Proceedings of 1993 International Joint Conference on Neural Networks

ADDITIONAL PERSPECTIVES ON FEEDFORWARD NEURAL-NETS AND THE FUNCTIONAL-LINK

Boris Igelnik   Yoh-Han Pao

Electrical Engineering and Computer Science
Case Western Reserve University
Cleveland, OH 44106-7221, USA

ABSTRACT

It has been proved that multilayer feedforward neural-nets with as few as a single hidden layer can serve as universal approximators of functions mapping multidimensional space R^S to one-dimensional space R. Our prior experience had provided us with ample pragmatic evidence that even that model can be simplified, with use of functional-links which need not be learned. In this paper, we prove theorems which provide a theoretical justification for use of the highly efficient functional-link approach.

Keywords: feedforward neural net, mapping, universal approximator, functional-link

1. INTRODUCTION

From a geometry point of view, in multilayer feedforward neural-net, the computations carried out in going from one layer of nodes to another constitute a nonlinear co-ordinate transformation; a vector described in one co-ordinate system is subsequently described in another co-ordinate system. The entity remains unchanged, but the representation is changed.

This view is particularly interesting in the case of the multilayer net with only a single hidden layer and a single output node. There is one nonlinear co-ordinate transformation on going from the input layer to the hidden layer. The subsequent mapping from the hidden layer to the output node is achieved through quadratic optimization; that part of the net is, in fact, a linear net. We note, in passing, that there is no loss in generality by focusing on one output node at a time.

It has been proposed by Pao and his collaborators (Pao 1989; Klassen and Pao 1989; Pao and Takefuji 1991) that the first stage nonlinear co-ordinate transformation need not be `learned' but can be implemented with use of predetermined `functional-links'. In practice, there has been abundant experience (Pao, Phillips, and Sobajic 1992; among other) to indicate the functional-link net, indeed, can serve as accurate approximators of functional mappings; but, until now there has been no rigorous examination of the theoretical basis for that capability.

In this paper, we present the essentials of the proofs of two theorems which show why it is possible to disentangle the parameter (weight) learning processes at the two successive node levels; and that it is not necessary to `learn' the parameters in the first transition. The functional-links which transform the input representation into that of the hidden layer can be prescribed and need not be `learned' laboriously.

In part, our first theorem duplicates the results of Hornik, Stinchcombe, and White (1989), of Funahashi (1989), and others but in a different manner. In other words, we do prove that the neural-net representation $f(\underline{x}) = \sum a_i \sigma(\underline{w}_i \underline{x} + b_i)$ is, indeed, a universal approximator; but we also prove that the *internal* parameters $\underline{w}_i$, b_i depend only on σ and do not require any learning, while *external* parameters a_i depend on f and can be learned and optimized on the training set. The second theorem is based on Stancu's (1970) method of function approximation and gives quantitative information about implementation of such neural nets. Estimators of approximation accuracy depend explicitly on two factors: one featuring f, and another featuring σ.

In this space-limited discussion, we state the theorems and provide outlines of the proofs. Detailed proofs will be published in due time. It is our intent that these outlines provide clear and sufficient delineation of the line of reasoning and demonstrates which form the basis of the proofs.

2. MAIN RESULTS

THEOREM 1.

Let f: $R^s \to R$ is a function, continuous on $I^s = [0,1]^s$. For a fixed sigmoid, continuous on R function σ: $R \to R$ and any $\in\ > 0$ there exist $n \in N, a_0, a_1, \ldots a_n, b_0, b_1, \ldots b_n \in R$, $\overline{w}_0, \overline{w}_1, \ldots \overline{w}_n \in R^s$ such that

a. $$\left| f(\overline{x}) - \sum_{i=0}^{n} a_i \sigma(\overline{w}_i \overline{x} + b_i) \right| < \in \text{ for any } \overline{x} \in I^s$$

b. choice of $\{\overline{w}_i, b_i, i = 0, \ldots n\}$ can be separated from the choice of $\{a_i, i = 0, \ldots n\}$.

Sketch of proof.

1. We prove the theorem for s = 1 by construction $\tilde{f}(x) = \sum_{i=0}^{n} a_i \sigma(w_i x + b_i)$ and give explicit formulas for w_i, b_i, $i = 1, \ldots n$.

2. For arbitrary $s \in N$ we represent the neural net in the form

$$\tilde{f}(\overline{x}) = \sum_{k_1=0}^{\tilde{n}} \cdots \sum_{k_s=0}^{\tilde{n}} a_{k_1 \ldots k_s} \prod_{i=1}^{s} \left[\sigma\left(w_{k_i} x_i + b_{k_i}\right) - \sigma\left(w_{k_i+1} x_i + b_{k_i+1}\right) \right]$$

with $w_{k_i}, b_{k_i}, k_i = 0, \ldots \tilde{n}+1, i = 1, \ldots s$ independent of f.

3. Applying theorem 2.4, Hornik et al. (1989) to $\prod_{i=1}^{s} \left[\sigma\left(w_{k_i} x_i + b_{k_i}\right) - \sigma\left(w_{k_i+1} x_i + b_{k_i+1}\right) \right]$ we get the final result.

For simplicity of notations, we give Theorem 2 for s = 1. Generalization on arbitrary natural s is straightforward. Let

$$\rho_{k\ell}(x,y) = \sigma\left(f_{k\ell}(x,y)\right) - \sigma\left(f_{k+1,\ell}(x,y)\right) - \sigma\left(f_{k,\ell+1}(x,y)\right) + \sigma\left(f_{k+1,\ell+1}(x,y)\right)$$

$$\rho_{k\cdot}(x,y)=\sum_{\ell=0}^{n}\rho_{k\ell}(x,y),\ \rho_{\cdot\ell}(x,y),=\sum_{k=0}^{n}\rho_{k\ell}(x,y)$$

$$k=0,\dots n,\ell=0,\dots n$$

For arbitrary $\sigma: R\to R, f_{k\ell}: I^2\to R$ one can easily see that

$$\sum_{k=0}^{n}\sum_{\ell=0}^{n}\rho_{k\ell}(x,y)=1,\sum_{k=0}^{n}\rho_{k\cdot}(x,y)=1,\sum_{\ell=0}^{n}\rho_{\cdot\ell}(x,y)=1$$

Denote

$$\sigma_1^{(n)}(x,y)=\sum_{k=0}^{n}\rho_{k\cdot}(x,y)\left(\frac{k}{n}-x\right)^2,\sigma_2^{(n)}(x,y)=\sum_{\ell=0}^{n}\rho_{\cdot\ell}(x,y)\left(\frac{\ell}{n}-x\right)^2$$

$$\tilde{f}(x,y)=\sum_{k=0}^{n}\sum_{\ell=0}^{n}f\left(\frac{k}{n},\frac{\ell}{n}\right)\rho_{k\ell}$$

THEOREM 2

a. If $\rho_{k\ell}(x,y)\geq 0$ for any $n\in N$, k, $\ell=0,\dots n,(x,y)\in I^2$ and $\sigma_1^n(x,y)$, $\sigma_2^{(n)}(x,y)\xrightarrow[n\to\infty]{}0$ uniformly on I^2, then

(1) $\tilde{f}(x,y)\xrightarrow[n\to\infty]{}f(x,y)$ uniformly on I^2,

(2) $\left|\tilde{f}(x,y)-f(x,y)\right|\leq\left(1+\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right)\omega\left(f:\alpha_1\sigma^{(n)},(x,y),\alpha_2\sigma_2^{(n)}(x,y)\right)$, where α_1, α_2, > 0 are arbitrary positive numbers, $\omega(f;.,.)$ is modulus of continuity of the function f;

b. if σ is sigmoid function and $f_{k\ell}(x,y)=w_k x+w_\ell y+b_{k\ell}$ with explicit formulas for $w_k,w_\ell,b_{k\ell},k,\ell=0,\dots n$ then (1) and (2) are true;

c. under conditions b and if $\sigma(x)=1/\left(1+e^{-x}\right)$ then

(3) $\sigma_1^{(n)}(x,y)n/x\xrightarrow[n\to\infty]{}1,\frac{x}{n},\sigma_2^{(n)}(x,y)n/y\xrightarrow[n\to\infty]{}1$

Sketch of proof.

a. is a simple corollary from Stancu's (1970) Theorem.

b. construction of the $\rho_{k\ell}(x,y)$ and special choice of $w_k,w_\ell,b_{k\ell}$ allows us to prove that

$\rho_{k\ell}(x,y) \geq 0$ and $\sigma_1^{(n)}(x,y), \sigma_2^{(n)}(x,y) \xrightarrow[n\to\infty]{} 0$

c. $\sigma_1^{(n)}(x,y), \sigma_2^{(n)}(x,y)$ depends only on one-dimensional distributions $\{\rho_{k\cdot}\}$ and $\{\rho_{\cdot\ell}\}$ which allows us for the popular choice of σ to obtain asymptotics (3).

REFERENCES

1. Pao, Y.H., 1989. *Adaptive Pattern Recognition and Neural Networks.* Addison-Wesley Publishing Company, Inc., Reading, MA.

2. Klassen, M.S. and Y.H. Pao, 1988. Characteristics of the functional-link net: A higher order delta rule net, *IEEE Proceedings of 2nd Annual International Conference on Neural Networks*, San Diego, CA.

3. Pao, Y.H. and Y. Takefuji, 1991. Functional-link net computing: Theory, system architecture, and functionalities, Special issue of *Computer* on Computer Architectures for Intelligent Machines, IEEE Computer Society Press, Vol. 3, pp. 76-79, also published in *Readings in Computer Architectures for Intelligent Systems* (expanded version), J. Herath (ed.), in press.

4. Pao, Y.H., S.M. Phillips, and D.J. Sobajic, 1992. Neural-net computing and the intelligent control of systems, Special Intelligent Control issue of *International Journal of Control,* Vol. 56, No. 2, pp. 263-290, August 1992, Taylor and Francis, Ltd., London, U.K.

5. Funahashi, K., 1989. On the approximate realization of continuous mappings by neural networks, *Neural Networks*, Vol. 2, pp. 183-192.

6. Hornik, K., M. Stinchcombe, and H. White, 1989. Multilayer feedforward network are universal approximators, *Neural Networks*, Vol. 2, pp. 359-366.

7. Stancu, D.D., 1970. Probabilistic methods in the theory of approximation of functions of several variables by linear positive operators, in *Approximation Theory*, A. Talbot (ed.), Academic Press, London and New York.