

Maria MRÓWCZYŃSKA*
Uniwersytet Zielonogórski

SIECI NEURONOWE O RADIALNYCH FUNKCJACH BAZOWYCH W ZASTOSOWANIU DO APROKSYMACJI RZEŻBY TERENU

Streszczenie. W artykule podjęto próbę przedstawienia wybranych zagadnień dotyczących aproksymacji rzeźby terenu za pomocą sieci neuronowych o radialnych funkcjach bazowych. Uniwersalna zgodność sieci z bazą funkcji radialnych jest w istotny sposób zależna od doboru wektorów c_1, \dots, c_K , nazywanych centrami, oraz od doboru parametrów $\sigma_1, \dots, \sigma_K$ kształtu funkcji bazowych. Praca zawiera opis struktury sieci, metodę wyznaczania wag z wykorzystaniem pojęcia pseudoodwrotności macierzy Greena oraz kilka sposobów doboru centrów. Rezultat podjętej pracy stanowi przykład liczbowy funkcjonowania sieci w zależności od proponowanych metod uczenia.

THE APPLICATION OF THE RADIAL BASIS NETWORKS TO APPROXIMATION OF THE TOPOGRAPHIC PROFILE

Summary. The article discusses the chosen issues concerning the approximation of the topographic profile which have been presented by means of the radial basis networks. The universal conformity of the networks to the radial functions base depends essentially on the choice of vectors c_1, \dots, c_K , which are called centres, and also depends on the choice of parameters $\sigma_1, \dots, \sigma_K$ of the basis functions shape. The paper contains the method of observations weights determination using the notion of pseudoinverse of Green's matrixes and some ways of selecting the centres.

1. Wstęp

Sieci neuronowe sigmoidalne wielowarstwowe z matematycznego punktu widzenia pełnią rolę aproksymacji stochastycznej funkcji wielu zmiennych, odwzorowując zbiór zmiennych wejściowych $x \in R^N$ w zbiór zmiennych wyjściowych $y \in R^M$ [2]. Odwzorowanie zbioru wejściowego w zbiór wyjściowy jest wykonywane poprzez dopasowanie funkcji

aproxymującej wielu zmiennych do wartości zadanych, czyli rozciągnięcie nad zbiorem uczącym hiperpłaszczyzny wielowymiarowej najlepiej dopasowującej się do wektora zadanego.

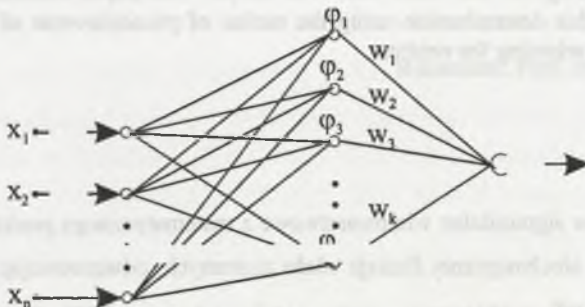
Sieci neuronowe, w których neuron ukryty realizuje funkcję zmieniającą się wokół wybranego centrum c , są to sieci o radialnych funkcjach bazowych $\varphi = (\|x - c\|)$. Rola neuronu ukrytego sprowadza się wówczas do radialnego odwzorowania przestrzeni wokół jednego zadanego punktu lub grupy takich punktów stanowiących klastery.

2. Struktura sieci neuronowej radialnej

Udowodniono [2], że przyjęcie dostatecznie dużej liczby neuronów ukrytych reprezentujących funkcje radialne $\varphi_i(x)$ zapewnia rozwiązanie przy użyciu jedynie dwóch warstw sieci: warstwy ukrytej odzwierciedlającej wektor funkcji $\varphi(x)$ i warstwy wyjściowej o jednym neuronie, którego sygnał jest funkcją liniowego sumatora ważonego.

Najprostsza sieć neuronowa typu radialnego (rys. 1) działa na zasadzie wielowymiarowej interpolacji, polegającej na przyjęciu p neuronów ukrytych typu radialnego i określeniu takiej funkcji odwzorowania $F(x)$, dla której są spełnione warunki interpolacji

$$F(x_i) = d_i. \quad (1)$$



Rys. 1. Ogólna postać sieci radialnej
Fig. 1. The general radial network form

Własności generalizacyjne sieci ulegają degradacji z chwilą przyjęcia liczby funkcji radialnych równej liczbie danych uczących. Przy takim toku postępowania, czyli przy założeniu, że liczba centrów jest równa liczbie współrzędnych wektora wejściowego ($c_i = x_i, i = 1, \dots, p$), model dopasowuje się do danych uczących, bowiem liczba stopni swobody układu (liczba równań – liczba niewiadomych)

$$\begin{bmatrix} \varphi_{11} & \varphi_{12} & \dots & \varphi_{1K} \\ \varphi_{21} & \varphi_{22} & \dots & \varphi_{2K} \\ \dots & \dots & \dots & \dots \\ \varphi_{p1} & \varphi_{p2} & \dots & \varphi_{pK} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_K \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \dots \\ d_p \end{bmatrix} \tag{2}$$

wynosi zero. W związku z powyższym, własności generalizacyjne sieci uzyskuje się przy spełnieniu warunku $K < p$, gdzie K oznacza liczbę centrów $c_i (i = 1, 2, \dots, K)$, a p oznacza liczbę wzorców uczących.

Zadanie aproksymacji polega na dobraniu odpowiedniej liczby parametrów funkcji radialnych $\varphi(\|x - c\|)$ i na takim doborze wag w_i , aby zminimalizować funkcję celu w postaci

$$E = \sum_{i=1}^p \left[\sum_{j=1}^K w_j \varphi(\|x_i - c_j\| - d_i) \right]^2 \tag{3}$$

W równaniu (3) symbol K reprezentuje liczbę neuronów ukrytych, a symbol p liczbę par uczących (x, d) . Wektor x jest wektorem wejściowym, natomiast wektor d jest wektorem zadany. Przy założeniu znajomości parametrów funkcji radialnych minimalizacja funkcji celu sprowadza się do rozwiązania układu równań liniowych względem wektora wag w w relacji

$$Gw = d, \tag{4}$$

gdzie G jest macierzą, zawierającą wartości radialnych funkcji bazowych, nazywaną macierzą Greena w postaci:

$$G = \begin{bmatrix} \varphi = (\|x_1 - c_1\|) & \varphi = (\|x_1 - c_2\|) & \dots & \varphi = (\|x_1 - c_K\|) \\ \varphi = (\|x_2 - c_1\|) & \varphi = (\|x_2 - c_2\|) & \dots & \varphi = (\|x_2 - c_K\|) \\ \dots & \dots & \dots & \dots \\ \varphi = (\|x_p - c_1\|) & \varphi = (\|x_p - c_2\|) & \dots & \varphi = (\|x_p - c_K\|) \end{bmatrix}$$

Wektor wag w wyznacza się wykorzystując pojęcie pseudoodwrotności G^+ macierzy G jako

$$w = G^+ d. \quad (5)$$

Jednym ze sposobów wyznaczenia pseudoodwrotności G^+ macierzy G jest metoda ortogonalizacji Grama – Schmidta, w której macierz G jest rozkładana na macierze Q i R [1]. Macierz Q jest macierzą o kolumnach ortonormalnych o wymiarach $p \times K$, natomiast macierz R jest macierzą trójkątną górną o wymiarach $K \times K$. Stosując tę metodę, pseudoodwrotność wyznacza się na podstawie zależności

$$G^+ = R^{-1} Q^T \quad (6)$$

Architektura sieci radialnych ma strukturę analogiczną do struktury wielowarstwowej sieci neuronowej o jednej warstwie ukrytej i liniowym neuronie wyjściowym. Argumentem funkcji radialnej jest odległość sygnału wejściowego x_i od centrum c_i , natomiast rolę neuronów ukrytych odgrywają radialne funkcje bazowe. Najczęściej stosowaną funkcją bazową w sieciach radialnych jest funkcja bazowa Gaussa, która przy przyjęciu centrum w punkcie c_i ma postać

$$\varphi(x) = \varphi(\|x - c_i\|) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right) \quad (7)$$

gdzie δ_i jest parametrem decydującym o szerokości funkcji bazowej.

3. Metody uczenia sieci neuronowych radialnych

Dla k radialnych funkcji bazowych rozwiązanie zadania aproksymacji można przedstawić w postaci:

$$F(x) = \sum_i^k w_i \varphi(\|x - c_i\|) \quad (8)$$

Skuteczny model dowolnej funkcji utworzonej przez radialne funkcje bazowe wymaga doboru odpowiedniej liczby parametrów c_i i σ_i funkcji radialnych oraz wag w_i ($1, 2, \dots, K$).

Ustalonym położeniem centroidów c_i , należy przyporządkować wartości parametrów σ_i , które decydują o kształcie funkcji i wielkości pola recepcyjnego. Liczba centroidów c_i oraz właściwy dobór parametrów σ_i , pozostaje najbardziej istotnym zagadnieniem w aspekcie dokładności aproksymacji.

Można wyróżnić następujące metody doboru centrów:

- metoda K – means,
- metoda hybrydowa,
- metoda oparta na propagacji wstecznej błędu.

W algorytmie samoorganizacji proces uczenia rozpoczynamy od podziału zbioru uczącego na klastry. Podziału tego można dokonać metodą K – means. Adaptacja centrów w n -tej iteracji odbywa się zgodnie z zasadą:

$$c_i(n+1) = \frac{1}{M_i} \sum_{j=1}^{M_i} x_j(n) \quad (9)$$

gdzie M_i oznacza liczbę wektorów $x_j(n)$ przydzielonych do i -tego centrum w n -tej iteracji. Adaptacja wartości wszystkich centrów odbywa się równocześnie, a po ustaleniu ostatecznego ich położenia odbywa się dobór wartości parametrów σ_i , odpowiadających poszczególnym funkcjom bazowym.

Proces uczenia przy zastosowaniu algorytmu hybrydowego dzieli się na dwa etapy. W pierwszym z nich następuje dobór parametrów liniowych sieci (wag warstwy wyjściowej) przy zastosowaniu techniki pseudoinwersji (por. wzór (5)). W drugim etapie adaptacji podlegają parametry nieliniowe funkcji radialnych przy zastosowaniu metody największego spadku, zgodnie z zasadą

$$c_i(n+1) = c_i(n) - \eta \nabla E(c_i) \quad (10)$$

$$\sigma_i(n+1) = \sigma_i(n) - \eta \nabla E(\sigma_i). \quad (11)$$

Korekta parametrów nieliniowych sieci kończy jeden cykl uczący, a wielokrotne powtórzenie obu etapów pozwala na szybkie nauczenie sieci, zwłaszcza wówczas gdy początkowe wartości c_i oraz σ_i były bliskie optymalnym.

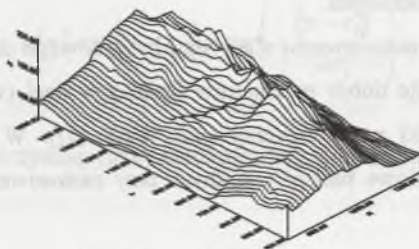
W metodach opartych na algorytmie propagacji wstecznej błędu do optymalizacji funkcja celu, definiowanej jako

$$E = \sum_{i=1}^p \left[\sum_{j=1}^k w_j \varphi_j(x_i) - d_i \right]^2 \quad (12)$$

stosowane są metody gradientowe. Niezależnie od przyjętej metody optymalizacji istotnym elementem procedury jest wyznaczenie współrzędnych wektora gradientu funkcji celu względem wszystkich parametrów sieci. Gradient funkcji celu może być wyznaczony za pomocą metody grafów przepływowych, która pozwala na określenie dowolnej składowej gradientu funkcji za pomocą grafu oryginalnego i dołączonego [2,3]. Wyznaczenie składowych gradientu funkcji celu pozwala skorzystać z dowolnej gradientowej metody optymalizacji przy aktualizacji parametrów sieci c , oraz σ , w kolejnych cyklach uczenia.

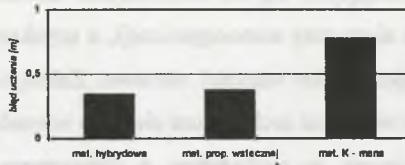
4. Przykład liczbowy

Odwzorowanie powierzchni terenu przeprowadzono na podstawie zbioru 2000 punktów (x,y,z) , generowanych jako punkty rozproszone (rys. 2), których wysokości zawierały się w granicach od 80 ÷ 210 m.



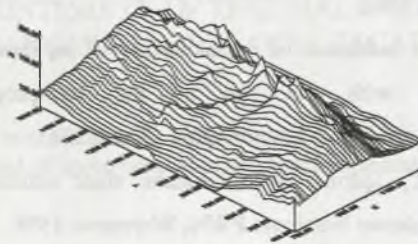
Rys. 2. Obraz powierzchni terenu uzyskany na podstawie pomiaru tachymetrycznego
Fig. 2. The picture of area surface obtained by the tachymetrial surveying

Wyniki uczenia sieci neuronowej o radialnych funkcjach bazowych w zależności od zastosowanych metod uczenia przedstawia rys. 3.



Rys. 3. Wyniki uczenia sieci radialnej w zależności od zastosowanej metody uczenia
 Fig. 3. The results of the radial network learning depending on the applied learning method

Najlepszy wynik uczenia sieci radialnej uzyskano przy zastosowaniu algorytmu hybrydowego, dla którego błąd łączny wyniósł $E=0,34$ m, natomiast najgorszy $E=0,67$ m przy wykorzystaniu metody probabilistycznego doboru centrów. Zbiór uczący składał się 1500 punktów, pozostałe 500 punktów to punkty testowe, liczba neuronów w warstwie ukrytej wyniosła 20. Obraz powierzchni terenu uzyskany za pomocą metody hybrydowej przedstawiono na rys. 4.



Rys. 4. Obraz powierzchni terenu uzyskany metodą hybrydową
 Fig. 4. The picture of the area surface obtained by the hybrid method

5. Podsumowanie

Przy rozwiązywaniu zagadnień za pomocą sieci neuronowych radialnych podstawowym problemem jest odpowiedni dobór parametrów startowych. Zastosowanie losowego wyboru tych parametrów zwiększa prawdopodobieństwo zatrzymania się procesu uczenia w minimum lokalnym. Dlatego korzystniejszy jest wybór parametrów startowych przy zastosowaniu procedur opierających się na informacjach zawartych w zbiorze uczącym, wykorzystując do tego celu algorytmy samoorganizacji, a uzyskane w ten sposób parametry funkcji radialnych przyjmujemy jako wartości startowe. Zaletą sieci radialnych jest prosty

zastosowaniu procedur opierających się na informacjach zawartych w zbiorze uczącym, wykorzystując do tego celu algorytmy samoorganizacji, a uzyskane w ten sposób parametry funkcji radialnych przyjmujemy jako wartości startowe. Zaletą sieci radialnych jest prosty algorytm uczący oraz ściśle określona architektura sieci, co warunkuje, że punkt startowy jest bliżej rozwiązania optymalnego w porównaniu z algorytmem stosowanym w sieciach sigmoidalnych. Ponadto, hybrydowe podejście wyznaczania parametrów funkcji radialnych oraz wektora wag upraszcza i przyspiesza zbieżność do rozwiązania zadania aproksymacji.

LITERATURA

1. Kiełbasiński A., Schwetlick H.: Numeryczna algebra liniowa. Wydawnictwo Naukowo – Techniczne, Warszawa 1992.
2. Osowski S.: Sieci neuronowe w ujęciu algorytmicznym. Wydawnictwo Naukowo – Techniczne, Warszawa 1996.
3. Mrówczyńska M.: The influence of normalisation of an input vector on the effect of numerical procedure with the use of the backpropagation error method. VI Międzynarodowa Konferencja Doktorantów, Brno 2004.
4. 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