

Stanisław PŁACZEK*

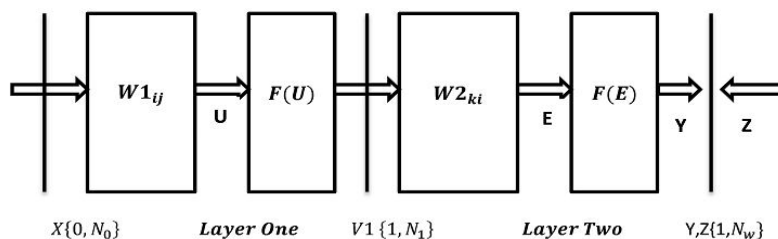
DEKOMPOZYCJA HIERARCHICZNEJ STRUKTURY SZTUCZNEJ SIECI NEURONOWEJ I ALGORYTM KOORDYNACJI

W artykule zaproponowano przeprowadzenie dekompozycji struktury sieci na dwie warstwy. W warstwie I poziomu znajduje się N_1 niepowiązanych podsieci. Natomiast w warstwie II poziomu (nadrzędnej) znajduje się podsieć warstwy ukrytej. Warstwy te powiązane są sygnałami V_1 , V_2 , które pozwalają na zastosowanie niezależnych algorytmów uczenia dla warstwy I oraz II. Prosty algorytm koordynacji umożliwia obliczenie wartości sygnałów między warstwowymi, a tym samym osiągnięcie minimum globalnej funkcji celu.

SŁOWA KLUCZOWE: Sztuczne sieci neuronowe, algorytm uczenia, dekompozycja, koordynacja, hierarchia

1. DEKOMPOZYCJA STRUKTURY SZTUCZNEJ SIECI NEURONOWEJ

Rozwiązując praktyczne zadania klasyfikacji, aproksymacji, predykcji, identyfikacji, a nawet sterowania procesami dynamicznymi stosuje się Sztuczne Sieci Neuronowe (SSN) z jedną lub dwoma warstwami ukrytymi.



Rys. 1. Uproszczony schemat standardowej struktury SSN z jedną warstwą ukrytą

Czas uczenia sieci powinien być maksymalnie krótki. Dlatego też stosuje się różne modyfikacje algorytmu wstecznej propagacji błędów.

Również przyspieszenie obliczeń można osiągnąć poprzez maksymalne zrównoleżenie wewnętrznej struktury algorytmów i wykorzystać dostępne

* Akademia Finansów i Biznesu Vistula w Warszawie.

narzędzia programistyczne związane z wielowątkowością. Z algorytmicznego punktu widzenia uczenie SSN to mnożenie macierzy dla procesu w przód i obliczenia błędu uczenia SSN $\psi(W1, W2)$. Dla procesu wstecz oblicza się macierze pochodnych błędu względem wag macierzy W1 oraz W2 (rys. 1).

Rozpatrzono SSN z warstwą wejściową, z jedną warstwą ukrytą oraz wyjściową. Błąd uczenia sieci oblicza się poprzez wielokrotne mnożenie macierzy [2].

$$U = W1 * X \quad (1)$$

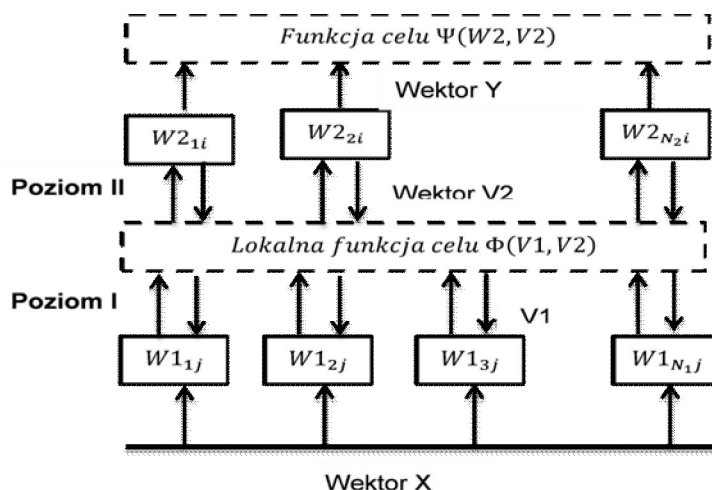
$$V = F(U) \quad (2)$$

$$E = W2 * V \quad (3)$$

$$Y = F(E) \quad (4)$$

$$\psi = \frac{1}{2} * (Y - Z)^T * (Y - Z) \quad (5)$$

Oznaczenia wszystkich zmiennych i macierzy pokazano na rys. 2.



Rys. 2. Schemat dekompozycji SSN na dwie warstwy

Wstępna analiza struktury sieci wskazuje na istnienie struktury hierarchicznej. Można wydzielić warstwę pierwszą zawartą pomiędzy wektorem X a wektorem V . Wektor X reprezentuje warstwę wejściową, w której można dokonać różnych przekształceń pierwotnego wektora X , jak filtrację, normalizację i inne. Warstwa druga, to warstwa ukryta reprezentowana przez macierz $W2$.

Dekompozycja polega na odseparowaniu warstwy drugiej, zwanej warstwą drugiego poziomu, od warstwy pierwszej (warstwa dolnego poziomu) poprzez wprowadzenie dwóch dodatkowych wektorów $V1$ oraz $V2$. Wprowadzono więc dwa poziomy sieci.

Dodatkowo, analizując strukturę warstw I oraz II poziomu z punktu widzenia wyjścia każdej warstwy, można rozbić warstwę I poziomu na N_1 podsieci ze wspólnym

wektorem wejściowym X. Warstwę drugiego poziomu, można rozbić na N_2 podsieci, ze wspólnym wektorem wejścia V2 (Rys. 2). Jednocześnie zdefiniowano dwie funkcje celu. Dla warstwy II jest to standardowa funkcja celu procesu uczenia, czyli błąd średnio kwadratowy ψ . Dla warstwy drugiej, również zdefiniowano błąd średnio kwadratowy ϕ dla wektorów V2, V1. W celu powiązania powstałych w ten sposób podsieci, wprowadzono dwie funkcje wektorowe G i H wiążące wektory V1, V2.

Dla procesu w przód

$$V2 = G(V1) \tag{6}$$

Dla procesu wstecz

$$V1 = H(V2) \tag{7}$$

Globalną i lokalną funkcje celu zdefiniowano jako

$$\psi(W2, V2) = \sum_{k=1}^{k=N_2} \psi_k = \sum_{k=1}^{k=N_2} (y_k - z_k)^2 \tag{8}$$

gdzie:

$$y_k = f \left[\sum_{i=0}^{i=N_1} W2_{ki} * V2_i \right] \tag{9}$$

$$k = 1, 2, \dots, N_2 \tag{10}$$

$$\phi = \sum_{i=1}^{i=N_1} \phi_i = \sum_{i=1}^{i=N_1} (v1_i - v2_i)^2 \tag{11}$$

gdzie:

$$v1_i = f \left[\sum_{j=0}^{j=N_0} W1_{ij} * x_j \right] \tag{12}$$

$$i = 1, 2, \dots, N_1 \tag{13}$$

W zdekomponowanej SSN, w każdej z warstw można wydzielić lokalne funkcje celu podsieci zgodnie z wzorami (8) i (11). We wzorach tych $z_k, v2_i$ występują jako parametry zadane z zewnątrz. Funkcje celu I oraz II poziomu mają postać addytywną. Ta zależność umożliwi dalsze uproszczenie algorytmu uczenia sieci.

2. STRUKTURA ZDEKOMPONOWANEGO ALGORYTMU UCZENIA

Adekwatnie do zdekomponowanej struktury SSN na dwie warstwy, można również zdekomponować standardowy algorytm uczenia oparty na wstecznej propagacji błędu i wydzielić[1]:

Zadanie II poziomu

W zadaniu tym poszukiwano minimum globalnej funkcji celu ψ . Ponieważ funkcja ta ma postać addytywną, to zgodnie z (8), jedno zadanie II poziomu można rozbić na N_2 podzadań poszukiwania minimum ψ_k .

Zadania te nie posiadają żadnych ograniczeń nierównościowych.

$$\min \psi_k = \left(f \left[\sum_{i=0}^{i=N_1} W_{2ki} * V_{2i} \right] - z_k \right) \quad (14)$$

Pochodne funkcji celu względem parametrów W_{2ki} , V_{2i}

$$\frac{\partial \psi_k}{\partial W_{2ki}} = (y_k - z_k) * \frac{\partial f}{\partial e_k} * v_{2i} \quad (15)$$

$$\frac{\partial \psi_k}{\partial v_{2i}} = \sum_{k=1}^{k=N_2} (y_k - z_k) * \frac{\partial f}{\partial e_k} * w_{2ki} \quad (16)$$

Nowe wartości w kolejnej iteracji algorytmu II poziomu

$$W_{2ki}(n+1) = w_{2ki}(n) - \alpha_2 * \frac{\partial \psi_k}{\partial W_{2ki}} \quad (17)$$

$$v_{2i}(n+1) = v_{2i}(n) - \gamma_2 * \frac{\partial \psi_k}{\partial v_{2i}} \quad (18)$$

Po obliczeniu v_{2i} przez wszystkie procedury II poziomu, wektor V_2 jako parametr jest przekazywany do koordynatora. Koordynator mając zdefiniowaną funkcję H oblicza nową wartość V_2 , którą przekazuje do podzadań I poziomu.

Zadanie I poziomu

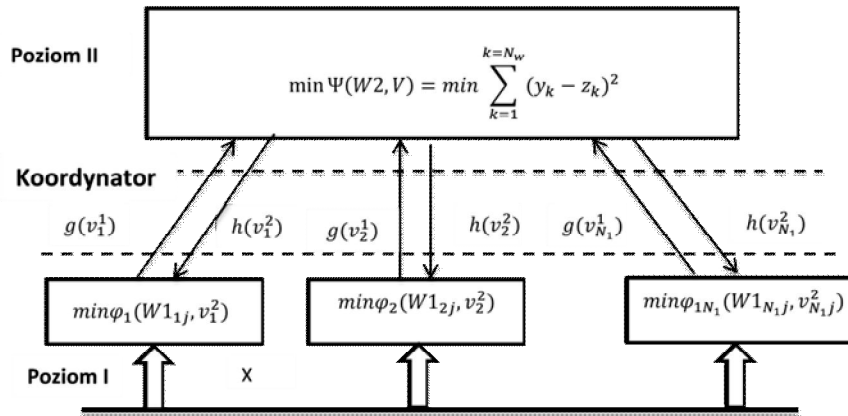
W zadaniu tym poszukiwano minimum lokalnej funkcji celu (11). Funkcja ta również ma postać addytywną, co pozwala na sformułowanie N_1 podzadań I poziomu.

$$\min \phi_i = \left(f \left[\sum_{j=0}^{j=N_0} W_{1ij} * x_j \right] - v_{2i} \right)^2 \quad (19)$$

Pochodne oraz nowe wartości W_{1ij} obliczono wg wzorów:

$$\frac{\partial \phi_i}{\partial W_{1ij}} = (v_{1i} - v_{2i}) * \frac{\partial f}{\partial e_k} * x_j \quad (20)$$

$$w_{1ij}(n+1) = w_{1ij}(n) - \alpha_1 * \frac{\partial \phi_i}{\partial W_{1ij}} \quad (21)$$



Rys. 3. Schemat procesu koordynacji algorytmu uczenia sieci typu on – line

W podzadaniach I poziomu składowe wektora V_2 otrzymane od koordynatora są parametrem dla lokalnej funkcji celu. Schemat procesu koordynacji przedstawiono na (rys. 3).

3. DEKOMPOZYCJA I KOORDYNACJA W ZADANIU KLASYFIKACJI

Zadanie polega na nauczaniu sieci rozpoznawanie funkcji XOR. Nie jest to funkcja bardzo skomplikowana lecz naszym celem jest poznanie wewnętrznych charakterystyk procesu uczenia dla podsieci I oraz II poziomu.



Rys. 4. Rezultat uczenia sieci

Zastosowano SSN z dwoma neuronami wejściowymi, trzy neurony w warstwie ukrytej i jeden w wyjściowej. Próbką ucząca zawierała cztery składowe wektorów X i Y .

Na pierwszym i drugim poziomie sieci przyjęto sigmoidalną funkcję aktywacji. Zastosowano algorytm typu on – line. Uaktualnianie wag zarówno na I oraz II poziomie odbywało się po każdorazowej prezentacji próbki uczącej. Główne parametry procesów uczenia to:

α - współczynnik uczenia się dla macierzy W1, W2.

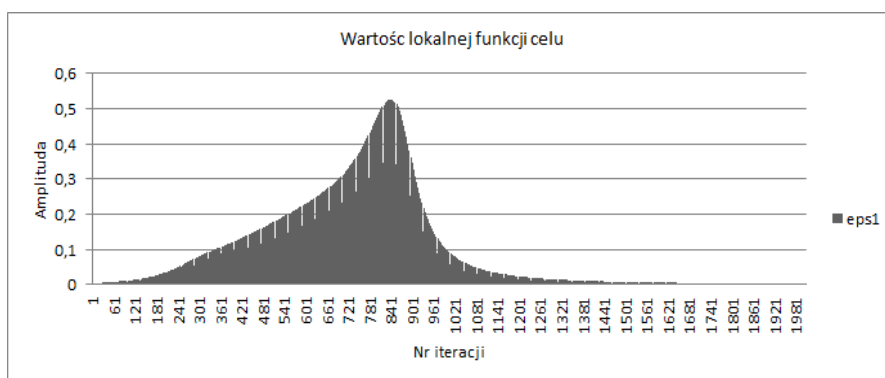
β - parametr regulatyzacyjny.

γ - współczynnik uczenia dla wektora V2 wejścia podsieci II poziomu.

ϵ_1 - wartość lokalnej funkcji celu.

ϵ_2 - wartość globalnej funkcji celu.

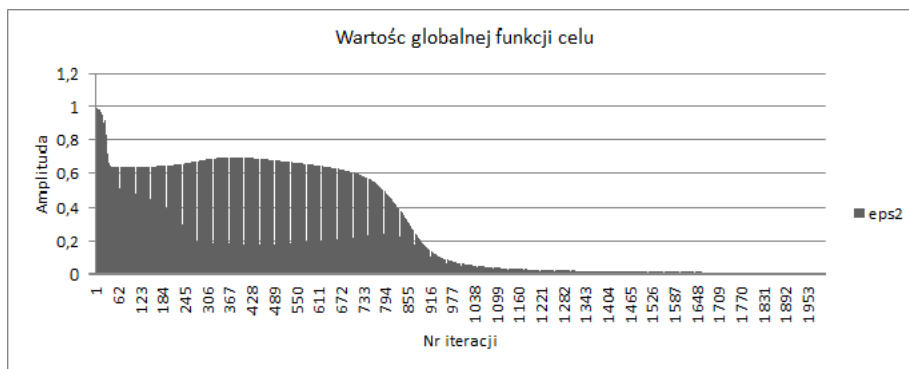
λ_1, λ_2 - parametry koordynatora dla funkcji G, H.



Rys. 5. Zależność wartości lokalnej funkcji celu od nr iteracji dla $\alpha = 0,7$, $b = 0,01$, $\gamma = 0,6$

Na rysunku 4 przedstawiono wynik uczenia sieci w oparciu o dane testowe. Ciemnym kolorem pokazano teoretyczne wartości funkcji XOR, natomiast jasny obrazuje wynik uczenia. Obarczony jest minimalnym błędem. Natomiast (rys. 5.) przedstawia zależność wartości lokalnej funkcji celu I poziomu od nr iteracji dla uczenia metodą on-line. Widoczne są maksymalne jak i minimalne wartości (ciemne pole) w poszczególnych iteracjach. Na pierwszym etapie podsieć II poziomu wymusza na funkcji celu I poziomu szybki wzrost. Wymusza w ten sposób przyspieszenie uczenia podsieci I poziomu, czyli obliczenie nowej optymalnej wartości macierzy W1. Następny etap procesu uczenia to asymptotyczne dążenie do wartości minimalnej synchronicznie z wartością globalnej funkcji celu (rys. 6.)

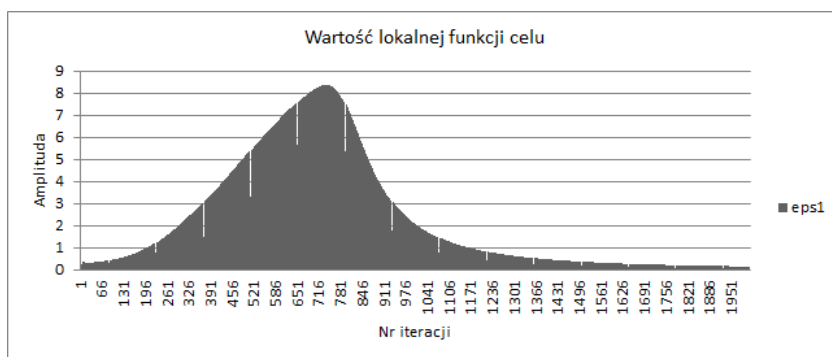
Jakościowa charakterystyka procesu uczenia podsieci II poziomu jest inna. Posiada również dwie wartości amplitudy uczenia - minimalną i maksymalną (ciemne pole). Jednak jakościowo proces uczenia dla II poziomu, różni się w sposób zasadniczy od procesu uczenia I poziomu. Maksymalna wartość przez pewien okres nie maleje, jakby czekała na podsieci I poziomu, które muszą zmniejszyć wartość lokalnej funkcji celu. Następnie asymptotycznie i synchronicznie z I poziomem podsieci, maleje do minimalnej wartości.



Rys. 6. Zależność wartości globalnej funkcji celu od nr iteracji dla $\alpha = 0,7$, $b = 0,01$, $\gamma = 0,6$

4. PODSUMOWANIE I WNIOSKI

W artykule przedstawiono rezultaty uzyskane w wyniku uczenia zdekomponowanej dwuwarstwowej SSN wykorzystując algorytm on-line. W algorytmie koordynacji zastosowano funkcje G i H zerowego rzędu. Natomiast zmieniano wartości α oraz γ .



Rys. 7. Zależność wartości lokalnej funkcji celu od nr iteracji dla $\alpha = 0,3$, $b = 0,01$, $\gamma = 4$

Na rysunku 7 przedstawiono charakterystykę uczenia podsieci I poziomu dla dużej wartości $\gamma = 4$. Charakterystyka procesu pokazuje, że jest przeregulowany, osiągając duże wartości lokalnej funkcji celu ($\max \epsilon_1 = 8$). W tabeli 1 przedstawiono porównanie różnych parametrów wraz z jakościową charakterystyką procesu uczenia. Dobór optymalnych wartości α , β , γ , λ_1 , λ_2 dla danego zagadnienia jest tematem do dalszych badań i analiz. Pokazano rezultaty dla algorytmu typu on – line. Intuicyjnie wydaje się, że część charakterystyk powinna ulegnie wygładzeniu dla procesu off – line.

Tabela 1. Parametry ilościowo – jakościowe charakterystyk uczenia SSN

α	β	γ	λ_1	λ_2	$\epsilon_1 \max$	$\epsilon_2 \max$	Uwagi
0,7	0,01	0,6	0	0	0,45	0,7	b. dobrze
0,5	0,01	0,6	0	0	0,45	0,7	b. dobrze
0,5	0	0,8	0	0	0,7	0,6	dobrze
0,3	0	8	0	0	18	0,4	przeregulowany
0,3	0	4	0	0	4	0,5	akceptowalny

LITERATURA

- [1] W. Findeisen, J. Szymanowski, A. Wierzbicki, Teoria i metody obliczeniowe optymalizacji. Państwowe Wydawnictwo Naukowe, Warszawa, 1977.
- [2] S. Osowski, Sieci Neuronowe do Przetwarzania Informacji. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2006.
- [3] Ch. M. Bishop, Pattern Recognition and Machine Learning, Springer Science + Business Media, LLC, 2006.
- [4] Marciniak A., Korbicz J., Kus J., *Wstępne przetwarzanie danych*, Sieci Neuronowe tom 6, Akademicka Oficyna Wydawnicza EXIT, 2000.
- [5] Płaczek S., *Sztuczne Sieci Neuronowe*. Miesięcznik Ubezpieczeniowy tom 7, nr 9, wrzesień, 2010.

DECOMPOSITION OF HIERARCHICAL STRUCTURE OF ARTIFICIAL NEURAL NETWORK AND COORDINATION ALGORITHM

The article presents decomposition of Artificial Network Structure into two layers. Layer one (lower one) consist of N_1 independent sub layers. The second layer (upper one) is a hidden layer. Vectors V_1 and V_2 are introduced as coordinator between two layers. The coordinator uses different algorithms connecting vectors V_1 and V_2 . In this way, the coordinator is able to coordinate two independent learning algorithms for each layer. The coordination algorithm was described and final learning results are presented. Presented results of an on – line learning algorithm were used for both, the first and the second layer. For the future study, an off-line learning algorithm will be used.