Journal of Mathematical Research & Exposition
Jan., 2009, Vol. 29, No. 1, pp. 124–128
DOI:10.3770/j.issn:1000-341X.2009.01.016
Http://jmre.dlut.edu.cn

$L^p(K)$ Approximation Problems in System Identification with RBF Neural Networks

NAN Dong, LONG Jin Ling
(Applied Mathematics Department, Dalian University of Technology, Liaoning 116024, China)
(E-mail: Nandong2004@yahoo.com.cn; jinling_long@sina.com)

Abstract L^p approximation problems in system identification with RBF neural networks are investigated. It is proved that by superpositions of some functions of one variable in $L^p_{\text{loc}}(\mathbb{R})$, one can approximate continuous functionals defined on a compact subset of $L^p(K)$ and continuous operators from a compact subset of $L^{p_1}(K_1)$ to a compact subset of $L^{p_2}(K_2)$. These results show that if its activation function is in $L^p_{\text{loc}}(\mathbb{R})$ and is not an even polynomial, then this RBF neural networks can approximate the above systems with any accuracy.

Keywords RBF neural networks; system identification; L^p-approximation; continuous functionals and operators.

Document code A
MR(2000) Subject Classification 41A65; 92B20
Chinese Library Classification O174.41

1. Introduction

One of the most important theoretical questions for radial basis function (RBF) neural networks is their approximation capabilities: Under what conditions on the activation function and the structure of neural networks (the quantity of hidden neurons and the choice of weights), can the neural networks approximate to any accuracy a function or a continuous operator? There have been many papers related to this topic, Pinkus[6], Chen[1−3,9], Leshno[4], Mhaskar[5] and Jiang[10,11], among many others.

In this paper, we apply the method used in [9], [10]. Combining with one of our latest results, we make a progress in system approximation by RBF neural networks. Our main results improve the existing result in [10], which needs an extra requirement $g(\cdot) \in \mathcal{S}'(\mathbb{R}^n)$.

The rest of the paper is organized as follows. Section 2 presents a few important lemmas. Some of the lemmas in Section 2 have been proved[1,7,9−11]. Our main results and their proofs are given in Section 3.

2. Lemmas

Lemma 2.1[10] *V is a compact set in $L^p(K)$ if and only if*

1) V is a closed set in $L^p(K)$;

Received date: 2007-01-03; **Accepted date**: 2007-05-26
Foundation item: the National Natural Science Foundation of China (No. 10471017).

2) There exists a constant $A > 0$, such that $\|f\|_{L^p(K)} \le A$ for all $f \in V$;

3) If $h \to 0$, $\|f_h - f\|_{L^p(K)}$ converges to 0 uniformly for all $f \in V$,

where $f_h(x) = \frac{1}{m(B(x,h))}\int_{B(x,h)\cap K} f(t)\mathrm{d}t$, $B(x,h)$ is a spheroid with center at x and radius h, and $m(B(x,h))$ is the volume of $B(x,h)$.

Lemma 2.2[10] *Suppose V is a compact set in $L^p(K)$, $V_h = \{f_h : f \in V\}$. Then V_h is a compact set of $C(K)$ for fixed $h > 0$.*

Lemma 2.3[10] *Suppose that V is a compact set in $L^p(K)$, F is a continuous functional on V, and $g(t) \in C(\mathbb{R}^1_+)$ is not an even polynomial. Then for any $\varepsilon > 0$, there exist N, $M \in \mathbb{N}$, $c_i \in \mathbb{R}^1$, n, $\lambda_i \in \mathbb{R}^1_+$, $\theta_i \in \mathbb{R}^M$ $(i = 1, 2, \ldots, N)$ and $x_1, \ldots, x_M \in K$ (K is a compact set in $\mathbb{R}^n$), such that $|F(u) - \sum_{i=1}^N c_i g(\lambda_i \|u^M_{\frac{1}{n}} + \theta_i\|)| < \varepsilon$ for all $u \in V$, where $u^M_{\frac{1}{n}} = (u_{\frac{1}{n}}(x_1), \ldots, u_{\frac{1}{n}}(x_M))'$, $'$ means transposition of a vector in Euclidean space, $u_{\frac{1}{n}}(x) = \frac{1}{m(B(x,\frac{1}{n}))}\int_{B(x,\frac{1}{n})\cap K} u(t)\mathrm{d}t$.*

Lemma 2.4[13] *Suppose $\sigma(x) \in L^p_{\mathrm{loc}}(\mathbb{R}^n)$. Then $\{\sum_{i=1}^N c_i\sigma(\lambda_i x + \theta_i)\}$ is dense in $L^p(K)$ if and only if σ is not a polynomial in R^n, where $1 \le p < \infty$, K is a compact set in $\mathbb{R}^n$, $N \in \mathbb{N}$, c_i, $\lambda_i \in \mathbb{R}^1$, x, $\theta_i \in \mathbb{R}^n$.*

Lemma 2.5[13] *Suppose that $g : \mathbb{R}^1_+ \to \mathbb{R}^1$ and $g(\|x\|_{\mathbb{R}^n}) \in L^p_{\mathrm{loc}}(\mathbb{R}^n)$. Then $\{\sum_{i=1}^N c_i g(\lambda_i\|x - y_i\|_{\mathbb{R}^n})\}$ is dense in $L^p(K)$ if and only if g is not an even polynomial in R^1, where $1 \le p < \infty$, K is a compact set $\mathbb{R}^n$, $i = 1, 2, \ldots, N$, $N \in \mathbb{N}$, $c_i \in \mathbb{R}^1$, $\lambda_i \in \mathbb{R}^1_+$, x, $y_i \in \mathbb{R}^n$.*

3. Main result and proof

Theorem 3.1 *Suppose that K and V are compact sets in $\mathbb{R}^n$ and $L^p(K)$, respectively, and $\sigma(x) \in L^p_{\mathrm{loc}}(\mathbb{R}^n)$ is not a polynomial $(1 \le p < \infty)$. Then for any $\varepsilon > 0$, there exist $N \in \mathbb{N}$, $\lambda_i \in \mathbb{R}^1$, $b_i \in \mathbb{R}^n$ which are independent of f, and constants $c_i(f)$ $(i = 1, 2, \ldots, N)$ depending on f, such that*

$$\Big\|f(x) - \sum_{i=1}^N c_i(f)\sigma(\lambda_i x - b_i)\Big\|_{L^p(K)} < \varepsilon$$

holds for all $f \in V$. Especially all $c_i(f)$ are continuous functionals on V.

Proof For any $f \in V$ and any $\varepsilon > 0$, there exists $h_0 > 0$, such that

$$\|f_h(x) - f(x)\|_{L^p(K)} < \frac{\varepsilon}{4} \tag{1}$$

holds when $0 < h < h_0$ by Lemma 2.1. Let $\Phi(x) = ce^{-\|x\|}$ with c being a constant satisfying $\int_{\mathbb{R}^n}\Phi(x)\mathrm{d}x = 1$. We define $\Phi_\delta(x) = \delta^{-n}\Phi(\delta^{-1}x)$ $(\delta \in \mathbb{R}^1\backslash 0)$. Obviously, we have $\int_{\mathbb{R}^n}\Phi_\delta(x)\mathrm{d}x = 1$ for any given δ. Define $f_h \star \Phi_\delta(x)$ by

$$f_h \star \Phi_\delta(x) = \int_K f_h(t)\Phi_\delta(x - t)\mathrm{d}t.$$

For any given h satisfying $0 < h < h_0$, according to Lemma 2.2 and Minkowski inequality, there exists δ_0, such that

$$\|f_h \star \Phi_\delta(x) - f_h(x)\|_{L^p(K)} < \frac{\varepsilon}{4} \tag{2}$$

holds for all $f_h \in V_h$ when $0 < \delta < \delta_0$. Next, we write $f_h \star \Phi_\delta(x)$ into Riemann sum $\sum_{j=1}^{M} f_h(t_j)\Phi_\delta(x-t_j)m(\Delta t_j)$, where $\bigcup_{j=1}^{M}\Delta t_j$ is a finite segmentation of K, $t_j \in \Delta t_j$. Then we have

$$\begin{aligned}&\int_K f_h(t)\Phi_\delta(x-t)\mathrm{d}t - \sum_{j=1}^{M} f_h(t_j)\Phi_\delta(x-t_j)m(\Delta t_j)\\ &= \sum_{j=1}^{M}\int_{\Delta t_j}(f_h(t)-f_h(t_j))\,\Phi_\delta(x-t)\mathrm{d}t + \sum_{j=1}^{M} f_h(t_j)\int_{\Delta t_j}(\Phi_\delta(x-t)-\Phi_\delta(x-t_j))\,\mathrm{d}t\\ &= I_1 + I_2. \end{aligned} \tag{3}$$

It follows from Minkowski inequality that

$$\begin{aligned}\|I_1\|_{L^p(K)} &= \Big(\int_K \Big|\sum_{j=1}^{M}\int_{\Delta t_j}(f_h(t)-f_h(t_j))\Phi_\delta(x-t)\mathrm{d}t\Big|^p \mathrm{d}x\Big)^{\frac{1}{p}}\\ &\le \Big(\int_K \sum_{j=1}^{M}\int_{\Delta t_j}|(f_h(t)-f_h(t_j))|^p|\Phi_\delta(x-t)|^p\mathrm{d}t\mathrm{d}x\Big)^{\frac{1}{p}}. \end{aligned} \tag{4}$$

Since Φ_δ is a continuous function, there exists $Q > 0$, such that $\int_K\int_K |\Phi_\delta(x-t)|^p \mathrm{d}x\mathrm{d}t < Q$. Note that $f_h \in V_h$ and that V_h is a compact set in $C(K)$ by Lemma 2.2. According to Arzeta-Ascoli Theorem on compact set, there exists $\eta_1 > 0$, such that $|f_h(t) - f_h(t_j)| < \frac{\varepsilon}{8Q}$ when $\max_{j=1,\dots,M}\{\mathrm{diam}(\Delta t_j)\} < \eta_1$. Then we can obtain

$$\|I_1\|_{L^p(K)} < \frac{\varepsilon}{8}. \tag{5}$$

Similarly, we have

$$\|I_2\|_{L^p(K)} < \frac{\varepsilon}{8}. \tag{6}$$

It follows from equations (1), (2), (3), (5) and (6) that

$$\Big\|f(x) - \sum_{j=1}^{M} f_h(t_j)m(\Delta t_j)\Phi_\delta(x-t_j)\Big\|_{L^p(K)} < \frac{3\varepsilon}{4}. \tag{7}$$

According to Lemma 2.4, for fixed j $(1 \le j \le M)$, there exist $N_j \in \mathbb{N}$, c_{ij}, $\lambda_{ij} \in \mathbb{R}^1$, $\theta_{ij} \in \mathbb{R}^n$ $(i = 1, 2, \dots, N)$, such that

$$\Big\|\Phi_\delta(x-t_j) - \sum_{i=1}^{N_j} c_{ij}\sigma(\lambda_{ij}x + \theta_{ij})\Big\|_{L^p(K)} < \frac{\varepsilon}{4L}, \tag{8}$$

where $L = \sup_{f\in V}\sum_{j=1}^{M}|f_h(t_j)|m(\Delta t_j)$. Theorem 3.1 follows readily from equations (7) and (8). □

Theorem 3.2 *Suppose that K and V are compact set in $\mathbb{R}^n$ and $L^p(K)$, respectively, $g(\cdot): \mathbb{R}^1_+ \to \mathbb{R}^1$, and $g(\|x\|) \in L^p_{\mathrm{loc}}(\mathbb{R}^n)$ is not an even polynomial $(1 \le p < \infty)$. Then for any $\varepsilon > 0$, there exist $N \in \mathbb{N}$, $\lambda_i \in \mathbb{R}^1_+$, $b_i \in \mathbb{R}^n$, $i = 1, 2, \dots, N$, which are independent of f, and constants*

$c_i(f)$ depending on f, such that

$$\left\| f(x) - \sum_{i=1}^{N} c_i(f) g(\lambda_i \|x - b_i\|) \right\|_{L^p(K)} < \varepsilon$$

holds for all $f \in V$. Especially all $c_i(f)$ are continuous functionals on V.

Proof Similar to the proof of Theorem 3.1, we have equation (7) holds for any $\varepsilon > 0$ and any $f \in V$. It follows from Lemma 2.5 that for any fixed j $(1 \le j \le M)$, there exist $N_j \in \mathbb{N}$, $c_{ij} \in \mathbb{R}^1$, $\lambda_{ij} \in \mathbb{R}^1_+$, $y_{ij} \in \mathbb{R}^n$, $i = 1, 2, \ldots, N$, such that

$$\left\| \Phi_\delta(x - t_j) - \sum_{i=1}^{N_j} c_{ij} g(\lambda_{ij} \|x - y_i\|_{\mathbb{R}^n}) \right\|_{L^p(K)} < \frac{\varepsilon}{4L}, \tag{9}$$

where $L = \sup_{f \in V} \sum_{j=1}^{M} |f_h(t_j)| m(\Delta t_j)$. Equations (7) and (9) lead to Theorem 3.2. □

Theorem 3.3 *Suppose that $g_1(\cdot) \in C(\mathbb{R}^1_+)$, $g_2 : \mathbb{R}^1_+ \to \mathbb{R}^1$, and $g_2(\|x\|) \in L^p_{\rm loc}(\mathbb{R}^{d_2})$, g_1, g_2 are not even polynomials, K_1, K_2 and V are compact sets in $\mathbb{R}^{d_1}$, $\mathbb{R}^{d_2}$ and $L^{p_1}(K_1)$, respectively, and $T : V \to L^{p_2}(K_2)$ is a continuous operator $(1 \le p_1, p_2 < \infty)$. Then, for any $\varepsilon > 0$, there exist N, N_i, $M_i \in \mathbb{N}$, n_i, η_i, $\lambda_{ik} \in \mathbb{R}^1_+$, $b_i \in \mathbb{R}^{d_2}$, $c_{ik} \in \mathbb{R}^1$, $\theta_{ik} \in \mathbb{R}^{M_i}$, $x_{ij} \in K_2$ $(i = 1, \ldots, N$, $k = 1, \ldots, N_i$, $j = 1, \ldots, M_i)$, such that*

$$\left\| T(u)(y) - \sum_{i=1}^{N} \sum_{k=1}^{N_i} c_{ik} g_1(\lambda_{ik} \|x^{M_i} - \theta_{ik}\|) g_2(\eta_i \|y - b_i\|) \right\|_{L^{p_2}(K_2)} < \varepsilon$$

holds for all $u \in V$, where $x^{M_i} = (u_{\frac{1}{n_i}}(x_{i1}), \ldots, u_{\frac{1}{n_i}}(x_{iM_i}))'$, $u_h(x) = \frac{1}{m(B(x,h))} \int_{B(x,h) \cap K_2} u(t) \mathrm{d}t$, $'$ means transposition of a vector in Euclidean space.

Proof Since T is a continuous operator, $T(V) = \{T(u) : u \in V\}$ is a compact set in $L^{p_2}(K_2)$. According to Theorem 3.2, we see that for any $\varepsilon > 0$, there exist $N \in \mathbb{N}$, $\eta_i \in \mathbb{R}^1_+$, $b_i \in \mathbb{R}^{d_2}$ which are independent of $T(u)$, and $\sigma_i(T(u))$ depending on $T(u)$, $i = 1, 2, \ldots, N$, such that

$$\left\| T(u)(y) - \sum_{i=1}^{N} \sigma_i(T(u)) g_2(\eta_i \|y - b_i\|) \right\|_{L^{p_2}(K_2)} < \frac{\varepsilon}{2} \tag{10}$$

holds for all $u \in V$. We observe that $\sigma_i(T(u))$ is a continuous functional on $T(V)$. It follows from the assumption and Lemma 2.3 that there exist M_i, $N_i \in \mathbb{N}$, $c_{ik} \in \mathbb{R}^1$, n_i, $\lambda_{ik} \in \mathbb{R}^1_+$, $\theta_{ik} \in \mathbb{R}^{M_i}$ and $x_{ij} \in K_2$ $(k = 1, \ldots, N_i$, $j = 1, \ldots, M_i)$, such that

$$\left| \sigma_i(T(u)) - \sum_{k=1}^{N_i} c_{ik} g_1(\lambda_{ik} \|x^{M_i} - \theta_{ik}\|) \right| < \frac{\varepsilon}{2L} \tag{11}$$

holds for any i, where $x^{M_i} = (u_{\frac{1}{n_i}}(x_{i1}), \ldots, u_{\frac{1}{n_i}}(x_{iM_i}))'$, $L = \sum_{i=1}^{N} \|g_2(\eta_i \|y - b_i\|)\|_{L^{p_2}(K_2)}$. Theorem 3.3 follows from equations (10) and (11). □

References

[1] CHEN Tianping, CHEN Hong. *Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks* [J]. IEEE Trans. Neural Networks, 1995, **6**(4):

904–910.

[2] CHEN Tianping, CHEN Hong. *Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems* [J]. IEEE Trans. Neural Networks, 1995, **6**(4): 911–917.

[3] CHEN Tianping. *Approximation problems in system identification with neural networks* [J]. Sci. China Ser. A, 1994, **37**(4): 414–421.

[4] LESHNO M, LIN Y V, PINKUS A. et al. *Multilayer feedforward networks with a non-polynomial activation function can approximate any function* [J]. Neural Networks, 1993, **6**: 861–867.

[5] MHASKAR H N, MICCHELLI C A. *Approximation by superposition of sigmoidal and radial basis functions* [J]. Adv. in Appl. Math., 1992, **13**(3): 350–373.

[6] PINKUS A. *TDI-subspace of* $C(\mathbb{R}^d)$ *and some density problems from neural networks* [J]. J. Approx. Theory, 1996, **85**(3): 269–287.

[7] RUDIN W. *Functional Analysis* [M]. McGraw-Hill, New York, 1973.

[8] WU Wei, FENG Guorui, LI Xin. *Training multilayer perceptrons via minimization of sum of ridge functions* [J]. Adv. Comput. Math., 2002, **17**(4): 331–347.

[9] CHEN Tianping. *Approximate problem in system identification with neural networks* [J]. Sci. China Ser. A, 1994, **24**(1): 1–7. (in Chinese)

[10] JIANG Chuanhai. *Approximation problems in neural networks* [J]. Chinese Ann. Math. Ser. A, 1998, **19**(3): 295–300. (in Chinese)

[11] JIANG Chuanhai, GUO Hongbin. *Approximation by neural networks in* $L^p(\mathbb{R}^n)$ *and system identification* [J]. Chinese Ann. Math. Ser. A, 2000, **21**(4): 417–422. (in Chinese)

[12] CHEN Minde, DENG Donggao, LONG Ruilin. *Real Analysis* [M]. Beijing: Higher Eudcation Press, 1993. (in Chinese)

[13] NAN Dong, WU Wei, LONG Jinling. et al. L^p *Approximation Capability of Radial Basis Functions* [J]. Acta Math. Sinica, 2008, **24**(9): 1533–1540.