzymanych wyników i wnioski końcowe.

Dokonując syntezy zaprezentowanych w niniejszej pracy wyników badań nad zastosowaniem sieci neuronowych i algorytmów genetycznych w zagadnieniu generowania decyzji inwestycyjnych, należy zwrócić uwagę na szereg czynników warunkujących efektywność zaproponowanej metodologii, a także wymagających uwzględnienia w przypadku ewentualnych praktycznych jej zastosowań. Poniżej omówiono najistotniejsze z nich, oraz sformułowano szereg wniosków, mogących również służyć jako przesłanki dla dalszych badań w tej dziedzinie.

Z badań zaprezentowanych w rozdziale 5 jednoznacznie wynika, że przy wykorzystaniu zaproponowanej metodologii można otrzymać sieci neuronowe mogące służyć jako efektywne narzędzie w procesie generowania sygnałów transakcyjnych dla zakupu i sprzedaży indeksu giełdowego WIG 20, to znaczy dla realizacji strategii opartej o prognozy dynamiki całego rynku. Z kolei w rozdziale 6 przedstawiono skuteczną metodę aktywnego zarządzania portfelem akcji w oparciu o sygnały sieci neuronowej, reprezentujące prognozy zmian kursu akcji w stosunku do wartości indeksu WIG. Otrzymane rezultaty pozwalają wyrazić generalny pogląd (często zresztą akcentowany w literaturze), iż SN można z powodzeniem stosować w takich zagadnieniach, jak prognozowanie dynamiki rynków finansowych (np. rynku akcji), oraz generowanie strategii inwestycyjnych wykorzystujących te prognozy.

Analizując rezultaty zawartych w niniejszej pracy badań dotyczących doboru istotnych parametrów metodologii, oraz ich wpływu na końcowe efekty jej zastosowania, można wywnioskować, iż na efektywność systemu decyzyjnego niewątpliwie wpływ mają takie czynniki jak:

- **Optymalny dobór wag początkowych SN.** Chociaż w pracy nie podano rozwiązania dla tego problemu, poza metodą wielokrotnego startu uczenia od losowo wygenerowanych punktów w przestrzeni wag i każdorazowej oceny sieci, to jednak wykazano istotny wpływ tego parametru na efektywność sieci. Zagadnienie to wymaga niewątpliwie przeprowadzenia dalszych prac badawczych.
- **Dobór odpowiedniej funkcji błędu podczas procesu uczenia sieci.** W pracy wykazano przydatność klasy funkcji opartych na sumie potęg modułów poszczególnych błędów (formuła 2.11) oraz zaobserwowano, że istotnym parametrem może być wykładnik dla tych potęg, decydujący o wrażliwości oceny sieci na duże błędy (znaczące „pomyłki” sieci). Pokazano, iż wykładnik mniejszy od powszechnie przyjmowanej wartości 2 (kwadrat błędu) może dawać lepsze rezultaty.
- **Postać i parametry funkcji aktywacji neuronów.** Przeprowadzone badania dla antysymetrycznej funkcji sigmoidalnej pozwoliły na potwierdzenie hipotezy, że przeciwdziedzina tej funkcji powinna być nieco większa od zakresu wartości znormalizowanych wyjściowych wartości wzorcowych dla sieci. Pokazano też znaczący wpływ parametru odpowiedzialnego za nachylenie funkcji aktywacji i podano zakres jego optymalnych wartości w analizowanym problemie.
- **Wartości współczynników uczenia i bezwładności, oraz kryterium zatrzymania procesu uczenia.** Chociaż nie przeprowadzono bezpośrednich badań dotyczących doboru tych wartości, to jednak zaproponowano heurystyczną metodę ich doboru i modyfikacji podczas uczenia, wraz z graficzną analizą procesu uczenia.
- **Wielkość warstwy ukrytej SN.** Przeprowadzone badania pokazały, że dobre efekty może przynosić zastosowanie sieci posiadających stosunkowo niewielki rozmiar warstwy ukrytej; nadmierne zwiększanie tej warstwy nie przynosi na ogół poprawy efektywności, a jedynie niepotrzebnie zwiększa rozmiary sieci, powodując wydłużenie procesu uczenia i ryzyko wystąpienia efektu przeuczenia. Wydaje się jednak, że optymalna liczba neuronów w warstwie ukrytej może być

uzależniona od długości ciągu uczącego (hipoteza ta wymaga potwierdzenia w trakcie dalszych badań).

W wyniku przeprowadzonych w pracy badań zaproponowano wprowadzić najkorzystniejsze wielkości wybranych parametrów dla rozważanego problemu dotyczącego poszukiwania aktywnej strategii decyzyjnej dla indeksu WIG 20, jednak należy przy tym pamiętać, iż w innym zagadnieniu wykorzystania SN na rynku finansowym optymalne wartości tych parametrów mogą być inne.

Podczas realizacji badań stwierdzono, że bardzo ważnym elementem budowy systemu decyzyjnego wykorzystującego SN jest kwestia wstępnego przetwarzania danych dla sieci. W trakcie opisywanych badań zwrócono szczególną uwagę na wykorzystanie **analizy głównych składowych** w celu redukcji i dekorelacji zmiennych wejściowych, oraz przedstawiono wyniki badań określających skutki zastosowania sieci PCA w rozważanym zagadnieniu, wraz z propozycją wybranych parametrów tej metody. Jakkolwiek wyniki analiz nie wskazują jednoznacznie na konieczność stosowania przekształcenia PCA, to jednak metoda ta może stanowić niezwykle efektywne narzędzie wstępnej transformacji danych, szczególnie w przypadku dużej ilości istotnie skorelowanych zmiennych wejściowych.

Analizując zastosowanie metody PCA zaobserwowano, że dobre efekty funkcjonowania sieci typu *perceptron* można uzyskać stosując transformację danych wejściowych, polegającą na eliminacji ostatnich (tzn. tych o najmniejszej wariancji) składowych wektora po przekształceniu PCA, a uwzględniającą jedynie około 60 % początkowych, najbardziej informatywnych głównych składowych. Pokazano przy tym możliwość zastosowania osobnej sieci neuronowej typu PCA, uczonej według reguły Sangera, która już po stosunkowo niewielkiej liczbie epok uczenia może dobrze spełniać zadanie realizacji transformacji danych wejściowych. Nie należy jednak wykluczyć, że w wielu problemach skuteczniejsza (szybsza) może się okazać klasyczna metoda analizy głównych składowych, oparta na obliczeniu wektorów własnych macierzy kowariancji.

Z zaprezentowanych w pracy badań efektywności omawianej metodologii wynika generalne spostrzeżenie, że sieci neuronowe, jako podstawowe elementy modeli

decyzyjnych generujących sygnały transakcyjne dla indeksu WIG 20, wykazują z reguły wysoką skuteczność w okresie IN-S (zawierającym dane użyte do uczenia i walidacji sieci), ale efektywność ta na ogół istotnie spada w okresie OUT-S, obejmującym dane związane z rzeczywistym (praktycznym) wykorzystaniem gotowego modelu opartego o nauczoną SN. Wykazano jednak, że istnieją takie sieci, które utrzymują stale wysoką efektywność, również w okresie OUT-S, przy czym dość nieoczekiwanym wynikiem badań jest fakt, iż próby „poprawiania” modelu poprzez douczanie sieci w oparciu o nowe wzorce, lub poprzez wielokrotne ponawianie procesu uczenia od innych wartości początkowych wag mogą tą efektywność znacznie pogorszyć. Biorąc pod uwagę powyższy fakt można sformułować wniosek, że pewne sprawdzone modele, zbudowane dzięki zaprezentowanej w niniejszej pracy metodologii, mogą niekiedy utracić swoje właściwości generowania efektywnych sygnałów transakcyjnych wskutek próby korekty takiego modelu (np. poprzez modyfikację wag).

Analizując zagadnienie doboru kryterium dla prawidłowej, wiarygodnej oceny końcowej sieci neuronowej należy ogólnie stwierdzić, że problem ten nie posiada oczywistego, jednoznacznego rozwiązania. Przyjęta jako ocena sieci wartość funkcji błędu obliczona dla ciągu walidacyjnego i ewentualnie dla ciągu uczącego tylko w pewnym stopniu wyjaśnia rzeczywistą efektywność SN, mierzoną zyskiem z generowanej przez nią strategii inwestycyjnej (por. podrozdział 5.3). Celowe jest zatem podjęcie badań nad poszukiwaniem innych czynników mających wpływ na tą efektywność, albo też badań nad nową, zmodyfikowaną metodyką uczenia SN zapewniającą silniejszy związek pomiędzy dokonywaną oceną sieci a jej praktyczną efektywnością.

Dość obiecująca - w zagadnieniu doboru kryterium oceny sieci - może okazać się zaprezentowana w niniejszej pracy metoda, polegająca na ocenie efektywności strategii inwestycyjnej generowanej przez badaną sieć w oparciu o zbiór danych nie wykorzystywanych bezpośrednio do uczenia sieci. Ponadto przydatne może być dodatkowe przetestowanie nauczonej już sieci w możliwie długim okresie OUT-S, najlepiej obejmującym możliwie aktualne dane, w celu weryfikacji skuteczności sieci w aktualnej sytuacji panującej na rynku. Być może konieczne jest jednak

poszukiwanie zupełnie nowego paradygmatu, określającego metody efektywnej konstrukcji i oceny sieci neuronowych dla zastosowań finansowych, innego niż powszechnie przyjęty w nauce zajmującej się SN.

Rozważając aspekt potencjalnego praktycznego zastosowania modeli opartych o SN należy wziąć pod uwagę fakt, iż nawet najlepiej przetestowany model nigdy nie będzie generował 100% poprawnych sygnałów transakcyjnych, toteż przy jego wykorzystywaniu należy liczyć się z występowaniem strat (por. analiza zagadnienia maksymalnej straty **dd**), a także okresów, w których strategia aktywna okaże się gorsza od pasywnej (zob. np. tabela 5-11). Wynikają stąd następujące wnioski:

- w celu testowania, oceny i dokonania badań porównawczych efektywności aktywnych strategii (np. przed ewentualnym jej zastosowaniem praktycznym) należy wykorzystywać stosunkowo **długi** okres czasu OUT-S,
- sprawdzona (przetestowana) strategia aktywna wymaga konsekwentnej realizacji wszystkich sygnałów decyzyjnych, nawet jeżeli chwilowo generują one straty.

Niewątpliwie jednak należy podkreślić, iż znalezienie efektywnych SN i opartych o te sieci aktywnych strategii (przy wykorzystaniu omówionej tutaj metodologii) może prowadzić (w dłuższym okresie) do istotnej poprawy efektywności procesu inwestowania na danym rynku finansowym.

O efektywności praktycznych implementacji narzędzi opartych na sieciach neuronowych niewątpliwie decyduje w dużym stopniu zastosowanie odpowiednio szerokiego spektrum informacji wejściowej, obejmującego wszystkie potencjalnie istotne dane mogące mieć jakikolwiek wpływ na dokładność prognozy. Fakt, iż w niniejszej pracy wykorzystano jedynie dane stanowiące elementy analizy technicznej nie wynika z przekonania autora o ponadprzeciętnej skuteczności tego typu analizy, ale z charakteru tych danych (możliwa jest tu bowiem obserwacja i interpretacja dziennych zmian poszczególnych wielkości, a ilość wszystkich tego typu danych okazała się podczas realizacji badań wystarczająco duża dla przeprowadzenia konstrukcji i testowania opartych na sieciach neuronowych modeli dla polskiego rynku finansowego). Informacja innego typu, np. dane o charakterze fundamentalnym, mogły okazać się do tej pory mniej przydatne ze względu na

specyfikę okresu transformacji polskiej gospodarki do warunków wolnorynkowych, a także z uwagi na stosunkowo krótki okres czasu, z którego pochodzą, w stosunku do dość długiego interwału pomiędzy kolejnymi momentami napływu tych danych (minimum miesiąc). Jednak stabilizacja sytuacji gospodarczej oraz duża ilość stale napływających nowych danych z poszczególnych spółek, sektorów, a także danych makroekonomicznych, pozwoli wkrótce na podjęcie badań efektywności zaproponowanej w niniejszej pracy metodologii z uwzględnieniem tych właśnie informacji, co także umożliwi przeprowadzenie eksperymentów jej wykorzystania przy założeniu dłuższego horyzontu prognozy. Badania takie, wykorzystujące obszerniejszy zakres informacji wejściowej dla systemu decyzyjnego, mogą wykazać jeszcze wyższą efektywność zaproponowanej metodologii, również przy założeniu istnienia kosztów transakcyjnych.

Dokonując podsumowania badań zawartych w niniejszej pracy można generalnie stwierdzić, że uzyskane rezultaty, chociaż jeszcze być może dość odległe od potencjalnych zastosowań praktycznych, są na tyle obiecujące, że stanowią przesłanki do dalszych badań oraz testów modeli skonstruowanych w oparciu o zaproponowaną tutaj metodologię. Jeżeli w odpowiednio długim okresie czasu wyniki testów określonych modeli (sieci neuronowych), opartych na właściwie dobranych zestawach danych uczących i parametrów, będą dobre, może to przesądzić o podjęciu prób praktycznego wykorzystania metody.

Zakończenie.

W pracy przedstawiono dwie metody sztucznej inteligencji: sieci neuronowe i algorytmy genetyczne, a następnie zaproponowano metodologię konstrukcji systemów decyzyjnych dla generowania aktywnych strategii inwestycyjnych, opartą na połączeniu tych metod. Następnie przeprowadzono badania efektywności tej metodologii w wybranych problemach dotyczących aktywnego inwestowania w oparciu o indeks giełdowy WIG 20, oraz aktywnego zarządzania portfelem wybranych papierów wartościowych. Zrealizowano także szereg badań dotyczących doboru istotnych parametrów zaproponowanej metody, a także celowości wykorzystania podczas konstrukcji systemu pewnych form transformacji danych (np. analizy głównych składowych). Na podstawie wyników przeprowadzonych badań sformułowano szereg wniosków, posiadających zarówno charakter badawczy (stanowiące punkty oparcia dla dalszych badań w tej dziedzinie), jak i aplikacyjny (posiadające wymiar potencjalnych zastosowań praktycznych).

Analizując zaprezentowane w pracy rezultaty badań należy uznać, iż postawiona na wstępie teza, stwierdzająca, że celowe jest wykorzystanie metod sztucznej inteligencji (do których należą omówione w niniejszej pracy sieci neuronowe i algorytmy genetyczne) w zagadnieniach modelowania rynków finansowych pod kątem predykcji ich dynamiki oraz generowania efektywnych strategii inwestycyjnych, została empirycznie uzasadniona. Wykazano bowiem istnienie efektywnych systemów decyzyjnych (poprzez ich efektywne zbudowanie), opartych na sieciach neuronowych i algorytmach genetycznych, które pozwalają na realizację skutecznych strategii decyzyjnych na rynku akcji. Niewątpliwie jednak stopień tej skuteczności zależy w dużej mierze od prawidłowego doboru zestawu opcji i

parametrów zdefiniowanych w ramach rozważanej metodologii, co z kolei może wymagać szeregu dalszych badań o charakterze eksperymentalnym.

Również wyrażony na wstępie pogląd, iż efektywność metodologii opartej na narzędziach sztucznej inteligencji będzie sukcesywnie wzrastać w miarę wzrostu dostępnej mocy obliczeniowej (rozwoju technologii komputerowej) wydaje się, zdaniem autora, w pełni uzasadniony. Istotnym ograniczeniem zaproponowanej metodologii jest bowiem jej wysoka złożoność obliczeniowa, wymagająca olbrzymiego nakładu obliczeń komputerowych w celu uzyskania zadowalających efektów. Ten czynnik ograniczający będzie jednak tracił na znaczeniu w miarę obserwowanego obecnie gwałtownego wzrostu mocy dostępnych komputerów. Ponadto możliwe będzie wykorzystanie szerszych zakresów wartości określonych parametrów systemu (ograniczonych dotychczas właśnie przez czas obliczeń), lub zastosowanie bardziej złożonych obliczeniowo modyfikacji omawianej metodologii.

Dokonując syntetycznego ujęcia niniejszej pracy, należy również zwrócić uwagę na szereg zawartych w niej elementów nowatorskich. Do najważniejszych z nich należą:

- ◆ oryginalna propozycja zmodyfikowanego algorytmu genetycznego dla doboru zmiennych wejściowych i wybranych parametrów sieci neuronowej,
- ◆ szereg nowych, zaproponowanych po raz pierwszy w tej pracy procedur wstępnego przetwarzania informacji pochodzącej z danego rynku finansowego,
- ◆ niektóre metody wstępnego przetwarzania danych (preprocessingu),
- ◆ szereg rozwiązań dotyczących uczenia sieci, jej walidacji i testowania (oceny), a także propozycje doboru wybranych parametrów dla tych operacji,
- ◆ niektóre propozycje dotyczące doboru typu zmiennej wyjściowej dla sieci,
- ◆ oryginalna metoda końcowej oceny sieci (wykorzystywanej na etapie realizacji algorytmu genetycznego), polegająca na zbadaniu efektywności określonej strategii inwestycyjnej opartej o sygnały nauczanej sieci,
- ◆ propozycja aktywnej metody zarządzania portfelem inwestycyjnym i badania jej efektywności.

Prezentowana praca stanowi także swoiste uzasadnienie celowości kontynuacji badań w dziedzinie zastosowania najnowszych metod informatyki w modelowaniu rynków finansowych i podejmowaniu prób praktycznych aplikacji systemów decyzyjnych opartych na tych modelach. W szczególności rozważając zaprezentowaną w niniejszej pracy tematykę, można wyznaczyć następujące kierunki dalszych prac:

- ◆ badania porównawcze modeli wykorzystujących metody sztucznej inteligencji (w szczególności SN i AG) z modelami opartymi na metodologiach klasycznych,
- ◆ badania związane z optymalizacją doboru parametrów systemów decyzyjnych,
- ◆ włączenie dalszych strumieni informacji wejściowej dla systemu, a więc nie tylko bezpośrednich, napływających z rynku, wartości finansowych szeregów czasowych, ale także danych o charakterze mikro- i makroekonomicznym, fundamentalnym, jakościowym, i innych informacji mogących mieć zasadniczy wpływ na poprawę efektywności generowanych przy użyciu omówionej metodologii modeli,
- ◆ badania nad przydatnością innych typów sieci neuronowych (np. sieci ze sprzężeniem zwrotnym), innych modyfikacji algorytmów genetycznych, sposobów transformacji danych wejściowych, itp.
- ◆ udoskonalenie zaproponowanej w rozdziale 6 aktywnej metody zarządzania portfelem inwestycyjnym, uwzględniające dodatkowo m.in. ryzyko inwestycyjne i jego minimalizację przy zastosowaniu metod statystycznych, stanowiących elementy klasycznej teorii portfela.

Literatura.

- Anderson J.A., Rosenfeld E., (eds.) 1988. *Neurocomputing: Foundations of Research*. Cambridge, MIT Press.
- Azoff E.M., 1994. *Neural network time series forecasting of financial markets*. New York, Wiley.
- Bäck T., Fogel D.B., Michalewicz Z., 1996. *Handbook of Evolutionary Computation*. University Oxford Press, New York.
- Bassi D.F., 1996. „Stock price predictions by recurrent multilayer neural network architectures”. W książce (Refenes et al., 1996).
- Bauer R. J. Jr, 1994. „An Introduction to Genetic Algorithms: A Mutual Fund Screening Example”, *Neurovest Journal*, vol.2, no.4, July/August 1994.
- Bauer R., 1994 - „*Genetic Algorithms and Investment Strategies*”, John Wiley & Sons, Inc, New York.
- Bierman H. Jr, Bonini C.P., Hausman W.H., 1986 - *Quantitative analysis for business decisions*. Homewood, Irwin.
- Black F., Scholes M., 1973. „The pricing of options and corporate liabilities”. *Journal of Political Economy*. May-June 1973.
- Bodie Z., Kane A., Marcus A.J., 1993. *Investments* 2nd ed. Boston, Irwin.
- Bortoli M., 1996. „Nonstationary state space models for multivariate financial time series: An introduction”. *Neurovest Journal*. July/August 1996.
- Cherkassky V., Lari-Najafi H., 1991. „Constrained topological mapping for nonparametric regression analysis”. *Neural Networks*, vol. 4.
- Colby R. W., Meyers T. A., 1988 - „*The Encyclopedia of Technical Market Indicators*”, Dow Jones - Irwin; Homewood, Illinois.
- Cytowski J., 1996 - „*Algorytmy genetyczne, podstawy i zastosowania*”, Akademicka Oficyna Wydawnicza PLJ, Warszawa.

- Czekaj J., 1994. *Kursy akcji notowanych na Gieldzie Papierów Wartościowych w Warszawie w świetle współczesnej teorii rynku kapitałowego*. Kraków, Akademia Ekonomiczna w Krakowie.
- Davis L. (editor), 1991 - „*Handbook of Genetic Algorithms*”, Van Nostrand Reinhold.
- De Jong K.A. (editor), 1993 - *Evolutionary Computation*. MIT Press.
- Elton E.J., Gruber M.J., 1991. *Modern portfolio theory and investment analysis*. New York, Wiley.
- Engle R.F., 1993. „Statistical models for financial volatility”. *Financial Analysts Journal*. Jan.-Feb. 1993.
- Fama E.F., 1970. „Efficient capital markets: A review of theory and empirical work”. *Journal of Finance*. No. 25, May 1970.
- Fama E.F., 1991. „Efficient capital markets: II”. *Journal of Finance*. No. 46, December 1991.
- Farrell J.L. Jr., 1993. „Systematic portfolio management: evolution, current practice and future direction”. *Financial Analysts Journal*. September-October 1993.
- Goldberg D. E., 1994 - „Genetic and Evolutionary Algorithms Come of Age”, *Communications of the ACM*, vol.37, No.3, March 1994.
- Goldberg D. E., 1995 - „*Algorytmy genetyczne i ich zastosowania*”, WNT Warszawa.
- Grabowski M., 1997. *Sieci neuronowe w analizie danych społeczno-ekonomicznych*. Praca doktorska, Akademia Ekonomiczna w Krakowie.
- Greene D.P., Smith S.F., 1987 - „A genetic system for learning models of consumer choice”, *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*, pp. 217-223.
- Guo Z., Uhrig R.E., 1992. „Use of genetic algorithms to select inputs for neural networks”. *Proc. of the International Workshop on Combinations of Genetic Algorithms and Neural Networks*, COGANN-92.
- Haugen R.A., 1996. *Teoria nowoczesnego inwestowania*. Warszawa, WIG Press.

- Haykin S., 1994. *Neural networks. A comprehensive foundation*. New York, Macmillan College Publishing Company.
- Hebb D.O., 1949. *The organization of behavior: A neuropsychological theory*. New York, Wiley.
- Hecht-Nielsen R., 1990. *Neurocomputing*. Reading, MA. Addison-Wesley.
- Hertz J., Krogh A., Palmer R.G., 1993. *Wstęp do teorii obliczeń neuronowych*. Warszawa, WNT.
- Holland J. H., 1975 - „Adaptation in Natural and Artificial Systems”, *University of Michigan Press*, Ann Arbor, MI.
- Hopfield J.J., 1982. „Neural networks and physical systems with emergent collective computational abilities”. *Proceedings of the National Academy of Sciences of the USA*. No 79.
- Jajuga K., 1990. *Statystyczna teoria rozpoznawania obrazów*. Warszawa, PWN.
- Jajuga K., 1995. „O sieciach neuronowych w klasyfikacji i analizie danych”. *Sekcja Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego - Zeszyt 2 „Klasyfikacja i analiza danych - problemy teoretyczne”*. Jelenia Góra - Wrocław - Kraków.
- Jajuga K., Jajuga T., 1993. *Jak inwestować w papiery wartościowe*. Warszawa, PWN.
- Jajuga K., Jajuga T., 1996. *Inwestycje: instrumenty finansowe, ryzyko finansowe, inżynieria finansowa*. Warszawa, PWN.
- Jašić T., 1998. „Testing for nonlinearity and deterministic chaos in monthly Japanese stock market data”. *Zagreb International Review of Economics & Business*, vol. I, no. 1, May, 1998.
- Katz J., McCormick D., 1994 - „Neurogenetics and its use in Trading System Development”, *Neurovest Journal*, vol.2, no.4, July/August 1994.
- Kester G.W., 1990. „Market timing with small versus large-firm stocks: Potential gains and required predictive ability”. *Financial Analysts Journal*. Sept.-Oct. 1990.

- Kohonen T., 1995. *Self-organizing maps*. Berlin, Springer-Verlag.
- Koza J., 1992 - *Genetic Programming: On the programming of computers by means of natural selection*. MIT Press.
- Kryzanowski L., Galler M., Wright D., 1993 - „Using Artificial Neural Networks to Pick Stocks”, *Financial Analysts Journal*, August 1993.
- Lee C., Rahman S., 1990. „Market timing, selectivity, and mutual fund performance: An empirical investigation”. *Journal of Business*. Vol. 63, No 2.
- Lee J.K., Trippi R.R., Chu S.C., Kim H.S., 1990. „K-FOLIO: Integrating the Markowitz model with a knowledge-based system”. *The Journal of Portfolio Management*. Fall, 1990.
- Leinweber D. J., Arnott R. D., 1995 - „Quantitative and Computational Innovation in Investment Management”, *The Journal of Portfolio Management*, Winter 1995.
- Levin R.I., Rubin D.S., Stainson J.P., 1986. *Quantitative approaches to management*. New York, McGraw-Hill.
- Levy H., Sarnat M., 1972. *Investment and portfolio analysis*. New York, Wiley.
- Majewski J., 1981. *Komputerowe modelowanie neuronu*. Kraków, AGH, Praca doktorska.
- Markowitz H.M., 1952. „Portfolio selection”. *Journal of Finance*. March 1952.
- Martinez Abascal E., Pregel G., 1994. *Weak form market efficiency: A comparison between the Spanish and the U.S. stock markets*. Research paper No 273, IESE, University of Navarra.
- McCulloch W.S., Pitts W., 1943. „A logical calculus of the ideas immanent in nervous activity”. *Bulletin of Mathematical Biophysics*, No 5, pp. 115 - 133.
- Michalewicz Z., 1996 - „Algorytmy genetyczne + struktury danych = programy ewolucyjne”, WNT Warszawa.
- Minsky M., Papert S., 1969. *Perceptrons*. Cambridge, MIT Press.

- Morajda J., 1991. „System modelowania sieci neuronowych dla mikrokomputera Commodore”. *Elektrotechnika*. Tom 10, zesz. 1, 1991.
- Morajda J., 1996. „Modelowanie sieci neuronowych”. *Zeszyty naukowe Akademii Ekonomicznej w Krakowie*. Nr 474.
- Morajda J., 1997a. „Wybrane możliwości zastosowań sieci neuronowych w ekonomii i zarządzaniu”. *Zeszyty naukowe Akademii Ekonomicznej w Krakowie*. Nr 493.
- Morajda J., 1997b. „Algorytmy genetyczne oraz możliwości ich zastosowań w systemach decyzyjnych” - *Materiały z XXXIII Konferencji Statystyków, Ekonometryków, Matematyków Polski Południowej*, Prace Naukowe nr 771 AE we Wrocławiu.
- Neuman von J., 1958. *The computer and the brain*. New Haven, Yale Univ. Press.
- Osowski S., 1996. *Sieci neuronowe w ujęciu algorytmicznym*. Warszawa, WNT.
- Packard N., 1990 - „A genetic learning algorithm for the analysis of complex data”, *Complex Sys.* 4, pp. 573-586.
- Pawłowski Z., 1982. *Zasady predykcji ekonometrycznej*. Warszawa, PWN.
- Peters E. E., 1997. *Teoria chaosu a rynki kapitałowe*. Warszawa, WIG-Press.
- Plummer T., 1995. *Psychologia rynków finansowych*. Warszawa, WIG-Press.
- Pring M. J., 1998. *Podstawy analizy technicznej*. Warszawa, WIG-Press.
- Ramón y Cajál S., 1911. *Histologie du système nerveux de l'homme et des vertébrés*. Paris, Maloine.
- Refenes A.P. (ed.), 1995. *Neural networks in the capital markets*. Chichester, Wiley.
- Refenes A.P., Abu-Mostafa Y., Moody J., Weigend A. (editors), 1996. „Neural Networks in Financial Engineering”. *Proc. of the 3rd International Conference on Neural Networks in the Capital Markets*, London, 11 - 13 October 1995.
- Reichenstein W., Rich S.P., 1994. „Predicting long-horizon stock returns: Evidence and implications”. *Financial Analysts Journal*. January-February 1994.

- Rosenblatt F., 1958. „The perceptron: A probabilistic model for information storage and organization in the brain”. *Psychological Review*, No 65.
- Rögnvaldsson T.S., Pi H., 1996. „A neural network approach to futures trading”. W książce (Refenes et al., 1996).
- Rumelhart D.E., Hinton G.E., Williams R.J., 1986. „ Learning internal representations by error propagation”. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (D.E. Rumelhart & J.L. McClelland, eds.), Vol. 1, Chapter 8, Cambridge MA, MIT Press.
- Rumelhart D.E., McClelland J.L. (editors), 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1, Cambridge MA, MIT Press.
- Rutkowska D., Piliński M., Rutkowski L., 1997. *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*. Warszawa, PWN.
- Rutkowski L. i inni (praca zbiorowa), 1996. *Sieci neuronowe i neurokomputery*. Częstochowa, Wyd. Politechniki Częstochowskiej.
- Schaffer J.D., Whitley L., Eshelman J., 1992. „Combinations of Genetic Algorithms and Neural Networks: A Survey of the State of the Art.”. *Proc. of the International Workshop on Combinations of Genetic Algorithms and Neural Networks*, COGANN-92.
- Selecting the right neural network tool, third edition, 1996. *Neurovest Journal*, September/October 1996.
- Sharpe W.F., 1963. „A simplified model for portfolio analysis”. *Management Science*. No. 19, pp. 277 - 293.
- Sharpe W.F., 1994. „The Sharpe ratio”. *The Journal of Portfolio Management*. Fall 1994.
- Siriopoulos C., Markellos R.N., Sirlantzis K., 1996. „Application of artificial neural networks in emerging financial markets”. W książce (Refenes et al., 1996).
- Tadeusiewicz R., 1985 - „Rozpoznawanie obrazów w zastosowaniach ekonomicznych”, skrypt Akademii Ekonomicznej w Krakowie.

- Tadeusiewicz R., 1988. *Biocybernetyka*. Wrocław, Ossolineum.
- Tadeusiewicz R., 1991. „Sieci neuronowe - przewodnik problemowy”. *Elektrotechnika*. Tom 10, zesz. 2, 1991.
- Tadeusiewicz R., 1993. *Sieci neuronowe*. Warszawa, Akademicka Oficyna Wydawnicza RM.
- Tadeusiewicz R., 1994a. *Problemy biocybernetyki*. Warszawa, PWN.
- Tadeusiewicz R., 1994b. „Neural networks in pattern recognition”. *Proceedings of „Intelligent Information Systems”*. IPI PAN, Warszawa 1994, pp. 410 - 426.
- Tadeusiewicz R., 1995. „Sieci neuronowe w prognozowaniu procesów gospodarczych”. *Mat. konferencyjne „Sztuczna inteligencja i infrastruktura informatyczna*. Siedlce 1995, ss. 325 - 342.
- Tadeusiewicz R., Mikrut Z., 1994. „Sieci neuronowe rozpoznające obrazy”. *Materiały konferencyjne I Krajowej Konferencji „Sieci neuronowe i ich zastosowania”*. Kule (woj. Częstochowskie), 12 - 15.04.1994, t. I, ss. 52 - 70.
- Taylor W.K., 1960. *Computers and the nervous system. Models and analogues in biology*. Cambridge, Cambridge Univ. Press.
- Taylor W.R.L., Yoder J.A., 1994. „Mutual fund trading activity and investor utility”. *Financial Analysts Journal*. May-June 1994.
- Thomason M.R., 1996. „Neural network input variable selection (revisted)”, *Neurovest Journal*, May/June 1996.
- Thompson B., Thompson B., 1986 - „Evolving knowledge from data”, *Computer Language*, 3 (11) pp. 23-26.
- Trippi R. R., DeSieno D., 1992 - „Trading Equity Index Futures with a Neural Network”, *The Journal of Portfolio Management*, Fall 1992.
- Werbos P.J., 1974. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. Ph.D. thesis, Harvard University, Cambridge, MA.
- Whitley G. (editor), 1992 - „*Foundations of Genetic Algorithms*”, Morgan Kaufmann, San Mateo, CA.

Wong F., 1994 - „Neurogenetic Computing Technology”, *Neurovest Journal*, vol.2, no.4, July/August 1994.

Žurada J., 1992. *Introduction to artificial neural systems*. West Publishing Co.