



WYŻSZA SZKOŁA EKONOMII I INNOWACJI W LUBLINIE

WYDZIAŁ TRANSPORTU I INFORMATYKI
SPECJALNOŚĆ: SYSTEMY I SIECI KOMPUTEROWE

ADAM ZUSIN

NR ALBUMU: 14822

Praca inżynierska napisana pod kierunkiem

prof. dr hab. Grzegorza Marcina Wójcika

Zastosowanie jednowarstwowych
sieci neuronowych
do rozpoznawania banknotów
Narodowego Banku Polskiego

Lublin 2013

Spis treści

Wstęp	3
1. Mózg i neuron	5
1.1. Budowa mózgu	5
1.1.1 Mózgowie	6
1.1.2 Funkcje mózgu	7
1.2 Budowa i działanie neuronu	7
1.2.1 Budowa komórki nerwowej	8
1.2.2 Działanie neuronu	8
2. Model sieci neuronowej	11
2.1 Modelowanie neuronów	11
2.1.1 Model sztucznego neuronu McCullocha-Pittsa	12
2.1.2 Model neuronu Hebba	13
2.1.3 Model neuronu chaotycznego	14
2.2 Sztuczne sieci neuronowe	14
2.2.1 Architektura sztucznych sieci neuronowych	15
2.2.2 Struktura i schemat sieci wielowarstwowej	17
2.2.3 Perceptron	17
3. Metody uczenia sztucznych sieci neuronowych	19
3.1 Uczenie nadzorowane	19
3.2 Uczenie nienadzorowane	19
3.3 Algorytm wstecznej propagacji błędów	20
4. Projekt prostej sieci do rozpoznawania banknotów	25
4.1 Konstrukcja sztucznej sieci neuronowej	25
4.1.1 Tworzenie warstw	25
4.1.2 Łączenie warstw	28
4.1.3 Inicjalizacja wag	29
4.2 Proces uczenia	30
4.2.1 Wczytywanie wzoru	30
4.2.2 Funkcja uczenia Backprop-Momentum	32
4.3 Wyniki uczenia	33
4.3.1 Wzór dla 20 złotych	34
4.3.2 Wzór dla 50 złotych	34
4.3.3 Wzór dla 100 złotych	35
4.3.4 Wzór dla 200 złotych	35
Zakończenie	36
Literatura	37

Wstęp

Sieci neuronowe to w dzisiejszych czasach popularny temat. Mają one szerokie zastosowania w dziedzinach takich jak finanse, medycyna, technika, geologia albo fizyka. Sieci neuronowe mogą być zastosowane wszędzie tam, gdzie pojawiają się zadania związane z predykcją, klasyfikacją czy sterowaniem. Olbrzymi sukces metody sieci neuronowych związany jest z kilkoma czynnikami. Pierwszym z nich jest moc. Sieci neuronowe są bardzo wyrafinowaną techniką modelowania, zdolną do odwzorowywania nadzwyczaj złożonych funkcji.

W szczególności sieci neuronowe mają charakter nieliniowy (termin ten zostanie szczegółowo omówiony w dalszej części), co istotnie wzbogaca możliwości ich zastosowań.

Rozwiniemy nieco ten temat, żeby wskazać na wagę formułowanych tu wniosków. Przez wiele lat powszechnie stosowaną techniką matematycznego opisywania różnych obiektów i procesów było modelowanie liniowe. Takie postępowanie jest z powodzeniem stosowane także obecnie, przynosząc korzystne rezultaty, głównie z uwagi na dobrze znaną strategię optymalizacji stosowaną przy konstrukcji modeli tego typu. Jednak wszędzie tam, gdzie nie ma podstaw do aproksymacji liniowej występujących zjawisk i procesów (a przypadki takie są w praktyce dość częste), modele liniowe nie sprawdzały się, prowadząc niekiedy do formułowania niesłusznych opinii o całkowitym braku możliwości matematycznego opisywania takich czy innych systemów. W takich przypadkach, przy rozwiązywaniu tych trudnych i kłopotliwych zagadnień odwołanie się do modeli tworzonych z wykorzystaniem sieci neuronowych (a więc modeli, które bez trudu mogą odwzorować zależności nieliniowe) może być najszybszym i najwygodniejszym rozwiązaniem problemu.

Zalety sieci neuronowych nie ograniczają się jedynie do tego, że umożliwiają one swobodne i łatwe (bez konieczności samodzielnego formułowania przez użytkownika skomplikowanych hipotez) tworzenie modeli nieliniowych. Sieci umożliwiają także kontrolę nad złożonym problemem wielowymiarowości, który przy stosowaniu innych metod znacząco utrudnia próby modelowania funkcji nieliniowych z dużą liczbą zmiennych niezależnych (tzw. funkcji wektorowych).

Drugim czynnikiem jest łatwość użycia. Sieci neuronowe w praktyce same konstruują potrzebne użytkownikowi modele, ponieważ automatycznie uczą się na podanych przez niego przykładach. Odbywa się to w taki sposób, że użytkownik sieci gromadzi reprezentatywne dane pokazujące, jak manifestuje się interesująca go zależność, a następnie uruchamia algorytm uczenia, który ma na celu automatyczne wytworzenie w pamięci sieci potrzebnej struktury danych. Opierając się na tej samodzielnie stworzonej strukturze danych sieć realizuje potem wszystkie funkcje związane z eksploatacją utworzonego modelu. Chociaż, zatem użytkownik potrzebuje pewnej, w głównej mierze empirycznej wiedzy dotyczącej sposobu wyboru i przygotowania danych stanowiących przykłady, a także wyboru właściwego rodzaju sieci neuronowej oraz sposobu interpretacji rezultatów, to jednak poziom wymaganej od użytkownika wiedzy teoretycznej, niezbędnej do skutecznego zbudowania modelu, przy stosowaniu sieci neuronowych jest znacznie niższy niż w przypadku stosowania tradycyjnych metod statystycznych.

Ciekawa jest również ta własność sieci neuronowych, że naśladują one działanie ludzkiego umysłu. Wprawdzie sieci oparte są na bardzo prostym modelu, przedstawiającym wyłącznie najbardziej podstawową istotę działania biologicznego systemu nerwowego, jednak ich działanie wzbudza ciekawość jako jedna z prób przeniknięcia istoty działania ludzkiego mózgu.

Niektórzy w związku z tym sądzą, że w przyszłości rozwój modelowania neurobiologicznego może doprowadzić do powstania prawdziwych komputerów inteligentnych.¹

W rozdziale pierwszym zostaną przedstawione biologiczne podstawy sztucznych sieci neuronowych. W rozdziale kolejnym, będą zaprezentowane przykładowe modele matematyczne sztucznego neuronu i sieci neuronowych. W rozdziale trzecim autor skupi się na wykazaniu metod uczenia sztucznych sieci neuronowych ze szczególnym uwzględnieniem metody wstecznej propagacji błędów. W ostatnim, czwartym rozdziale zostanie zademonstrowany proces tworzenia i uczenia sieci do rozpoznawania banknotów, a przedstawione zostaną wyniki tej symulacji komputerowej wykonanej w darmowym programie JNNS.

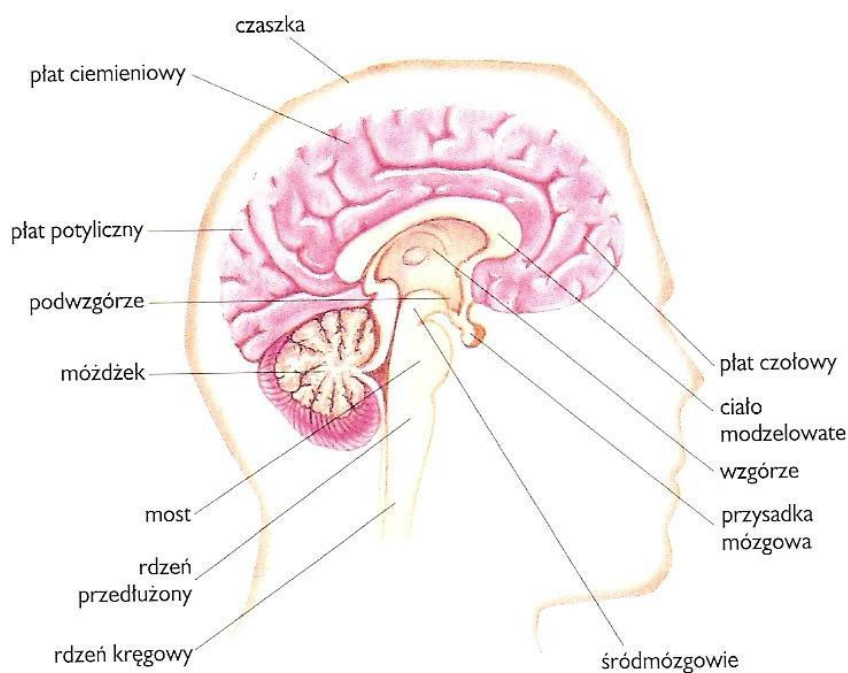
¹ StatSoft, Inc., Internetowy podręcznik statystyki, 1984-2011
http://www.statsoft.pl/textbook/stathome_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstnet.html

1. Mózg i neuron

Najważniejszym organem człowieka jest mózg. Kontroluje on pracę wszystkich narządów człowieka a także umysł. Decyduje o osobowości i świadomości. Uszkodzenie mózgu prowadzi do zaburzeń umysłowych, utraty pamięci, zdolności, świadomości. Śmierć mózgu oznacza śmierć całego organizmu. Mózg człowieka osiągnął najwyższy poziom rozwoju w porównaniu z mózgami pozostałych istot żywych. Od niego zależy inteligencja i kreatywność jednostki ludzkiej, a także jest źródłem takich emocji jak miłość, zazdrość, wybaczenie, złość.² U wszystkich ssaków przez środek mózgu przechodzi głęboka podłużna szczelina, która dzieli go na dwie niemal symetryczne części – zwane półkulami mózgowymi.³ O wydolności mózgowia (największa część mózgu) decyduje przede wszystkim liczba zachodzących między nimi połączeń.⁴

1.1 Budowa mózgu

Mózg znajduje się w czaszce i składa się z wielu elementów, przedstawionych na rysunku 1. Mózg tworzą dwie półkule mózgowe, których części składowe to kora, istota biała oraz jądra podstawy, czyli skupiska istoty szarej, dającej się zaobserwować wśród włókien nerwowych (istoty białej), łączących poszczególne obszary kory z innymi ośrodkami.⁵



Rysunek 1. Elementy mózgu.⁶

² D. Benzaia, S. Carleton, C. Caruthers : Rodzinna encyklopedia zdrowia, Reader's Digest Przegląd, Sp. z o.o, Warszawa 1999 s. 13.

³ A. Grabowska, W. Buhodowska: Dwie półkule – jeden mózg, Wiedza Powszechna, Warszawa 1994, s. 10.

⁴ L. Bożydar, J. Kaczmarek: Mózg, język, zachowanie, Wydawnictwo UMCS 1998 Lublin, s. 29.

⁵ Tamże, s.35.

⁶ D. Benzaia, S. Carleton, C. Caruthers : Rodzinna encyklopedia zdrowia, Reader's Digest Przegląd, Sp. z o.o, Warszawa 1999 s. 13

Pień mózgu obejmuje międzymózgowie, śródmózgowie, oraz most i rdzeń przedłużony.⁷ Rdzeń przedłużony jest połączony z rdzeniem kręgowym. Tworzą go przede wszystkim nerwy idące z rdzenia kręgowego do pozostałych części mózgu. Most tworzy wybrzuszenie nad przednim odcinkiem pnia mózgu. Łączy różne części mózgu. Śródmózgowie znajduje się ponad mostem. Stanowi ośrodek odruchów wzrokowych i słuchowych.

Kora mózgowa to ułożona w fałdy zwiększające jej powierzchnię; czynnościowo dzieli się na korę motoryczną, sensoryczną i asocjacyjną.

Istota biała – składa się z mielinowych aksonów neuronów łączących różne rejony mózgu; aksony te są ułożone w wiązki (drogi); łączy neurony w obrębie tej samej kuli, prawą i lewą półkulę, mózdzek z innymi częściami mózgu i rdzenia kręgowego.⁸

1.1.1 Mózgowie

Mózgowie człowieka waży około 1500 gramów, z tym, że jego waga u normalnego osobnika waha się w granicach od 820 do 2800 g. Trzeba jednak zauważyć, że nie ma bezpośredniego związku między inteligencją danej osoby a wielkością tego organu. Dla przykładu mózgowie Turgieniewa ważyło 2021 gramów, podczas gdy waga mózgowia innego wybitnego pisarza Anatola France’a wynosiła zaledwie 1100 gramów.⁹ Mózgowie i rdzeń są zawieszone w czaszce i kanale kręgowym w szczególny sposób. Otoczone jest trzema błonami- oponą twardą, pajęczynówką i oponą miękką. Pierwsza z nich jest zrośnięta z wewnętrzną okostną czaszki a trzecia otula szczelnie mózgowie i jest przyrośnięta do niego. Pajęczynówka przylega na zewnątrz w sposób umożliwiający przesuwanie się, do twardówki, a z drugiej strony jest za pomocą włókienek łącznotkankowych przytwierdzona do opony miękkiej. Przestrzeń podpajęczynówkowa jest wypełniona płynem mózgowo-rdzeniowym.¹⁰

Mózgowie człowieka to przede wszystkim dwie połączone ze sobą półkule mózgu. Od góry mają kształt jajowaty, a od dołu są spłaszczone i tworzą powierzchnię podstawy mózgu, która spoczywa na podstawie czaszki. W linii środkowej półkule oddzielone są od siebie głęboką szczeliną podłużną mózgu. Na dnie tej szczeliny znajduje się wielkie spoidło mózgu utworzone z bardzo licznych włókien nerwowych, łączących między sobą wszystkie ośrodki obydwu półkul. Szczelina podłużna ku tyłowi dochodzi do szczeliny poprzecznej, która oddziela mózg od mózdzku. Powierzchnię półkul pokrywa kora mózgu. Jest to warstwa tkanki nerwowej, szarej barwy, wykazująca silne pofałdowanie.

Poszczególne fałdy kory, nazywane zakrętami mózgu, są rozgraniczone przez bruzdy i szczeliny. Grupy zakrętów tworzą płaty: czołowy, skroniowy, ciemieniowy i potyliczny. Na podstawie mózgu znajduje się skrzyżowanie nerwów wzrokowych, opuszki węchowe i przysadka mózgowa. Z podstawy mózgu wychodzi 12 par nerwów czaszkowych.¹¹ Mózgowie i rdzeń kręgowy są strukturami zbudowanymi z naczyń krwionośnych, neuronów i komórek gleju.¹²

⁷ M. Krause: Człowiek i jego układ nerwowy, Śląsk Sp. z o.o Wydawnictwo Naukowe, Katowice 2002, s. 15.

⁸ Eldra P. Solomon, Linda R. Berg, Diana W. Martin :Biologia wg VII wydania amerykańskiego, MULTICO Oficyna Wydawnicza, Warszawa 2007 s.770

⁹ L. Bożydar, J. Kaczmarek: Mózg, język, zachowanie, Wydawnictwo UMCS 1998 Lublin, s. 29.

¹⁰ M. Krause: Człowiek i jego układ nerwowy, Śląsk Sp. z o.o Wydawnictwo Naukowe, Katowice 2002, s. 25.

¹¹ M. Barczyński :Domowy poradnik medyczny, Państwowy Zakład Wydawnictw Lekarskich, Warszawa 1991 s. 15.

¹² M. Krause: Człowiek i jego układ nerwowy, Śląsk Sp. z o.o Wydawnictwo Naukowe, Katowice 2002, s. 22.

1.1.2 Funkcje mózgu

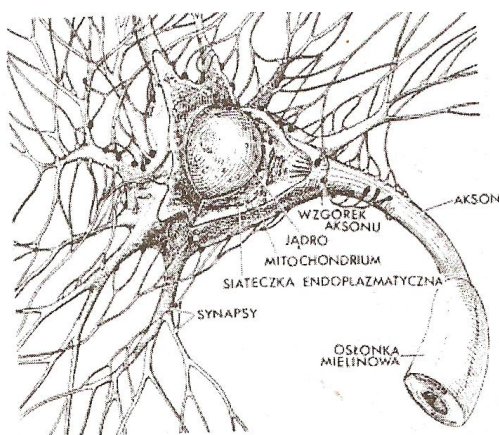
Główną funkcją mózgu jest przetwarzanie i wymiana informacji dostarczanych przez inne narządy. Praca mięśni również jest kontrolowana przez mózg, dzięki czemu człowiek może się np. poruszać. Ponadto to dzięki mózgowi człowiek jest w stanie zachować równowagę. Bez mózgu niemożliwe byłoby uczenie się i zapamiętywanie, jak również mowa. Mózg kontroluje pracę narządów, dzięki czemu możliwe są czynności zautomatyzowane, jak np. oddychanie czy sen. Wszelkie odruchy obronne także zawdzięczamy mózgowi. Dzięki mózgowi człowiek posiada umysł, a więc potrafi myśleć, jest także zdolny do uczuć, np. przyjaźń, miłość, nienawiść. Wykorzystując mózg człowiek potrafi planować, przestrzegać norm etycznych, tworzyć sztukę.¹³

Lewa półkula zazwyczaj kontroluje mowę, odpowiada za myślenie logiczne, potrzebne do zrozumienia np. nauk ścisłych. Włókna nerwowe idące od lewej półkuli przechodzą na stronę prawą rdzenia kręgowego, skutkiem czego mięśnie jednej strony ciała pozostają pod kontrolą ośrodków nerwowych strony przeciwnej (np. uszkodzenie lewej półkuli mózgu powoduje porażenie mięśni po stronie prawej).

Prawa półkula kontroluje myślenie artystyczne i twórcze. Włókna nerwowe idące od prawej półkuli przechodzą na stronę lewą rdzenia kręgowego, skutkiem czego mięśnie jednej strony ciała pozostają pod kontrolą ośrodków nerwowych strony przeciwnej (np. uszkodzenie prawej półkuli mózgu powoduje porażenie mięśni po stronie lewej).

1.2 Budowa i działanie neuronu

Neuron (komórka nerwowa) jest podstawowym elementem systemu nerwowego. W mózgu człowieka jest ok. 100 miliardów neuronów, które oddziałują na siebie przez ogromną liczbę połączeń. Każdy neuron posiada dokładnie jeden akson, za pomocą którego wysyła impulsy do wielu innych neuronów. Informacje przekazywane są do innych neuronów dzięki złączom nerwowym zwanych synapsami. Fragment neuronu biologicznego znajduje się na rys. 2.



Rysunek 2. Fragment neuronu biologicznego.¹⁴

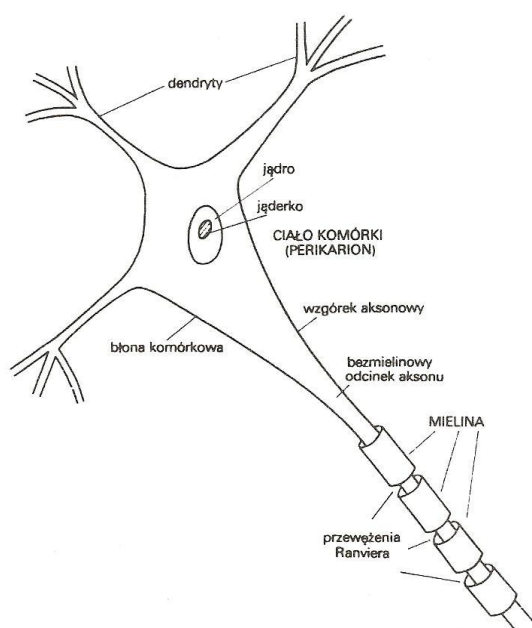
¹³ A. Grabowska, W. Buhodowska: Dwie półkule – jeden mózg, Wiedza powszechna, Warszawa 1994, s. 15.

¹⁴ Fragment neuronu z dendrytami i początkowym odcinkiem aksonu. (Źródło: A. Grabowska, W. Buhodowska: Dwie półkule – jeden mózg, Wiedza powszechna, Warszawa 1994, s. 15)

1.2.1 Budowa neuronu

Komórka nerwowa jest jednostką morfologiczną i czynnościową. Składa się z somy (ciała komórkowego), protoplazmatycznych wypustek przewodzących impulsy do ciała komórki i wypustki osiowej zwanej neurylem, przewodzącej impulsy z ciała komórki na obwód. Dendryty są przeważnie krótkie i rozgałęzione, neurylety są zaś zawsze pojedyncze i przeważnie długie.¹⁵

Akson jest wypustką neuronu, długości od kilku mikrometrów do ponad 1m. Pełni funkcję „nadajnika” ,przez który wysyła sygnały do innych neuronów. Każdy neuron ma tylko jeden akson. Dendryt (gr. *dendros* – drzewo) jest również wypustką neuronu, jednak w odróżnieniu od aksonu, służy neuronowi do odbierania sygnałów od innych neuronów. Każdy neuron ma wiele dendrytów. Istotnym elementem neuronu są synapsy – jest to miejsce komunikacji błony kończącej akson z błoną komórkową drugiej komórki nerwowej. Synapsy umożliwiają więc komunikację między komórkami nerwowymi.



Rysunek 3. Budowa neuronu. Przewężenia Ranviera umożliwiają skokowe przewodzenie bodźców.¹⁶

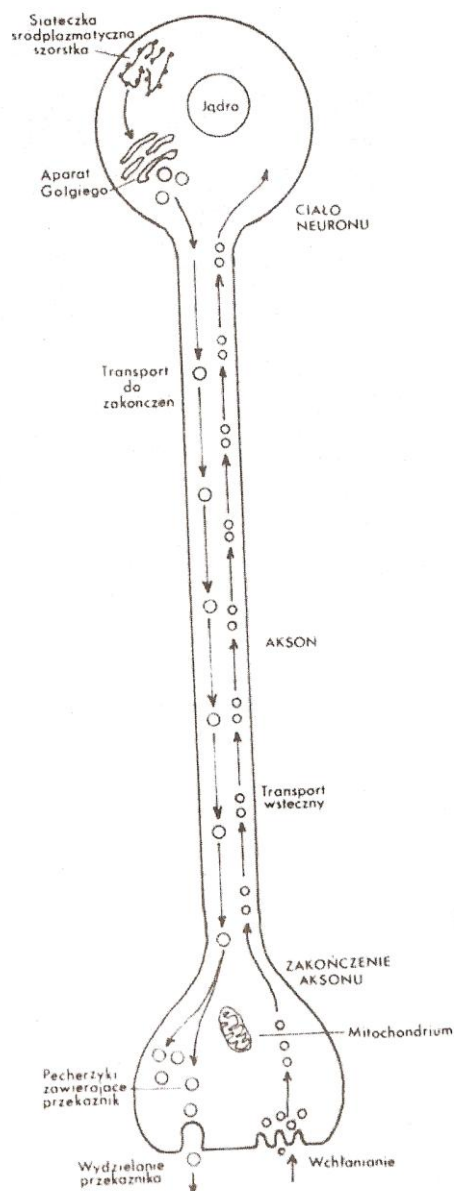
1.2.2 Działanie neuronu

Część sygnałów docierających do neuronu wpływa pobudzająco, a część hamująco na neuron, który sumuje impulsy pobudzające i hamujące. Jeżeli ich suma algebraiczna przekracza pewną wartość progową, to sygnał na wyjściu neuronu jest przesyłany – poprzez akson – do innych neuronów. W obrębie synapsy impuls elektryczny wywołuje związek chemiczny zwany przekaźnikiem (mediator) i uwalnia go do przestrzeni międzyneuronalnej, nazywanej szczeliną synaptyczną. Przekaźnik ten staje się bodźcem dla następnego neuronu. W ten sposób są przekazywane informacje nerwowe i nerwowo-mięśniowe. Synapsy mogą mieć charakter pobudzający lub hamujący, zależnie od rodzaju przekaźnika uwalnianego do szczeliny synaptycznej.¹⁷

¹⁵ M. Barczyński :Domowy poradnik medyczny, Państwowy Zakład Wydawnictw Lekarskich, Warszawa 1991 s. 12-13.

¹⁶ Tamże s. 12-13

¹⁷ Tamże s. 13.



Rysunek 4. Schematyczny rysunek neuronu, pokazujący transport wewnątrzaksonalny.¹⁸

Poprzez dendryty i synapsy sygnał w postaci impulsów elektrycznych o amplitudzie ok. 100mV i czasie trwania ok 1ms dostaje się do jądra neuronu. Tam następuje proces obliczeniowy, po którym generowany jest sygnał wynikowy. Sygnał wynikowy poprzez akson przesyłany jest dalej do wejść innych neuronów w sieci. Z cybernetycznego punktu widzenia neuron biologiczny traktuje się jako specyficzny przetwornik sygnałów, który posiada wiele wejść i tylko jedno wyjście.

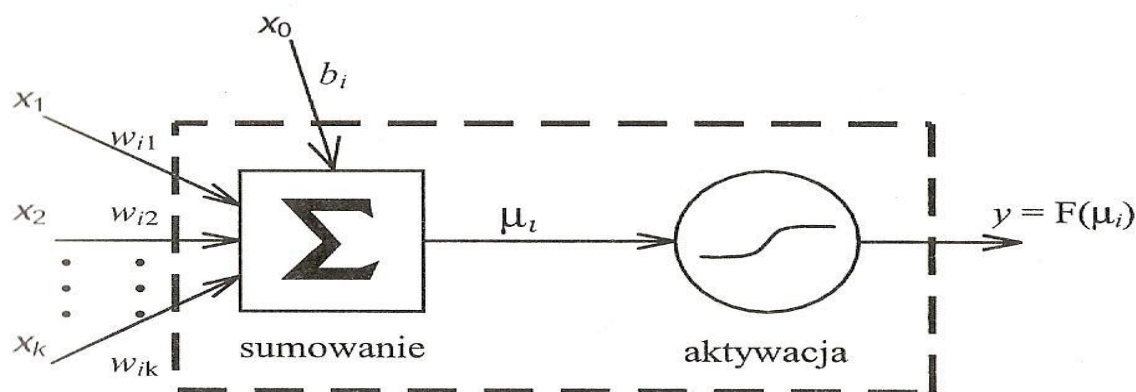
¹⁸ M. Barczyński :Domowy poradnik medyczny, Państwowy Zakład Wydawnictw Lekarskich, Warszawa 1991 s. 13

2. Model sieci neuronowych

2.1 Modelowanie neuronów

Neuron jest jednostką przetwarzającą informację, która ma podstawowe znaczenie dla funkcjonowania sieci neuronowej. Na rysunku 5 jest pokazany model neuronu. Wyróżniamy trzy podstawowe elementy w tym modelu.

Pierwszym elementem jest zbiór synaps albo połączeń wejściowych neuronu, zwanych wagami synaptycznymi. Kolejnym elementem jest sumator, sumujący wejściowe sygnały przemnożone przez wagi synaptyczne. Ostatni element to blok aktywacji, przetwarzający zsumowany sygnał. Aktywacja dokonuje się przez odwzorowanie zadaną funkcją. Funkcja aktywacji ogranicza zakres sygnału wejściowego (odpowiedzi) do zadanego zakresu. Najczęściej znormalizowana amplituda odpowiedzi jest zawarta w przedziale domkniętym $[0,1]$ lub $[-1,1]$.¹⁹



Rysunek 5. Model sztucznego neuronu.²⁰

W modelu sztucznego neuronu wprowadzono następujące zmienne i parametry: wektor wejścia, oznaczany jako $\mathbf{X} = \{x_1, x_2, \dots, x_k\}$, gdzie x_1, x_2, \dots, x_k oznaczają kolejne sygnały wejściowe; wektor wag synaptycznych $\mathbf{W} = \{w_{i1}, w_{i2}, \dots, w_{ik}\}$, gdzie $w_{i1}, w_{i2}, \dots, w_{ik}$ oznaczają kolejne wagi synaptyczne; wartość progowa (bias) $b_i = w_{i0}$; potencjał sztucznego neuronu $\mu_i = \sum w_{ij} + b_i$ oraz funkcja aktywacji $F(\mu_i)$.²¹

Każdy neuron jest pewnego rodzaju procesorem: sumuje z odpowiednimi wagami sygnały wejściowe pochodzące z innych neuronów, tworzy nieliniową (progową) funkcję sumy i przekazuje tę wartość do innych neuronów powiązanych z nim.

¹⁹ L. Ziemiański: Sieci neuronowe w dynamice konstrukcji- wybrane zagadnienia, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów 1999, s. 16

²⁰ Tamże, s. 16

²¹ Tamże, s. 16

W pierwszych modelach neuronu w związku z obowiązującą zasadą: ‘wszystko, albo nic’, przyjmowano sygnał wyjściowy w postaci binarnej (0 lub 1). Wartość 1 odpowiada pobudzeniu neuronu większemu niż próg zadziałania, a wartość 0 – pobudzeniu mniejszemu niż próg.²² Zadaniem każdego neuronu w sieci jest odbieranie sygnałów otrzymywanych od innych (niekoniecznie wszystkich) neuronów, przetworzenie sumarycznego sygnału przy zastosowaniu tzw. funkcji aktywacji, a następnie wysłanie przetworzonego sygnału do innych neuronów.²³ Neuron taki może istnieć jako element w modelu teoretycznym, bądź w programie numerycznym stanowiącym sztuczną sieć neuronową (element programowy). Taki matematyczny model często jest nazywany *modelem formalnym*. Sztuczny neuron może być też zrealizowany w postaci elementu sprzętowego, np. w układzie scalonym.²⁴

Działanie sztucznej sieci neuronowej polega na tym, że sygnały pobudzające zawarte w wektorze wejściowym podawane na wejścia sieci, przetwarzane są w poszczególnych neuronach. Po tej projekcji na wyjściach sieci otrzymuje się wartości liczbowe, które stanowią odpowiedź sieci na pobudzenie i stanowią rozwiązanie postawionego problemu. Jednak aby takie rozwiązanie uzyskać, należy przejść żmudną drogę uczenia sieci. Jedną z metod to uczenie metodą wstecznej propagacji błędów.

Funkcja aktywacji może występować w różnych postaciach, jednak każda funkcja, aby mogła pełnić tę rolę musi być ciągła i łatwo różniczkowalna. W praktyce stosuje się najczęściej następujące funkcje: liniową, logistyczną, tangens hiperboliczny (funkcja sigmoidalna), sinus oraz signum.

Jedną z najczęściej wykorzystywanych funkcji aktywacji jest funkcja sigmoidalna. Opisać ją można za pomocą wzoru:

$$y = \frac{1}{1 + \exp(2 * (-2 * s * r))} \quad (1)$$

$$d = 2 * s * y * (1 - y) \quad (2)$$

Funkcja rozpatrywana jest w przedziale $0 < y < 1$, gdzie: x to wejście funkcji aktywacji, y – wyjście funkcji, s to stromizna funkcji, a d – pochodna funkcji.

2.1.1 Model neuronu McCullocha-Pittsa

Jednym z ważniejszych modeli neuronu jest *neuron McCullocha-Pittsa* wprowadzony w 1943 roku. Stan takiego neuronu o numerze i -tym jest w tym modelu opisany następującymi zależnościami:

$$\sigma_i(t) = f[h_i(t) - T_i], \quad (3)$$

$$h_i(t) = \sum_{j=1}^n J_{ij} \sigma_j(t-1), \quad (4)$$

²² S. Osowski: Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2006, s. 12

²³ J. Mańdziuk: Sieci neuronowe typu Hopfielda. Teoria i przykłady zastosowań, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2000, s. 4

²⁴ R. A. Kosiński: Sztuczne sieci neuronowe. Dynamika nieliniowa i chaos, Wydawnictwa Naukowo-Techniczne, Warszawa 2002-2006, s. 33

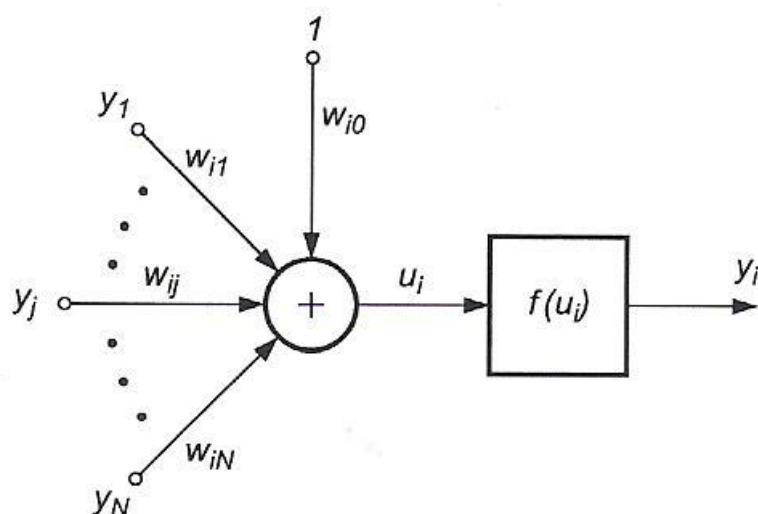
gdzie: $\sigma_i(t)$ jest stanem neuronu i -tego w czasie (chwili) t , przy czym czas jest wielkością dyskretną $t = 0, 1, 2, 3, \dots, f$ – funkcją aktywacji neuronu, $h_i(t)$ jest nazywane polem lokalnym działającym na i -ty neuron w chwili t , pochodzącym od wszystkich n neuronów połączonych z i -tym neuronem, T_i – progiem zadziałania i -tego neuronu J_{ij} – siłą (wagą) połączenia synaptycznego między i -tym a j -tym neuronem, $\sigma_j(t-1)$ – stanem j -tego neuronu w chwili wcześniejszej ($t-1$).²⁵ Model McCullocha-Pittsa jest modelem dyskretnym, w którym stan neuronu w chwili ($k+1$) określany na podstawie stanu sygnałów wejściowych neuronów w chwili poprzedniej k . Przyjęcie modelu dyskretnego jest uzasadnione występowaniem zjawiska refrakcji w rzeczywistej komórce nerwowej, które powoduje, że neuron może zmieniać swój stan z pewną ograniczoną częstotliwością, przy występowaniu stref nieczułości, regulowanych przez częstotliwość próbkowania.²⁶

2.1.2 Model neuronu Hebba

D. Hebb zaprezentował teorię uczenia (doboru wag w_{ij} neuronów) w zastosowaniu do pamięci asocjacyjnych. Wykorzystał przy tym obserwacje, że waga połączeń międzyneuronowych jest wzmacniana przy stanach uaktywnienia obu neuronów. W modelu Hebba przyrost wagi Δw_{ij} w procesie uczenia jest proporcjonalny do iloczynu sygnałów wyjściowych y_i oraz y_j neuronów połączonych wagą w_{ij}

$$w_{ij}(k+1) = w_{ij}(k) + \eta y_j(k) \eta y_i(k) \quad (5)$$

w którym k oznacza kolejny cykl, a η jest współczynnikiem uczenia.²⁷



Rysunek 6. Model neuronu Hebba.²⁸

²⁵ R. A. Kosiński: Sztuczne sieci neuronowe. Dynamika nieliniowa i chaos, Wydawnictwa Naukowo-Techniczne, Warszawa 2002-2006, s. 35)

²⁶ S. Osowski: Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2006, s. 13

²⁷ S. Osowski: Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2006, s. 13

²⁸ R. A. Kosiński: Sztuczne sieci neuronowe. Dynamika nieliniowa i chaos, Wydawnictwa Naukowo-Techniczne, Warszawa 2002-2006, s. 33-34

2.1.3 Model neuronu chaotycznego

Przykładem bardziej złożonego modelu neuronu, w którym istnieje pamięć o jego poprzednich stanach, sięgająca tak daleko w przeszłość, jak to ustalimy odpowiednim parametrem, jest model neuronu używany w pracach Aihary. Jest on opisany zależnością

$$S_1(t+1) = f\left[\sum_{j=1}^N J_{ij} S_j - \sum_{r=0}^t k^r S_i(t-r) - T_i\right] \quad (6)$$

gdzie: f jest funkcją aktywacji, która ma na ogół postać $f(x) = \frac{1}{1+\exp(-gx)}$,

g – parametrem wzmocnienia, k – czynnikiem tłumiącym, T_i – progiem. Jak wynika z (4) na bieżący stan neuronu $S_i(t+1)$ wpływają jego poprzednie stany z r poprzednich kroków czasowych, przy czym im taki wcześniejszy stan jest bardziej odległy od bieżącego stanu neuronu, tym jego wpływ jest mniejszy. Jest to opisane wykładniczym czynnikiem k^r (k jest liczbą ułamkową). Sieć neuronowa złożona z neuronów tego typu ma bardzo ciekawe właściwości fizyczne, m.in. dość łatwo w takiej sieci można uzyskać chaotyczne zachowania.²⁹

2.2 Sztuczne sieci neuronowe

Sztuczna sieć neuronowa to zbiór połączonych ze sobą jednostek wejściowo-wyjściowych. Z każdym połączeniem skojarzona jest waga, która może zostać zmieniona w trakcie uczenia.

Przystępując do konstruowania sztucznej sieci neuronowej należy określić liczbę neuronów N , z których będzie się składała ta sieć. Wielkość N zależy oczywiście od zastosowania sieci – największe sieci mogą mieć liczbę neuronów rzędu 10^9 .

W przypadku sprzętowych implementacji sieci liczba neuronów jest na ogół mniejsza, jednak, ze względu na wielość zastosowań, trudno tutaj podać jakąś typową liczbę. Na przykład układ nerwowy do rozpoznawania pewnej klasy zapachów może składać się z kilkudziesięciu neuronów, a sieć komórkowa do rozpoznawania obrazów TV może się składać z ok. 250 000 neuronów, co odpowiada ilości pikseli w obrazie kamery TV.

Następnym krokiem w konstruowaniu sieci jest ustalenie połączeń synaptycznych J_{ij} między neuronami. To również jest zdeterminowane przeznaczeniem sieci. Połączenia synaptyczne mogą być symetryczne bądź asymetryczne – ten ostatni przypadek odpowiada jednokierunkowemu działaniu synapsy w sieci żywej. Każde połączenie synaptyczne ma określoną wartość J_{ij} , będącą na ogół liczbą rzeczywistą z przedziału $[-1, +1]$, chociaż używa się też sieci, w których mogą one przyjmować wartości dyskretne. Ujemne wartości odpowiadają połączeniom synaptycznym hamującym w biologicznych sieciach neuronowych, a dodatnie – połączeniom pobudzającym.³⁰ Sieci neuronowe, zarówno biologiczne, jak i sztuczne są przykładem sieci złożonych. Mają one właściwości sieci typu *małego świata*³¹ oraz są bezskalowe.

²⁹ R. A. Kosiński: Sztuczne sieci neuronowe. Dynamika nieliniowa i chaos, Wydawnictwa Naukowo-Techniczne, Warszawa 2002-2006, s. 35

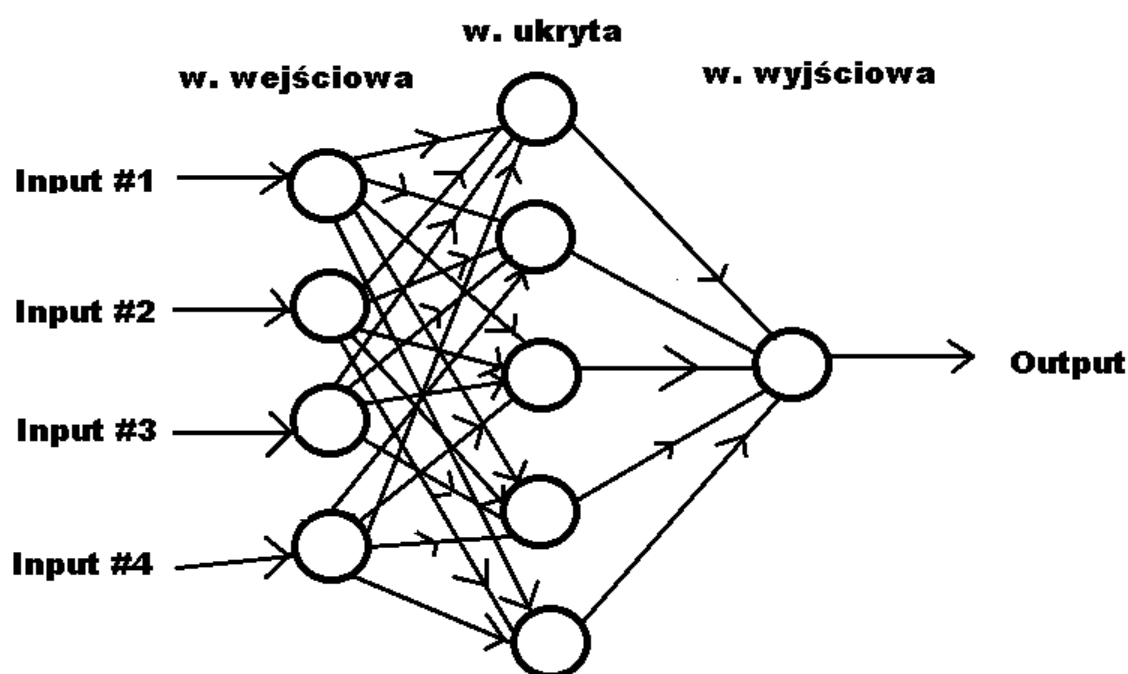
³⁰ R. A. Kosiński: Sztuczne sieci neuronowe. Dynamika nieliniowa i chaos, Wydawnictwa Naukowo-Techniczne, Warszawa 2002-2006, s. 40

³¹ z ang. *small world* – mały świat

Oznacza to, że średnia najkrótsza droga L między dwoma neuronami jest stosunkowo krótka oraz prawdopodobieństwo P , iż neuron ma k sąsiednich neuronów, z którymi jest połączony, jest proporcjonalna do wartości k^γ , gdzie stała wartość $2 < \gamma < 3$ – zależy od rodzaju sieci neuronowej.³²

2.2.1 Architektura sztucznych sieci neuronowych

Ze względu na fakt, iż pojedynczy neuron posiada niewielkie możliwości w zakresie przetwarzania informacji, tworzy się strukturę nazywaną **siecią neuronową** powstałą w wyniku połączenia grup neuronów w **warstwy**. Stosowana jest przy tym zasada łączenia każdego neuronu warstwy poprzedniej z każdym neuronem warstwy następnej. Powstające w ten sposób grupy możemy podzielić na warstwę wejściową, gdzie wprowadzane są dane wejściowe, warstwy ukryte, na które propagowane są dane oraz warstwę wyjściową, gdzie otrzymywane są wartości wyjściowe.



Rysunek 7. Schemat budowy sieci wielowarstwowej. (Źródło: opracowanie własne)

Sposób, w jaki neurony łączą się w sieć neuronową decyduje o architekturze tej sieci oraz algorytmie uczenia zastosowanym do jej trenowania. Architektury sieci neuronowych można podzielić na cztery główne rodzaje:

1. jednowarstwowe sieci jednokierunkowe
2. wielowarstwowe sieci jednokierunkowe
3. sieci rekurencyjne
4. sieci komórkowe.³³

Neurony wchodzące w skład sieci tworzą warstwy, z których pierwsza nosi nazwę *warstwy wejściowej*, ostatnia *warstwy wyjściowej*, zaś wszystkie warstwy znajdujące się pomiędzy nimi określane są jako *warstwy ukryte*. Wartości wejściowe sieci wprowadzane są na wejścia neuronów warstwy wejściowej.

³² L. Ziemiański: Sieci neuronowe w dynamice konstrukcji- wybrane zagadnienia, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów 1999, s. 15

³³ Tamże, s. 19

Następnie, poprzez istniejące połączenia, wartości wyjściowe neuronów jednej warstwy przekazywane są na wejścia elementów przetwarzających kolejnej warstwy. Wartości uzyskane na wyjściach neuronów ostatniej warstwy są jednocześnie wartościami wyjściowymi sieci.

Sieć jednowarstwową tworzą neurony ułożone w jednej warstwie. Każdy neuron ma polaryzację (waga w_{i0} prowadząca od sygnału jednostkowego) oraz wiele wag w_{ij} prowadzących od sygnałów wejściowych x_j .

W **wielowarstwowch** sieciach neuronowych muszą istnieć co najmniej dwie warstwy: wejściowa i wyjściowa. Między nimi natomiast mogą znajdować się warstwy ukryte. Jeżeli sieć zawiera tylko dwie warstwy, to warstwę wejściową utożsamiamy z warstwą ukrytą.³⁴ Sieci te zostaną szerzej opisane w dalszej części pracy.

Sieci rekurencyjne to struktury, w których występuje zjawisko sprzężenia zwrotnego, polegające na tym, że sygnał otrzymany na wyjściu sieci trafia powtórnie na jej wejście, nazywamy sieciami rekurencyjnymi. Jednorazowe pobudzenie struktury ze sprzężeniem zwrotnym może generować całą sekwencję nowych zjawisk i sygnałów, ponieważ sygnały z wyjścia sieci trafiają ponownie na jej wejście, generując nowe sygnały, aż do ustabilizowania się sygnałów wejściowych. Takiemu przebiegowi towarzyszą często tłumienia, oscylacje, gwałtowne narastanie lub opadanie sygnałów.³⁵

W **sieciach komórkowych** neurony są ułożone na płaszczyźnie kwadratu bądź prostokąta, są to więc sieci z wewnętrzną strukturą geometryczną. Położenie każdego neuronu, podobnie jak położenie elementu macierzy, jest opisane dwoma indeksami i, j , określającymi numer wiersza i kolumny. Połączenia synaptyczne łączą każdy neuron tylko z neuronami w sąsiedztwie o określonym promieniu r . Tak więc dla $r=1$ każdy neuron ma 8 sąsiadów, dla $r=2$ ma 24 sąsiadów itd. Neurony leżące blisko krawędzi mają odpowiednio obniżoną liczbę sąsiadów. Podstawową cechą sieci komórkowych jest to, że dla każdego neuronu w sieci układ połączeń synaptycznych w obrębie jego sąsiedztwa jest taki sam (co do położenia i wartości połączeń synaptycznych). Ten powtarzający się układ połączeń synaptycznych jest nazywany szablonem sieci. Neurony leżące blisko krawędzi sieci i mające obniżoną liczbę sąsiadów mają również odpowiednio zmieniony szablon. Można więc stwierdzić, że sieci komórkowe są przypadkiem sieci z rozrzedzeniem połączeń synaptycznych.³⁶

Topologia sieci z połączeniami typu "każdy z każdym" to rozwiązanie, w którym każdy neuron ukryty i każdy neuron wyjściowy połączony jest z każdym neuronem z warstwy poprzedniej.³⁷ Jest to reguła dość kosztowna obliczeniowo – w przypadku gdy neuronów łączonych jest n , wtedy aktywnych połączeń jest $(n-1)n/2$, tak więc aby zapisać wartości wszystkich występujących w tego typu połączeniach współczynników wagowych potrzebny jest komputer o bardzo dużym zasobie pamięci.

³⁴ Rutkowski: Metody i techniki sztucznej inteligencji, Wydawnictwo Naukowe PWN, Warszawa 2006 s. 179

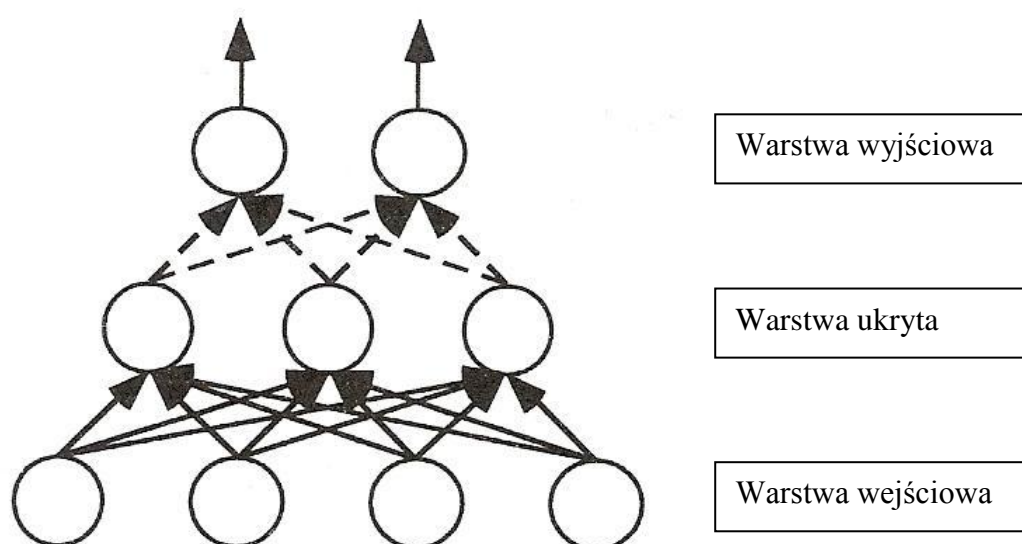
³⁵ Tamże, s. 199

³⁶ R. A. Kosiński: Sztuczne sieci neuronowe. Dynamika nieliniowa i chaos, Wydawnictwa Naukowo-Techniczne, Warszawa 2002-2006, s. 75

³⁷ Statsoft, Wprowadzenie do sieci neuronowych, Statsoft Polska Sp. z o.o, Kraków 2001, s. 9

2.2.2 Struktura i schemat sieci wielowarstwowej

Jak już wspominałem, sieć wielowarstwowa jednokierunkowa składa się ze zbioru neuronów, logicznie rozmieszczonych w dwóch lub więcej warstwach. Istnieją warstwa wejściowa i warstwa wyjściowa, zawierające co najmniej po jednym neuronie. Neurony w warstwie wejściowej są hipotetyczne w tym sensie, że nie mają własnych wejść i nie wykonują żadnego przetworzenia. Ich pobudzenie (sygnały wejściowe) jest określone przez wejście sieci. Istnieje zwykle jedna lub więcej warstw ukrytych, umieszczonych między warstwami wejściową i wyjściową. Określenie „jednokierunkowe” oznacza, że informacja przepływa tylko od wejścia do wyjścia, czyli, że sygnały wejściowe dochodzące do neuronów w każdej warstwie pochodzą wyłącznie od sygnałów wejściowych neuronów z warstw poprzednich, a sygnały wyjściowe z tych neuronów dochodzą wyłącznie do neuronów w następnych warstwach. Sieć wielowarstwowa składa się z co najmniej jednej warstwy pośredniej. Każda warstwa pośrednia w sieciach jednokierunkowych nazywana jest warstwą ukrytą z uwagi na to, że jej działalność nie może być obserwowana bezpośrednio ani na wejściu sieci ani na jej wyjściu. Schemat tego typu sieci pokazuje rys. 8.



Rysunek 8. Wielowarstwowa sieć jednokierunkowa.³⁸

2.2.3 Perceptron

Perceptron, skonstruowany przez Rosenblatta w 1958 roku był pierwszą sztuczną siecią neuronową³⁹. Był on oparty na modelu biologicznym i posiadał zdolność uczenia się. Niestety, jego możliwości były bardzo ograniczone. Rosenblatt wprowadził wiele odmian perceptronu; ten najprostszy składa się z trzech warstw. Pierwszą z nich jest „siatkówka” wejściowa, połączona z warstwą drugą, złożoną z tak zwanych „jednostek kojarzących”, które pełnią funkcję detektora cech.

³⁸ L. Ziemiański: Sieci neuronowe w dynamice konstrukcji- wybrane zagadnienia, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów 1999, s. 20

³⁹ Rosenblatt, Frank (1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6, pp. 386–408

Wreszcie ta warstwa jest połączona z warstwą odpowiedzi wyjściowej. Jednostki (neurony) tego modelu generują sygnał wyjściowy równy sumie ważonej sygnałów wejściowych z uwzględnieniem progu. Nieliniowość progowa decyduje o właściwościach większości sieci neuronowych. Można wykazać, że gdyby odpowiedź była funkcją liniową, warstwa środkowa byłaby zbędna. Niektóre modele zawierają połączenia hamujące między neuronami wyjściowymi w celu uniknięcia pobudzenia więcej niż jednego neuronu jednocześnie.⁴⁰

⁴⁰ T. Masters: Sieci neuronowe w praktyce. Programowanie w języku C++, Wydawnictwa Naukowo Techniczne, Warszawa 1993, s. 18

3. Metody uczenia sztucznych sieci neuronowych

Istnieje kilka sposobów uczenia sieci neuronowych. Najczęstsze z nich to *uczenie pod nadzorem*, zwane także *uczeniem z nauczycielem* oraz *uczenie bez nadzoru* (bez nauczyciela). Metoda uczenia z nauczycielem znajduje zastosowanie w zadaniach modelowania i programowania, natomiast metoda uczenia bez nadzoru jest przydatna przy uczeniu sieci samoorganizujących się (samouczących się) do których należą m.in. sieci Kohonena. Obie metody zostaną szerzej opisane w kolejnych podrozdziałach.

3.1 Uczenie nadzorowane

Uczenie pod nadzorem, zwane także *uczeniem z nauczycielem* polega ona na tym, że zbiera się kilka próbek, które posłużą za przykłady. Wszystkie dane określające wejścia i wyjścia potrzebne do prezentacji tych zbiorów są określone przez każdą z próbek zbioru uczącego. Następnie każde próbki podzbioru zbioru uczącego podaje się na wejście sieci. Sygnał wejściowy każdej próbki porównuje się z sygnałem wyjściowym jaki chce się otrzymać. Następnie koryguje się wagi łączące neurony w sieci. Celem tej korekcji jest zmniejszanie się błędu działania sieci. Prezentację podzbioru próbek uczących wraz z odpowiednią korekcją wag nazywa się *epoką*, zaś *rozmiar epoki* to liczba próbek podzbioru. Jeśli rozmiar epoki jest mniejszy od całego zbioru uczącego to należy za każdym razem wybierać podzbiór przypadkowo, gdyż w przeciwnym wypadku mogą wystąpić niepożądane oscylacje. Epoki uczenia powtarza się do osiągnięcia dobrej jakości przetwarzania sieci. Dobór wag następuje w procesie uczenia sieci, czyli dopasowania sygnałów wyjściowych y_i do wartości zadanych d_i . Miara dopasowania jest wartość funkcji celu, zwanej również funkcją kosztu. Przy p wektorach uczących (x, d) i M neuronach wyjściowych, funkcję celu można zdefiniować, wykorzystując metrykę euklidesową

$$E = \frac{1}{2} \sum_{k=1}^p \|y^{(k)} - d^{(k)}\|^2 = \frac{1}{2} \sum_{k=1}^p \sum_{i=1}^M (y_i^{(k)} - d_i^{(k)})^2 \quad (7)$$

gdzie wartości sygnałów wyjściowych y_i neuronu są funkcjami wag sieci w_{ij} , które podlegają adaptacji w procesie uczenia, prowadząc do minimum funkcji celu.⁴¹

3.2 Uczenie nienadzorowane

W większości zastosowań sieci neuronowe uczą się z nauczycielem. Dla każdej próbki wejściowej zawartej w zbiorze uczącym znane są pożądane wyjścia. Sieć uczy się kojarzyć prezentowane wejścia i wyjścia. Jednak istnieją pewne zastosowania, dla których „prawidłowe” wyjścia nie są znane.

Metodą przeciwną jest uczenie bez nadzoru (bez nauczyciela). Również jak w metodzie z nadzorem wykorzystujemy zbiór próbek sygnałów wejściowych, jednak nie wprowadza się na wyjście sieci.

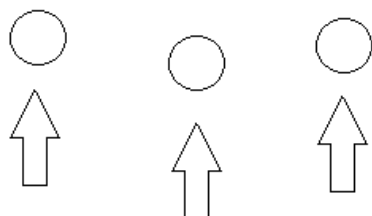
Jednym z najbardziej znanych modeli sieci neuronowej mogących uczyć się bez nadzoru jest sieć Kohonena. Jest to sieć dwuwarstwowa, wykorzystująca uczenie

⁴¹ S. Osowski: Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2006, s. 41

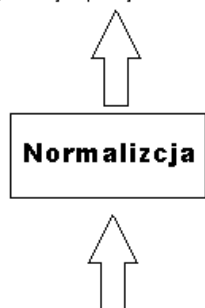
konkurencyjne. Polega to na tym, że neurony rywalizują ze sobą o przywilej uczenia się. Tylko nieliczne będą mogły modyfikować swoje wagi w odpowiedzi na prezentację na wejściu. Ponieważ sieć wymaga normalizacji wejść, niektórzy badacze uważają ją za sieć trójwarstwową, wyróżniając warstwę normującą wejścia, większość jednak uważa normalizację tylko za krok wstępnego przygotowania danych, nie zasługujący na przyznanie mu specjalnej warstwy.⁴²

Niektóre rodzaje normalizacji wejścia wiążą się z generacją dodatkowego sygnału wejściowego zwanym wejściem syntetycznym. Jeżeli próbka pomiarowa jest wektorem o składowych $=n$, to warstwa wejściowa sieci będzie się składać z $n+1$ neuronów. Aktywacje neuronów wejściowych od 0 do $(n-1)$ -szego będą równe n rzeczywistym wejściom pomnożonym przez współczynnik skalowania. Aktywacja n -tego neuronu wejściowego, odpowiadająca wejściu syntetycznemu, będzie funkcją wszystkich rzeczywistych wejść. Jeżeli normalizacja będzie wymagała wejścia syntetycznego, to wciąż będziemy wykorzystywali $n + 1$ neuronów wejściowych, aby zachować spójność oznaczeń. Będziemy przyjmowali, że w tym przypadku wejście syntetyczne będzie tożsamościowo równe 0.⁴³

m neuronów Kohonena tworzy warstwę wyjściową



n skalowanych wejść plus jedno wejście syntetyczne



n wejść rzeczywistych

Rysunek 9. Sieć Kohonena.⁴⁴

3.3 Algorytm wstecznej propagacji błędów

Metoda wstecznej propagacji błędów to najbardziej rozpowszechniona technika uczenia sieci wielowarstwowych.⁴⁵ Nazwa tego algorytmu pochodzi od sposobu jego realizacji – błąd jest „cofany” (propagowany) od warstwy wyjściowej do wejściowej. Do wyprowadzenia tego algorytmu konieczna jest definicja miary błędu.

⁴² T. Masters: Sieci neuronowe w praktyce. Programowanie w języku C++, Wydawnictwa Naukowo Techniczne, Warszawa 1993, s. 301

⁴³ Tamże

⁴⁴ L. Rutkowski: Metody i techniki sztucznej inteligencji, Wydawnictwo Naukowe PWN, Warszawa 2006 s. 181

⁴⁵ Tamże

Będzie to funkcja, w której w rolach zmiennych występują wszystkie wagi wielowarstwowej sieci neuronowej.⁴⁶ Oznacza się ją przez $Q(\mathbf{w})$, gdzie \mathbf{w} oznacza wektor wszystkich wag sieci. W trakcie uczenia należy dążyć do znalezienia minimum funkcji Q względem wektora \mathbf{w} .

Schemat przebiegu algorytmu wstecznej propagacji błędów. Przebieg algorytmu dla wszystkich elementów ciągu uczącego nazywa się epoką.

Jak już wspomniano wyżej, sieć jest uczona poprzez odpowiednie zmiany wartości wag na wejściach poszczególnych neuronów. Aby można tego dokonać potrzebne jest określenie wartości błędów na wyjściach poszczególnych neuronów. Oblicza się je według następujących wzorów:

dla warstwy $k=N$, ostatniej:

$$S_i(n) = \varepsilon_i(n) f^1(S_i(n)), \quad (8)$$

gdzie $\varepsilon_i(n) = d_i(n) - y_i(n)$ jest błędem sygnału wyjściowego poszczególnego neuronu,

$$S_i(n) = \sum_{j=0}^{N_{i-1}} w_{ij}(n) * x_j(n), \quad (9)$$

jest sygnałem wyjściowym części liniowej neuronu

$d_i(n)$ - sygnał wzorcowy i -tego neuronu

Dla każdej innej warstwy k , błąd wylicza się według następującego wzoru:

$$\delta_i^{(k)}(n) = f^1(S_i^{(k)}(n)) * \sum_{m=1}^{N_{i+1}} \delta_i^{(k+1)}(n) w_{mi}^{(k+1)}(n) \quad (10)$$

gdzie y_i oznacza wyjście neuronu, $w_{ij}(k)$ oznacza wagę wejścia j w i -tym neuronie w warstwie k i n to numer okresu przebiegu algorytmu.

Wyznaczanie wartości błędów odbywa się w kierunku od warstwy wyjściowej wstecz do warstwy wejściowej. Gdy błędy są wyliczone, przystępuje się do modyfikacji wartości wag, według wzoru:

$$w_{ij}^{(k)}(n+1) = w_{ij}^{(k)}(n) + \eta \delta_i^{(k)} x_j^{(k)} \quad (11)$$

⁴⁶ L. Rutkowski: Metody i techniki sztucznej inteligencji, Wydawnictwo Naukowe PWN, Warszawa 2006 s. 181

Algorytm można więc przedstawić następująco:⁴⁷

$$w_{ij}^{(k)}(n+1) = w_{ij}^{(k)}(n) + \eta \varepsilon_i^{(k)}(n) f^1 \left(S_i^{(k)}(n) \right) x_j^{(k)} \quad (12)$$

Algorytm rozpoczyna się od podania wzorca uczącego na wejście sieci. Neurony warstwy pierwszej przetwarzają go (wyznaczają sygnał wejściowy dla każdego neuronu w danej warstwie). Sygnały otrzymane w ten sposób zostają wejściami dla neuronów następnej warstwy. Procedurę tę powtarza się wyznaczając wartości sygnałów na wyjściach neuronów i przekazując je dalej, aż do warstwy ostatniej. Jeżeli znamy sygnał wyjściowy ostatniej warstwy i sygnał wzorcowy z ciągu uczącego to możemy obliczyć błąd na wyjściu sieci. Za pomocą reguły delta można zmodyfikować wagi neuronów ostatniej warstwy. Jest to proces, od którego w oczywisty sposób zależy powodzenie użycia sztucznej sieci neuronowej do rozwiązywania postawionego problemu. Projektowanie sieci neuronowej zaczyna się już na poziomie analizy formułowanego zagadnienia. Mówiąc wprost, określenie jakie i ile danych chcemy lub możemy podać na wejścia sieci zdeterminuje wielkość warstwy wejściowej. Ilość wyjść sieci natomiast określi rodzaj oczekiwanej odpowiedzi czyli to, co spodziewamy się otrzymać na wyjściu. Pozostanie zatem do określenia liczba warstw ukrytych i neuronów w tych warstwach. Jest to najtrudniejszy moment tego etapu pracy. Przyjmuje się, że sieć z jedną warstwą ukrytą powinna nauczyć się rozwiązywania większości postawionych problemów. Nie znane są problemy wymagające do rozwiązania sieci z więcej niż trzema warstwami ukrytymi. Nie ma natomiast dobrej recepty na dobór właściwej ilości neuronów w warstwach ukrytych.

Dla jednej takiej warstwy można próbować według wzoru:

$$N_{wu} = \sqrt{N_{wwe} * N_{wwy}} \quad (13)$$

gdzie:

N_{wu} - liczba neuronów w warstwie ukrytej (zaokrąglona oczywiście do wartości całkowitej)

N_{wwe} - liczba neuronów w warstwie wejściowej

N_{wwy} - liczba neuronów w warstwie wyjściowej

⁴⁷ StatSoft, Inc., Internetowy podręcznik statystyki, 1984-2011
http://www.statsoft.pl/textbook/stathome_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstneunet.html

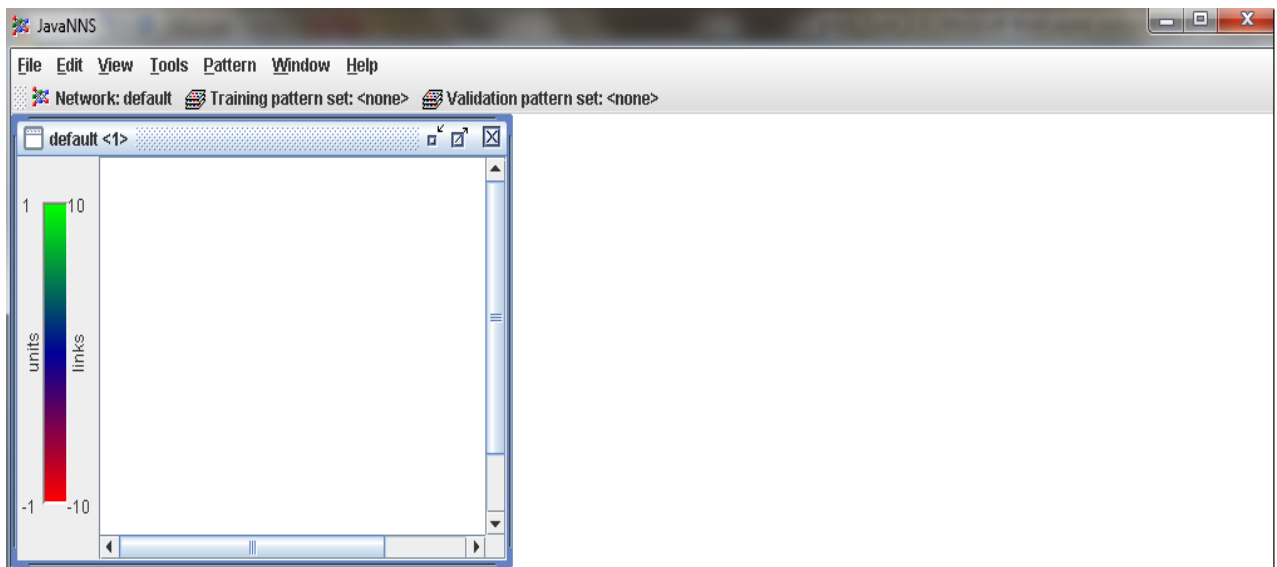
Dla wielu warstw ukrytych przydatna może być metoda *piramidy geometrycznej*, która zakłada, że liczba neuronów w kolejnych warstwach tworzy kształt piramidy i maleje od wejścia do wyjścia. Sytuacja taka została przedstawiona na rysunku 4. Należy jednak podkreślić, że są to metody tylko przybliżone. Ogólnie jednak uczenie rozpoczyna się z małą ilością neuronów ukrytych a następnie, obserwując postępy tego procesu, doświadczalnie zwiększa się ich ilość. Ale skąd może być wiadomo ostatecznie, jaka ilość neuronów w warstwach ukrytych jest optymalna? 5, 10 a może 63? Otóż podczas uczenia należy obserwować błąd sieci i czas jej uczenia. Po dodaniu kolejnego neuronu można często zaobserwować moment prawie skokowej zmiany jakości uczenia. Objawia się to poprzez np. zdecydowane skrócenie czasu uczenia. W celu zoptymalizowania sieci należy zatem badać otoczenie tego punktu. Dodanie jednego neuronu lub zmiana wartości parametru o 0.05 powoduje zasadnicze polepszenie własności sieci.

4. Projekt sieci neuronowej do rozpoznawania znaków na banknotach

W tym rozdziale przedstawiona zostanie tworzenie sieci neuronowej od rozpoznawania znaków na banknotach. Następnie przeprowadzone zostanie uczenie sieci z wykorzystaniem algorytmu Backprop-Momentum.

4.1 Konstrukcja sztucznej sieci neuronowej

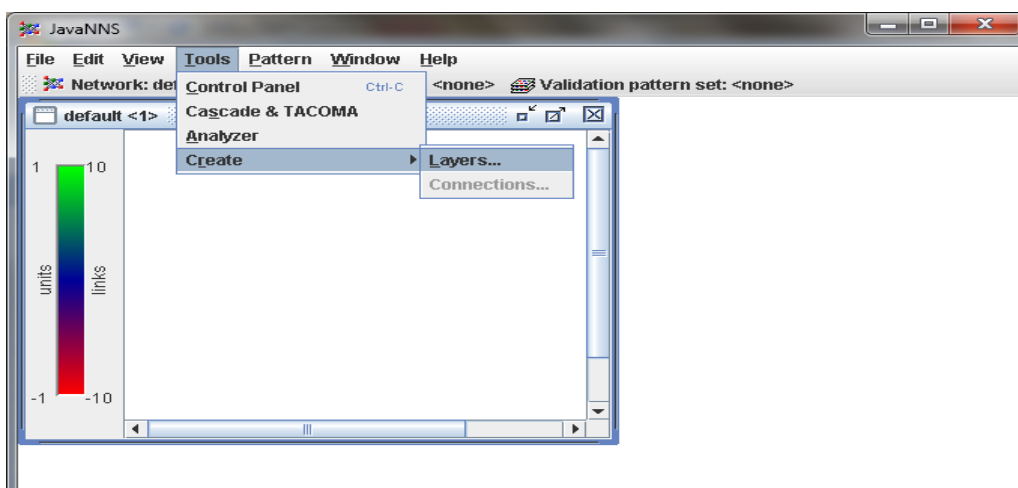
W programie Java Neural Network Simulator (dawniej Stuttgart Neural Network Simulator) przedstawione zostanie tworzenie banknoty.net.



Rysunek 10. Okno startowe JNNS (Źródło: opracowanie własne)

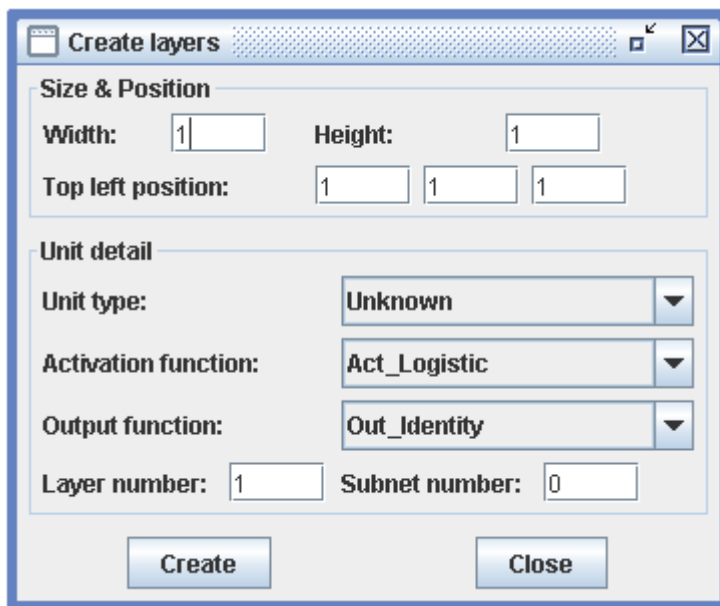
4.1.1 Tworzenie warstw

Aby utworzyć sieć neuronową należy najpierw stworzyć warstwy tej sieci. W tym celu z menu głównego należy wybrać Tools (Narzędzia) następnie Create (Utwórz) > Layers (Warstwy).



Rysunek 11. Tworzenie warstw. (Źródło: opracowanie własne)

Pojawia się okno:



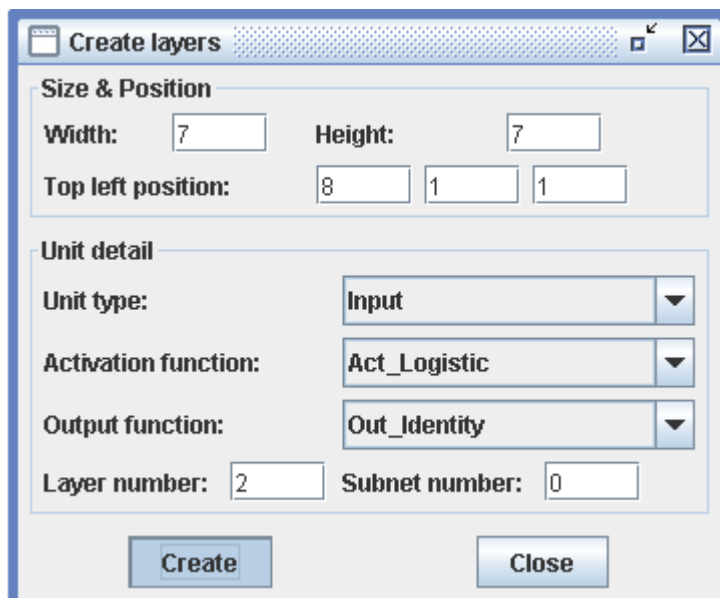
The 'Create layers' dialog box is shown. It has two main sections: 'Size & Position' and 'Unit detail'. In the 'Size & Position' section, the 'Width' is set to 1, 'Height' is set to 1, and 'Top left position' is set to 1, 1, 1. In the 'Unit detail' section, 'Unit type' is set to 'Unknown', 'Activation function' is set to 'Act_Logistic', 'Output function' is set to 'Out_Identity', 'Layer number' is set to 1, and 'Subnet number' is set to 0. At the bottom, there are 'Create' and 'Close' buttons.

Rysunek 12. Tworzenie warstw. (Źródło: opracowanie własne)

Teraz można przystąpić do tworzenia warstw. Po ustaleniu szerokości (Width) i wysokości (Height) oraz pozycji (Top left position) warstwy należy wybrać jej typ (Unit type). Do wyboru jest warstwa wejściowa (Input), warstwa ukryta (Hidden) oraz warstwa wyjściowa (output). Następnie należy ustalić funkcję aktywacji (Activation function) oraz funkcję wyjściową (Output function).

Ostatnie z opcji to numer warstwy (Layer number) i numer podsieci (subnet number).

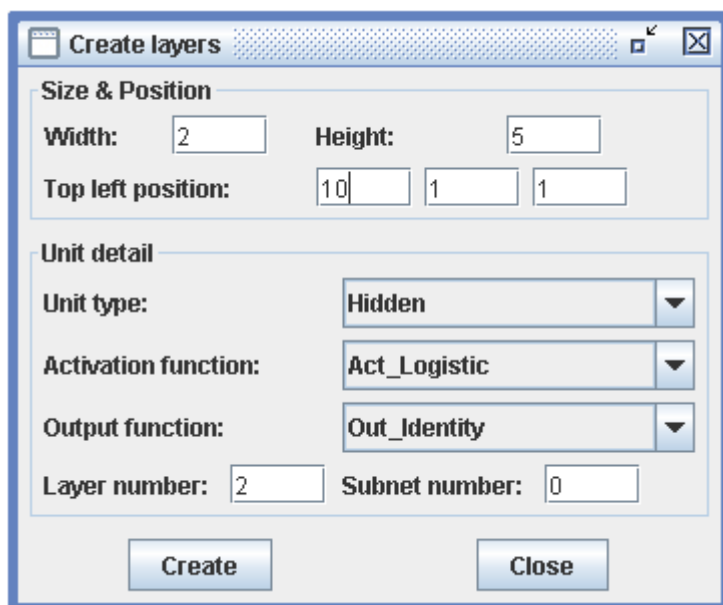
Najpierw stworzymy warstwę wejściową o parametrach 7/7, pozycji 8/1/1



The 'Create layers' dialog box is shown. In the 'Size & Position' section, 'Width' is set to 7, 'Height' is set to 7, and 'Top left position' is set to 8, 1, 1. In the 'Unit detail' section, 'Unit type' is set to 'Input', 'Activation function' is set to 'Act_Logistic', 'Output function' is set to 'Out_Identity', 'Layer number' is set to 2, and 'Subnet number' is set to 0. At the bottom, there are 'Create' and 'Close' buttons.

Rysunek 13. Tworzenie warstwy wejściowej. (Źródło: opracowanie własne)

Następnie warstwę ukrytą 2/5, pozycji 10/1/1



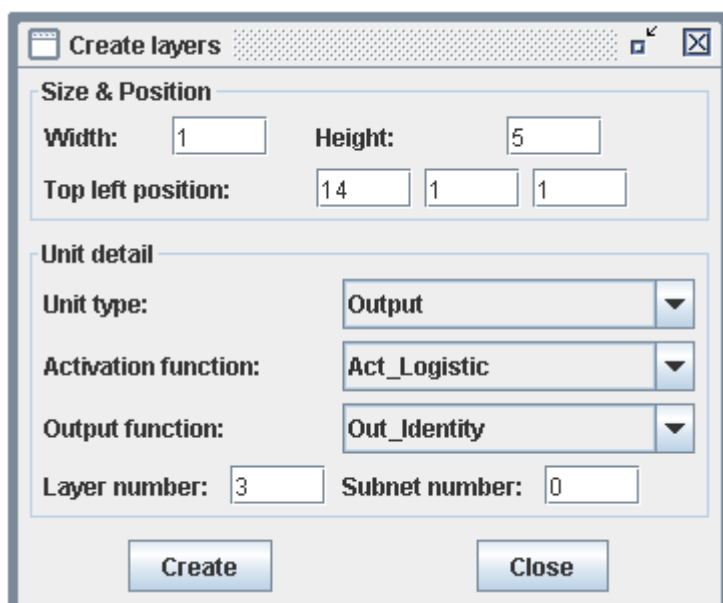
The 'Create layers' dialog box is shown with the following settings:

- Size & Position:**
 - Width: 2
 - Height: 5
 - Top left position: 10, 1, 1
- Unit detail:**
 - Unit type: Hidden
 - Activation function: Act_Logistic
 - Output function: Out_Identity
 - Layer number: 2
 - Subnet number: 0

Buttons: Create, Close

Rysunek 14. Tworzenie warstwy ukrytej. (Źródło: opracowanie własne)

A na koniec warstwę wyjściową 1/5, pozycji 14/1/1



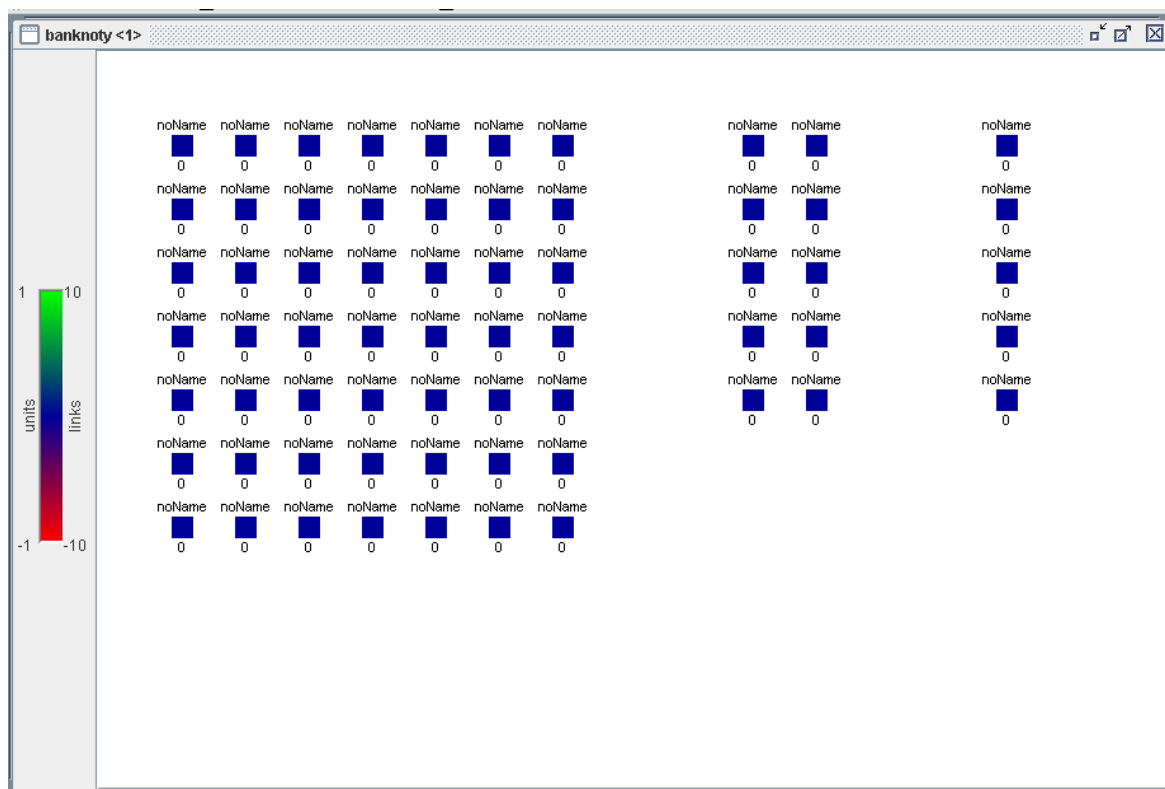
The 'Create layers' dialog box is shown with the following settings:

- Size & Position:**
 - Width: 1
 - Height: 5
 - Top left position: 14, 1, 1
- Unit detail:**
 - Unit type: Output
 - Activation function: Act_Logistic
 - Output function: Out_Identity
 - Layer number: 3
 - Subnet number: 0

Buttons: Create, Close

Rysunek 15. Tworzenie warstwy wyjściowej. (Źródło: opracowanie własne)

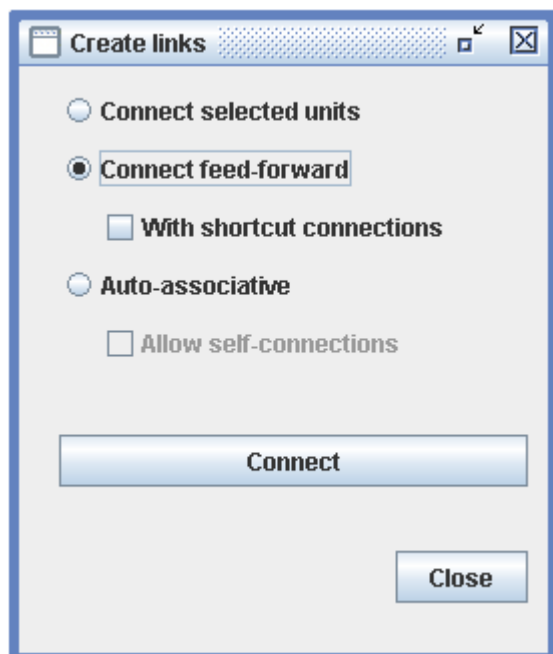
W oknie głównym programu pojawiają się nasze warstwy:



Rysunek 16. Warstwy- od lewej: wejściowa, ukryta, wyjściowa. (Źródło: opracowanie własne)

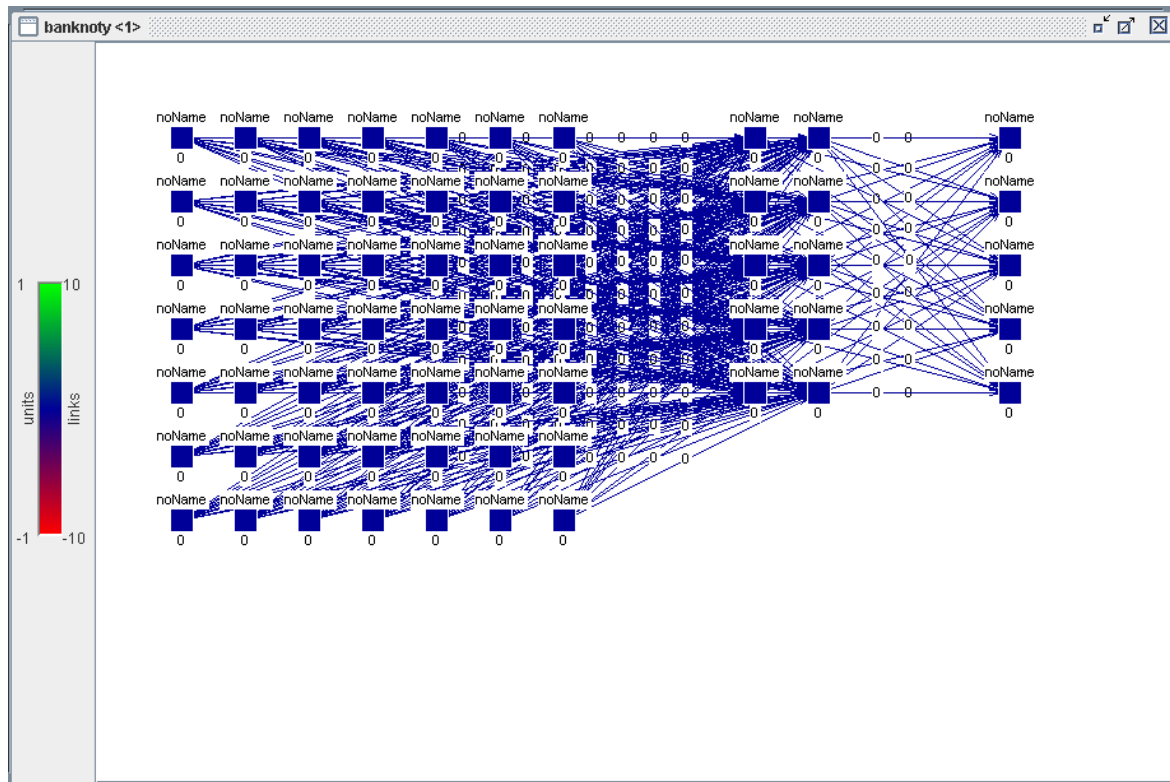
4.1.2 Łącznie warstw

Po utworzeniu warstw należy je połączyć. W tym celu z menu Tools (Narzędzia) należy wybrać Create (Utwórz) > Connections (Połączenia). Pojawia się okno:



Rysunek 16. Łączenie warstw jednokierunkowe. (Źródło: opracowanie własne)

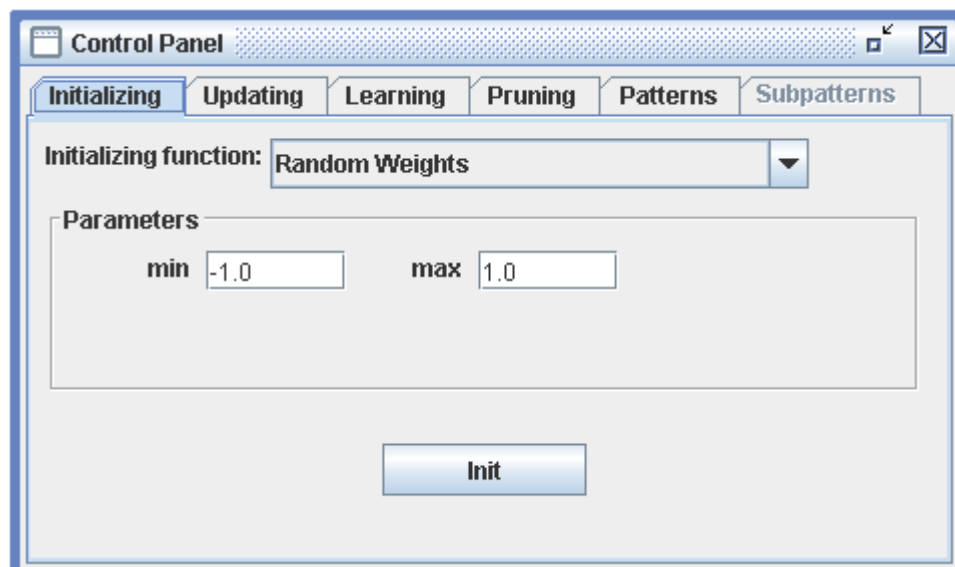
Dla swojej sieci wybrałem opcję „Connect feed-forward” (Połącz jednokierunkowe). Po kliknięciu „Connect” (Połącz) warstwy łączą się, obrazuje to poniższy rysunek:



Rysunek 17. Warstwy połączone jednokierunkowo. (Źródło: opracowanie własne)

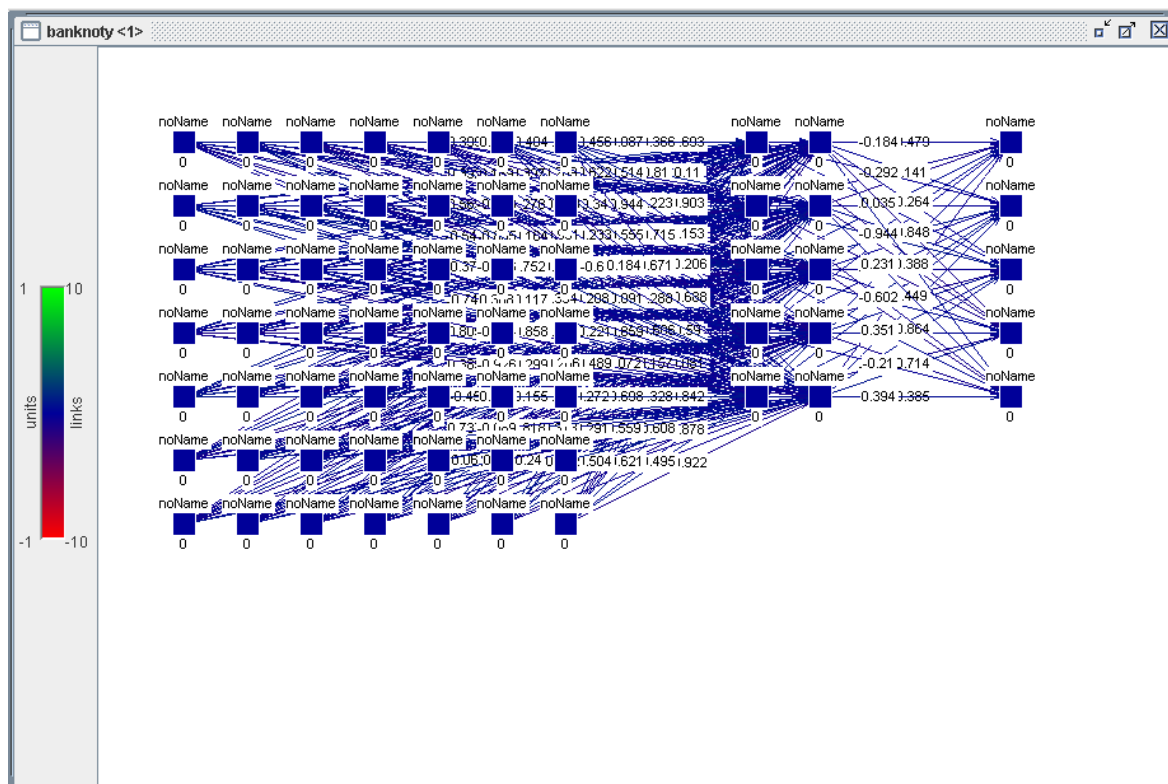
4.1.3 Inicjalizacja wag

Teraz należy dobrać wagi. Dobierzemy losowe wagi, wchodząc do menu Tools (Narzędzia) > Control Panel (Panel Sterowania). W zakładce Initializing (Inicjalizacja) należy wybrać Initialization Function (Funkcję Inicjalizacji) typu Random Weights (Losowe Wagi). Operację zatwierdzamy przyciskiem Init (Inicjalizuj)



Rysunek 18. Inicjalizacja wag. (Źródło: opracowanie własne)

Otrzymaliśmy losowe wagi:



Rysunek 19. Warstwy połączone jednokierunkowo po losowej inicjalizacji wag. (Źródło: opracowanie własne)

4.2 Proces uczenia sieci.

W tym podrozdziale zostanie opisany proces wczytania wzoru a następnie uczenia sieci za pomocą tego wzoru.

4.2.1 Wczytywanie wzoru

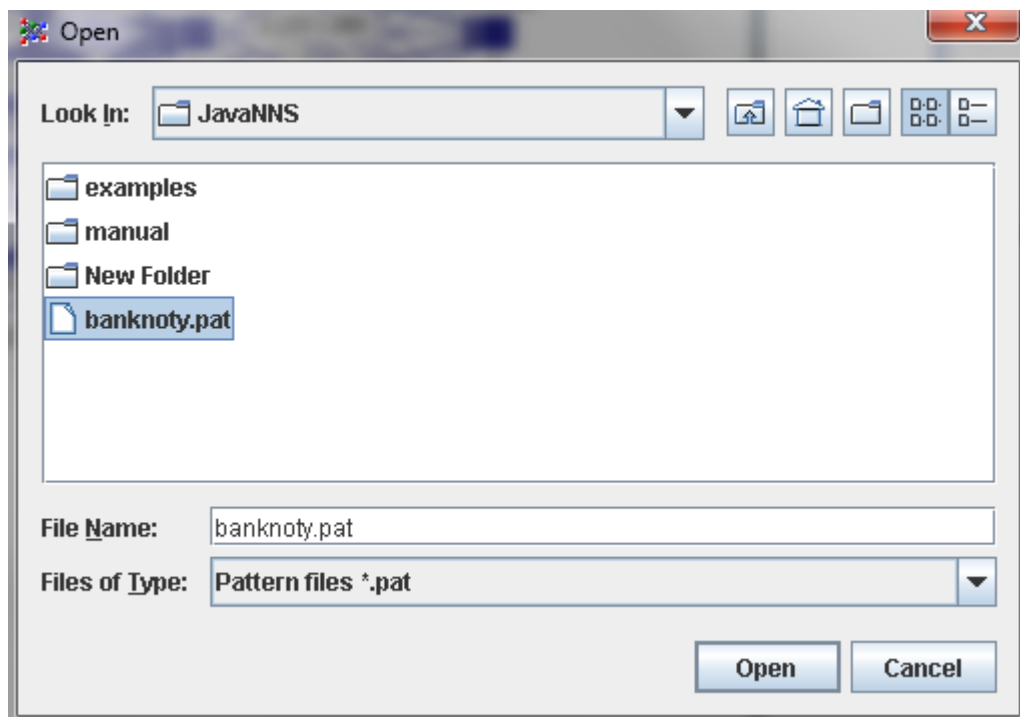
W celu nauczania sieci należy wczytać przygotowany wcześniej plik typu .pat (pattern⁴⁸). Struktura pliku pattern w wersji tekstowej wygląda tak jak na rysunku 20.

⁴⁸ z ang. wzór

```
banknoty2.pat
5 No. of patterns : 5
6 No. of input units : 49
7 No. of output units : 5
8
9 # Input pattern 1:
10 1 1 1 1 1 1 1
11 1 0 0 0 0 0 1
12 1 0 0 0 0 0 1
13 1 0 0 0 0 0 1
14 1 0 0 0 0 0 1
15 1 0 0 0 0 0 1
16 1 1 1 1 1 1 1
17 # Output pattern 1:
18 1 0 0 0 0
19
20 # Input pattern 2:
21 0 0 1 1 1 0 0
22 0 1 0 0 0 1 0
23 1 0 0 0 0 0 1
24 1 0 0 0 0 0 1
25 1 0 0 0 0 0 1
26 0 1 0 0 0 1 0
27 1 0 1 1 1 0 0
28 # Output pattern 2:
29 0 1 0 0 0
30
31 # Input pattern 3:
32 0 0 0 1 0 0 0
33 0 0 1 0 1 0 0
34 0 1 0 0 0 1 1
35 1 0 0 0 0 0 1
36 0 1 0 0 0 1 0
37 0 0 1 0 1 0 0
38 0 0 0 1 0 0 0
39 # Output pattern 3:
40 0 0 1 0 0
41
42 # Input pattern 4:
43 0 0 0 1 0 0 0
44 0 0 0 1 0 0 0
45 0 0 0 1 0 0 0
46 1 1 1 1 1 1 1
47 0 0 0 1 0 0 0
48 0 0 0 1 0 0 0
49 0 0 0 1 0 0 0
50 # Output pattern 4:
51 0 0 0 1 0
```

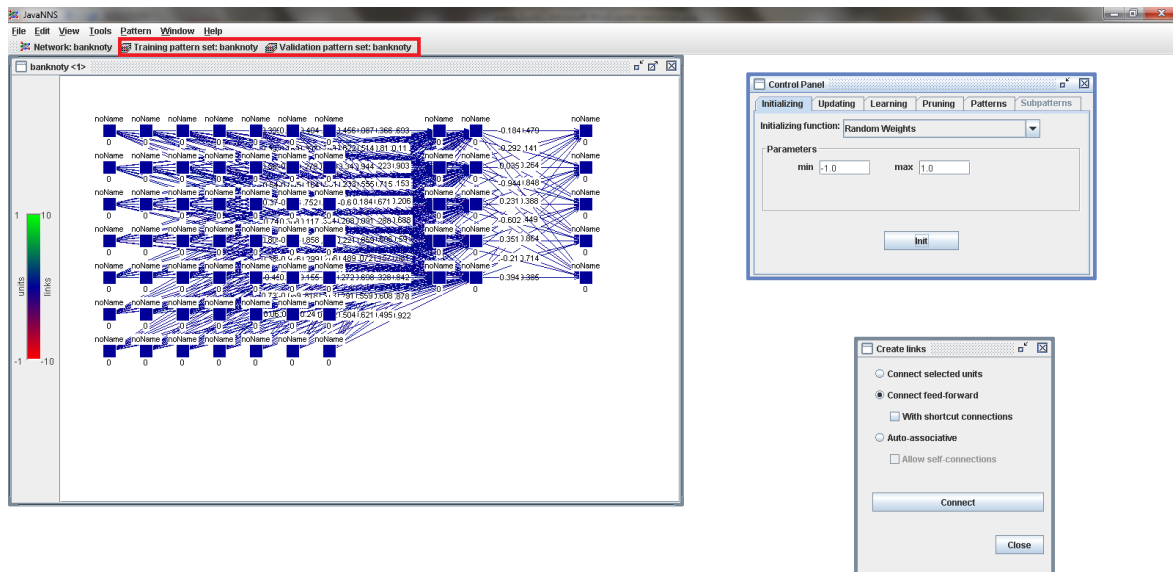
Rysunek 20. Struktura pliku pattern. (Źródło: opracowanie własne)

Z menu File > Open wystarczy wybrać żądany plik i program JNNS go wczyta.



Rysunek 21. Wczytywanie pliku pattern. (Źródło: opracowanie własne)

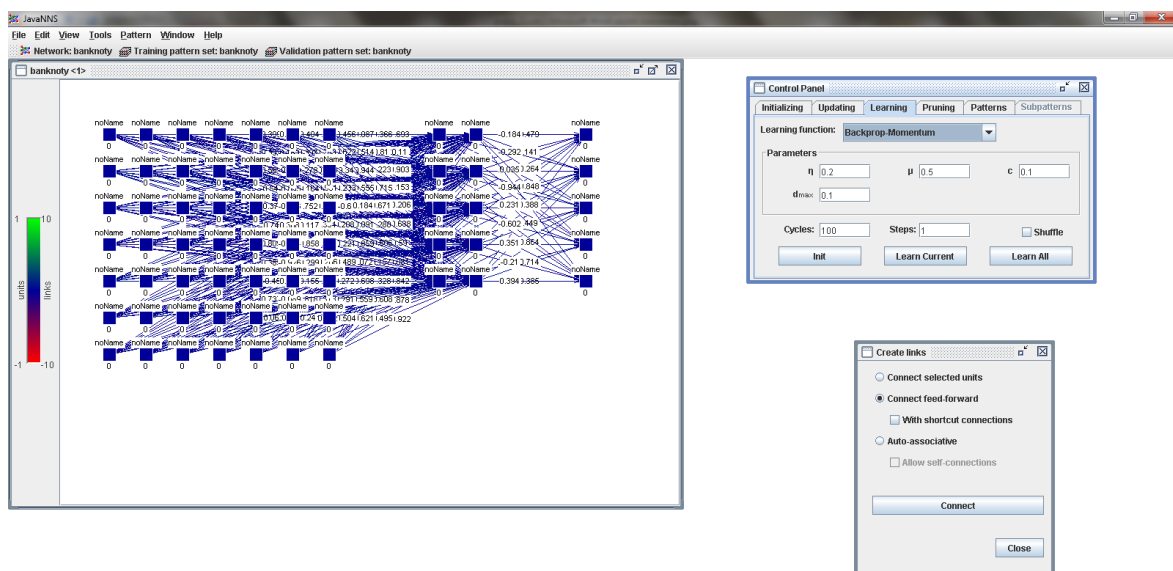
Wzór został wczytany:



Rysunek 22. Struktura programu po wczytaniu pliku pattern. (Źródło: opracowanie własne)

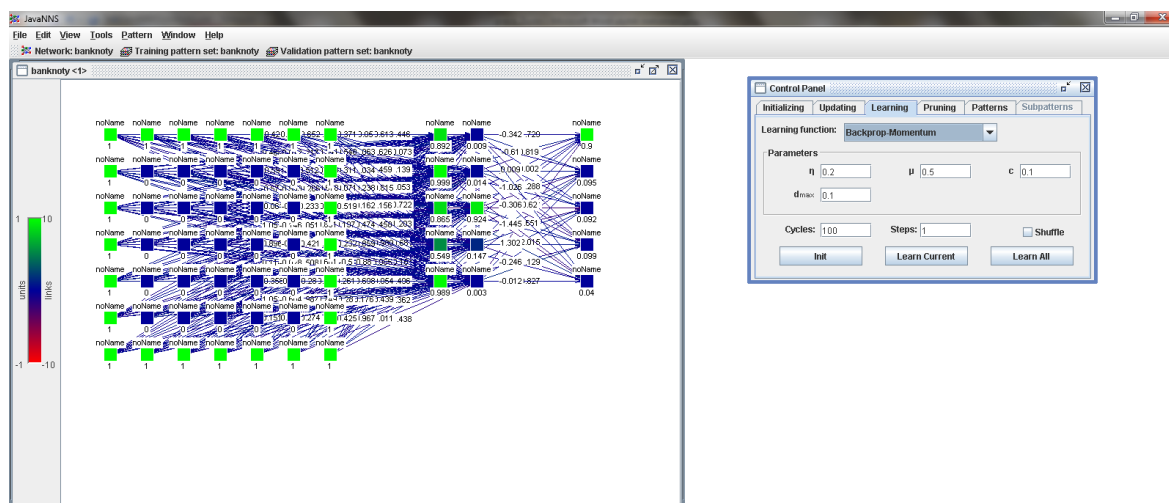
4.2.2 Funkcja uczenia Backprop-Momentum.

Teraz pozostaje tylko nauczyć sieć. W Control Panel (Panelu Sterowania) wybieramy zakładkę Learning (Uczenie). Jako Learning Function (Funkcję Uczenia) wybierzemy Backprop-Momentum. Następnie klikamy Learn Current (Naucz Obecny) aby rozpocząć proces uczenia.



Rysunek 23. Rozpoczynanie procesu uczenia. (Źródło: opracowanie własne)

Od razu widać efekty – sieć rozpoznała pierwszy ze znaków: kwadrat występujący na banknotach dziesięciozłotowych (wzór 1).

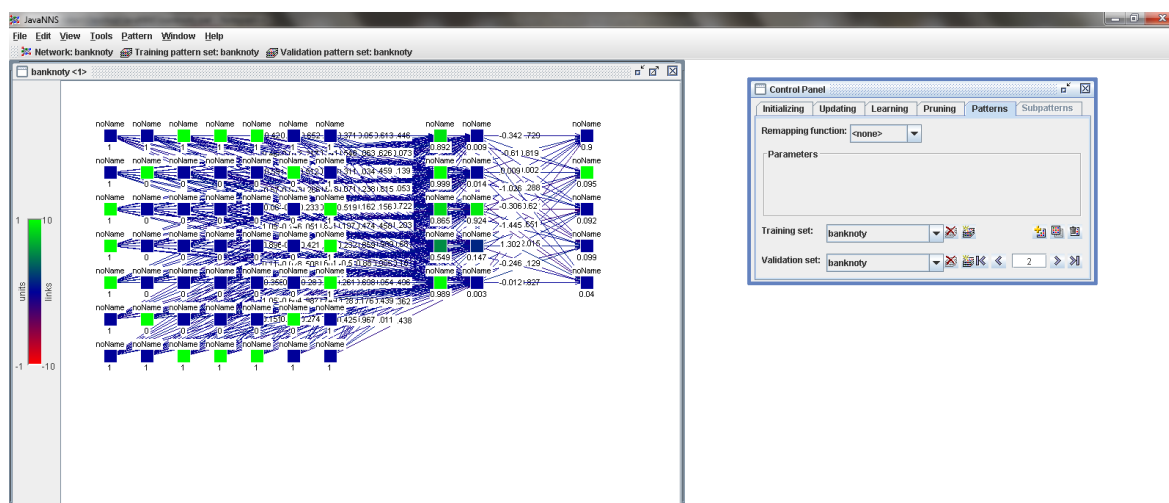


Rysunek 24. Wzór dla banknotu dziesięciozłotowego. (Źródło: opracowanie własne)

4.3 Wyniki uczenia

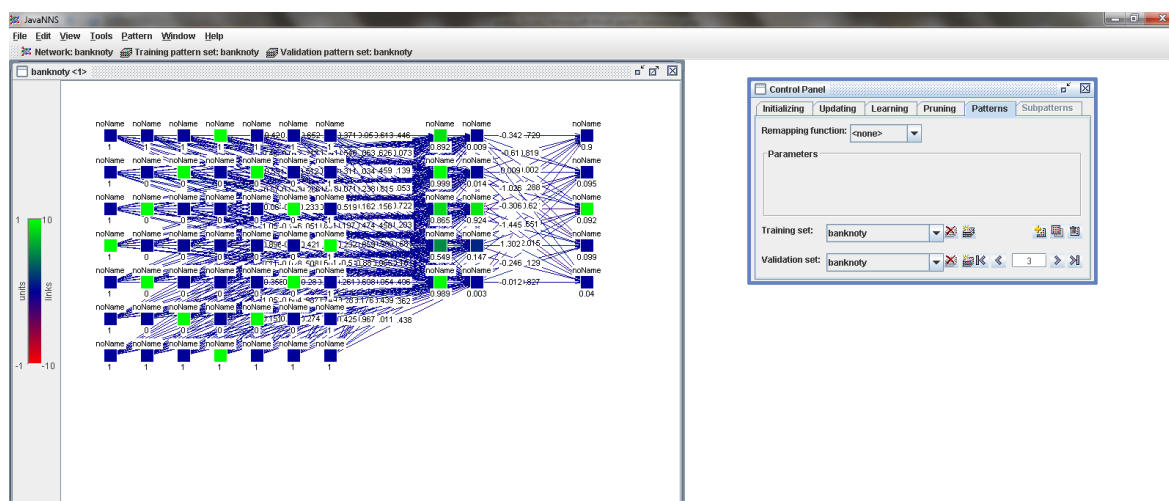
W zakładce Patterns (wzory) możemy zaobserwować kolejne rozpoznawanie znaków przez sieć: okrąg dla banknotów dwudziestozłotowych (wzór 2), romb dla banknotów pięćdziesięciuzłotowych (wzór 3), krzyż dla banknotów stu złotych (wzór 4) oraz trójkąt równoramienny dla banknotów o nominale dwieście złotych (wzór 5)

4.3.1 Wzór dla 20 złotych



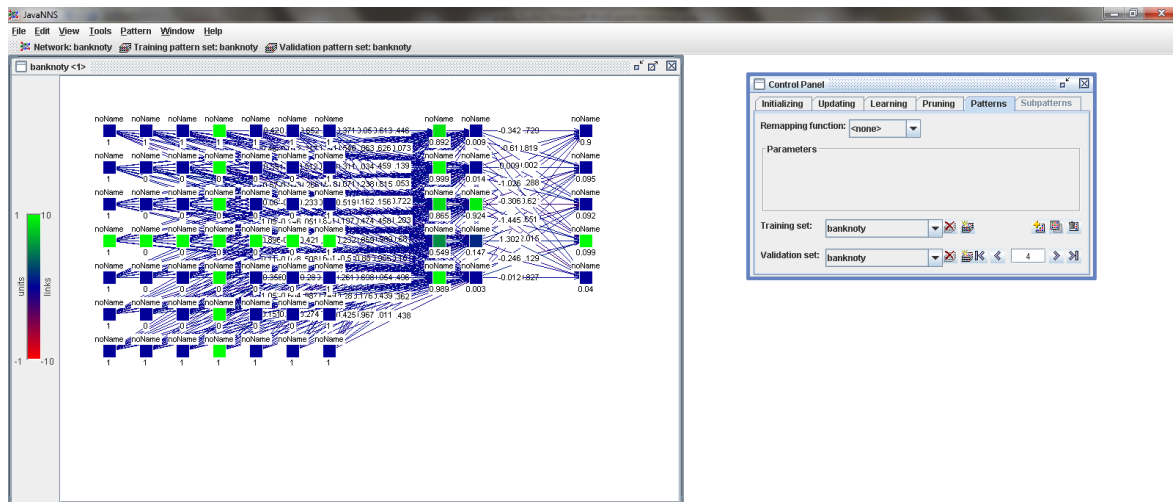
Rysunek 25. Wzór dla banknotu dwudziestozłotowego. (Źródło: opracowanie własne)

4.3.2 Wzór dla 50 złotych



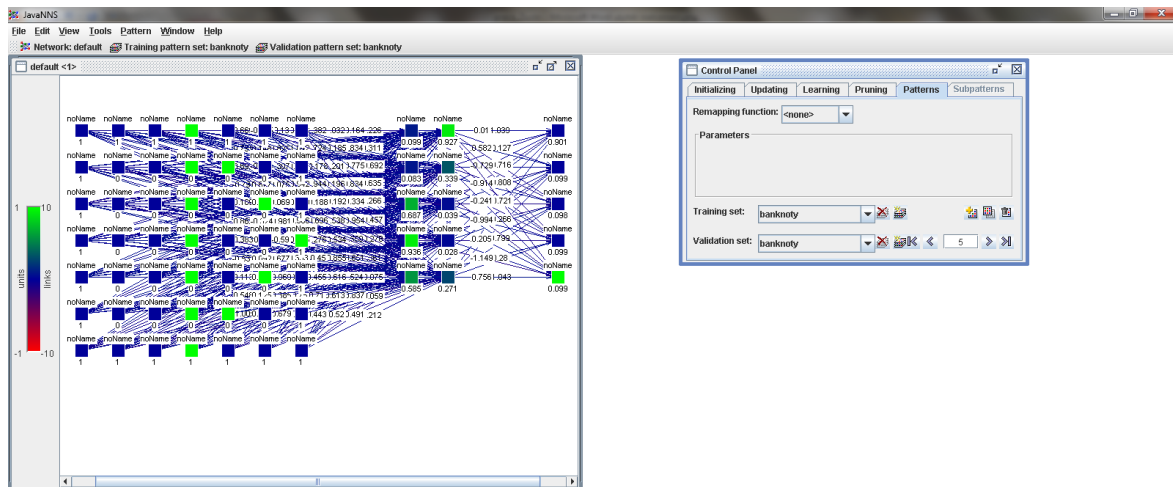
Rysunek 26. Wzór dla banknotu pięćdziesięciuzłotowego (Źródło: opracowanie własne)

4.3.3 Wzór dla 100 złotych



Rysunek 27. Wzór dla banknotu stużłotowego (Źródło: opracowanie własne)

4.3.4 Wzór dla 200 złotych



Rysunek 28. Wzór dla banknotu o nominalie dwieście złotych. (Źródło: opracowanie własne)

Zakończenie

W rozdziale pierwszym przedstawione zostały biologiczne podstawy sztucznych sieci neuronowych. W rozdziale drugim zaprezentowano przykładowe modele matematyczne sztucznego neuronu i sieci neuronowych. W kolejnym rozdziale autor skupił się na wykazaniu metod uczenia sztucznych sieci neuronowych ze szczególnym uwzględnieniem metody wstecznej propagacji błędów. W ostatnim, czwartym rozdziale w kolejnych podrozdziałach zademonstrowano proces tworzenia sieci neuronowej, a następnie przebieg uczenia tej sieci rozpoznawania znaków na banknotach oraz wyniki uczenia. Czynności te zostały wykonane w programie Java Neural Networks Simulator.

Literatura

- M. Barczyński :Domowy poradnik medyczny, Państwowy Zakład Wydawnictw Lekarskich, Warszawa 1991
- D. Benzaia, S. Carleton , C. Caruthers : Rodzinna encyklopedia zdrowia, Reader's Digest Przegląd, Sp. z o.o, Warszawa 1999
- L. Bożydar, J. Kaczmarek: Mózg, język, zachowanie, Wydawnictwo UMCS Lublin 1998
- A. Grabowska, W. Buhodowska: Dwie półkule – jeden mózg, Wiedza powszechna, Warszawa 1994
- R. A. Kosiński. Wydawnictwa Naukowo-Techniczne, Warszawa 2002-2006
- M. Krause: Człowiek i jego układ nerwowy, Śląsk Sp. z o.o Wydawnictwo Naukowe, Katowice 2002
- J. Mańdziuk: Sieci neuronowe typu Hopfielda. Teoria i przykłady zastosowań, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2000
- T. Masters –„Sieci neuronowe w praktyce: programowanie w języku C++ , Wydawnictwa Naukowo Techniczne, Warszawa 1993
- S. Osowski: Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2006
- B. Pincon –„Wprowadzenie do Scilaba”
<http://www.iecn.u-nancy.fr/~szulc/docpl.pdf> (27.05.2007)
- L. Rutkowski: Metody i techniki sztucznej inteligencji, Wydawnictwo Naukowe PWN, Warszawa 2006
- Eldra P. Solomon, Linda R. Berg, Diana W. Martin: Biologia wg VII wydania amerykańskiego, MULTICO Oficyna Wydawnicza, Warszawa 2007
- Statsoft, Wprowadzenie do sieci neuronowych, Statsoft Polska Sp. z o.o, Kraków 2001
- StatSoft, Inc., Internetowy podręcznik statystyki, 1984-2011
http://www.statsoft.pl/textbook/stathome_stat.html?http%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstneunet.html
- L. Ziemiański: Sieci neuronowe w dynamice konstrukcji- wybrane zagadnienia, Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów 1999