ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2003
Seria: BUDOWNICTWO z. 101 Nr kol. 1595

Maria MRÓWCZYŃSKA*
Uniwersytet Zielonogórski

EFEKTYWNOŚĆ ALGORYTMU PROPAGACJI WSTECZNEJ BŁĘDU Z ZASTOSOWANIEM NIELINIOWYCH FUNKCJI AKTYWACJI DO APROKSYMACJI RZEŹBY TERENU

Streszczenie. W artykule przestawiono zagadnienia związane z procesem uczenia sieci neuronowych dwuwarstwowych metodą propagacji wstecznej błędu w zastosowaniu do aproksymacji powierzchni terenu. Podstawowym celem podjętej pracy jest badanie dokładności odwzorowania rzeźby terenu w zależności od zastosowanych nieliniowych funkcji aktywacji. Rozpatrywano również zagadnienia związane z doborem architektury sieci i parametrów uczenia oraz dokonano porównania wybranych metod przyspieszania zbieżności.

THE EFFICIENCY OF ALGORITHM OF BACK PROPAGATION WITH THE APPLICATION OF NONLINEAR ACTIVATION FUNCTIONS TO THE APPROXIMATION OF THE TOPOGRAPHIC PROFILE

Summary. The paper presents the issue of learning of the unidirectional and two-layer net by the means of the method of reverse propagation of the error applied to approximation of area surface, represented by a point pattern, which was determined in the configuration of space polar coordinates. The problem connected with the convergence, the speed of learning process convergence and the selection of the network architecture has been discussed in the paper as well.

1. Wstęp

Zagadnienie aproksymacji powierzchni terenu za pomocą sieci neuronowej polega na przybliżeniu funkcji wielu zmiennych $\varphi(x)$, określonej na dyskretnym zbiorze punktów, funkcją ciągłą $f(x,w)$, gdzie $x=[x_1,x_2,\ldots,x_n]^T$ jest wektorem wejściowym w postaci danych wynikających z pomiaru, zaś $w=[w_1,w_2,\ldots,w_m]^T$ ($w_1,\ldots,w_n \in \{0,1\}$), jest wektorem

* Opiekun naukowy: Prof. dr hab. inż. Józef Gil.

współczynników (wag). Rozwiązanie zadania polega na wyznaczeniu takiego wektora w^*, aby był spełniony warunek

$$d[f(w^*,x),\varphi(x)] \le d[f(w,x),\varphi(x)] \tag{1}$$

gdzie odległość $d[f(w,x),\varphi(x)]$ jest wskaźnikiem jakości przybliżenia funkcji $f(x,w)$ do $\varphi(x)$ w określonym przedziale $(a \le x \le b)$.

W niniejszej pracy zostanie przedstawiona próba aproksymacji powierzchni terenu metodą propagacji wstecznej błędu, za pomocą sieci neuronowej dwuwarstwowej przy zastosowaniu wybranych, ciągłych funkcji aktywacji.

2. Model sieci neuronowej

Podstawowym elementem sieci neuronowej, pełniącej w rozpatrywanym zagadnieniu rolę uniwersalnego aproksymatora funkcji wielu zmiennych, jest sztuczny neuron, złożony z wejść synaptycznych, węzła przetwarzającego oraz jednego wyjścia. Sygnał wyjściowy neuronu jest pewną funkcją liniowego sumatora ważonego, wyrażonego zależnością [6]

$$y = f(w_i^t x) \qquad (i = 1,2,\ldots,n) \tag{2}$$

gdzie $f(w_i^t x)$ jest nieliniową oraz ciągłą funkcją aktywacji, której dziedziną jest zbiór łącznych pobudzeń neuronu. Założenie ciągłej funkcji aktywacji pozwala przyjąć strategię doboru wag na podstawie gradientowych metod optymalizacji.

Funkcja celu, która stanowi podstawę algorytmu, jest definiowana jako suma kwadratów różnic pomiędzy wartościami na wyjściu: oczekiwaną d_k i aktualną z_k (k=1,2,...,K), czyli

$$E = \frac{1}{2}\sum_{k=1}^{K}(d_k - z_k)^2 \tag{3}$$

Biorąc pod uwagę sposób oszacowania dokładności wyników rozwiązywanych zadań w dziedzinie geodezji, dokładność aproksymacji wyrazimy w postaci odchylenia standardowego

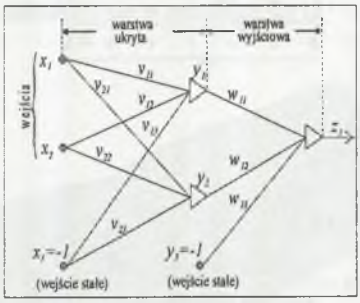
$$\sigma = \sqrt{\frac{[(d_k - z_k)^2]}{K-m}} \tag{4}$$

gdzie m oznacza ogólną liczbę wag połączeń synaptycznych.

Dla sieci dwuwarstwowej (rys. 1) sygnał wyjściowy wynosi

$$z = \Gamma[\mathbf{W}\mathbf{y}] = \Gamma[\mathbf{W}\Gamma[\mathbf{V}\mathbf{x}]] \tag{5}$$

gdzie **W** jest wektorem wag w warstwie wyjściowej, a **V** jest wektorem wag w warstwie ukrytej.



Rys. 1. Sieć dwuwarstwowa
Fig. 1. Two-layer lattice

3. Funkcje aktywacji

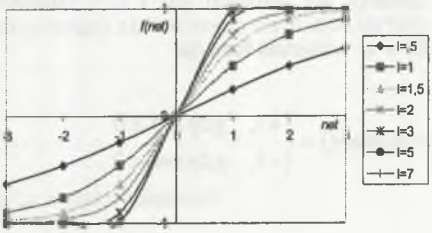
Do najczęściej stosowanych funkcji aktywacji należą [1,6]:

- bipolarna funkcja ciągła (rys. 2):

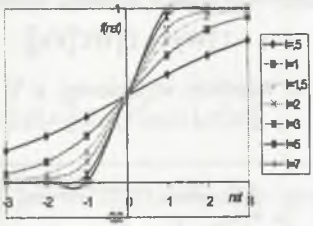
$$f(net) = \frac{2}{1+\exp(-\lambda net)} - 1 \tag{6}$$

- unipolarna funkcja ciągła (rys. 3):

$$f(net) = \frac{1}{1+\exp(-\lambda net)} \tag{7}$$



Rys. 2. Bipolarna funkcja ciągła
Fig. 2. Bipolar continuous function



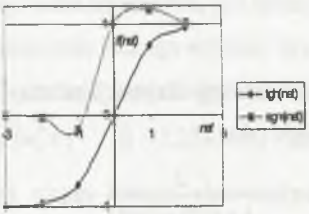
Rys. 3. Unipolarna funkcja ciągła
Fig. 3. Unipolar continuous function

- tangens hiperboliczny (rys. 4):

$$f(net) = \frac{\exp(net) - \exp(-net)}{\exp(net) + \exp(-net)} \tag{8}$$

- sigmoida (krzywa logistyczna) (rys. 4):

$$f(net) = \frac{1}{1 + \exp(-2net)}. \tag{9}$$



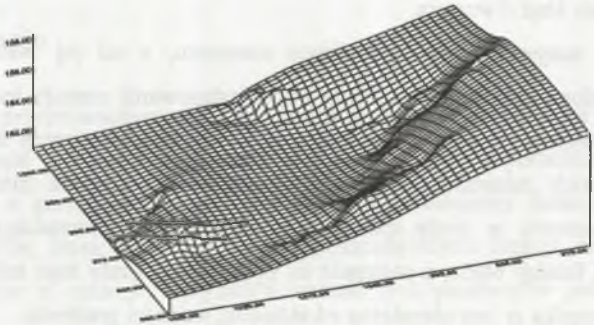
Rys. 4. Funkcje: tangens hiperboliczny i sigmoida
Fig. 4. Hyperbolic tangent and sigmoid functions

Parametr λ $(\lambda > 0)$, zawarty we wzorach (6) i (7), charakteryzuje nachylenie funkcji aktywacji f(net); jeżeli $\lambda \to \infty$ wówczas funkcja

$$f(net) = \mathrm{sgn}(net) = \begin{cases} +1, & \text{gdy } net \geq 0 \\ -1, & \text{gdy } net < 0 \end{cases}. \tag{10}$$

4. Przykład liczbowy

Odwzorowanie powierzchni terenu przeprowadzono na zbiorze uczącym, składającym się ze 177 punktów (x,y,z), generowanych jako punkty rozproszone w obrębie pewnego obszaru (rys. 5). Wysokości punktów zawierały się w granicach od 151 ÷ 158 m.



Rys. 5. Obraz powierzchni terenu uzyskany na podstawie pomiaru tachimetrycznego
Fig. 5. The picture of the area surface obtained on the basis of the tachymetry survey

Uczenie sieci metodą propagacji wstecznej błędu przeprowadzono stosując wymienione wcześniej funkcje aktywacji, przy czym aktualizacja wag następowała w sposób przyrostowy w kierunku ujemnego gradientu funkcji celu. Dane zostały sprowadzone do postaci bezwymiarowej z uwzględnieniem jednolitego zakresu zmienności według wzoru:

$$x_{norm} = \frac{x}{x_{max}} \tag{11}$$

W celu ograniczenia liczby iteracji zastosowano dwie metody przyspieszenia zbieżności, a mianowicie:

- metodę momentu,
- metodę różnicową.

Korektę wag metodą momentu wyraża zależność

$$\Delta w(k) = -\eta \nabla E_k + \alpha \Delta w(k-1) \tag{12}$$

gdzie:

η - współczynnik korekcji,

∇E_k - gradient funkcji błędu,

α - stała dodatnia $(0 < \alpha < 1)$,

k-1, k - poprzedni i aktualny krok uczenia.

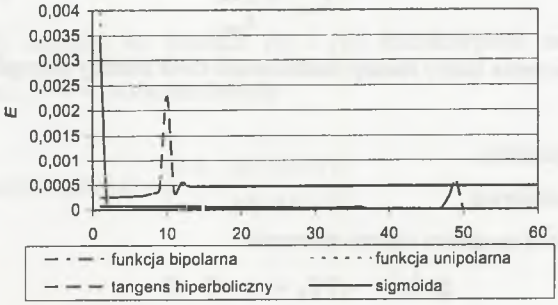
Współczynnik korekcji η dobierano eksperymentalnie do każdego problemu. Ogólnie, wartość współczynnika η powinna mieścić się w zakresie od 10^{-3} do 10. W procesie uczenia współczynnik η należy zwiększać o stałą wartość, gdy błąd E maleje, oraz zmniejszać geometrycznie, gdy błąd E wzrasta.

Stała α jest nazywana współczynnikiem momentu, a od jej wartości zależy siła oddziaływania członu momentu na korektę wag. Zastosowanie metody momentu powoduje nie tylko przyspieszenie procesu uczenia, ale ma jednocześnie korzystny wpływ na problem wystąpienia płytkich minimów lokalnych. W strefie przyciągania minimum lokalnego współczynnik momentu α może spowodować zmianę wag, prowadzącą do chwilowego wzrostu wartości funkcji błędu, a następnie do opuszczenia strefy tego minimum, ponieważ wartość współczynnika α jest niezależna od aktualnej wartości gradientu.

Drugą dość skuteczną metodą przyspieszania zbieżności jest metoda różnicowa, która polega na zmianie w niektórych iteracjach (np. wielokrotność liczby 3) kierunku $-\nabla E_k$ na kierunek

$$\Delta w = w_k - w_{k-2}. \tag{13}$$

Decyzja tego rodzaju wynika stąd, że kierunek przeciwny do kierunku gradientu funkcji błędu może być najlepszym kierunkiem lokalnym, ale nie musi być najlepszym kierunkiem globalnym. Wyniki uczenia dla poszczególnych funkcji aktywacji zostały przedstawione w postaci wartości błędów łącznych E zamieszczonych w tabeli 1 i zilustrowanych graficznie na rys. 6.



Rys. 6. Zmiany wartości błędu łącznego w zależności od zastosowanych funkcji aktywacji
Fig. 6. The changes of values of associative error depending on the applied activation functions

Tabela 1

Lp.	Zastosowana funkcja aktywacji	Wartość błędu łącznego σ [m]
1	bipolarna funkcja ciągła	0,50
2	tangens hiperboliczny	0,50
3	unipolarna funkcja ciągła	0,66
4	sigmoida (krzywa logistyczna)	0,67

5. Podsumowanie

Na podstawie przeprowadzonych badań można stwierdzić, że najlepsze efekty uzyskamy stosując w procesie uczenia jako funkcję aktywacji tangens hiperboliczny oraz funkcję bipolarną ciągłą, a gorsze efekty otrzymamy, jeżeli zastosujemy funkcję unipolarną lub funkcję sigmoidalną. Poza tym zastosowanie poszczególnych funkcji aktywacji wymaga doboru parametrów η oraz λ. Optymalna wartość tych parametrów jest uzależniona od rozwiązywanego zagadnienia i jest różna dla różnych funkcji aktywacji. Zdaniem autora, dla funkcji unipolarnej parametr λ powinien być dwukrotnie większy aniżeli w przypadku stosowania funkcji bipolarnej, a stosując funkcję aktywacji o postaci sigmoidalnej o połowę mniejszy.

W celu zwiększenia efektywności uczenia zastosowano metody przyspieszania zbieżności do rozwiązania. Z omówionych wcześniej metod najlepsze wyniki otrzymano stosując metodę momentu. Zastosowanie tej metody zmniejsza dwukrotnie liczbę iteracji w porównaniu z liczbą iteracji z pominięciem wymienionych metod realizacji numerycznej postawionego zadania.

Na zakończenie należy dodać, że wyniki uczenia byłyby o wiele lepsze, gdyby zbiór danych wejściowych został poddany klasyfikacji.

LITERATURA

1. Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R.: Sieci neuronowe. Akademicka Oficyna Wydawnicza, Warszawa 2000.
2. Gaździcki J.: Systemy informacji przestrzennej. PPWK, Warszawa 1990.

3. Mrówczyńska M.: Efektywność i zbieżność metody propagacji wstecznej błędu w zastosowaniu do aproksymacji powierzchni terenu. VI Konferencja Naukowo - Techniczna „Problemy automatyzacji w geodezji". Warszawa 27 – 28 marca 2003.
4. Osowski S.: Sieci neuronowe do przetwarzania informacji. Oficyna Wydawnicza PW, Warszawa 2000.
5. Urbański J.: Zrozumieć GIS. Wydawnictwo Naukowe PWN, Warszawa 1997.
6. Żurada J., Barski M., Jędruch W.: Sztuczne sieci neuronowe. Podstawy teorii i zastosowania. Wydawnictwo Naukowe PWN, Warszawa 1996.

Recenzent: Prof. dr hab. inż. Andrzej Majde

Abstract

On the basis of investigations it was found that the best results in learn process was obtained with a hyperbolic tangent and a bipolar continuous function as an activation function. By unipolar or sigmoidaln function results was worse. An application of anyone of activation functions requires parameters choice η and λ. An optimal value of this parameters is conditional on solving problem and is different for different activation functions. In authors opinion, for unipolar function λ parameter should be twice greater than for bipolar function and for sigmoidaln function twice lower.