COMPARISON OF ACTIVATION FUNCTIONS IN MULTILAYER NEURAL NETWORK FOR PATTERN CLASSIFICATION

Kazuyuki HARA† Kenji NAKAYAMA‡ †
†Graduate School of Nat. Sci. & Tech., Kanazawa Univ.
‡Faculty of Tech., Kanazawa Univ.

2–40–20 , Kodatuno , Kanazawa , 920 JAPAN
E–mal : nakayama@haspnn1.ec.t.kanazawa-u.ac.jp

1 INTRODUCTION

Advantage of multilayer neural networks trained by the back-propagation (BP) algorithm is to extract common properties, features or rules, which can be used to classify data included in several groups [1]. Especially, when it is difficult to analyze the common features using conventional methods, the supervised learning, using combinations of the known input and output data, becomes very useful.

We studied multi-frequency signal classification using multilayer neural network[2][3]. The following advantages over conventional methods were confirmed. The neural network can classify the signals using a small number of samples and a short observation period with which Fourier transform can not classify. The number of calculation is sufficiently smaller than the convolution calculation, required in digital filters.

In this paper, classification performance of the three activation functions, the sigmoid, the sinusoidal and the Gaussian functions will be investigated. Furthermore, noise suppression capability of the network is investigated. For this purpose, white noise is used. The effects of training with noisy signal is compared with pure signals training.

2 MULTI-FREQUENCY SIGNALS

Multi–frequency signals are defined by

$$x_{pm}(n) = \sum_{r=1}^{R} A_{mr} \sin(\omega_{pr} nT + \phi_{mr}), n = 1 \sim N \ , \ \omega_{pr} = 2\pi f_{pr} \tag{1}$$

M samples of $x_{pm}(n), m = 1 \sim M$, are included in the group X_p as follows.

$$X_p = \{x_{pm}(n), m = 1 \sim M\}, p = 1 \sim P \tag{2}$$

P signal groups, $X_p, p = 1 \sim P$, are assumed. T is a sampling period. The signals have N samples. In one group, the same frequencies are used.

$$F_p = [f_{p1}, f_{p2}, \ldots, f_{pR}]Hz, p = 1 \sim P \tag{3}$$

Amplitude A_{mr} and phase ϕ_{mr} are different for each frequency in the same group. They are generated as random numbers, uniformly distributed in following ranges.

$$0 < A_{mr} \leq 1 \qquad 0 \leq \phi_{mr} < 2\pi \tag{4}$$

3 MULTILAYER NEURAL NETWORK

A two-layer neural network is taken into account. N samples of the signal $x_{pm}(n)$ are applied to the input layer in parallel. The nth input unit receives the sample at nT. Connection weight from the nth input to the jth hidden unit is denoted by w_{nj}. The input of the jth hidden unit is given by

$$net_j = \sum_{n=0}^{N-1} w_{nj} x_{pm}(n) \tag{5}$$

In the hidden layer, one of three activation functions, sigmoid, sinusoidal and Gaussian functions is used as follows:

Sigmoid function: $$y_j = f_{sig}(net_j) = \frac{1}{1 + e^{-(net_j + \theta)}} \tag{6}$$

Sinusoidal function: $$y_j = f_{sine}(net_j) = \sin(\pi net_j) \tag{7}$$
Gaussian function: $$y_j = f_g(net_j) = e^{-net_j^2} \tag{8}$$

Letting the connection weight from the jth hidden unit to the kth output unit be w_{jk}, the input of the kth output unit is given by

$$net_k = \sum_{j=0}^{J-1} w_{jk} y_j \tag{9}$$

The activation function of the output layer is the sigmoid function of Eq.(6).

The number of output units is equal to that of the signal groups P. The neural network is trained so that a single output unit responds to one of the signal groups.

3.1 Training and Classification

Sets of signals are categorized into training and untraining sets, denoted by X_{T_p} and X_{U_p} whose elements are expressed by $x_{T_{pm}}(n)$ and $x_{U_{pm}}(n)$, respectively.

The neural network is trained by using $x_{T_{pm}}$(n), m = 1 ~ M_T, for the pth group. Here, M_T is the number of training data. After the training is completed , the untraining signals $x_{U_{pm}}$(n) are applied to the neural network , and the output is calculated. For the input signal $x_{U_{pm}}$(n), if the pth output y_p has the maximum value, then the signal is exactly classified. Otherwise , the network fails in classification .

4 CLASSIFICATION WITHOUT NOISE

4.1 Multi-frequency Signals

The number of frequency components R is 3, and the signal groups P is 2, respectively. The frequency components are located alternately between the groups as follows: F_1 = [1, 2, 3] Hz for group 1 (#1) and F_2 = [1.5, 2.5, 3.5] Hz for group 2 (#2). The sampling frequency is 10 Hz, that is T = 0.1 sec. The number of samples N is 10. Therefore, the observation interval is 1 sec.

4.2 Neural Network Classification

$x_{T_{pm}}$(n), m = 1 ~ 200 and $x_{U_{pm}}$(n), m = 1 ~ 1800 are used. Simulation results are shown in table 1. Training converged three hidden units for all activation function. In the case of the Gaussian and the sinusoidal function , training almost converged with one hidden unit. Detailed discussion will be provided in Sec. 6.

Table 1:Classification rates by three functions[%]

Activation Function	Hidden Unit	Training		Untraining	
		#1	#2	#1	#2
Sigmoid	1	44.5	100	47.9	100
	3	100	100	97.4	100
Sinusoidal	1	86.0	99.0	79.8	99.0
	3	100	100	92.6	100
Gaussian	1	99.5	100	98.1	100
	3	100	100	99.1	99.9

5 CLASSIFICATION WITH NOISE

Suppression capability for white noise, that is uncorrelated interference, is investigated. It can be expected that the neural network can be insensitive after training using a large number of signals including uncorrelated interference. White noise is added to both the training and the untraining signals.

5.1 White Noise

White noise, *noise(n)* is generated as random number and added to the signal x_{pm}(n). Noisy signal x'_{pm}(n) is given by

$$x'_{pm}(n) = x_{pm}(n) + noise(n) \tag{10}$$

This kind of noise occur in the electric components as thermal noise.

5.2 Training Using Signals with White Noise

The multi-frequency signals with and without noise are used for training the neural network. N is 10 and M is 200 for each group. After training, untraining signals with white noise are applied, and classification rates are evaluated. White noise is uniformly distributed in the range ±0.5. The results are shown in Table 2. Symbols(A) and (B) indicate the training signals without and with white noise, respectively. From these results, it can be confirmed that trining using noisy signals is useful to achieve robustness.

Table 2: Classification rates using training signals without (A) and with (B) white noise [%]

Activation Function	Hidden Unit	(A)		(B)	
		#1	#2	#1	#2
Sigmoid	1	47.0	52.9	92.8	28.5
	3	97.3	8.4	82.6	78.0
Sinusoidal	1	80.2	20.9	61.7	87.7
	3	65.9	36.2	79.9	82.7
Gaussian	1	98.2	4.8	71.7	65.9
	3	85.3	46.3	79.8	70.2

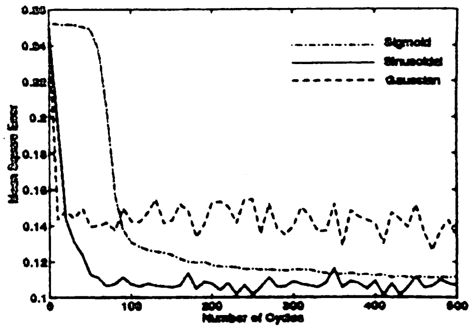

Figure 1: Learning curve

Convergence Rates: Figure 1 shows learning curve of the three hidden units networks. Within a error range of 0.14, Gaussian decreases the error faster than other two networks, but it does not decreases the error any more. Sinusoidal also decreases the error faster, and decreases the error to the error range of 0.1. The sigmoid function decreases to the range of 0.1, but it decrease the error slowly. So, the sinusoidal is the best activate function on the convergence rate.

6 COMPARISON OF TREE ACTIVATION FUNCTIONS

6.1 Convergence Property Using One Hidden Unit

The obtained neural networks are further investigated by hidden unit output distribution. Figure 2 illustrates the hidden unit output distribution of the sigmoid (a1), the sinusoidal (b2) and Gaussian (c1). No noise is added to the training data.

In case of the Gaussian function, #2 located at near the peak where the differential coefficient is large. Then #2 can be located in this area very fast. Most of the #1 data distributed both sides. So, data nature can be reflected into the network directly. In case of sinusoidal function, #2 located at near the peak at zero input. The sinusoidal function have large coefficient except peak. Then #2 can locate at near the peak fast. #1 can locate every where which output is less than 0.5. In case of sigmoid function, #1 and #2 have to locate one of the sides. Then the network have to modify the weight to separate data into two regions. From this result, it can be concluded as follows. The data distribution of the multi-frequency signals is that #2 is concentrated on a point and #1 is distributed widely.

6.2 Convergence Property Using Three Hidden Units

Figured 3, 4 and 5 show distributions of the output which has the hidden unit of the sigmoid, the sinusoidal and the Gaussian functions. For each figure, (a), (b) and (c) correspond to one of the hidden unit. (a1), (b1) and (c1) are the response for #1 and (a2), (b2) and (c2) are for #2. From these figures, one of the groups located at near the peak of the functions and the other distributed widely as like the case of one hidden unit. In case of three hidden units, not only each hidden unitd output distribution is important but also combination of three hidden units outputs. So, it can be concluded that three hidden units form the combination to make the output unit input linear separable.

6.3 Convergence Property Using Noisy Signals

Figure 2 also shows the hidden unit output distributions which applied noisy data. Corresponding functions of (a2), (b2) and (c2) are as same as Sec.6.1. In the case of the Gaussian, #2 data can not stay at the peak no more and distributed arounds. Because stable area is very narrow. The sinusoidal, #2 also widely distributed all along the function curve. These are the reasons of poor accuracy using the noisy data.

7 CONCLUSION

Effect of the activation function for multi-frequency signal classification has been discussed using multilayer neural network supervised by BP algorithm. In the case of without noise, The Gaussian function is the best solution.

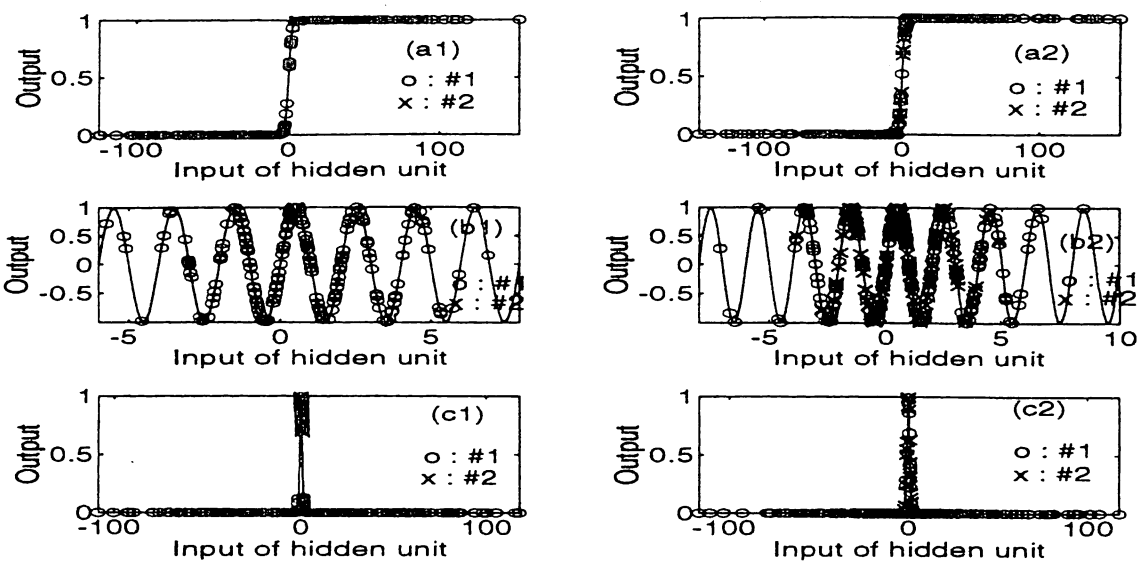

Figure 2: Hidden unit output distributions

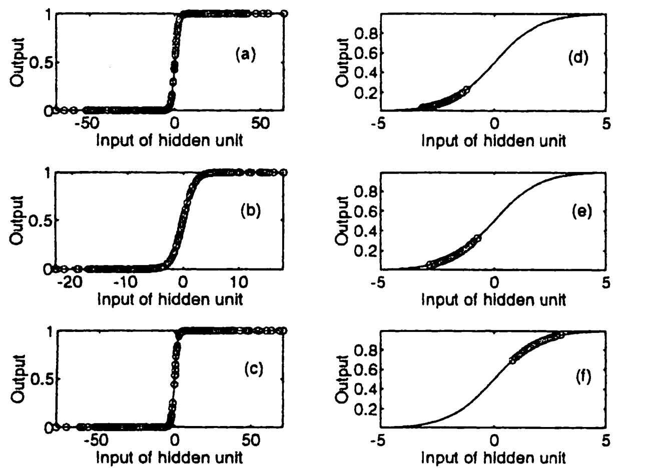

Figure 3: Distribution of sigmoid hidden unit outputs

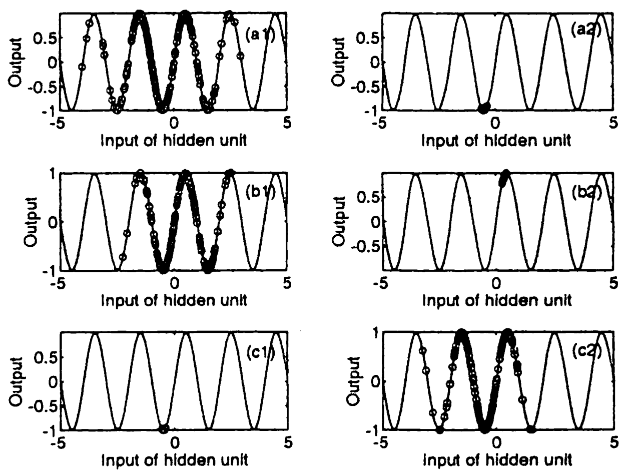

Figure 4: Distribution of sinusoidal hidden unit outputs

It converged the fastest among three functions and training converged with one hidden unit. In the case of with noise, the sigmoid and the sinusoidal convergence rates are better than the Gaussian. The training using the sinusoidal converged faster than the sigmoid.

References

[1] D.E.Rumelhart and J.L.McCelland et al (1986) Parallel Distributed Processing, MIT Press, 1986.

[2] K.Hara and K.Nakayama (1992) "Multi-frequency signal classification using multilayer neural network trained by backpropagation algorithm (in Japanese)", Tech., Rep. IEICE, NC92-75, 47–54

[3] K.Hara and K.Nakayama (1993) "High resolution of multi-frequencies using multilayer networks trained by back-propagation algorithm", Proc. WCNN'93, IV, 675–678

[4] K.Hara and K.Nakayama (1993) Classification of multi-frequency signals with random noise using multilayer neural networks, I, 601–604

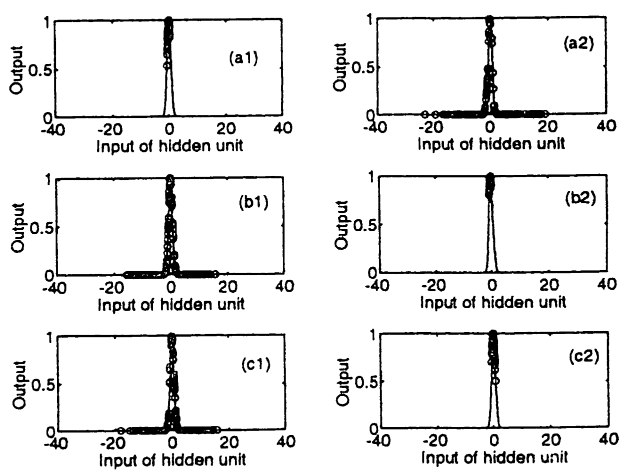

Figure 5: Distribution of Gaussian hidden unit outputs