Geometric Interpretation of Nonlinear Approximation Capability for Feedforward Neural Networks*

Bao-Gang Hu[1,2], Hong-Jie Xing[1,2], and Yu-Jiu Yang[1,2]

[1] National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
[2] Beijing Graduate School, Chinese Academy of Sciences
P.O. Box 2728, Beijing, 100080 China
{hubg,hjxing,yjyang}@nlpr.ia.ac.cn

Abstract. This paper presents a preliminary study on the nonlinear approximation capability of feedforward neural networks (FNNs) *via* a geometric approach. Three simplest FNNs with at most four free parameters are defined and investigated. By approximations on one-dimensional functions, we observe that the Chebyshev-polynomials, Gaussian, and sigmoidal FNNs are ranked in order of providing more varieties of nonlinearities. If neglecting the compactness feature inherited by Gaussian neural networks, we consider that the Chebyshev-polynomial-based neural networks will be the best among three types of FNNs in an efficient use of free parameters.

1 Introduction

Machine learning through input-output data from examples can be considered as approximations of unknown functions. Two cases can be found in the nonlinear approximations. One is having a certain degree of knowledge about the nonlinear functions investigated. The other is in the case that *a priori* information is unavailable in regards to the degree of nonlinearity of the problem. The last case presents more difficulty in handling, but often occurs in real world problems. In this work, we will investigate feedforward neural networks (or **FNNs**) as the nonlinear approximators for the last case.

Significant studies have been reported on that FNNs are universal approximators with various basis (or activation) functions [1-3]. However, the fundamental question still remains: *Among the various basis functions, which one provides the most efficiency in approximations of arbitrary functions*? This efficiency can be evaluated by the approximation accuracy over a given number of free parameters employed by its associated FNN. Some numerical investigations have shown that the radial basis functions usually afford the better efficiency than the sigmoidal

* This work is supported by Natural Science of Foundation of China (#60275025, #60121302).

F. Yin, J. Wang, and C. Guo (Eds.): ISNN 2004, LNCS 3173, pp. 7–13, 2004.

functions [4]. Hence, further question arises: *What are natural reasons for some basis function exhibiting better efficiency than the others*?

In this paper, we attempt to answer the two basic questions above *via* a geometric approach. The interpretations from the approach seem simply and preliminary at this stage, but we believe that a geometric approach does provides a unique tool for understanding the nature insights of nonlinear approximation capabilities of universal approximators. This paper is organized as follows. Section 2 proposes a new methodology. Three simplest FNNs are defined and examined in Section 3. A nonlinearity domain analysis is made with respect to their availability of nonlinearity components for the three FNNs in Section 4. Finaly, some remarks are given in Section 5.

2 Proposed Methodology

In the studies of approximation capability, the conventional methodology used in FNNs is generally based on the performance evaluations from approximation errors. Two common methods are employed in the selections of basis functions. One is on the estimation of error bonds, and the other is on the examination of numerical errors to the specific problems. Few studies related the basis functions to the approximation capability using a geometric approach. In this work, we propose a new methodology from the following aspects.

2.1 Nonlinearity Domain Analysis

Nonlinearity domain analysis is a novel concept. There is no existing and explicit theory for such subject. We propose this concept in order to characterize a given nonlinear function by its nonlinearity components similar to a frequency domain analysis. Here are some definitions:

Definition 1. *Free parameters, linear parameters, and nonlinear parameters.*

Any nonlinear function can be represented in a form as:

$$\mathbf{y} = f(\mathbf{x}, \theta) = f(\mathbf{x}, \theta_L, \theta_{NL}), \tag{1}$$

where $\mathbf{x} \in R^N$ and $\mathbf{y} \in R^M$ are input and output variables, respectively; θ, θ_L and θ_{NL} are free parameter set, linear and nonlinear parameter sets respectively. The behaviors and properties of nonlinear functions are controlled by free parameters. If it can change the shape or orientation of nonlinear function, this parameter will fall into a nonlinear parameter set. Otherwise, it is a linear parameter (also called location parameter).

Definition 2. *Nonlinearity domain, nonlinearity components and nonlinearity variation range.*

Nonlinearity domain is a two dimensional space used for characterization of nonlinearity of functions. Nonlinearity components are a set of discrete points with an infinite number along the horizontal axis of nonlinearity domain. The vertical axis represents the variable of nonlinearity variation range. The plot in Fig. 1 can be called "*Nonlinearity Spectrum*", which reveals two sets of information. First, for a given function, how many nonlinearity components could be generated by changing the free parameters. Each component represents a unique class of nonlinear functions, say, NC_j for the jth nonlinearity component, which could be an "S-type curve". Second, for each included component, what is its associated nonlinearity variation range. This range exhibits the admissible range realized by the given function. A complete range is normalized within [0,1]. If $NV_j = 0.5$, it indicates that the given function can only span a half space in the jth nonlinearity component.

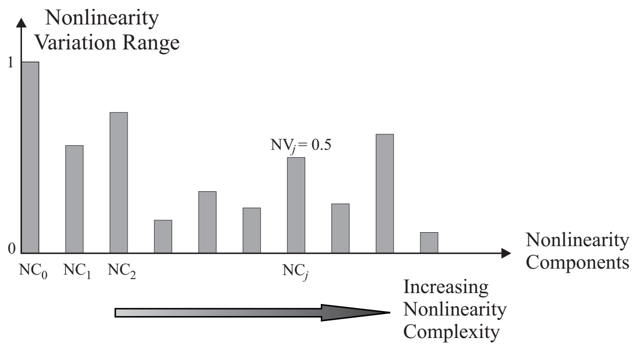

Fig. 1. Nonlinearity Domain Analysis

2.2 Definition of the Simplest-Nonlinear FNNs

The general form of FNNs is a nonlinear mapping: $f : R^N \rightarrow R^M$. In the nonlinearity domain analysis, it will be a complex task if the high dimensionality of FNNs is involved. In order to explore the nature insights of FNNs, one has to make necessary simplifications, or assumptions. In this work, we define the simplest-nonlinear FNNs for the nonlinearity domain analysis.

Definition 3. *Simplest-nonlinear FNNs.*

The simplest-nonlinear FNNs present the following features in their architectures: I. A single hidden layer. II. A single hidden node (but more for polynomial-based FNNs). III. A single-input-single-output nonlinear mapping, $f_s : R \rightarrow R$. IV. Governed by at most four free parameters:

$$y = f_s(x, \theta) = f_s(x, a, b, c, d), \tag{2}$$

where $a, b, c, d \in R$. Further classification of the four parameters is depending on the basis function applied. We will give discussions later about the reason of choosing four free parameters, and call FNNs in eq (2) the simplest FNNs.

When we define the simplest FNNs, one principle should be followed. The conclusions or findings derived from the simplest nonlinear FNNs can be extended directly to judge approximation capability of the general FNNs. For example, we only study the nonlinearity of curves. However, the nature insights obtained from this study can also be effective to the FNNs that construct hypersurfaces.

2.3 Classification for Nonlinearity Components

In the nonlinearity domain analysis, all nonlinearity components are classified according to the geometric properties from nonlinear functions. However, there exist various geometric features for classification. These include continuity, monotonicity, symmetry, periodicity, compactness, boundness, singularity, *etc.* In this work, we restrict the studies within the smooth functions. Therefore, the geometric properties in related to the continuity and monotonicity features will be used in the classification. In this work, we consider the following aspects:

- G1. Monotonic increasing or decreasing.
- G2. Convexity or concavity.
- G3. Number of inflection points.
- G4. Number of peaks or valleys.

Therefore, each nonlinearity component should represent a unique class of nonlinear functions with respect to the above aspects. After the classification, we usually arrange the components, NC_j, along the axis in an order of increasing nonlinearity complexity. In this work, we only consider the one-dimensional nonlinear functions. Then, we immediately set $\{NC_0 : y = c\}$ and $\{NC_1 : y = ax+c\}$ to be constant and linear components, respectively. Although these two components are special cases for the zero degree of nonlinearity, both of them cannot be missed for the completeness of nonlinearity components. The next will start from simple nonlinear curves, say, "C-type" and "S-type" curves. We will give more detailed classification examples later.

3 Examination of Parameters on Three Simplest FNNs

In this section, we will examine the parameters on the simplest FNNs with three different basis functions, *i.e.*, sigmoidal, Gaussian, and Chebyshev-polynomials. Their mathematic representations are given in the following forms.

The simplest sigmoidal FNNs:

$$y = \frac{a}{1 + \exp(bx + c)} + d \tag{3}$$

The simplest Gaussian FNNs:

$$y = a \exp\left[-\frac{(x-c)^2}{b^2}\right] + d \tag{4}$$

The simplest Chebyshev-polynomial FNNs:

$$y = aT_3(x) + bT_2(x) + cT_1(x) + d \tag{5}$$

where $\mathrm{T}_i(x)$ are the Chebyshev polynomials (see [5] and their architectures of FNNs).

Both simplest sigmoidal and Gaussian FFNs apply at most four free parameters. In order to make a fair comparison, we set the cubic forms for the simplest Chebyshev-polynomial FNNs. The first analysis of the three types of FNNs is the classification of linear and nonlinear parameters. We catalogue two sets of parameters for the reason that linear parameters do not change the geometric properties (say, G1-G4 in Section 2.3) of nonlinear functions. Table 1 presents the two parameter sets for the three simplest FNNs. All linear parameters play a "shifting" role to the curves; but only nonlinear parameters can change the shape, or curvatures, of functions. We conclude that the Chebyshev shows better features over the others on the following aspects:

I. The Chebyshev presents a bigger set of nonlinear parameters, which indicates that it can form a larger space for nonlinearity variations.
II. The nonlinear parameters in the Chebyshev can be changed into linear parameters. This flexibility feature is not shared by the others.
III. The nonlinear parameters in both sigmoidal and Gaussian can produce a "scaling" effect to the curves. This will add a dependency feature to the nonlinear parameters and thus reduce the approximation capability if similarity is considered. The Chebyshev, in general, does not suffer this degeneration problem.

Table 1. Comparisons of linear and nonlinear parameters for the three simplest FNNs

	Linear parameters	Nonlinear parameters
Sigmoidal	c, d	a, b
Gaussian	c, d	a, b
Chebyshev	d	a, b, c

4 Nonlinearity Domain Analysis on Three Simplest FNNs

In this section we will conduct nonlinearity domain analysis on the three simplest FNNs. Without losing the generality, a smooth function in a compact interval, $\{f(x); x \in [0,1]\}$, will be approximated. Therefore, any function can be approximated by the linear combinations of several simplest FNNs. The approximation will allocate the proper segmentation range (by using linear parameters) from the given basis function (by using nonlinear parameters for the proper shapes). Therefore, each simplest FNN can provide the different nonlinear curves (see Fig. 2 for the sigmoidal function).

In this work, we summarize the nonlinearity components in their availability for the three simplest FNNs in Table 2, in which each NC_j is given graphically in Fig. 3. All NC_j represent unique classes of nonlinear, but smooth, functions according to the geometric properties. One can observe that the Chebyshev is

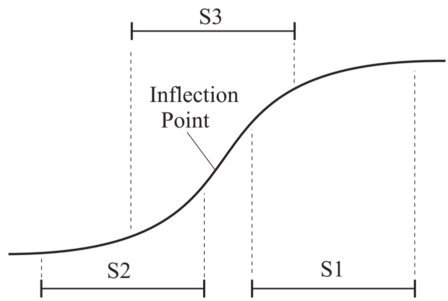

Fig. 2. Segmentation of a sigmoidal function and different nonlinear curves. (S1, S2 and S3 correspond to the C-, Inverse C- and S-curves, respectively.)

the best again in producing the most nonlinearity components. However, the nonlinearity variations range for each FNNs is not given and will be a future work.

Table 2. Comparisons of nonlinearity components in their availability for the three simplest FNNs. (The sign "$\surd$" indicates the availiabilty of its current component, otherwise it is empty).

	NC_0	NC_1	NC_2	NC_3	NC_4	NC_5	NC_6	NC_7	NC_8	NC_9	NC_{10}
Sigmoidal	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$		$\surd$		
Gaussian	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$		$\surd$		$\surd$
Chebyshev	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$

5 Final Remarks

In this work, we investigate the FNNs with three commonly used basis functions. A geometric approach is used for interpretation of the nature in FNNs. We conclude that the Chebyshev-polynomial FNNs are the best type in comparing with the sigmoidal and Gaussian FNNs by including more nonlinearity components. However, a systematic study on the nonlinearity domain analysis

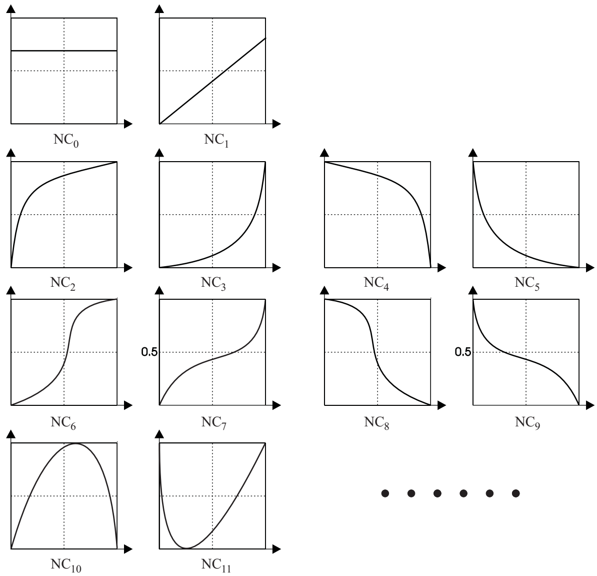

Fig. 3. Nonlinear curves and their associated nonlinearity components NC_j

is necessary to reach overall conclusions for each type of FNNs. For example, the compactness feature of the Gaussian is more efficiency in approximation of a local behavior of nonlinear functions. On the other hand, we believe that both performance-based and function-based evaluation approaches [6] will provide a complete study to ease the difficulty of "*trial and error*" in designs of universal approximators, such as fuzzy systems and neural networks.

References

1. Hornik, K., Stinchcombe, M., White, H.: Multilayer Feedforward Networks are Universal Approximators. Neural Networks. **2** (1989) 359–366
2. Park, J., Sandberg, J.W.: Universal Approximation Using Radial Basis Functions Network. Neural Computation. **3** (1991) 246-257
3. Chen, T-P., Chen, H.: Universal Approximation to Nonlinear Operators by Neural Networks with Arbitrary Activation Functions and Its Application to Dynamical Systems. IEEE Trans. on Neural Networks. **6** (1995) 911-917
4. Park, J.-W., Harley, R. G., Venayagamoorthy, G. K.: Indirect Adaptive Control for Synchronous Generator: Comparison of MLP/RBF Neural Networks Approach With Lyapunov Stability Analysis. IEEE Trans. on Neural Networks. **15** (2004) 460-464
5. Lee, T.-T., Jeng, J.-T.: The Cyebyshev-Polynomials-Based Unified Model Neural Net-works for Function Approximation. IEEE Trans. on Sys. Man Cyber. **28B** (1998) 925-935
6. Hu, B.-G., Mann, G.K.I., Gosine, R.G.: A Systematic Study of Fuzzy PID Controllers—Function-based Evaluation Approach. IEEE Trans. on Fuzzy Systems. **9** (2001) 699-712