Proceedings of 1993 International Joint Conference on Neural Networks

APPROXIMATION CAPABILITY TO FUNCTIONS OF SEVERAL VARIABLES, NONLINEAR FUNCTIONALS AND OPERATORS BY RADIAL BASIS FUNCTION NEURAL NETWORKS

Tianping Chen
Dept. of Mathematics
Fudan University
Shanghai
P. R. China

Hong Chen
Dept. of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556
U.S.A.

ABSTRACT

The purpose of this paper is to explore the representation capability of radial basis function (RBF) neural networks. The main results are: (1) The ***necessary and sufficient*** condition for a function of one variable to be qualified as an activation function in RBF network is that the function is not an even polynomial. (2) The capability of approximating nonlinear ***functionals*** and ***operators*** by RBF networks is revealed, using sample data either in frequency domain or in time domain.

1 Introduction

There have been several recent works concerning the representation capabilities of multilayer feedforward neural networks. For example, in 1989 several papers (see [1-3] and many others) related to this topic have appeared. They all claimed that a three-layered neural network with sigmoid units on the hidden layer can approximate continuous or other kinds of functions defined on compact sets in $\mathbf{R}^n$. In many of those papers, sigmoidal functions need to be assumed to be continuous or monotone. Recently [8], we pointed out that the ***boundedness*** of the sigmoidal function plays an essential role for its being an activation function in the hidden layer, *i.e.*, instead of continuity or monotonity, the boundedness of sigmoidal functions ensures the network capability.

In addition to sigmoidal functions, many other functions can be used as activation functions of neural networks. For example, in [4] Hornik proved that any bounded non-constant continuous functions is qualified to be an activation function.

Recently [10] we proved that for a continuous function to be an activation function in feedforward neural networks, the necessary and sufficient condition is that the function is not a polynomial.

The above papers are all significant advances towards solving the problem of whether a function is qualified as an activation function in neural networks. However, they only dealt with affine-basis-function neural networks (also called multilayer perceptrons (MLP)) and the goal there is to approximate continuous functions by the family

$$\sum_{i=1}^{N} c_i\, g(y_i \cdot x + \theta_i)$$

where $y_i \in \mathbf{R}^n$, $c_i, \theta_i \in \mathbf{R}^1$ and $y_i \cdot x$ denotes the inner product of y and x.

Among the various kinds of promising neural networks currently under active research, there is another type called ***radial-basis-function (RBF) networks*** [7] (also called *localized receptive field network*), where the activation functions are radially symmetric and produce a localized response to input stimulus. A block diagram of an RBF network is shown in Fig. 1. One of the special basis functions that are commonly used is a Gaussian kernel function. Using the Gaussian basis function, RBF networks is capable of forming an arbitrarily close approximation to any continuous functions, as shown by [5].

More generally, the goal here is to approximate functions of a finite number of real variables by

$$\sum_{i=1}^{N} c_i\, g(\lambda_i \|x - x_i\|_{\mathbf{R}^n})$$

where $x \in \mathbf{R}^n$ and $\|x - x_i\|_{\mathbf{R}^n}$ is the distance between x and x_i in $\mathbf{R}^n$. Here, the activation function g is not necessarily Gaussian. In this direction, several results concerning RBF neural networks were obtained [6] [7]. In [6], Sandberg proved the following theorem:

Let $K : \mathbf{R}^n \to \mathbf{R}$ be a radial symmetric, integrable, bounded function such that K is continuous almost everywhere and $\int_{\mathbf{R}^n} K(x)dx \neq 0$, then the

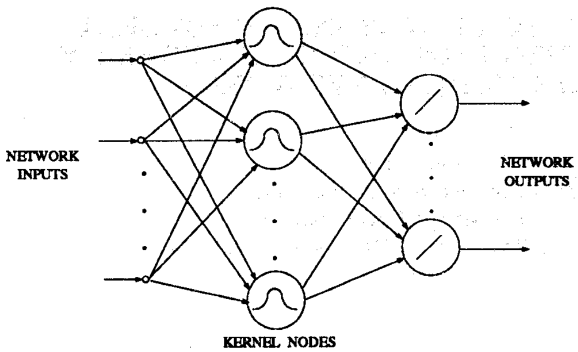

Figure 1: A Radial Basis Function Network

family

$$\sum_{i=1}^{N} c_i\, g(\frac{\|x - x_i\|}{\sigma})$$

is dense in $L^p(\mathbf{R}^n)$, where $g(\|x\|) = K(x)$.

In [7], Lippmann used smooth technique and approximated the convolution by Riemann sum.

It is natural to raise the following questions: (1) What is the *necessary* and *sufficient* condition for a function to be qualified as an activation function in RBF neural networks? (2) How to approximate nonlinear *functionals* by RBF neural networks? (3) How to approximate nonlinear *operators* (eg. the output of a dynamical system) by RBF neural networks, using sample data in frequency (phase) domain or in time (state) domain?

The purpose of this paper is to give strong results in answering those questions.

The paper is organized as follows. In Section 2, we review some definitions and notations. In Section 3, we show that the necessary and sufficient condition for a continuous function to be qualified as an activation function in RBF networks is that it is not an even polynomial. In Section 4, we show the capability of RBF neural networks to approximate nonlinear functionals and operators on some Banach space as well as on some compact set in $C(K)$, where K is a compact set in any Banach space. Furthermore, we establish the capability of neural networks to approximate nonlinear *operators* from $C(K_1)$ to $C(K_2)$. Approximations using samples in both frequency domain and time domain are discussed. Examples are given, which includes the use of wavelets to the approximation. It is also pointed out that the main results in Section 4 has a direct application to computing nonlinear dynamical systems. We conclude this paper with Section 5.

The proofs for the theorems in this paper are omitted here due to length restriction. They can be found in [11].

2 Notations and Definitions

Definition 1. A function $\sigma : \mathbf{R}^1 \to \mathbf{R}^1$ is called a sigmoidal function, if it satisfies

$$\begin{cases} \lim_{x\to-\infty} \sigma(x) = 0\ , \\ \lim_{x\to\infty} \sigma(x) = 1\ . \end{cases} \tag{1}$$

Definition 2. A polynomial of one variable $p(x)$ is called an even polynomial, if it is an even function, namely $p(-x) = p(x)$.

Definition 3. Suppose that X is a Banach space, $V \subseteq X$ is called a compact set in X, if for every sequence $\{x_n\}_{n=1}^{\infty}$ with all $x_n \in V$, there is a subsequence $\{x_{n_k}\}$ which converges to some element $x \in V$.

It is well known that if $V \subseteq X$ is a compact set in X, then for any $\delta > 0$, there is a δ-net $N(\delta) = \{x_1, \ldots, x_{n(\delta)}\}$, with all $x_i \in V$, $i = 1, \ldots, n(\delta)$, *i.e.* for every $x \in X$, there is some $x_i \in N(\delta)$ such that $\|x_i - x\|_X < \delta$.

Definition 4. Let X be a Banach space with norm $\|\cdot\|_X$. If there are elements $x_n \in X$, $n = 1, 2, \ldots$, such that for every $x \in X$ there is a unique real number sequence $a_n(x)$, such that

$$x = \sum_{n=1}^{\infty} a_n(x) x_n\ ,$$

where the series converges in X, then $\{x_n\}_{n=1}^{\infty}$ is called a Schauder basis in X and X is called a Banach space with Schauder basis.

In the sequel, we will often use the following notations.

X: some Banach space with norm $\|\cdot\|_X$.

$\mathbf{R}^n$: Euclidean space of dimension n with norm $\|\cdot\|_{\mathbf{R}^n}$.

K: some compact set in a Banach space.

$C(K)$: Banach space of all continuous functions defined on K, with norm $\|f\|_{C(K)} = \max_{x\in K} |f(x)|$.

V: some compact set in $C(K)$.

$S(\mathbf{R}^n)$: Schwartz functions in distribution theory.

$S'(\mathbf{R}^n)$: All the distributions defined on $S(\mathbf{R}^n)$.

$C^\infty(\mathbf{R}^n)$: All infinitely differentiable function with compact support in $\mathbf{R}^n$.

3 Characteristics of Continuous Functions as Activation in RBF Networks

In this section, we show that for a continuous function to be qualified as an activation function in RBF networks, the *necessary* and *sufficient* condition is that it is not an even polynomial and we prove two approximation theorems by RBF networks. More precisely, we prove

Theorem 1 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$, then the family*

$$\sum_{i=1}^{N} c_i\, g(\lambda_i \|x - y_i\|_{\mathbf{R}^n})$$

is dense in $C(K)$, if and only if g is not an even polynomial, where K is a compact set in $\mathbf{R}^n$, $y_i \in \mathbf{R}^n$, $c_i, \lambda_i \in \mathbf{R}^1$, $i = 1, \ldots, N$.

Theorem 2 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$ and is not an even polynomial, K is a compact set in $\mathbf{R}^n$, V is a compact set in $C(K)$, then for any $\epsilon > 0$, there are an integer N, $\lambda_i \in \mathbf{R}^1$, $y_i \in \mathbf{R}^n$, $i = 1, \ldots, N$, which are all independent of $f \in V$ and constants $c_i(f)$ depending on f, $i = 1, \ldots, N$, such that*

$$|f(x) - \sum_{i=1}^{N} c_i(f)\, g(\lambda_i \|x - y_i\|_{\mathbf{R}^n})| < \epsilon \qquad (2)$$

holds for all $x \in K$, $f \in V$. Moreover, every $c_i(f)$ is a continuous functional defined on V.

Remark 1. It is worth noting that λ_i and y_i are all independent of f in V and $c_i(f)$ are continuous functionals, which will play an important role in approximation to nonlinear operators by RBF networks.

4 Approximation to Nonlinear Functionals and Operators by RBF Neural Networks

In this section, we show some results concerning capability of RBF neural networks in approximating nonlinear functionals and operators defined on some Banach space, which can be used to approximate outputs of dynamical systems using sample data in either frequency (phase) domain or time (state) domain. Our main results are:

Theorem 3 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$ and is not an even polynomial, X is a Banach space with Schauder basis $\{x_n\}_{n=1}^{\infty}$, $K \subseteq X$ is a compact set in X, f is a continuous functional defined on K. Then for any $\epsilon > 0$, there are positive integers M, N, $y_i^M \in \mathbf{R}^M$, constants $c_i, \lambda_i \neq 0$, $i = 1, \ldots, N$, such that*

$$|f(x) - \sum_{i=1}^{N} c_i\, g(\lambda_i \|x^M - y_i^M\|_{\mathbf{R}^M})| < \epsilon \qquad (3)$$

holds for all $x \in K$, where $x^M = (a_1(x), \ldots, a_M(x)) \in \mathbf{R}^M$, $x = \sum_{n=1}^{\infty} a_n(x) x_n$.

To illustrate the applications of Theorem 3, we now give some examples.

Example 1. Let H be a Hilbert space, $K \subseteq H$ be a compact set in H. As we showed in [12], K can be considered as a compact set in a Hilbert space H_1 with countable basis $\{x_n\}_{n=1}^{\infty}$. Thus, Theorem 3 can be applied and we conclude that every continuous functional can be arbitrarily well approximated by the sum

$$\sum_{i=1}^{N} c_i\, g(\lambda_i \|x^M - y_i^M\|_{\mathbf{R}^M})$$

Example 2. Let $X = L^2[0, 2\pi]$, then $\{1, \{\cos nx\}_{n=1}^{\infty}, \{\sin nx\}_{n=1}^{\infty}\}$ is a Schauder basis and $a_n(x)$ are just the Fourier coefficients correspondingly. Thus, we can approximate every nonlinear functional defined on some compact set in $L^2[0, 2\pi]$ by RBF neural networks using sample data in *frequency domain.*

Example 3. Let $X = L^2(\mathbf{R}^1)$ and $\{\psi_{j,k}\}_{j,k=1}^{\infty}$ be wavelets in $L^2(\mathbf{R}^1)$. Then we can approximate continuous functionals defined on a compact set in $L^2(\mathbf{R}^1)$ by RBF neural network using *wavelet* coefficients as sample data (also in frequency domain).

Remark 2. Since in Theorem 3 we only require that $\{x_n\}_{n=1}^{\infty}$ is a Schauder basis (no orthogonal requirement is imposed), therefore we don't require that $\{\psi_{j,k}\}_{j,k=1}^{\infty}$ be orthogonal wavelets. This is a property of significant advantage, for non-orthogonal wavelets are much easier to construct than orthogonal wavelets.

The following theorem shows the possibility of approximating functionals by RBF neural networks using sample data in *time* (or state) domain.

Theorem 4 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$ is not an even polynomial, X is a Banach space, $K \subseteq X$ is a compact set, V is a compact set* *in $C(K)$, f is a continuous functional defined on V, then for any $\epsilon > 0$, there are positive integers N, M, $x_1, \ldots, x_M \in K$ and $\lambda_i, c_i \in \mathbf{R}^1$, $\xi_i = (\xi_{i1}, \ldots, \xi_{iM}) \in \mathbf{R}^M$, $i = 1, \ldots, N$ such that*

$$|f(u) - \sum_{i=1}^{N} c_i\, g(\lambda_i \|u^M - \xi_i\|_{\mathbf{R}^M})| < \epsilon \quad (4)$$

for all $u \in V$, where $u^M = (u(x_1), \ldots, u(x_M)) \in \mathbf{R}^M$.

To conclude this section, we construct an RBF neural network to approximate nonlinear operators. More precisely, we prove

Theorem 5 *Suppose that $g \in C(\mathbf{R}^1) \cap S'(\mathbf{R}^1)$ and is not an even polynomial, X is a Banach space, $K_1 \subseteq X$, $K_2 \subseteq \mathbf{R}^n$ are two compact sets in X and $\mathbf{R}^n$ respectively, V is a compact set in $C(K_1)$, G is a nonlinear continuous operator, which maps V into $C(K_2)$, then for any $\epsilon > 0$, there are positive integers M, N, m, constants $c_i^k, \lambda_k, \mu_i \in \mathbf{R}^1$, $k = 1, \ldots, N$, $i = 1, \ldots, M$, m points $x_1, \ldots, x_m \in K_1$, $\omega_1, \ldots, \omega_N \in \mathbf{R}^n$, such that*

$$|G(u)(y) - \sum_{i=1}^{M} \sum_{k=1}^{N} c_i^k\, g(\mu_i \|u^m - \xi_{ik}^m\|_{\mathbf{R}^m}) \\ g(\lambda_k \|y - \omega_k\|_{\mathbf{R}^n})| < \epsilon \quad (5)$$

for all $u \in V$ and $y \in K_2$, where $u^m = (u(x_1), \ldots, u(x_m))$, $\xi_k^m = (\xi_{k1}^m, \ldots, \xi_{k,m}^m)$, $k = 1, \ldots, N$.

Remark 3. Theorem 5 shows the capability of RBF networks in approximating nonlinear operators using sample data in *time* (or state) domain. Likewise, by Theorem 3, we can construct RBF networks using sample data in *frequency* (or phase) domain.

Remark 4. We can also construct neural networks, where affine basis functions are mixed with radial basis functions. For example, we can approximate $G(u)(y)$ by

$$\sum_{k=1}^{N} \sum_{i=1}^{M} c_i^k\, g(\mu_i \|u^m - \xi_{ik}^m\|_{\mathbf{R}^m})\, g(\omega_k \cdot y + \zeta_k)\ .$$

The details are omitted here.

Remark 5. Since a dynamical system can be viewed as an operator between inputs and outputs, Theorem 5 thus shows the capability of computing the output of a dynamical system by RBF neural networks.

5 Conclusion

In this paper, the problem of approximating functions of several variables, functionals and nonlinear operators by radial basis function neural networks are studied. The necessary and sufficient condition for a continuous function to be qualified as an activation function in RBF networks is given. Results on using RBF neural networks for computing the output of dynamical systems by sample data in frequency domain or time domain are given.

References

[1] G. Cybenko, "Approximation by Superpositions of a Sigmoidal Function," in *Math. of Control, Signals and Systems*, Vol. 2, No. 4, pp. 303-314 (1989).

[2] S. M. Carroll and B. W. Dickinson, "Construction of Neural Nets using Radon Transform," in *IJCNN Proc. I*, pp. 607-611 (1989).

[3] K. Hornik, M. Stinchcombe and H. White, "Multi-layer Feedforward Networks are Universal Approximators," *Neural Networks*, Vol. 2, pp. 359-366 (1989).

[4] K. Hornik, "Approximation Capabilities of Multilayer Feedforward Networks," *Neural Networks*, Vol. 4, pp. 251-257 (1991).

[5] E.J. Hartman, J.D. Keeler and J.M. Kowalski, "Layered Neural networks with Gaussian Hidden Units as Universal Approximators," *Neural Computation*, Vol. 2(2) (1990).

[6] J. Park and I.W. Sandberg, "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation*, pp. 246-257, Vol. 3 (1991).

[7] R.P. Lippmann, "Pattern Classification using Neural Networks," *IEEE Commun. Magazine*, pp. 47-64, Vol. 27 (1989).

[8] T. Chen, H. Chen and R.-W. Liu, "A Constructive Proof of Cybenko's Approximation Theorem and Its Extensions," pp. 163 - 168 in *Computing Science and Statistics* (editors LePage and Page), Springer-Verlag, ISBN 0-387-97719-8 (1990).

[9] T. Chen and H. Chen, "Approximation to Continuous Functionals by Neural Networks with Application to Dynamic Systems," *IEEE Trans. on Neural Networks* (in press).

[10] T. Chen and H. Chen, "Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamic Systems," submitted for publication.

[11] T. Chen and H. Chen, "Approximation Capability to Functions of Several Variables, Nonlinear Functionals and Operators by Radial Basis Function Neural Networks," submitted for publication.

[12] T. Chen, "Approximation to Nonlinear Functionals in Hilbert Space by Superposition of Sigmoidal Functions," *Kexue Tongbao* (1992).