

**THE EFFECTIVENESS OF THE LEARNING ALGORITHM OF
RADIAL BASIS NETWORKS WITH RELATION TO
THE TRANSFER FUNCTIONS APPLIED
ON THE EXAMPLE OF MAPPING OF
THE LIE LAND OF ZIELONA GÓRA CITY**
*APMĀCĪBAS ALGORITMA EFEKTIVITĀTE PILSĒTAS VIDES
PLĀNOŠANAS UZDEVUMA GADĪJUMĀ*

MARIA MRÓWCZYŃSKA

University of Zielona Góra, Institute of Building Engineering
65-516 Zielona Góra, ul. Z. Szafrana 2, Poland
Phone: + 48 683282636, e-mail: m.mrowczyńska@ib.uz.zgora.pl

Abstract. *The article presents problems connected to the use of radial basis networks for the approximation of the ground surface. The main goal of this paper is to research into the precision of topographic profile representation with relation to the transfer functions applied. The paper contains a description of the structure of a radial basis network and a description of networks learning by means of the hybrid method with the use of the notion of the Green matrix pseudoinverse. Special attention was given to the problem of a choice of transfer functions: the Gauss function, the exponential function, the Hardy function, the spliced function of the third and fourth degree as well as bicentral functions with an independent slope and rotation. the result of this article is an example of the operation of a network with relation the transfer functions under discussion.*

Key words: *radial basis networks, transfer functions, topographic profile representation.*

Introduction

From the mathematical point of view multilayer sigmoidal neural networks play the part of the approximation of stochastic functions with several variables, which represent the set of input variables $\mathbf{x} \in R^N$ onto the set of output variables $\mathbf{y} \in R^M$ [3]. The representation of the input set onto the output set is effected by adjusting an approximation function of several variables to the values assigned i.e. stretching over the learning set of a multidimensional hyperplane, which best adapts to the vector assigned.

Basis networks, in which a hidden neuron realizes a function changing around a chosen center \mathbf{c} , are networks with radial basis functions $\varphi = (\|\mathbf{x} - \mathbf{c}\|)$. The role of the hidden neuron boils down to the radial representation of the space around one point assigned or a group of such points as a cluster.

The simplest radial network (fig. 1) operates on the basis of multidimensional interpolation, which consists in adopting p hidden neurons of the radial type and specifying a representation function $F(\mathbf{x})$ for which the conditions of interpolation are satisfied.

$$F(\mathbf{x}_i) = d_i \tag{1}$$

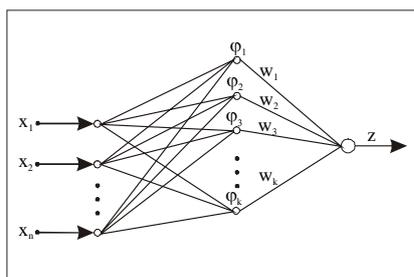


Fig. 1. The general radial network form

The generalization properties of a network are degraded when the number of radial functions is the same as the number of learning data. In this course of action i.e. with the assumption that the number of centres equals the number of coordinates of the input vector ($c_i = x_i, i = 1, \dots, p$), the model adapts to the learning data, because the number of extents of freedom for the system (the number of equations – the number of unknown quantities)

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

is zero. From the above, the generalization quantities of a network are obtained when the condition $K < p$ is fulfilled, where K is the number of centres $c_i (i = 1, 2, \dots, K)$, and p is the number of learning standards (\mathbf{x}, \mathbf{d}) . The vector \mathbf{x} is an input vector, and the vector \mathbf{d} is an assigned vector.

Materials and methods

It has been proved [3] that the adoption of a sufficient number of hidden neurons representing radial functions $\varphi(\mathbf{x})$ makes it possible to solve the task by means of only two network layers: the hidden layer realizing the function transfer vector $\varphi(\mathbf{x})$ of the i^{th} neuron and the output layer with one neuron, whose signal is the function of a linear weighted adder.

The architecture of radial networks is analogous in structure to the structure of a multiplayer neural network with one hidden layer and a linear output neuron. The argument of a radial function is the distance between the input signal x_i and the centre c_i , and the role of hidden neurons is played by radial basis functions.

The most frequently used radial function is the Gauss function. With the assumption that its centre is in point c_i the function has the form (reduced):

$$\varphi(\mathbf{x}) = \varphi(\|\mathbf{x} - c\|) = \exp\left(-\frac{\|\mathbf{x} - c_i\|^2}{2\delta_i^2}\right) \quad (3)$$

where δ_i is the parameter which determines the width of the function. Apart from the Gauss function a radial exponential function is used [3,5]

$$\varphi(\|\mathbf{x} - c\|) = \frac{1}{\sqrt{\|\mathbf{x} - c\|^2 + \sigma^2}} \quad (4)$$

and a radial linear function

$$\varphi(\|\mathbf{x} - c\|) = \|\mathbf{x} - c\|^{2n+1} \quad (n = 1, 2, \dots). \quad (5)$$

The diagrams of the abovementioned functions have been presented in figures 2a – 2c.

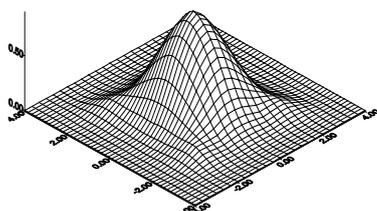


Fig. 2-a. The Gauss function (3)

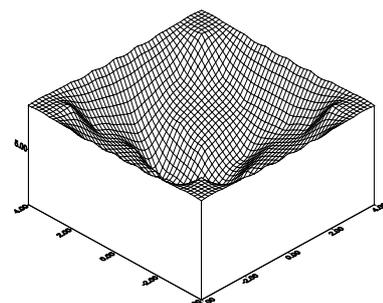


Fig. 2-b. The exponential function (4)

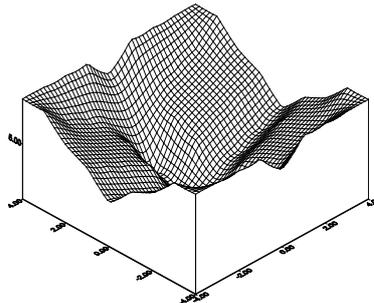


Fig. 2-c. The linear function (5)

Fig. 2. Diagrams of radial basis functions

Apart from the abovementioned radial basis functions, in particular cases other function definitions are used [1], such as for example a spliced function of the second degree

$$\varphi(\|x - c\|) = (\sigma\|x - c\|)^2 \ln(\sigma\|x - c\|) \quad (6)$$

a circular spliced function of the third degree

$$\varphi(\|x - c\|) = \frac{1}{4\sigma^2} \begin{cases} \sigma^3 + 3\sigma^2(\sigma - r) + 3\sigma(\sigma - r)^2 + 3(\sigma - r)^3 & \text{for } r \leq \sigma \\ (2\sigma - r)^3 & \text{for } \sigma < r \leq 2\sigma \\ 0 & \text{for } 2\sigma < r \end{cases} \quad (7)$$

and a circular spliced function of the fourth degree

$$\varphi(\|x - c\|) = \frac{1}{2\sigma^2} \begin{cases} -2r^2 + 3\sigma^2 & \text{for } r \leq \sigma \\ (2\sigma - r)^2 & \text{for } \sigma < r \leq 2\sigma \\ 0 & \text{for } 2\sigma < r \end{cases} \quad (8)$$

where $r = \|x - c\|$. Diagrams of the functions (6), (7), (8) have been presented in figures 3a – 3c.

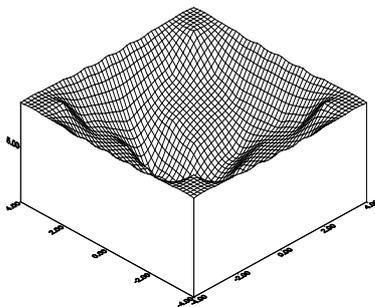


Fig. 3-a. The spliced function (6)

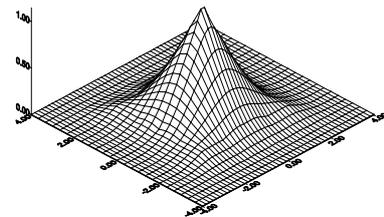


Fig. 3-b The spliced function of the third degree (7)

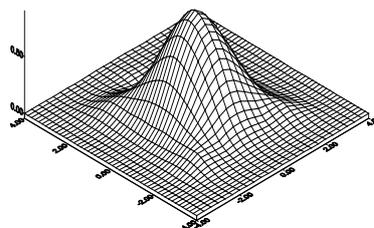


Fig. 3-c. The spliced function of the fourth degree (8)

Fig. 3. Diagrams of radial basis functions

Another group of functions used in the learning process of radial networks are bicentral functions [1], which result from a combination of two sigmoidal functions. The simplest bicentral function is defined as the product of two sigmoidal functions in the form

$$Bi(x, c, b, s) = \sum_{i=1}^n \delta(\exp(s_i) \times (x_i - c_i + \exp(b_i))) (1 - \delta(\exp(s_i) \times (x_i - c_i - \exp(b_i)))) \quad (9)$$

where:

$$\delta(x) = \frac{1}{(1 + \exp(-x))},$$

c – centre,

$$b = \frac{\max(x_i) - \min(x_i)}{2} \text{ - broadening,}$$

s – slope.

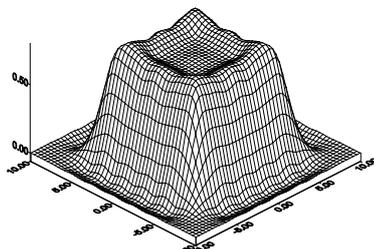


Fig. 4-a. $b = [5 \ 5]$ and $s = [3 \ 3]$

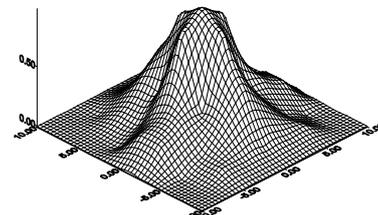


Fig. 4-b. $b = [5 \ 5]$ and $s = [1 \ 1]$

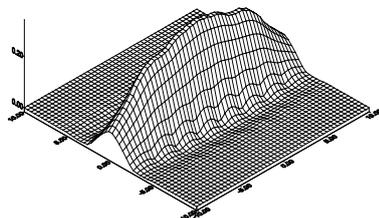


Fig. 4-c. $b = [5 \ 2]$ and $s = [0.3 \ 3]$

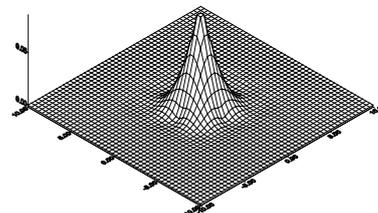


Fig. 4-d. $b = [1 \ 1]$ and $s = [0.5 \ 0.5]$

Fig 4. Examples of bicentral functions for the broadening value b and the slope of the transfer functions

The form of the bicentral function $B_i(x, c, b, s)$ can be changed by changing the position of the centre point c , changing the broadening b and specifying the slope s . Examples of bicentral functions with relation to the values assumed b and s have been presented in fig. 4a – 4d. The broadening b and the slope s as independent parameters make it possible to represent more complicated surface contours than do the Gauss functions [1]. An increase in the accuracy representation by means of bicentral functions is also possible because of the use of two centres $c_i + \exp(b_i)$ as well as $c_i - \exp(b_i)$, which is different in the case of radial base functions localised around one centre $\|x - c\|$. Another way of changing the form of a bicentral function is the use of two independent slopes s_i and s'_i , which increases the adaptation flexibility of the functions estimated. The adoption of two independent slopes s_i and s'_i leads to the definition of a bicentral transfer function as:

$$Bi2s(x, c, b, s) = \sum_{i=1}^n \delta(\exp(s_i) \times (x_i - c_i + \exp(b_i))) (1 - \delta(\exp(s'_i) \times (x_i - c_i - \exp(b_i)))) \quad (10)$$

One of the learning algorithms for neural networks with radial basis functions is the hybrid algorithm, where learning consists in choosing a suitable number of parameters of radial functions $\varphi(\|\mathbf{x} - \mathbf{c}\|)$ and weights w_j so that the goal function is minimized to the form

$$E = \sum_{i=1}^p \left[\sum_{j=1}^K w_j \varphi(\|\mathbf{x}_i - \mathbf{c}_j\| - d_i) \right]^2 \quad (11)$$

With the assumption that the parameters of radial functions are known the minimization of the goal function boils down to solving a set of linear equations with relation to the weight vector \mathbf{w} in the form

$$\mathbf{G}\mathbf{w} = \mathbf{d} \quad (12)$$

where \mathbf{G} is a matrix containing the values of radial basis functions called the Green matrix in the form:

$$\mathbf{G} = \begin{bmatrix} \varphi_{11} = (\|\mathbf{x}_1 - \mathbf{c}_1\|) & \varphi_{12} = (\|\mathbf{x}_1 - \mathbf{c}_2\|) & \cdots & \varphi_{1K} = (\|\mathbf{x}_1 - \mathbf{c}_K\|) \\ \varphi_{21} = (\|\mathbf{x}_2 - \mathbf{c}_1\|) & \varphi_{22} = (\|\mathbf{x}_2 - \mathbf{c}_2\|) & \cdots & \varphi_{2K} = (\|\mathbf{x}_2 - \mathbf{c}_K\|) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_{p1} = (\|\mathbf{x}_p - \mathbf{c}_1\|) & \varphi_{p2} = (\|\mathbf{x}_p - \mathbf{c}_2\|) & \cdots & \varphi_{pK} = (\|\mathbf{x}_p - \mathbf{c}_K\|) \end{bmatrix}$$

The first stage of the hybrid algorithm is the specification of the weight vector \mathbf{w} by means of the notion of the pseudoinverse \mathbf{G}^+ of the matrix \mathbf{G} as

$$\mathbf{w} = \mathbf{G}^+ \mathbf{d} \quad (13)$$

One of the ways of specifying the pseudoinverse \mathbf{G}^+ of the matrix \mathbf{G} is the Gram-Schmidt method of orthogonalization, where the matrix \mathbf{G} is decomposed into the matrices \mathbf{Q} and \mathbf{R} [2]. The matrix \mathbf{Q} is a matrix with orthonormal columns with the dimensions $p \times K$, and the matrix \mathbf{R} is an upper triangular matrix with the dimensions $K \times K$. According to this method the pseudoinverse is specified from the relation

$$\mathbf{G}^+ = \mathbf{R}^{-1} \mathbf{Q}^T \quad (14)$$

In the second stage, with frozen values of input weights, centres and widths of radial basis functions are adapted with the use of gradient optimization methods, most frequently the method of the greatest fall. For the goal function defined by the formula (11) the adaptation of centres c and widths of basis functions σ is achieved according to the formulas

$$c_{ij}(n+1) = c_{ij}(n) - \eta \frac{\partial E}{\partial c_{ij}} \quad (15)$$

$$\sigma_{ij}(n+1) = \sigma_{ij}(n) - \eta \frac{\partial E}{\partial \sigma_{ij}} \quad (16)$$

where n – present interaction, $(n+1)$ – subsequent interaction, η - learning coefficient.

Results and discussion

Radial basis functions including the transfer functions mentioned in this article have been used for describing topographic profile. The learning set consisted of 2000 points (x, y, z) , generated as dispersed points, whose heights were within the limit of 80÷210 m. The assessment of the quality of neural networks with radial basis functions with relation to the transfer functions used has been expressed by means of the learning error – *RMSE* (*Root Mean Squared error*) and presented in fig. 5.

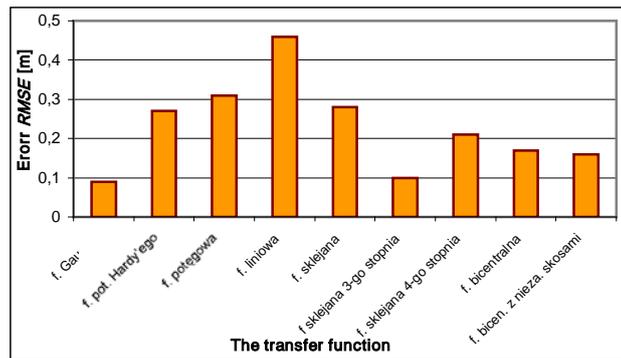


Fig. 5. The learning results of a radial network with relation to the transfer function used

The best learning result of a radial network was achieved when the Gauss function (1) was used as a transfer function, for which the learning error was $RMSE=0.09m$ (fig. 6), and the greatest error $RMSE=0.46m$ (fig. 7) occurred when a radial linear function was used (3). Learning was effected by means of the hybrid method in a learning set, which consisted of 2000 points and 500 test points, the number of neurons in the hidden layer was 20 [4].

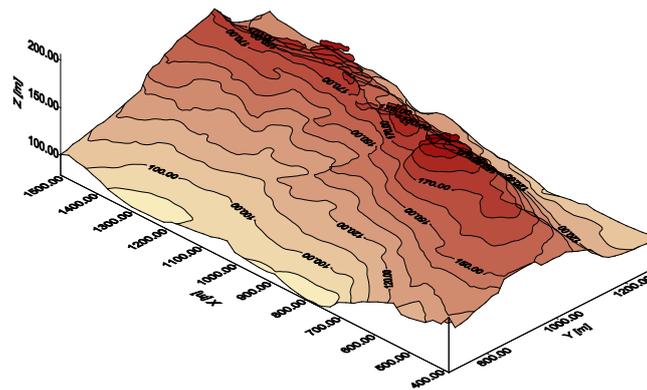


Fig. 6. A representation of the terrain surface by means of the Gauss function

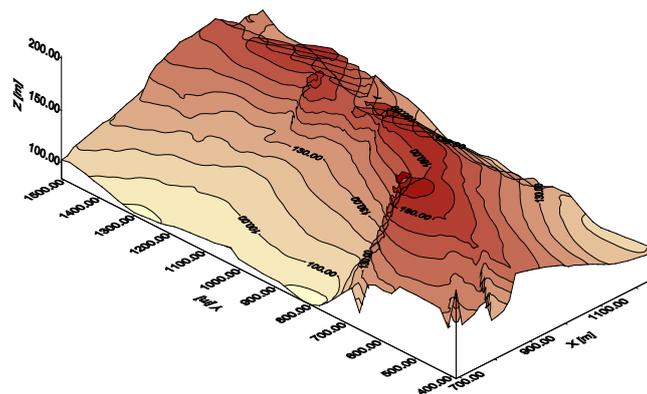


Fig. 7. A representation of the terrain surface by means of a linear function

Conclusions

While solving tasks by means of radial basis networks the basic problems are the right choice of initial parameters and as well as the most suitable transfer function for a particular problem. The use of a random choice of these parameters increases the probability of stopping the process of learning at a local minimum. Therefore, it is better to choose initial parameters by means of procedures based on information in the learning set, and parameters of radial functions obtained in this way are adopted as initial values. An advantage of radial networks is a simple learning algorithm and precise network architecture, which is a condition for the starting point to be closer to the optimum solution in comparison to the algorithm used in sigmoidal networks. Moreover, the hybrid approach to the specification of parameters of radial functions and the weight vector simplifies and quickens convergence to the solution of the task of approximation.

References

1. Jankowski N., Ontogeniczne sieci neuronowe. O sieciach zmieniających swoją strukturę. Akademicka Oficyna Wydawnicza EXIT, Warsaw 2003.
2. Kielbasiński A., Schwetlick H., Numeryczna algebra liniowa. Wydawnictwo Naukowo – Techniczne, Warsaw 1992.
3. Osowski S., Sieci neuronowe w ujęciu algorytmicznym. Wydawnictwo Naukowo – Techniczne, Warsaw 1996.
4. 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.
5. Żurada J., Barski M., Jędruch W., Sztuczne sieci neuronowe. Podstawy teorii i zastosowania, Wydawnictwo naukowe PWN, Warsaw 1996.