On the Radial Basis Function Networks with the Basis Function of q-Normal Distribution

Kotaro ECCYUYA and Masaru TANAKA
Department of Information & Computer Sciences,
Saitama University
255 Shimo-okubo,Saitama-shi, Saitama, 338-8570, Japan
Tel. +81-48-858-3957,Fax. +81-48-858-3716
E-mail: {kotaro,mtanaka}@mi.ics.saitama-u.ac.jp

Abstract: Radial Basis Function (RBF) networks is known as efficient method in classification problems and function approximation. The basis function of RBF networks is usual adopted normal distribution like the Gaussian function. The output of the Gaussian function has the maximum at the center and decrease as increase the distance from the center. For learning of neural network, the method treating the limited area of input space is sometimes more useful than the method treating the whole of input space. The q-normal distribution is the set of prbability density function include the Gaussian function. In this paper, we introduce the RBF networks with the basis function of q-normal distribution and actually approximate a function using the RBF networks.

1. Introduction

The Radial Basis Function(RBF) networks have turned out to be among the most powerful artificial neural network types, e.g. in the area of function approximation, pattern classification and data clustering.

The RBF networks are multilayer feedforward-type neural networks, using radially basis functions. While the multilayer perceptron (MLP) networks used Back-Propagation (BP) are known to be slow in convergence and often trapped in local minimum in the parameter spaces, the RBF networks are fast in convergence to the optimum point and have excellent fitting ability.

Other main difference between the architectures of MLP and RBF is the separation surfaces implemented in classification problems. The MLP networks separate classes by building hyperplanes in the input space. Another side, RBF networks divide the input space into some sub-spaces and only a few hidden RBF units represent each sub-space.

In the point of function approximation, the MLP networks approximate a function as the sum of sigmoid functions. On the other hands, function approximation of the RBF networks makes use of particular basis functions that affect the input-data in the only limited area of input space. When we approximate functions, the method treating limited area of input space is sometimes more useful than the method treating the whole of input space.

Many probability density functions for the basis function of RBF networks are proposed. Among them, the Gaussian function is generally used. The output of the Gaussian function has the maximum at the center and decrease as increase the distance from the center.

On the other hands, Masaru Tanaka proposed q-normal distribution(2002). The q-normal distribution is the set of probabililty density function include the Gaussian function as a particular state. Because the q-normal distribution include the Gaussian function, we can think that the q-normal distribution is more useful than the Gaussian function at least.

In this paper, we propose the RBF networks with the basis function of the q-normal distribution. And we carry out function approximation using the q-normal distributions for every q, and we investigate the results.

2. Radial Basis Function Networks

We introduce the Radial Basis Function(RBF) networks proposed by Moody and Darken (1989). The RBF networks are multilayer feedforward-type neural networks, using radially basis functions as the activation function of the hidden nodes. The RBF networks have shown to be universal approximators, which means that they can approximate any function to any desired degree of accuracy. They are fast in convergence to the optimum point and have excellent fitting ability as compared with the conventional multi-layer perceptron networks. The RBF networks are networks which outputs are calculated by carrying out linear combination of the output of the basis function. Fig.1 shows the general structure of the RBF networks with M inputs and P outputs. When the Gaussian RBF is used as the activation function (ϕ Fig. 1) of the hidden nodes, the basis function ϕ_m is represented mathematically as

$$\phi_m(x) = \exp\left\{\frac{-(x-\mu_m)^2}{\sigma^2}\right\}, \tag{1}$$

where x is an input data, μ is a point in the input space (for example, a center of the corresponding cluster), and σ is a variable which represents the extent of affected space of the basis function in the input space.

When the Gaussian function is used as the basis function, we obtain the output $f(x)$ of the network as linear combination of the outputs of the middle layer elements as follow:

$$f(x) = G(x)\sum_{m=1}^{M} w_{km}\phi_m(x), \tag{2}$$

where M is the number of the hidden nodes, w_{ki} is the weight between the kth output and ith input, and $G(x)$ is called normalization constant, which is the reciprocal of the linear combination of middle layers, like follow:

$$G(x) = \frac{1}{\sum_m \phi_m(x)}. \tag{3}$$

Among the many learning algorithms, the k-means clustering scheme for the center positions, the k-nearest neighbors for the width parameter, and the gradient-descent method for the weight parameters are very popular, which are adopted in this paper. The center positions and the width parameters are determined adaptively during the operation of the system. Based on this setup and the universal uniform approximation property of the RBF networks, the approximation error is small and satisfies the condition, provided that the number of layers of the RBF networks used is large enough and the learning time is long enough, which is guaranteed by theory. The RBF networks with the basis function of Gaussian distribution are the networks, which supports only the domain set the basis function, because the Gaussian function has the maximum at the center and decreases as increase the distance from the center.

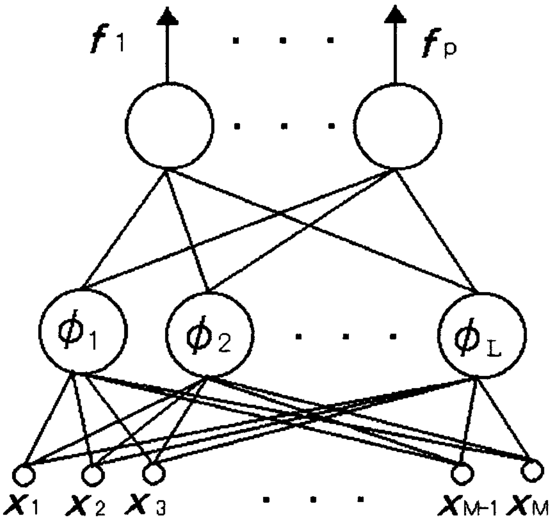


Fig.1: General structure of the RBF networks.

3. The q-normal distributions

The q-normal distribution is the set of probability density functions rely on the parameter q, and it include Gaussian function. Because the q-normal distributions is smooth on q and it represents the Gaussian function at $q = 1$, we can expect that we get more performance in the field adapted the q-normal distributions than Gaussian. The q-normal distributions is defined as follow:

$$P_q(x,y) = \frac{1}{Z_q}\left\{1 - \frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right\}^{\frac{1}{1-q}}, \tag{4}$$

where Z_q is normalization constant as

$$Z_q = \int dx \left\{1 - \frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right\}^{\frac{1}{1-q}}. \tag{5}$$

Note that the probability density functions must be non-negative. So we introduce cut-off x_- and x_+ into it under $q < 1$ as follow:

$$x_- = \mu - \sqrt{\frac{3-q}{1-q}}\sigma, \tag{6}$$

$$x_+ = \mu + \sqrt{\frac{3-q}{1-q}}\sigma. \tag{7}$$

In the q-normal distribution at some parameter q, we can find the probability density function that was discovered and used independently until now. Also, because q-normal distribution is smooth on q, it is able to connect their probability density function smoothly. We give the typical examples as follows.

The q-normal distribution represents Gaussian function at $q = 1.0$,

$$p_1(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

At $q = 1 + \frac{2}{n+1}$,

$$p_{1+\frac{2}{n+1}}(x) = \frac{1}{\sqrt{n}B(\frac{n}{2},\frac{1}{2})\sigma}\left\{1 + \frac{1}{n}\frac{(x-\mu)^2}{2\sigma^2}\right\}^{-\frac{n+1}{2}},$$

the probability density function is called t-distribution. In this equation, variance σ^2 is interpreted as a scale factor.
Moreover at $q = 2.0$,

$$p_2(x) = \frac{1}{B(\frac{1}{2},\frac{1}{2})\sigma}\left\{1 + \frac{(x-\mu)^2}{2\sigma^2}\right\}^{-1},$$

the probability density function represent Cauchy-distribution, and also in this equation variance σ^2 is interpreted as a scale factor.

In Fig.1 we show some graph of q-normal distribution on various parameter q.

4. Experiment

We experiment for the accuracy on q of the function approximation using the RBF networks with the basis function of the q-normal distribution. As input data, we adopt the data which quantized the function like Fig.3 and Fig.4.

First, we quantized the horizontal axis of this function for every 0.1 and use those quantized points input

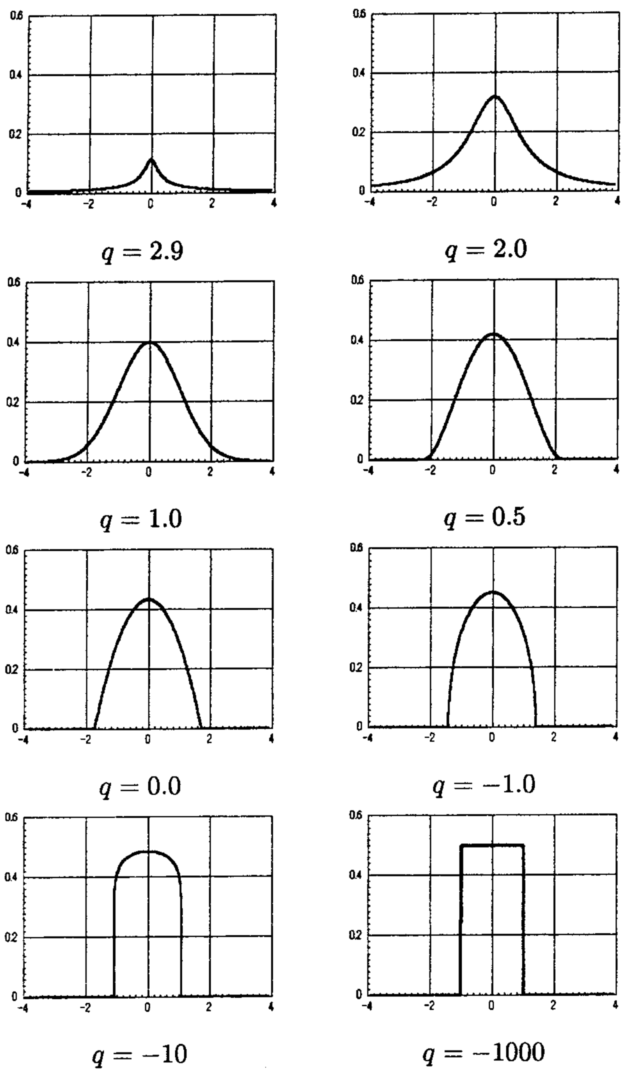


Fig. 2:The q-normal distributions for various q ($\mu = 0, \sigma = 1$).

data of the function approximate. Next, we simulate the function approximate until the sum total of a mean squared error with all input data became smaller than a fixed value. Last, we ask for the mean squared error between original functions and the approximated fnction for every q. And we seek the optimal parameter q.

In Fig.5, we denote the transition of mean square error between the output of the input function and the corresponding approximate function on q. Moreover in Fig6 and Fig7, we show the result of function approximate at some q. In Fig.5, the vertical axis expresses the mean squared error between original functions and the approximated function for every q, and the horizontal axis expresses value of parameter q. In Fig.6 and Fig.7, we drew the approximated function.

5. Conclusion and future works

In this paper, we proposed the RBF networks with the basis function of the q-normal distribution. And we carried out function approximation using the q-normal distributions.

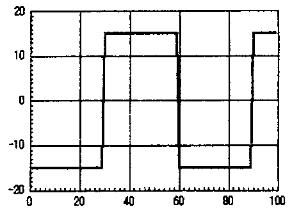


Fig.3 : input data A

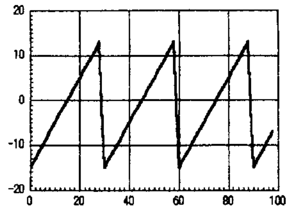
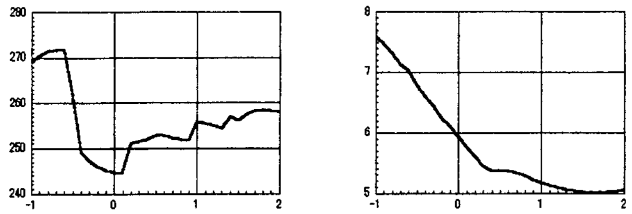


Fig.4: input data B



error value of function A for q

error value of function B for q

Fig.5 : the mean squared error between original functions and the approximated function for q.

It is seen that the approximated function by RBF networks becomes smooth as increase the parameter q of the q-normal distributions. But if it becomes larger than a fixed value, the mean square error between the outputs of the original function and the corresponding approximate function increase. We can guess that it is due to existence of the suitable value of the parameter q depending on an approximating function.

As future works, we are going to approximate more large number of functions, and determine the suitable value of the parameter q depending on an approximating function and the suitable number of the sampling points to recover the original function through the our approximation method to find the relation with the sampling theorem.

References

[1] Keun Bum Kim ,Jin Bae Park , Yoon Ho Choi , Guanrong Chen, “Control of chaotic dynamical systems using radial basis function network approximators,” *Information Sciences*,vol.130, pp.165-183 ,2000.

[2] S. A. Billing , G. L . Zheng, “Qualitative Validation of Radial Basis Function Networks,” *Mechanical Systems and Signal Processing*, vol.13(2), pp.335-349, 1999.

Table 1:the value of q which makes the minimum of the mean squared errors.

	function A	function B
q	0.1	1.6

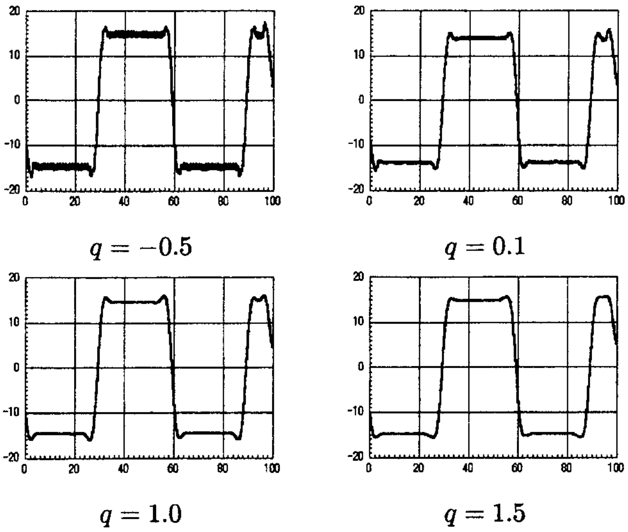


Fig.6 : The result of approximation function A on q.

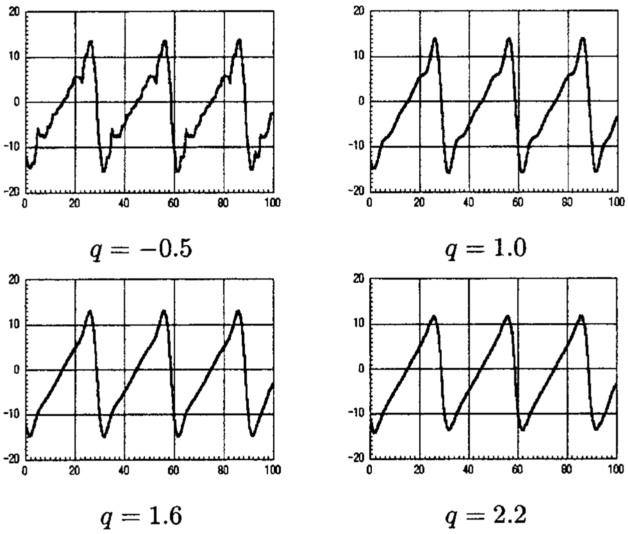


Fig.7 : The result of approximation function B on q.

[3] M. Marinaro , S. Scarpetta, "On-line learning in RBF neural networks: a stochastic approach," *Neural Networks*,vol.13, pp.719-729 ,2000.

[4] Masaru Tanaka, "A Consideration on A Family of q-Normal Distribution," *the paper of IEICE D-II*, 2002.