Neurocomputing 74 (2011) 2502–2510

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

MELM-GRBF: A modified version of the extreme learning machine for generalized radial basis function neural networks

Francisco Fernández-Navarro *, César Hervás-Martínez, Javier Sanchez-Monedero, Pedro Antonio Gutiérrez

Department of Computer Science and Numerical Analysis, University of Córdoba, Campus de Rabanales, Albert Einstein Building, 3rd floor, 14074 Córdoba, Spain

A R T I C L E I N F O

Available online 20 May 2011

Keywords:
Generalized radial basis functions neural networks
Extreme learning machine
Multi-classification
Generalized Gaussian distribution

A B S T R A C T

In this paper, we propose a methodology for training a new model of artificial neural network called the generalized radial basis function (GRBF) neural network. This model is based on generalized Gaussian distribution, which parametrizes the Gaussian distribution by adding a new parameter τ. The generalized radial basis function allows different radial basis functions to be represented by updating the new parameter τ. For example, when GRBF takes a value of $\tau = 2$, it represents the standard Gaussian radial basis function. The model parameters are optimized through a modified version of the extreme learning machine (ELM) algorithm. In the methodology proposed (MELM-GRBF), the centers of each GRBF were taken randomly from the patterns of the training set and the radius and τ values were determined analytically, taking into account that the model must fulfil two constraints: locality and coverage. An thorough experimental study is presented to test its overall performance. Fifteen datasets were considered, including binary and multi-class problems, all of them taken from the UCI repository. The MELM-GRBF was compared to ELM with sigmoidal, hard-limit, triangular basis and radial basis functions in the hidden layer and to the ELM-RBF methodology proposed by Huang et al. (2004) [1]. The MELM-GRBF obtained better results in accuracy than the corresponding sigmoidal, hard-limit, triangular basis and radial basis functions for almost all datasets, producing the highest mean accuracy rank when compared with these other basis functions for all datasets.

1. Introduction

Artificial neural networks (ANN) are largely used in applications involving classification or function approximation. Lately, it has been proved that several classes of ANN are universal function approximators [2–4]. Among them, we find radial basis function neural networks (RBFNNs) [5,6], multi-layer perceptrons (MLPs) [7] or product unit neural networks (PUNNs) [8,9]. All are multi-layered networks and can be considered as connectionist models. RBFNNs use, in general, hyper-ellipsoids to split the pattern space. This is different from MLPs which build their classifications on pseudo-hyper-planes, defined by a weighted sum [10].

RBFNNs use the value of the distance to estimate the response value, being functions of two arguments, $\mathbf{x}$ and $\mathbf{c}$, where $\mathbf{x} = (x_1, x_2, \ldots, x_K)^T$ is the vector of co-ordinates of a pattern of the dataset and $\mathbf{c} = (c_1, c_2, \ldots, c_K)^T$ are the location parameters to determine kernel positions. The characteristic feature of local RBFNNs is the fact that their response value decreases monotonically with the distance from the center $\mathbf{c}$ of the radial function.

RBFNNs are parametrized by a width denoted here by r. If the distance between $\mathbf{x}$ and $\mathbf{c}$ is small compared to the width of the kernel, the kernel value will be close to one. Large distances by contrast are mapped to values close to zero. The width of the RBFNNs in kernel-based methods must produce a correct balance between *covering*: the sum of all RBFs must have a high value in all patterns of the dataset; and *locality*: the RBF should provide a high value (close to one) for patterns that are close to the environment where the RBF is located, and low values (near zero) for patterns that are not located in the region of space where the RBF is centered.

The Gaussian RBFs are based on the Gaussian density function and are defined by a "center" position and a "width" parameter. The Gaussian function gives the highest output when the incoming variables are closest to the center position and decreases monotonically as the distance from the center increases. Gaussian distribution can be parametrized by a real parameter τ, resulting in generalized Gaussian distribution (GGD). The GGD may represent different forms of distribution function by changing a real parameter τ. We can highlight the impulsive, Laplacian, Gaussian and uniform distributions.

* Corresponding author. Tel.: +34 957 21 83 49; fax: +34 957 21 83 60.
E-mail address: i22fenaf@uco.es (F. Fernández-Navarro).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.11.032

Based on this probability distribution, we propose the generalized radial basis function (GRBF) by removing the constraints of a probability function. In this way, the generalized radial basis function (GRBF) is defined as

$$\phi(\mathbf{x};\mathbf{c},r,\tau)=\exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|^{\tau}}{r^{\tau}}\right). \quad (1)$$

Training of RBFNNs can be classified into two categories: quick learning and full learning. Quick learning usually involves a two-step process. First, the parameters governing the basis functions are determined by a relatively fast, unsupervised clustering [11] or vector quantization approach [12]. Next, the weights of the basis functions are determined using linear optimization techniques. A full learning scheme (for instance gradient-descent-based methods) optimizes all of the parameters in a supervised mode [6,13,14].

Gradient-descent-based algorithms may converge very slowly to the solution of the given problem if the learning rate is small. However, if the learning rate is large, they can be unstable or divergent. They may also easily get over-fitting or be stuck in local optima [15,16]. Moreover, most of the training algorithms based on gradient descent are still slow due to the many iterative steps required in the learning process. That is the reason why our proposal will be based on the first approach.

Recently, Huang et al. showed that a single hidden layer feedforward neural network (SLFN) can learn distinct observations with an arbitrary small error margin if the activation function is chosen properly [17–19]. An effective training algorithm for SLFNs called extreme learning machine (ELM) was also proposed by Huang et al. [20,21]. In ELM, the input weights of the hidden nodes are randomly chosen, and the output weights of SLFNs can be determined through the pseudo-inverse operation of the output matrix in the hidden layer. This algorithm can avoid many of the problems which occur in gradient-descent-based learning methods. For that reason, the GRBF proposed in this paper was trained by means of a modified version of the ELM algorithm (MELM-GRBF).

The main novelty introduced by the MELM-GRBF is in the determination of the GRBFs. While in the ELM-RBF algorithm [1], the centers and the radii of the RBFs are selected randomly, in the MELM-GRBF algorithm proposed, the centers are initialized by randomly selecting some patterns in the training dataset. The values of the radius and τ are determined analytically by solving two equations that ensure that the model fulfils two constraints: locality and coverage.

This paper is organized as follows: a brief analysis of the generalized Gaussian distribution is given in Section 2. The single layer feedforward GRBFNN is presented in Section 3. A methodology to optimize the GRBFNN parameters based on ELM is presented in Section 4. Section 5 explains the experiments that were carried out. Finally, Section 6 summarizes the conclusions of our work.

2. Generalized Gaussian distribution

In order to cope with some limitations of the Gaussian RBF [22–24], we need to use another model that can describe the statistical behaviors of the object and background classes in a multiclassification problem in the best possible way. A possible solution is to adopt a more general parametric model that should satisfy two main properties: (i) flexibility (i.e., it should be capable of modeling a large variety of statistical behaviors) and (ii) stability (i.e., it should not require the estimation of a large number of parameters). Motivated by the above observations, the present study proposes a new class of RBFs based on generalized Gaussian distribution (GGD).

The GGD requires only one additional parameter to be estimated compared to the Gaussian distribution, and it can approximate a large class of statistical distributions (e.g., impulsive, Laplacian, Gaussian, and uniform distributions). The analytical equation of the probability density function of the GGD is given by

$$p(\mathbf{x};\mathbf{c},r,\tau)=\frac{\tau}{2r\Gamma(1/\tau)}\exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|^{\tau}}{r^{\tau}}\right), \quad (2)$$

where $\mathbf{c}$, $r>0$ and $\tau>0$ are the parameters of the mean, the scale or width and the shape of the distribution, respectively. $\Gamma(z)$ is the Gamma function, an extension of the factorial function, which is defined as $\Gamma(z)=\int_0^{\infty} t^{z-1}e^{-t}\,dt$, for $z>0$. The scale parameter r that expresses the width of the distribution is related to the normal standard deviation by the equation:

$$r=\sigma\sqrt{\frac{\Gamma(1/\tau)}{\Gamma(3/\tau)}}, \quad (3)$$

where σ is the normal standard deviation. The shape parameter τ refines the decay rate of the density function. It is worth noting that $\tau=2$ yields Gaussian density and $\tau=1$ results in Laplacian density distribution. As limit cases, for $\tau\to 0$, the distribution becomes impulsive, whereas for $\tau\to\infty$ it approaches uniform distribution (Fig. 1). Then, the scale parameter models the width of the GGD peak and the shape parameter is inversely proportional to the decreasing rate of the peak.

The GGD model is intrinsically stable, since it is characterized by few parameters to be estimated. Compared to Gaussian distribution, thanks to an additional statistical parameter (i.e., the shape parameter), it is more flexible and can approximate a large class of statistical distributions.

In this paper, based on this probability distribution, we define a novel RBF, by removing the constraints of a probability function, called generalized radial basis function (GRBF) which is defined using the following expression (for a k-dimensional input space):

$$\phi_j(\mathbf{x};\mathbf{c}_j,r_j,\tau_j)=\exp\left(-\frac{\|\mathbf{x}-\mathbf{c}_j\|^{\tau_j}}{r_j^{\tau_j}}\right), \quad (4)$$

where $\mathbf{x}_i=(x_{i1},\ldots,x_{ik})^T$ is the vector of measurements, k is the number of inputs, r_j the width of the GRBF, $\mathbf{c}_j=(c_{j1},\ldots,c_{jk})^T$ the center and τ_j the shape parameter of the j-th GRBF.

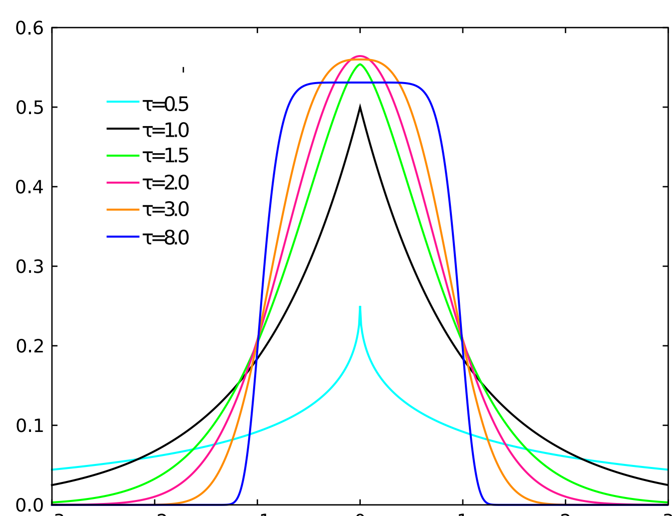

Fig. 1. Probability density function of the generalized Gaussian distribution (GGD) with different values of τ, $c=0$ and $r=1$.