Nonlinear Approximations Using RBF Neural Networks

RBF 신경망을 이용한 비선형근사

Jooyoung Park*
박 주 영*

이 논문은 1994년도 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었음.

ABSTRACT

In this paper, some fundamental problems concerning RBF(radial-basis-function) networks and approximation of functions are addressed. First, a comprehensive introduction to RBF networks is given with typical RBF networks classified into three classes. Next, sharp conditions are given under which continuous functions of a finite number of real variables can be approximated arbitrarily well by a certain class of RBF networks. Finally, a related result is given concerning the representation of functions in the form of distributed RBF networks.

요 약

본 논문은 RBF 신경망과 함수근사에 관한 기본 이론을 다룬다. 첫째, RBF 신경망에 관한 전반적인 소개가 전형적인 RBF 신경망의 분류와 함께 제공된다. 다음에는, 유한개의 실변수를 갖는 연속함수가 일정한 클래스의 RBF 신경망에 의해 충분히 가깝게 근사될 수 있기 위한 조건이 제시된다. 마지막으로, 분산형 RBF 신경망을 이용한 함수 표현에 관한 결과가 주어진다.

I. Introduction

There have been several recent studies concerning feedforward neural networks and the problem of approximating arbitrary functionals of finite number of real variables. Some of these studies deal with cases in which the nonlinear elements in hidden layers are not sigmoidal functions cascaded with affine maps. The so-called RBF(radial-basis-function) networks are one of the most promising alternatives along the lines. In [1]-[3], results are given corncerning the approximation of functions using RBF networks. The theorems given there provide conditions under which RBF networks are capable of approximation in the sense of the uniform metric on compact subsets of R^r or the L^p metric. The first objective of this paper is to present an extension of a part of these results. For the clear and convenient presentation of the main results, a comprehensive introduction to RBF networks is given in

*Dept. of Control and Instrumentation Engineering, Korea University

section II. The section also partitions typical RBF networks into three classes: $S_0(K)$, $S_1(K)$, and $E(K)$ based on the function form of the network responses. In section III, an approximation theorem, which relaxes the conditions of kernel functions for the universal approximation using RBF networks, is reported. The theorem is a stronger result along the lines of material in [2]. As observed in [4]-[7], approximation of functions using kernel-based networks has drawn strong interest in various areas of science and engineering. In [4], the pdf(probability density function) of a random variable is estimated from the observed input samples using the nonparametric approach based on Parzen window. In [5], it is shown that RBF networks with gaussian basis function are capable of universal approximation on a compact subset of finite dimensional Euclidean space. As will be clear after presentation of our main results, the main theorem of [5] can be shown to be a simple corollary of our theorem 1. In [6], the mapping capability is investigated for a network using kernel functions which have the properties of bounds and locality, and it is shown that the network is a universal approximator. As a recent work, [7] shows that the so-called regularized networks with special types of activation functions, under mild conditions, are capable of learning and approximating arbitrarily well any function belonging to a reproducing kernel Hilbert space, at an optimal nonparametric rate. The theoretical studies concerning the approximation capability of kernel-based networks usually assume weak conditions on the kernel function, which are similar-looking but different according to the distance functions and vector spaces under consideration. Of course, results of this kind are of interest in connection with the question of whether a satisfactory network solution can be obtained by using some member of a given family.

Another purpose of this paper is to consider a connection between the study of RBF networks and the study of wavelets(for an overview of wavelet theory, see [8]). In wavelet theory attention is often focused on a transform called the CWT(continuous wavelet transform). In section IV, an inversion formula is given in the form of a limit of integrals for recovering f defined on R^r from a multi-dimensional version of the CWT of f. Also the discretization of the inversion formula is discussed briefly in the last section, and a remarkable observation is obtained to the effect that approximate inverses given by the discretization become elements of RBF networks. Thus the material in the section bears on the matter of providing an algorithm for the construction of approximating functions using RBF networks. But except for a brief discussion in the concluding remarks, this is not fully pursued.

Throughout the paper, we use the following definitions and notation, in which N and R denote the set of natural numbers and the set of real numbers, respectively, and for any positive number r, R^r denotes the normed linear space of real r-vectors(viewed as row vectors) with the Euclidean norm $\|\cdot\|$. $L^p(R^r)$ denotes the usual space of R-valued maps f defined on R^r such that f is pth power integrable. S and S' denote the set of rapidly decreasing functions and the usual set of tempered distributions on R^r, respectively. A functional $\tilde{f} \in S'$ is said to be regular in S' if there exists an ordinary function f such that $\tilde{f}(\phi) = \int_{R^r} f(x)\phi(x)dx$ for each $\phi \in S$, and when this is the case we say that

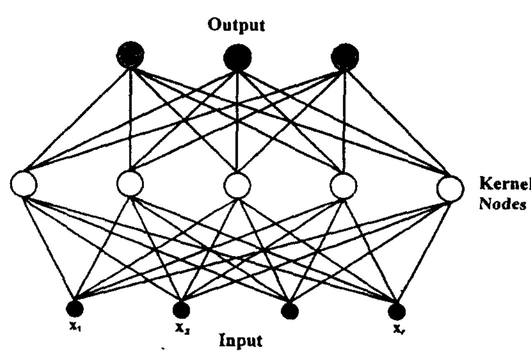

Fig. 1 A Radial Basis Function Network

$\tilde{f}$ and f are equal or equivalent and write $\tilde{f}=f$. The usual inner product in R^r and the Hilbert space $L^2(R^r)$ are both denoted by $\langle\cdot,\ \cdot\rangle$. With $W\subset R^r$, $C(W)$ denotes the space of continuous R-valued maps defined on W. The usual L^p and uniform norms are denoted by $\|\cdot\|_p$ and $\|\cdot\|_\infty$, respectively. The convolution operation is denoted by "$*$", and the Fourier transforms [9] of the function f and measure μ are written as $\hat{f}$ and $\hat{\mu}$, respectively. Finally, supp$f$ denotes the support of f, in which f can be a function or a tempered distribution.

Ⅱ. Classification of RBF networks

As shown in Figure 1, RBF networks can be regarded as a class of three-layered feedforward neural networks with single layer of hidden units. Each hidden node(often called kernel node) of an RBF network has its output derived from a single function indexed by two parameters, the centroid z and the smoothing factor σ. The centroid of a hidden node is the location of its center. Thus, the response of a hidden node $h: R^r\rightarrow R$ can be represented using a function translated by its centroid z:

$$h(x)=K(x-z;\sigma) \tag{1}$$

The smoothing factor of a hidden node changes the shape of the response of the hidden node through dilation. There are several ways that the shape of the response is controlled through dilation.

In its simplest form, the smoothing factor is a single positive number σ that is the same across the kernel nodes, in which case the equation (1) can be expressed as

$$K(x-z;\sigma)=K(\frac{x-z}{\sigma})$$

for a certain function $K: R^r\rightarrow R$ (which is called the kernel function). An output node of an RBF network with such hidden nodes is represented by $q: R^r\rightarrow R$ of the form

$$q(x)=\sum_{i=1}^{M} w_i K(\frac{x-z_i}{\sigma}) \tag{2}$$

where $M\in N$ is the number of hidden nodes in the hidden layer, $w_i\in R$ is the weight from the ith hidden node to the output node, x is an input vector(an element of R^r). Here $z_i\in R^r$ and $\sigma>0$ are the centroid and the smoothing factor of the ith hidden node, respectively. We call this family of RBF networks having $K: R^r\rightarrow R$ as the kernel function of their hidden nodes $S_0(K)$. Ordinarily K is radially symmetric with respect to a certain norm $\|\cdot\|$ (typically the Euclidean norm) in the sense that $\|x\|=\|y\|$ implies $K(x)=K(y)$.

Some networks in this class have an interesting interpretation in terms of smooth approximation of input-output data. In [10] a certain class of networks are derived using the so-called regularization technique. Given M data points $\{(x_i,\ y_i)\in R^r\times R: i=1,\cdots,\ M\}$, the problem of finding a function $f: R^r\rightarrow R$ such that $f(x_i)=y_i$ for all $i\in\{1,\ \cdots,\ M\}$ is ill-posed in the sense that it has an infinite number of solutions. A solution for this is to exploit a smoothness constraint to change the problem into a well-posed one. This is called the regularization technique [11]. The technique seeks a solution that minimizes a cost functional consisting of two terms, the first for the distance between the data and the solution f and the second for the cost associated with the deviation from smoothness. According to [10], under certain conditions, the solution of the minimization problem is given in the following form:

$$f(x)=\sum_{i=1}^{M} w_i G(x,\ z_i) \tag{3}$$

where $w_i\in R$, $z_i\in R^r$ and G is a Green's function of a certain self-adjoint differential operator. With a certain choice of the distance measures and some additional constraints on G, (3) can be expressed in the form of (2), in which case the minimizing solutions f

are elements of $S_0(K)$.

The class $S_0(K)$ has been widely used in many areas of engineering [12]-[15][1)] and results there show that it is very useful in handling highly nonlinear problems.

Families with a vector space structure are often particularly important. For example, networks are widely used in which the smoothing factors are positive real numbers as in $S_0(K)$, but can have different values across the hidden nodes. Given the kernel function K: $R^r \rightarrow R$, we call these vector spaces $S_1(K)$. The response at a general output node is represented by q: $R^r \rightarrow R$, where

$$q(x) = \sum_{i=1}^{M} w_i K(\frac{x - z_i}{\sigma_i})$$

in which $M \in N$, $\sigma_i > 0$, $w_i \in R$, and $z_i \in R^r$ for $i = 1, \cdots, M$. This family has been successfully applied in many studies [16]-[19]. Sometimes modifications are made to typical architectures of $S_1(K)$ for various purposes. In [20] an extended architecture is introduced by adding additional output nodes representing training data density and confidence limit on each network output. This extension attempts to not only fit functions but also provide information on the probability density of the input and confidence intervals for the predictions of the networks. In [21] a modified version of $S_1(K)$ is derived by minimizing a least square cost function with the mild constraint that the mutual information between input and output is maximized. This new architecture permits optimal generalization in a certain sense.

As pointed out in [17] and [22], relaxing the radial constraint and allowing a different scale for each dimension can often improve performance. The resulting architectures comprise the third family in our classification. Since a different scale is allowed for each dimension of the input, the smoothing factor can no longer be given as a real number. Instead, it is necessary that each input dimension has its own positive dilation variable. Given the kernel function $K: R^r \rightarrow R$, such networks have output nodes each of which is represented by a function $q: R^r \rightarrow R$ of the form

$$q(x) = \sum_{i=1}^{M} w_i K(\frac{x_1 - z_{i1}}{\sigma_{i1}}, \frac{x_2 - z_{i2}}{\sigma_{i2}}, \cdots, \frac{x_r - z_{ir}}{\sigma_{ir}})$$

where $M \in N$, $x \triangleq (x_1, \cdots, x_r) \in R^r$, $w_i \in R$, $z_{i1}, \cdots, z_{ir} \in R$, and $\sigma_{i1}, \cdots, \sigma_{ir} > 0$ for $i = 1, \cdots, M$. We call this family $E(K)$; it has a vector space structure like $S_1(K)$. Of course, when the input space is one dimensional, $E(K)$ is the same as $S_1(K)$. Applications of E(K) and its important variations are reported in [23]-[25].

Ⅲ. Approximation theorem

In [1]-[3] results are given concerning the approximation of functions using RBF networks. An improvement of a part of these results is given in this section. More precisely, the approximation capability is investigated for RBF networks in the class $S_1(K)$, and sharp conditions on the kernel function K are given under which the elements of $S_1(K)$ are capable of universal approximation on compact subsets of R^r. For the sake of clarity and convenience, only a one dimensional output space is considered here instead of outputs represented by multiple nodes as in Figure 1. The extension of our results to multi-dimensional output spaces is straightforward.

In Theorem 4 of [2], it is shown that if $K: R^r \rightarrow R$ is an integrable function such that K is continuous and such that $\hat{K}^{-1}(0)$ includes no proper cone, then $S_1(K)$ is dense in $C(W)$ with respect to the norm $\|\cdot\|_\infty$ for any compact subset W of R^r. Here the cone in R^r is defined as a set $C \subset R^r$ such that $x \in C$ implies that $\alpha x \in C$ for all $\alpha \geq 0$. By a proper cone we mean a cone that is neither empty nor the singleton $\{0\}$. In the following theorem, the two conditions on K for the universal approximation using $S_1(K)$ are relaxed: integrability and the cone condition.

1) Low order polynomials (e.g. a constant term [14] or an affine map [15]) are often added to the form (2) to improve performance.

Theorem 1: Suppose that $K: R^r \to R$ is a continuous regular functional in S' with the property that $\hat{K}$ is also a regular element of S' and, for any nonempty open set $U \subset R^r$, there exists a positive σ such that $K_\sigma: R^r \to R$ defined by $K_\sigma(x) = K(-x/\sigma)$ satisfies supp $\hat{K}_\sigma \cap U \neq \phi$. Then $S_1(K)$ is dense in $C(W)$ with respect to the norm $\|\cdot\|_\infty$ for any compact subset W of R^r.

Proof: Consider any compact subset W of R^r. Suppose that $S_1(K)$ is not dense in $C(W)$. Then by the Hahn-Banach theorem [26], there exists a bounded functional Λ on $C(W)$ such that

$$\Lambda \text{ (the closure of } S_1(K)) = \{0\}, \qquad \text{but } \Lambda(C(W)) \neq \{0\}.$$

By the Riesz representation theorem [26], the functional $\Lambda : C(W) \to R$ can be represented by a nonzero finite signed measure μ which is concentrated on W and which satisfies

$$\int_{R^r} K(\frac{x-z}{\sigma})\, d\mu(x) = \int_W K(\frac{x-z}{\sigma})\, d\mu(x) = 0 \tag{4}$$

for any $z \in R^r$ and $\sigma > 0$. Since $K_\sigma: R^r \to R$ is defined by $K_\sigma(x) = K(-x/\sigma)$, the equality (4) can be written in the following convolution form:

$$(K_\sigma * \mu)\ (z) = 0 \tag{5}$$

for any $z \in R^r$ and $\sigma > 0$. Clearly for each $\sigma > 0$, K_σ is an element of S', thus its Fourier transform is also well-defined in S'. Note that an element of S' is a convolute in S' if and only if its Fourier transform is a multiplier in S' [27]. Since μ is of bounded support, μ is a convolute in S' (p.140 of [27]). Thus $\hat{K}_\sigma \hat{\mu}$ is well-defined in S', and indeed by the equality (5) we have

$$\hat{K}_\sigma \hat{\mu} = 0, \qquad \forall \sigma > 0. \tag{6}$$

Note that both $\hat{K}_\sigma$ and $\hat{\mu}$ are regular, thus equation (6) is equivalent to $\forall \sigma > 0$, $\hat{K}_\sigma(x)\hat{\mu}(x) = 0$ for almost every x in R^r.

Consider a subset U of R^r defined by $U \triangleq \hat{\mu}^{-1}(R \backslash \{0\})$. Since $\hat{\mu}$ is continuous(Theorem 14.2 of [27]) and is not identically zero over R^r, U is a nonempty open subset of R^r. Thus by the supposition of the theorem, there exists $\sigma > 0$ such that

$$\text{supp } \hat{K}_\sigma \cap U = \text{supp } \hat{K}_\sigma \cap \hat{\mu}^{-1}(R \backslash \{0\}) \neq \phi.$$

By the definition of support for tempered distributions, the equation (6) means that there exists a test function h on R^r such that supp $h \subset U$ and h is not in the nullity of the regular functional $\hat{K}_\sigma \in S'$; i.e.

$$\int_U \hat{K}_\sigma(x)\, h(x)\, dx \neq 0. \tag{7}$$

Since $\hat{\mu}(x) \neq 0$ for every $x \in U$, h/μ is a well-defined function on U. Therefore we have

$$\int_U \hat{K}_\sigma(x)\, \hat{\mu}(x)\, \frac{h(x)}{\hat{\mu}(x)}\, dx = \int_U \hat{K}_\sigma(x)\, h(x)\, dx \neq 0. \tag{8}$$

Finally, note that equation (8) contradicts the fact obtained in equation (6). Thus the proof is completed. □

Theorem 1 extends the previous result(Theorem 4 of [2]) in two directions: First, the integrability condition for the kernel function K is relaxed, and elements K of S' such that both K and $\hat{K}$ are regular in S' are considered instead. Since the indicated condition is satisfied by not only integrable functions but also many functions outside $L^1(R^r)$ (for example, locally integrable functions $K: R^r \to R$ belonging to the class K [27]), the set of kernel functions K that can guarantee the approximation capability of $S_1(K)$ is enlarged. Next, the cone condition used in Theorem 4 of [2] turns out to be not necessary for the universal approximation, and a weaker condition takes the place of it.

Ⅳ. The RBF as a wavelet in multi-dimensional spaces

The theorem given in section III provides conditions under which functions can be approximated by elements of $S_1(K)$. Since the members of $S_1(K)$ are finite linear combinations of RBFs, a natural related question is whether it is possible to obtain an exact integral representation(not an approximation) in the form of "distributed" RBF networks for a large class of functions. In this section, an affirmative answer is obtained for the integral representation problem from the perspective of wavelet theory.

Wavelet theory is a body of results concerning the study of signals in which interest is centered on the so-called wavelet transforms. Roughly speaking, the transform provides a signal decomposition onto a set of basis functions which are generated by a single function through the operations of dilations and translations. When a function ψ in $L^2(R)$ satisfies the so-called "admissibility" condition

$$0 < c_\psi \triangleq \int_{-\infty}^{\infty} \frac{|\hat{\psi}(w)|^2}{|w|}\, dw < \infty$$

ψ is called a basic(or mother) wavelet. Consider the following family of functions $\psi^{(a,b)}: R \rightarrow R$ defined by

$$\psi^{(a,b)}(x) = \frac{1}{\sqrt{|a|}}\,\psi\left(\frac{x-b}{a}\right), \quad a \in R\backslash\{0\},\ b \in R.$$

The CWT(continuous wavelet transform) of f with respect to the basic wavelet ψ is the function $W_\psi f: (R\backslash\{0\}) \times R \rightarrow R$ defined by

$$(W_\psi f)(a, b) = \int_{-\infty}^{\infty} f(x)\,\psi^{(a,b)}(x)\,dx.$$

If ψ satisfies the admissibility condition, then for any $f \in L^2(R)$ and each $x \in R$ at which f is continuous, an inversion formula exists for $r = 1$ [28]:

$$f(x) = \frac{1}{c_\psi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} (W_\psi f)(a, b)\,\psi^{(a,b)}(x)\,\frac{da}{a^2} \right] db \tag{9}$$

In [3], Park and Sandberg give a limit-in-the-mean representation(Theorem 6 of [3]) of arbitrary $f \in L^1(R^r) \cap L^2(R^r)$ closely related in form to the right side of (9). For the readers' convenience we state the integral representation with some additional definitions. For $w \triangleq (w_1, \cdots, w_r) \in R^r$, $\triangle(w)$ denotes $w_1 \cdots w_r$. For any $K: R^r \rightarrow R$ and any ordered pair

$$(a, b) \triangleq ((a_1, \cdots, a_r), (b_1, \cdots, b_r)) \in (R/\{0\})^r \times R^r$$

$K^{(a,b)}: R^r \rightarrow R$ is defined by

$$K^{(a,b)}(x) = \frac{1}{\sqrt{|\triangle(a)|}}\, K\left(\frac{x_1 - b_1}{a_1}, \cdots, \frac{x_r - b_r}{a_r}\right).$$

Finally, U_ε denotes $(-\infty, -\varepsilon) \cup (\varepsilon, \infty)$ for $\varepsilon > 0$. Then we have the following [3]: If $K: R^r \rightarrow R$ is a square-integrable function that satisfies

$$0 < c_K \triangleq \int_{R^r} \frac{|\hat{K}(w)|^2}{|\triangle(w)|}\, dw < \infty,$$

every f in $L^1(R^r) \cap L^2(R^r)$ satisfies

$$\lim_{\varepsilon \rightarrow 0} \left\| f - \frac{1}{c_K} \int_{R^r_\varepsilon} \int_{R^r} \langle f, K^{(a,b)} \rangle K^{(a,b)}(\cdot)\, db\, \frac{1}{\triangle^2(a)}\, da \right\|_2 = 0. \tag{10}$$

As shown in the above, the limit-in-the-mean formula (10) involves a certain form of integration of the basis function of the class $E(K)$. Similar result can be obtained concerning the represention of functions as a limit-in-the-mean of integrals involving the basis function of $S_1(K)$ which is radially symmetric. If $K \in L^2(R^r)$ is radially symmetric, its Fourier transform $\hat{K}$ is also radially symmetric. Thus, for such K, there exists a corresponding complex-valued function η_K defined on $[0, \infty)$ with the property that $\hat{K}(w) = \eta_K(\|w\|)$ almost everywhere in R^r. Using the radial symmetry of K and $\hat{K}$, functions in $L^1(R^r) \cap L^2(R^r)$ can now be represented in a "distributed" version of RBF networks $S_1(K)$, which constitutes our second theorem. The radial symmetry of the kernel function K plays a critical role in the proof of the following theorem. Note that a closely related study [29] contains a similar looking integral representation of L^2 functions. The difference between these two approaches is well discussed in [3].

Theorem 2: Let $K \in L^2(R^r)$ be radially symmetric and satisfy

$$0 < c_K \triangleq \int_0^\infty \frac{1}{\tau} \; |\eta_K(\tau)|^2 \, d\tau < \infty.$$

Then for any f in $L^1(R^r) \cap L^2(R^r)$, we have the following:

(i) The function defined by $b \mapsto \langle f, K(\frac{\cdot - b}{\tau}) \rangle K(\frac{x-b}{\tau})$ is integrable over R^r for each $\tau > 0$ and each $x \in R^r$.

(ii) The function defined by $\tau \mapsto \frac{1}{\tau^{2r+1}} \int_{R^r} \langle f, K(\frac{\cdot - b}{\tau}) \rangle K(\frac{x-b}{\tau})\, db$ is integrable on (ε, ∞) for each $\varepsilon > 0$ and $x \in R^r$.

(iii) The function $I_\varepsilon(x) \triangleq \int_\varepsilon^\infty \int_{R^r} \langle f, K(\frac{\cdot - b}{\tau}) \rangle K(\frac{x-b}{\tau})\, db \, \frac{1}{\tau^{2r+1}}\, d\tau$ belongs to $L^2(R^r)$ for each $\varepsilon > 0$.

(iv) $\lim_{\varepsilon \to 0} \| f - \frac{1}{c_K} \int_\varepsilon^\infty \int_{R^r} \langle f, K(\frac{\cdot - b}{\tau}) \rangle K(\frac{\cdot - b}{\tau})\, db \, \frac{1}{\tau^{2r+1}}\, d\tau \|_2 = 0.$

<u>Proof</u>: Consider any $f \in L^1(R^r) \cap L^2(R^r)$. Note that

$$\int_{R^r} | \langle f, K(\frac{\cdot - b}{\tau}) \rangle K(\frac{x-b}{\tau}) | \, db$$

$$\leq \int_{R^r} \int_{R^r} | f(y)\, K(\frac{y-b}{\tau}) K(\frac{x-b}{\tau}) | \, dy db$$

$$= \int_{R^r} \int_{R^r} | K(\frac{b-y}{\tau}) K(\frac{b-x}{\tau}) | \, db \, |f(y)| \, dy$$

$$\leq \tau^r \|K\|_2^2 \, \|f\|_1 < \infty. \qquad (11)$$

for any $\tau > 0$ and any $x \in R^r$. Thus (i) holds.

Let $\varepsilon > 0$. Note that by (11)

$$\int_\varepsilon^\infty \left| \int_{R^r} \langle f, K(\frac{\cdot - b}{\tau}) \rangle K(\frac{x-b}{\tau})\, db \right| \frac{1}{\tau^{2r+1}}\, d\tau$$

$$\leq \|K\|_2^2 \, \|f\|_1 \int_\varepsilon^\infty \frac{1}{\tau^{r+1}}\, d\tau = \|K\|_2^2 \, \|f\|_1 / (r\varepsilon^r) < \infty$$

for each $x \in R^r$, showing that (ii) is true.

Let g be an arbitrary element of $L^1(R^r) \cap L^2(R^r)$. Note that

$$\int_{R^r} |g(x)| \int_\varepsilon^\infty \int_{R^r} \int_{R^r} |f(y) K(\frac{y-b}{\tau})| \, dy \; |K(\frac{x-b}{\tau})| \, db \, \frac{1}{\tau^{2r+1}}\, d\tau dx$$

$$\leq \|g\|_1 \, \|K\|_2^2 \, \|f\|_1 / (r\varepsilon^r) < \infty.$$

By Fubuni's theorem and Parseval's identity, we have

$$\int_{R^r} g(x)\, I_\varepsilon(x)\, dx$$

$$= \int_{R^r} g(x) \int_\varepsilon^\infty \int_{R^r} \langle f, K(\frac{\cdot - b}{\tau}) \rangle K(\frac{x-b}{\tau})\, db \; \frac{1}{\tau^{2r+1}}\, d\tau dx$$

$$= \int_\varepsilon^\infty \int_{R^r} \left[\int_{R^r} f(y) K(\frac{y-b}{\tau})\, dy \right] \left[\int_{R^r} g(x) K(\frac{x-b}{\tau})\, dx \right] db \, \frac{1}{\tau^{2r+1}}\, d\tau$$

$$= \int_\varepsilon^\infty \int_{R^r} \left[\int_{R^r} \hat{f}(y) \{\hat{K}(\tau\omega)\}^* \exp(2\pi i b\omega)\, d\omega \right] \left[\int_{R^r} \{\hat{g}(\omega)\}^* \hat{K}(\tau\omega) \exp(-2\pi i b\omega)\, d\omega \right] db \, \frac{1}{\tau}\, d\tau.$$

Note that the first expression between brackets is the complex conjugate of the Fourier transform of $\{\hat{f}\}^* \hat{K}(\tau \cdot)$, and the second one is the Fourier transform of $\{\hat{g}\}^* \hat{K}(\tau \cdot)$. Since $f, g \in L^1(R^r) \cap L^2(R^r)$, $\hat{f}$ and $\hat{g}$ are both bounded. Thus, expressions between brackets are all square-integrable. By Parseval's identity and Fubini's theorem again, it follows that

$$\int_{R^r} g(x) I_\varepsilon(x)\, dx$$

$$= \int_\varepsilon^\infty \int_{R^r} \hat{f}(\omega)\{\hat{g}(\omega)\}^* |K(\tau\omega)|^2\, d\omega \frac{1}{\tau}\, d\tau$$

$$= \int_{R^r} [\int_\varepsilon^\infty \frac{1}{\tau} |\hat{K}(\tau\omega)|^2 d\tau] \{\hat{g}(\omega)\}^* f(\omega)\, d\omega.$$

$$= \int_{R^r} [\int_\varepsilon^\infty \frac{1}{\tau} |\eta_K(\tau\|\omega\|)|^2 d\tau] \{\hat{g}(\omega)\}^* f(\omega) |\, d\omega. \quad (12)$$

Hence, $| \int_{R^r} g(x) I_\varepsilon(x)\, dx |$ is bounded by

$$\int_{R^r} [\int_0^\infty \frac{1}{\tau} |\eta_K(\tau\|\omega\|)|^2 d\tau] | \{\hat{g}(\omega)\}^* f(\omega) |\, d\omega.$$

In case that ω is not zero, the expression between brackets remains the same when $\tau\|\omega\|$ is replaced by τ; i.e.

$$\int_\varepsilon^\infty \frac{1}{\tau} |\eta_K(\tau\|\omega\|)|^2 d\tau = \int_0^\infty \frac{1}{\tau} |\eta_K(\tau)|^2 d\tau = c_K \text{ for } \omega \neq 0.$$

Therefore, we have

$$\sup\{ | \int_{R^r} g(x) I_\varepsilon(x)\, dx | : g \in L^1(R^r) \cap L^2(R^r),\ \|g\|_2 = 1\}$$

$$\le c_K \|f\|_2 \|g\|_2 < \infty,$$

which implies that (iii) is true, i.e. $I_\varepsilon \in L^2(R^r)$.

Finally, by (12) we have

$$\langle f - \frac{1}{c_K} I_\varepsilon, g\rangle = \langle f, g\rangle - \frac{1}{c_K}\langle I_\varepsilon, g\rangle$$

$$= \int_{R^r} [1 - \frac{1}{c_K} \int_\varepsilon^\infty \frac{1}{\tau} |\eta_K(\tau\|\omega\|)|^2 d\tau] | \{\hat{g}(\omega)\}^* f(\omega) |\, d\omega.$$

Thus

$$\|f - \frac{1}{c_K} I_\varepsilon\|_2$$

$$= \sup_{g \in L^1 \cap L^2, \|g\|_2 = 1} | < f - \frac{1}{c_K} I_\varepsilon, g > |$$

$$\le \int_{R^r} [1 - \frac{1}{c_K} \int_\varepsilon^\infty \frac{1}{\tau} |\eta_K(\tau\|\omega\|)|^2 d\tau] | \{\hat{g}(\omega)\}^* f(\omega) |\, d\omega.$$

Note that the integrand of the right side of this inequality tends to 0 as ε goes to 0, and is bounded by $2 | \{\hat{g}(\omega)\}^* f(\omega) |$ which is integrable. Therefore, by Lebesgue's dominated convergence theorem, we have

$\lim_{\varepsilon \to 0} \|f - \frac{1}{c_K} I_\varepsilon\|_2 = 0$, which completes the proof. □

V. Concluding remarks

In this paper, we addressed questions concerning the approximation of functions in a certain general space using RBF networks and some related problems.

We provided an introduction to RBF networks with a partition of typical RBF networks. Theorem 1 of this paper gives sharp conditions under which real-valued functions of several real variables can be approximated arbitrarily well by a certain class of RBF networks. The theorem enlarges the set of RBF networks that have guaranteed approximation capability. Our second theorem provides a limit-in-the-mean formula for representing functions in $L^1(R^r) \cap L^2(R^r)$ in the form of distributed RBF networks. Provided that the kernel function K satisfies the conditions indicated and the integrand is sufficiently smooth, the formula in the theorem can be discretized to give an approximation of the inverse wavelet transformation

$$\lim_{\varepsilon \to 0} \frac{1}{c_K} \int_\varepsilon^\infty \int_{R^r} < f, K(\frac{\cdot - b}{\tau}) > K(\frac{\cdot - b}{\tau})\, db \frac{1}{\tau^{2r+1}}\, d\tau,$$

which will be an element of $S_1(K)$ approximating the original function f. Thus the integral representation in Theorem 2 can be a good starting point for the construction of an RBF network which approximates a given function within a specified error bound.

Finally, issues yet to be investigated in further studies include derivation of a concrete algorithm for the discretization mentioned above, and an investigation of relation between error bound and number of hidden nodes for each class of RBF networks.

References

1. J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," Neural Computation, vol. 3, pp. 246-257, 1991.
2. J. Park and I. W. Sandberg, "Approximation and radial-basis-function networks," Neural Computation, vol. 5, pp. 305-316, 1993.
3. J. Park and I. W. Sandberg, "Nonlinear approximation using elliptic basis function networks," Circuits, Systems, and Signal Processing, vol. 13, pp. 99-113, 1993.
4. E. Parzen, "On the estimation of a probability density function and mode," Annals of Mathematical Statistics, vol. 33, pp. 1065-1076, 1962.
5. E. J. Hartman, J. D. Keeler, and J. M. Kowalski, "Layered neural networks with gaussian hidden units as universal approximations," Neural Computation vol. 2, pp. 210-215, 1990.
6. R. M. Kil, "Function approximation based on a network with kernel functions of bounds and locality: an approach of non-parametric estimation," ETRI Journal, vol. 15, pp. 35-51, 1993.
7. V. Corradi and H. White, "Regularized neural networks: some convergence rate results," Neural Computation, vol. 7, pp. 1225-1244, 1995.
8. O. Rioul and M. Vetterli, "Wavelets and signal processing," IEEE Signal Processing Magazine, vol. 8, pp. 14-38, 1991.
9. E. M. Stein and G. Weiss, Introduction to Fourier analysis on Euclidean spaces, Princeton University Press, Priceton, New Jersey, 1971.
10. T. Poggio and F. Girosi, "Networks for approximation and learning," Proceedings of the IEEE, vol. 78, pp. 1481-1497, 1990.
11. V. A. Morozovc, Methods for solving incorrectly posed problems, Springer-Verlag, New York, 1984.
12. D. S. Broomhead and D. Lowe, "Multi-variable functional interpolation and adaptive networks," Complex Systems, vol. 2, pp. 321-355, 1988.
13. D. G. Bounds, P. J. Lloyd, and B. G. Mathew, "A comparison of neural network and other pattern recognition approaches to the diagnosis of low back disorder," Neural Networks, vol. 3, pp. 583-591, 1990.
14. C. Bishop, "Improving the generalization properties of radial basis function neural networks," Neural Computation, vol. 3, pp. 579-588, 1991.
15. A. I. Mees, M. F. Jackson, and L. O. Chua, "Device modelling by radial-basis-functions," IEEE Transactions on Circuits and Systems, vol. I-39, pp. 19-27, 1992.
16. J. Ghosh, S. Chakravarthy, Y. Shin, and C. Chu, "Adaptive kernel classifiers for short-duration oceanic signals," Proceedings of the IEEE Conference on Neural Networks for Ocean Engineering, pp. 41-48, 1991.
17. J. Ghosh, L. Deuser, and S. Beck, "A neural network based hybrid system for detection, characterization and classification of short-duration oceanic signals," IEEE Journal on Ocean Engineering, Oct. 1992.
18. J. A. Leonard and M. A. Kramer, "Radial basis function networks for classifying process faults," IEEE Control Systems Magazine, vol. 11, pp. 31-38, 1991.
19. K. J. Cious, R. E. Tjia, N. Liu, and R. A. Langenderfer, "Study of continuous ID3 and radial basis function algorithms for the recognition of glass defects," Proceedings of 1991 IEEE INNS International Joint Conference on Neural Networks, vol. 1, pp. 49-54, 1991.
20. L. A. Leonard, M. A. Kramer, and L. H. Ungar, "Using radial basis functions to approximate a function and its error bounds," IEEE Transactions on Neural Networks, 1992.
21. R. D. Jones, C. W. Barnes, Y. C. Lee, and W. C. Mead, "Information theoretic derivation of network architecture and learning algorithms," Proceedings of 1991 IEEE INNS International Joint Conference on Neural Networks, vol. 2, pp. 473-478, 1991.

22. E. J. Hartman and J. D. Keeler, "Predicting the future: Advantage of semilocal units," Neural Computation, vol. 3, pp. 566-578, 1991.
23. S. V. Chakravarthy, Capability of adaptive networks with gaussian basis functions, Master's thesis, The University of Texas at Austin, 1991.
24. N. Weymaere and J. Martens, "A fast and robust learning algorithm for feedforward neural networks," Neural Networks, vol. 4, pp. 361-370, 1991.
25. S. Lee and R. M. Kil, "A gaussian potential function network with hierarchically self-organizing learning," Neural Networks, vol. 4, pp. 207-224, 1991.
26. R. Rudin, Real and complex analysis, third edition, McGraw-Hill, New York, 1987.
27. D. C. Champeney, A handbook of Fourier theorems, Cambridge University Press, Cambridge, UK, 1987,
28. C. K. Chui, An introduction to wavelets, Academic Press, Boston, 1992.
29. I. Daubechies, Ten lectures on wavelets, SIAM, Philadelphia, 1992.

박 주 영(Jooyoung Park) 정회원

1961년 2월 13일생

1983년 2월 : 서울대 전기공학과 졸업(공학사)

1992년 : University of Texas at Austin 전기 및 컴퓨터 공학과 졸업(공학박사)

1993년~현재 : 고려대 제어계측공학과 조교수

주관심분야 : 신경망과 제어이론