Honam Mathematical J. **27**(2005), No. 2, pp. 225–232

DEGREE OF APPROXIMATION TO A SMOOTH FUNCTION BY GENERALIZED TRANSLATION NETWORKS

NAHMWOO HAHM, MEEHYEA YANG* AND BUM IL HONG

Abstract. We obtain the approximation order to a smooth function on a compact subset of $\mathbb{R}$ by generalized translation networks. In our study, the activation function is infinitely many times continuously differentiable function but it does not have special properties around ∞ and $-\infty$ like a sigmoidal activation function. Using the Jackson's Theorem, we get the approximation order. Especially, we obtain the approximation order by a neural network with a fixed threshold.

1. Introduction

A feedforward network with one hidden layer is of the form

$$\sum_{i=1}^{n} c_i \sigma(a_i x + b_i)$$

where the weight a_i, the threshold b_i and c_i are real numbers for $1 \le i \le n$ and σ is an activation function. In most density problems, a sigmoidal function is used as an activation function. Note that a sigmoidal function is a function $\sigma : \mathbb{R} \to \mathbb{R}$ such that

$$\lim_{t\to\infty} \sigma(t) = 1, \quad \lim_{t\to-\infty} \sigma(t) = 0. \tag{1.1}$$

Received April 28, 2005. Revised June 02, 2005.

2000 Mathematics Subject Classification : 41A10, 41A24, 41A29.

Key words and phrases : Generalized Translation Network, Jackson's Theorem, Approximation order.

*This paper was supported by UNIVERSITY OF INCHEON RESEARCH FUND.

Cybenko[4] proved that any feedforward neural network with a continuous sigmoidal activation function can approximate any continuous function to any degree of precision on compact subsets of $\mathbb{R}$ if an unlimited number of neurons is permitted. Chen and Chen[2] extended the results of Cybenko by proving that any feedforward network with bounded sigmoidal activation function can approximate any continuous function on a compact subset of $\mathbb{R}$ and their result is very good since the activation function need not be continuous. Leshno, Lin, Pinkus and Schocken[5] established a complete characterization of the activation functions that lead to universal approximation of continuous functions on compact subsets of $\mathbb{R}$. Other papers[8, 9] proved some density and complexity results without the differentiability of the activation function. In [6], Medvedeva used the infinitely differentiable activation function and this gave us the motivation of this paper.

There are some differentiable activation functions. One example is the squashing function which is defined as

$$\sigma(x) = (1 + e^{-x})^{-1} \tag{1.2}$$

and the other example is the Gaussian function which is defined as

$$\sigma(x) = e^{-x^2}. \tag{1.3}$$

In [7], Mhaskar and Hahm introduced the generalized translation networks. A generalized translation network with n neurons is of the form

$$\sum_{i=1}^{n} c_i \psi(a_i \cdot x + b_i) \tag{1.4}$$

where a_i, b_i and c_i are real numbers for $1 \leq i \leq n$ and ψ is a real-valued function defined on $\mathbb{R}$. For a fixed natural number n, $\Pi_{n,\psi}$ denotes the set of such functions. In the generalized translation network, the activation function ψ doesn't have a special properties around ∞ and $-\infty$ like a sigmoidal function.

Using this generalized translation network, we obtain the approximation order to a finitely many times continuously differentiable function on a compact subset of $\mathbb{R}$. Basically, we use the divided difference formula to get a polynomial approximation by the generalized translation network and get an approximation order using the Jackson's theorem for differentiable functions. In our approximation, the threshold in the superposition of an activation function is fixed.

2. Preliminaries

In order to get a main result, we need the following.

Definition 2.1. Let $f(x)$ be defined in a closed interval I and let

$$\omega(\delta) := \omega(f, \delta) = sup|f(x_1) - f(x_2)| \tag{2.1}$$

for $x_1, x_2 \in I$ with $|x_1 - x_2| \leq \delta$.

Note that the modulus of continuity satisfies the following properties.

(1) $\omega(t) \to 0$ as $t \to 0$

(2) $\omega(t)$ is a nondecreasing function

(3) $\omega(t_1 + t_2) \leq \omega(t_1) + \omega(t_2)$ for any $t_1, t_2 > 0$.

Note that if f is k times continuously differentiable on $[a, b]$, then $|f^{(k)}(x)| \leq M$ for each $x \in [a, b]$ and some $M > 0$ since $f^{(k)}(x)$ is continuous on a compact subset of $\mathbb{R}$. Thus $\omega(f^{(k)}, \delta) \leq 2M$ for any $\delta > 0$.

Throughout the paper, the domain of a target function is $[-1, 1]$ and so we use $||\cdot||_\infty$ instead of $||\cdot||_{[-1,1],\infty}$.

3. Main Results

First of all, we approximate polynomials on $[-1, 1]$ by the generalized translation networks.

Theorem 3.1. Let ψ be a real-valued function on $\mathbb{R}$ which is infinitely many times differentiable in some open interval $(b-\delta, b+\delta)$ in $\mathbb{R}$. Assume that for any k,

$$D^k\psi(b) \neq 0 \tag{3.1}$$

and let $P_n(x) = \sum_{k=0}^{n} c_k x^k$. Then for any $\epsilon > 0$, there exists a generalized translation network $G \in \Pi_{\frac{(n+1)(n+2)}{2},\psi}$ such that

$$||P_n - G||_\infty < \epsilon.$$

Proof. Consider x^k for $0 \leq k \leq n$. From the formula

$$\psi_k(a,x) := \frac{\partial^k}{\partial a^k}\psi(a \cdot x + b) = x^k \cdot \psi^{(k)}(a \cdot x + b),$$

we have

$$x^k = (\psi^{(k)}(b))^{-1} \cdot \psi_k(0,x).$$

For any $h > 0$, the formula

$$\Psi_{k,h} = \frac{1}{h^k}\sum_{i=0}^{k}(-1)^{(k-i)}\binom{k}{i}\psi(h \cdot i \cdot x + b)$$

represents a divided difference for $\psi_k(0,x)$. In addition, we have

$$||\Psi_{k,h} - \psi_k(0,\cdot)||_\infty \leq M \cdot h$$

where M is a constant depending on ψ and n, and $|h| < \delta/n$.

Now, we choose

$$h_0 := min\{\frac{\delta}{n}, \frac{\epsilon}{M \cdot \sum_{k=0}^{n}(\psi^{(k)}(b))^{-1} \cdot |a_k|}\}$$

Then, the generalized translation network $G \in \Pi_{\frac{(n+1)(n+2)}{2},\psi}$ defined by

$$G := \sum_{k=0}^{n} a_k(\psi^{(k)}(b))^{-1} \cdot \Psi_{k,h_0}(x)$$

satisfies

$$||P_n - G||_\infty < \epsilon.$$

This completes the proof. □

Next theorem is the Jackson's theorem for differentiable functions[10, page 261].

Theorem 3.2. Let f be a k times continuously differentiable function on $[-1,1]$ and $n > k$. Then, there exists a polynomial P_n with degree n such that

$$||f - P_n||_\infty \leq M_k \cdot (\frac{2}{n})^k \cdot \omega(f^{(k)}, \frac{2}{n})$$

where M_k is a constant depending only on k.

As we mentioned early, $||f^{(k)}||_\infty \leq D$ for some constant $D > 0$ and so we have $\omega(f^{(k)}, \frac{2}{n}) \leq 2D$. Thus

$$||f - P_n||_\infty \leq \frac{C}{n^k} \tag{3.2}$$

where C is a constant depending on k and f.

From Theorem 3.1 and 3.2, we get the following result.

Theorem 3.3. Let f be a k times continuously differentiable function on $[-1,1]$ and let ψ be a real-valued function on $\mathbb{R}$ which is infinitely many times differentiable in some open interval $(b-\delta, b+\delta)$ in $\mathbb{R}$. Assume that

$$D^k\psi(b) \neq 0$$

Then, there exists a generalized translation network $G \in \Pi_{\frac{(n+1)(n+2)}{2},\psi}$ such that

$$||P_n - G||_\infty \leq \frac{C}{n^k}$$

Proof. Let $\epsilon > 0$ be given. From (3.2), there exists a polynomial P_n of degree n such that

$$||f - P_n||_\infty \leq \frac{C}{n^k}$$

where C is a constant depending on k and f. For the polynomial P_n, there exists a generalized translation network $G \in \Pi_{\frac{(n+1)(n+2)}{2},\psi}$ such that

$$||P_n - G||_\infty < \epsilon$$

by Theorem 3.1. In all,

$$\begin{aligned} ||f - G||_\infty &\le ||f - P_n||_\infty + ||P_n - G||_\infty \\ &\le \frac{C}{n^k} + \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we get the result. □

4. Discussion

In most papers related to neural network approximation, only the continuity property is given to the target function. But, D. Chen[1] conjected the possibility that if an activation function σ is an infinitely differentiable sigmoidal function and the target function $f \in C^k[-1,1]$, then there is a neural network G with n neurons such that

$$||f - G||_\infty \le \frac{c}{n^k}.$$

In Theorem 3.3, we "almost" proved this conjecture since we get the result using $\frac{(n+1)(n+2)}{2}$ neurons instead of n neurons.

In most paper related to neural network approximation, the weights and the thresholds vary even for the density result. In [3], Chui and Li showed that a continuous neural networks with "integer weights" and "free thresholds" can approximate any continuous target function on a compact subset of $\mathbb{R}$. As we pointed out before, the threshold in the generalized translation network approximation is fixed and so this is remarkable.

Now, we are considering the possibility of simultaneous approximation by generalized translation networks. In fact, for any m with $0 \le m \le k$, it is computed inductively that

$$\begin{aligned} &(\Psi_{k,h})^{(m)}(x) \\ &= k \cdots (k-m+1)\frac{1}{h^{k-m}} \sum_{i=0}^{k} (-1)^{(k-m-i)} \binom{k-m}{i} \psi^{(m)}(h \cdot i \cdot x + b). \end{aligned}$$

In addition, we can easily compute that

$$P_n^{(m)}(x) = \sum_{i=m}^{n} \frac{i!}{m!} c_i \cdot x^{i-m}.$$

The study of simultaneous neural network approximation is very important in engineering, especially in the study of robot learning of smooth movement. We are trying to get the complexity results of simultaneous approximation by generalized translation networks. That is, we want to compute the approximation order related to

$$||f^{(m)} - G^{(m)}||_\infty$$

for $0 \leq m \leq k$ where f, G and k satisfy the conditions of Theorem 3.3. We will explore this in the future.

References

[1] D. Chen,*Degree of Approximation by Superpositions of a Sigmoidal function*, J. Approximation Theory and Appl. **9** (1993), 17-28.

[2] T. Chen and H. Chen, *Approximation Capability in $C(\mathbb{R}^n)$ by Multilayer Feedforward Networks and Related Problems*, IEEE Trans. Neural Networks, **6** (1997), 25-30

[3] C. K. Chui and X. Li, *Approximation by Ridge Functions and Neural networks with One Hidden layer*, J. Approximation Theory, **70** (1992), 131-141.

[4] G. Cybenko, *Approximation by Superposition of Sigmoidal Function*, Mathematics of Control, Signal and Systems, **2** (1989), 303-314.

[5] M. Leshno, V. Y. Lin, A. Pinkus and S. Schoken, *Multilayer Feedforward Networks with a Nonpolynomial Activation Function Can Approximate Any Function*, Neural Networks, **7** (1995), 861-867.

[6] M. V. Medvedeva, *On Sigmoidal Functions,* Moscow University Mathematics Bulletin, **53(1)** (1998), 16-19.

[7] H. N. Mhaskar and N. Hahm, *Neural Networks for Functional Approximation and System Identification,* Neural Computation, **9(1)** (1997), 143-159.

[8] N. Hahm and B. I. Hong, *Approximation Order to a Function in $\bar{C}(\mathbb{R})$ by Superposition of a Sigmoidal Function,* Applied Mathematics Letters, **15** (2002), 591 - 597.

[9] H. N. Mhaskar and C. A. Micchelli, *Degree of Approximation by Neural and Translation Networks with a Single Hidden Layer*, Advances in Applied Mathematics, **16** (1994), 151-183.

[10] A. F. Timan, Theory of Approximation of Functions of a Real Variable, Dover, 1994.

Nahmwoo Hahm
Department of Mathematics
University of Incheon
Incheon 402-749, Korea
E-mail: nhahm@incheon.ac.kr

Meehyea Yang
Department of Mathematics
University of Incheon
Incheon 402-749, Korea
E-mail: mhyang@incheon.ac.kr

Bum Il Hong
Department of Mathematics and Institute of Natural Sciences
Kyung Hee University
Yongin, Kyunggi 449-701, Korea
E-mail: bihong@khu.ac.kr