

Effect of Nonlinear Transformations on Entropy of Hidden Nodes

Sang-Hoon Oh

Department of Information Communication Engineering
Mokwon University, Daejeon, 302-729, Korea

ABSTRACT

Hidden nodes have a key role in the information processing of feed-forward neural networks in which inputs are processed through a series of weighted sums and nonlinear activation functions. In order to understand the role of hidden nodes, we must analyze the effect of the nonlinear activation functions on the weighted sums to hidden nodes. In this paper, we focus on the effect of nonlinear functions in a viewpoint of information theory. Under the assumption that the nonlinear activation function can be approximated piece-wise linearly, we prove that the entropy of weighted sums to hidden nodes decreases after piece-wise linear functions. Therefore, we argue that the nonlinear activation function decreases the uncertainty among hidden nodes. Furthermore, the more the hidden nodes are saturated, the more the entropy of hidden nodes decreases. Based on this result, we can say that, after successful training of feed-forward neural networks, hidden nodes tend not to be in linear regions but to be in saturated regions of activation function with the effect of uncertainty reduction.

Key words: Entropy, Hidden Nodes, Nonlinear Activation Function, Feed-Forward Neural Networks.

1. INTRODUCTION

When an input sample is presented to feed-forward neural networks (FNNs), it is processed through a series of weighted sums and nonlinear activation functions. It was proved that the FNNs with enough hidden nodes can approximately implement any function [1]-[4]. Herein, the weighted sums to hidden nodes are a sort of projections from the input space to a hidden feature space followed by element-wise nonlinear activation functions. There have been research results to understand the role of hidden nodes. Oh and Lee proved that the nonlinear function of hidden nodes has an effect of decreasing correlations among hidden nodes [5]. They also argued that FNNs are a special type of nonlinear whitening filter [5]. And, Shah and Poon investigated that hidden nodes with sigmoidal activation functions have the ability to produce linearly independent internal representations [6].

In neural network field, information theory provides many fruitful research results. Lee et al. reported that FNNs have a capability of information extraction for pattern classification through hierarchically keeping inter-class information while reducing intra-class variations [7]. Learning rules were proposed stemming from information theory [8]-[14]. The upper bound for the probability of error was derived based on Renyi's entropy [15]. Information theory can provide the construction strategy of neural networks [16]. Also, hidden information maximization and maximum mutual information

methods were proposed for feature extractions [17], [18]. In this paper, we focus on the nonlinear activation functions of hidden nodes in a viewpoint of information theory. Under the assumption that nonlinear activation function can be approximated piece-wise linearly, we derive that the entropy of hidden nodes decreases after piece-wise linear transformation. Based on the derivation, we can interpret the role of hidden nodes in a viewpoint of entropy or uncertainty.

2. NONLINEAR EFFECT ON THE ENTROPY OF HIDDEN NODES

In FNNs, inputs are processed through a series of weighted sums and nonlinear activation functions. When inputs or weights are perturbed randomly, the weighted sums to hidden nodes are approximately jointly Gaussian according to the central limit theorem [19]-[21]. Therefore, we analyze the effect of nonlinear function on jointly Gaussian random variables.

If $\mathbf{u}$ and $\mathbf{v}$ are jointly Gaussian random variables with zero means, then the joint entropy of $\mathbf{u}$ and $\mathbf{v}$ [22] is given by

$$h(\mathbf{u}, \mathbf{v}) = \log \left(2\pi e \sigma_u \sigma_v \sqrt{1 - r^2} \right). \quad (1)$$

Herein, σ_u and σ_v are the standard deviation of $\mathbf{u}$ and $\mathbf{v}$, respectively, and r is the correlation coefficient between $\mathbf{u}$ and $\mathbf{v}$.

Let's assume that $\mathbf{u}$ and $\mathbf{v}$ are transformed into $\mathbf{y}$ and $\mathbf{z}$ as follows:

* Corresponding author, Email: shoh@mokwon.ac.kr
Manuscript received Oct. 01, 2013; revised Jan. 02, 2014;
accepted Jan. 12, 2014

$$\mathbf{y} = \begin{cases} a\mathbf{u}, & \mathbf{u} > 0 \\ b\mathbf{u}, & \mathbf{u} \leq 0 \end{cases} \text{ and } \mathbf{z} = \begin{cases} c\mathbf{v}, & \mathbf{v} > 0 \\ d\mathbf{v}, & \mathbf{v} \leq 0 \end{cases} \quad (2)$$

where a, b, c , and d are nonzero real values. Then, the entropy of $\mathbf{y}$ is given by

$$h(\mathbf{y}) = - \int f_{\mathbf{y}}(y) \log f_{\mathbf{y}}(y) dy, \quad (3)$$

where $f_{\mathbf{y}}(y)$ is the probability density function (p.d.f.) of random variable $\mathbf{y}$. By substituting [19]

$$f_{\mathbf{y}}(y) = \begin{cases} \frac{1}{|a|} f_{\mathbf{u}}\left(\frac{y}{a}\right), & \mathbf{u} > 0 \\ \frac{1}{|b|} f_{\mathbf{u}}\left(\frac{y}{b}\right), & \mathbf{u} \leq 0 \end{cases}, \quad (4)$$

Eq. (3) can be derived by

$$\begin{aligned} h(\mathbf{y}) &= - \int_{-\infty}^0 f_{\mathbf{y}}(y) \log f_{\mathbf{y}}(y) dy - \int_0^{\infty} f_{\mathbf{y}}(y) \log f_{\mathbf{y}}(y) dy \\ &= - \int_{-\infty}^0 f_{\mathbf{u}}(u) \log f_{\mathbf{u}}(u) du + \int_{-\infty}^0 \log|b| f_{\mathbf{u}}(u) du \\ &\quad - \int_0^{\infty} f_{\mathbf{u}}(u) \log f_{\mathbf{u}}(u) du + \int_0^{\infty} \log|a| f_{\mathbf{u}}(u) du \\ &= h(\mathbf{u}) + \frac{1}{2} \log(|ab|) \end{aligned} \quad (5)$$

Here, $f_{\mathbf{u}}(u)$ is the p.d.f. of random variable $\mathbf{u}$ and

$$\int_{-\infty}^0 f_{\mathbf{u}}(u) du = \int_0^{\infty} f_{\mathbf{u}}(u) du = \frac{1}{2}, \quad (6)$$

since $\mathbf{u}$ is Gaussian with zero mean. Accordingly, the entropy of $\mathbf{z}$ is given by

$$h(\mathbf{z}) = h(\mathbf{v}) + \frac{1}{2} \log(|cd|) \quad (7)$$

The joint entropy after the piece-wise-linear transformations is defined by

$$h(\mathbf{y}, \mathbf{z}) = - \iint f_{\mathbf{yz}}(y, z) \log f_{\mathbf{yz}}(y, z) dy dz. \quad (8)$$

The joint p.d.f. of $\mathbf{y}$ and $\mathbf{z}$ can be separated into four quadrants as follows:

$$f_{\mathbf{yz}}(y, z) = \begin{cases} \frac{f_{\mathbf{uv}}\left(\frac{y}{a}, \frac{z}{c}\right)}{|ac|}, & \mathbf{u} > 0 \text{ and } \mathbf{v} > 0 \\ \frac{f_{\mathbf{uv}}\left(\frac{y}{b}, \frac{z}{c}\right)}{|bc|}, & \mathbf{u} \leq 0 \text{ and } \mathbf{v} > 0 \\ \frac{f_{\mathbf{uv}}\left(\frac{y}{b}, \frac{z}{d}\right)}{|bd|}, & \mathbf{u} \leq 0 \text{ and } \mathbf{v} \leq 0 \\ \frac{f_{\mathbf{uv}}\left(\frac{y}{a}, \frac{z}{d}\right)}{|ad|}, & \mathbf{u} > 0 \text{ and } \mathbf{v} \leq 0. \end{cases} \quad (9)$$

Therefore, Eq. (8) is also separated into four parts as

$$\begin{aligned} h(\mathbf{y}, \mathbf{z}) &= - \int_0^{\infty} \int_0^{\infty} f_{\mathbf{yz}}(y, z) \log f_{\mathbf{yz}}(y, z) dy dz \\ &\quad - \int_{-\infty}^0 \int_0^{\infty} f_{\mathbf{yz}}(y, z) \log f_{\mathbf{yz}}(y, z) dy dz \\ &\quad - \int_{-\infty}^0 \int_{-\infty}^0 f_{\mathbf{yz}}(y, z) \log f_{\mathbf{yz}}(y, z) dy dz \\ &\quad - \int_0^{\infty} \int_{-\infty}^0 f_{\mathbf{yz}}(y, z) \log f_{\mathbf{yz}}(y, z) dy dz. \end{aligned} \quad (10)$$

Since the joint p.d.f. of $\mathbf{u}$ and $\mathbf{v}$ is given by

$$\begin{aligned} f_{\mathbf{uv}}(u, v) &= \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-r^2}} \\ &\quad \exp\left[-\frac{1}{2(1-r^2)}\left(\frac{u^2}{\sigma_u^2} - 2r\frac{uv}{\sigma_u\sigma_v} + \frac{v^2}{\sigma_v^2}\right)\right], \end{aligned} \quad (11)$$

the first quadrant term in Eq. (10) is derived by

$$\begin{aligned} &- \int_0^{\infty} \int_0^{\infty} f_{\mathbf{yz}}(y, z) \log f_{\mathbf{yz}}(y, z) dy dz \\ &= A \int_0^{\infty} \int_0^{\infty} f_{\mathbf{yz}}(y, z) dy dz + \int_0^{\infty} \int_0^{\infty} B f_{\mathbf{yz}}(y, z) dy dz. \end{aligned} \quad (12)$$

Here,

$$A = \log\left(2\pi|ac|\sigma_u\sigma_v\sqrt{1-r^2}\right) \quad (13)$$

and

$$B = \frac{1}{2(1-r^2)}\left(\frac{y^2}{a^2\sigma_u^2} - 2r\frac{yz}{|ac|\sigma_u\sigma_v} + \frac{z^2}{c^2\sigma_v^2}\right). \quad (14)$$

The second term in the right side of Eq. (12) becomes

$$\begin{aligned} \int_0^{\infty} \int_0^{\infty} B f_{\mathbf{yz}}(y, z) dy dz &= \frac{1}{2(1-r^2)} \left[\frac{1}{2} \int_0^{\infty} \left(\frac{y^2}{a^2\sigma_u^2}\right) f_{\mathbf{y}}(y) dy + \right. \\ &\quad \left. \frac{1}{2} \int_0^{\infty} \left(\frac{z^2}{c^2\sigma_v^2}\right) f_{\mathbf{z}}(z) dz - \frac{2r}{|ac|\sigma_u\sigma_v} \int_0^{\infty} \int_0^{\infty} yz f_{\mathbf{yz}}(y, z) dy dz \right] \end{aligned} \quad (15)$$

Since $\mathbf{y} = a\mathbf{u}$ is symmetric and zero mean,

$$\begin{aligned} \int_0^{\infty} \left(\frac{y^2}{a^2\sigma_u^2}\right) f_{\mathbf{y}}(y) dy &= \frac{1}{a^2\sigma_u^2} \int_0^{\infty} y^2 f_{\mathbf{y}}(y) dy \\ &= \frac{1}{a^2\sigma_u^2} \frac{1}{2} \int_{-\infty}^{\infty} y^2 f_{\mathbf{y}}(y) dy \\ &= \frac{1}{a^2\sigma_u^2} \frac{1}{2} a^2 \sigma_u^2 \\ &= \frac{1}{2}. \end{aligned} \quad (16)$$

And according to the same procedure,

$$\int_0^{\infty} \left(\frac{z^2}{c^2\sigma_v^2}\right) f_{\mathbf{z}}(z) dz = \frac{1}{2}. \quad (17)$$

By Papoulis [19],

$$\begin{aligned} \int_0^{\infty} \int_0^{\infty} f_{\mathbf{yz}}(y, z) dy dz &= \int_{-\infty}^0 \int_{-\infty}^0 f_{\mathbf{yz}}(y, z) dy dz \\ &= \frac{1}{4} + \frac{1}{2\pi} \tan^{-1} \frac{r}{\sqrt{1-r^2}} \end{aligned} \quad (18)$$

and

$$\begin{aligned} \int_{-\infty}^0 \int_0^{\infty} f_{yz}(y, z) dy dz &= \int_0^{\infty} \int_{-\infty}^0 f_{yz}(y, z) dy dz \\ &= \frac{1}{4} - \frac{1}{2\pi} \tan^{-1} \frac{r}{\sqrt{1-r^2}}. \end{aligned} \quad (19)$$

Also by Oh [23],

$$\int_0^{\infty} \int_0^{\infty} yz f_{yz}(y, z) dy dz = \frac{|ac|\sigma_u\sigma_v}{2\pi} \left[\frac{r}{2\pi} + \sqrt{1-r^2} + r \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right] \quad (20)$$

$$\int_{-\infty}^0 \int_{-\infty}^0 yz f_{yz}(y, z) dy dz = \frac{|bd|\sigma_u\sigma_v}{2\pi} \left[\frac{r}{2\pi} + \sqrt{1-r^2} + r \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right] \quad (21)$$

$$\int_{-\infty}^0 \int_0^{\infty} yz f_{yz}(y, z) dy dz = \frac{|bc|\sigma_u\sigma_v}{2\pi} \left[\frac{r}{2\pi} - \sqrt{1-r^2} - r \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right] \quad (22)$$

and

$$\int_0^{\infty} \int_{-\infty}^0 yz f_{yz}(y, z) dy dz = \frac{|ad|\sigma_u\sigma_v}{2\pi} \left[\frac{r}{2\pi} - \sqrt{1-r^2} - r \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right] \quad (23)$$

By substituting Eqs. (18) and (20) into Eq. (12),

$$\begin{aligned} &A \int_0^{\infty} \int_0^{\infty} f_{yz}(y, z) dy dz + \int_0^{\infty} \int_0^{\infty} B f_{yz}(y, z) dy dz \\ &= A \left[\frac{1}{4} + \frac{1}{2\pi} \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right] \\ &\quad + \frac{1}{2(1-r^2)} \left[\frac{1}{2} - \frac{r}{\pi} \left(\frac{r}{2\pi} + \sqrt{1-r^2} + r \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right) \right]. \end{aligned} \quad (24)$$

Using the same procedure from Eq. (11) to (24), we can derive that

$$\begin{aligned} &-\int_{-\infty}^0 \int_0^{\infty} f_{yz}(y, z) \log f_{yz}(y, z) dy dz \\ &= \left[\frac{1}{4} - \frac{1}{2\pi} \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right] \log \left(2\pi |bc| \sigma_u \sigma_v \sqrt{1-r^2} \right) \\ &\quad + \frac{1}{2(1-r^2)} \left[\frac{1}{2} - \frac{r}{\pi} \left(\frac{r}{2\pi} - \sqrt{1-r^2} - r \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right) \right], \end{aligned} \quad (25)$$

$$\begin{aligned} &-\int_{-\infty}^0 \int_{-\infty}^0 f_{yz}(y, z) \log f_{yz}(y, z) dy dz \\ &= \left[\frac{1}{4} + \frac{1}{2\pi} \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right] \log \left(2\pi |bd| \sigma_u \sigma_v \sqrt{1-r^2} \right) \\ &\quad + \frac{1}{2(1-r^2)} \left[\frac{1}{2} - \frac{r}{\pi} \left(\frac{r}{2\pi} + \sqrt{1-r^2} + r \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right) \right], \end{aligned} \quad (26)$$

and

$$\begin{aligned} &-\int_0^{\infty} \int_{-\infty}^0 f_{yz}(y, z) \log f_{yz}(y, z) dy dz \\ &= \left[\frac{1}{4} - \frac{1}{2\pi} \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right] \log \left(2\pi |ad| \sigma_u \sigma_v \sqrt{1-r^2} \right) \\ &\quad + \frac{1}{2(1-r^2)} \left[\frac{1}{2} - \frac{r}{\pi} \left(\frac{r}{2\pi} - \sqrt{1-r^2} - r \tan^{-1} \frac{r}{\sqrt{1-r^2}} \right) \right]. \end{aligned} \quad (27)$$

By substituting Eqs. (24), (25), (26), and (27) into Eq. (10), we attain

$$\begin{aligned} h(\mathbf{y}, \mathbf{z}) &= \log \left(2\pi e \sigma_u \sigma_v \sqrt{|abcd| (1-r^2)} \right) \\ &= \log \left(2\pi e \sigma_u \sigma_v \sqrt{1-r^2} \right) + \log \sqrt{|abcd|}. \end{aligned} \quad (28)$$

Finally we take the derivation

$$h(\mathbf{y}, \mathbf{z}) = h(\mathbf{u}, \mathbf{v}) + \log \sqrt{|abcd|} \quad (29)$$

with the substitution of Eq. (1) into (28).

In FNNs, we usually use the $\tanh(\cdot)$ sigmoidal function (Fig. 1) as the nonlinear activation function of hidden node. Assuming that the nonlinear activation function can be approximated piece-wise linearly as in Eq. (2), then the parameters a , b , c , and d correspond to the slope of nonlinear activation function. As shown in Fig. 2, the slope is less than or equal to 0.5. Thus, $\log \sqrt{|abcd|} < 0$ and $h(\mathbf{y}, \mathbf{z}) < h(\mathbf{u}, \mathbf{v})$. Also, the less steep the slopes are, the more the joint entropy after the nonlinear function decreases.

Consequently, we can argue that the nonlinear activation function decreases the entropy or uncertainty of hidden nodes. The sigmoid activation function can be separated into linear region with steep slope and saturated region with gentle slope as shown in Fig. 1. When hidden node values are in the saturation region of sigmoid activation function, the entropy decreases much more. This coincides with the argument that hidden nodes tend to be saturated after successful training.

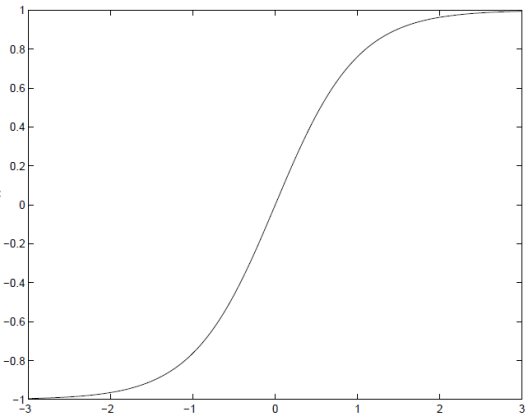


Fig. 1. The sigmoid activation function of hidden node

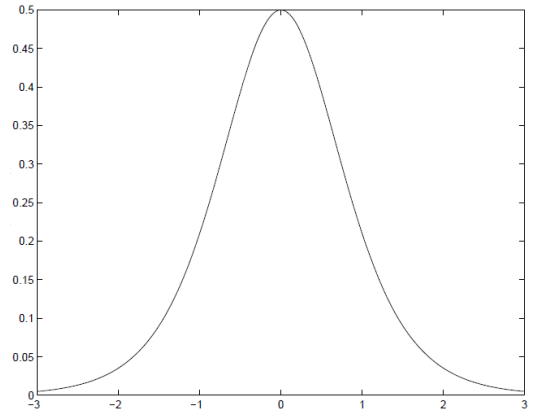


Fig. 2. The slope of the sigmoid activation function

Furthermore, the entropy of hidden nodes provides hierarchical understanding of information extraction capabilities acquired through learning. Lee et al. argued that input samples have the inter-class information as well as the intra-class variation [7]. The inter-class information is the information content that an input sample belongs to a specific class, and the intra-class variation is a measure of the average variations within the classes including noise contaminations. After learning, input samples are projected to hidden nodes through weighted sums and element-wise nonlinear transformations. In this paper, we proved that the entropy of hidden nodes decreases after the nonlinear transformations. The decreasing of hidden nodes' entropy is correspond to the decreasing of intra-class variations, as pointed out by Lee et al [7].

3. CONCLUSIONS

In this paper, we prove that the entropy of jointly Gaussian random variables decreases after piece-wise linear transformations. Also, the less steep the slopes are, the more the joint entropy decreases. Since the nonlinear activation function of hidden nodes can be approximated piece-wise linearly, we can argue that the nonlinear activation function decreases the uncertainty among hidden nodes. Furthermore, the entropy of hidden nodes decrease much more after successful training of FNNs.

REFERENCES

- [1] K. Hornik, M. Stincombe, and H. White, "Multilayer feed-forward networks are universal approximators," *Neural Networks*, vol. 2, 1989, pp. 359-366.
- [2] K. Hornik, "Approximation Capabilities of Multilayer Feedforward Networks," *Neural Networks*, vol. 4, 1991, pp. 251-257.
- [3] S. Suzuki, "Constructive Function Approximation by Three-layer Artificial Neural Networks," *Neural Networks*, vol. 11, 1998, pp. 1049-1058.
- [4] Y. Liao, S. C. Fang, and H. L. W. Nuttle, "Relaxed Conditions for Radial-Basis Function Networks to be Universal Approximators," *Neural Networks*, vol. 16, 2003, pp. 1019-1028.
- [5] S. H. Oh and Y. Lee, "Effect of Nonlinear Transformations on Correlation Between Weighted Sums in Multilayer Perceptrons," *IEEE Trans., Neural Networks*, vol. 5, 1994, pp. 508-510.
- [6] J. V. Shah and C. S. Poon, "Linear Independence of Internal Representations in Multilayer Perceptrons," *IEEE Trans., Neural Networks*, vol. 10, 1999, pp. 10-18.
- [7] Y. Lee and H. K. Song, "Analysis on the Efficiency of Pattern Recognition Layers Using Information Measures," *Proc. IJCNN'93 Nagoya*, vol. 3, Oct. 1993, pp. 2129-2132.
- [8] A. El-Jaroudi and J. Makhoul, "A New Error Criterion for Posterior probability Estimation with Neural Nets," *Proc. IJCNN'90*, vol. 3, June 1990, pp. 185-192.
- [9] M. Bichsel and P. Seitz, "Minimum Class Entropy: A maximum Information Approach to Layered Networks," *Neural Networks*, vol. 2, 1989, pp. 133-141.
- [10] S. Ridella, S. Rovetta, and R. Zunino, "Representation and Generalization Properties of Class-Entropy Networks," *IEEE Trans. Neural Networks*, vol. 10, 1999, pp. 31-47.
- [11] D. Erdogmus and J. C. Principe, "Entropy Minimization Algorithm for Multilayer Perceptrons," *Proc. IJCNN'01*, vol. 4, 2001, pp. 3003-3008.
- [12] K. E. Hild II, D. Erdogmus, K. Torkkola, and J. C. Principe, "Feature Extraction Using Information-Theoretic Learning," *IEEE Trans. PAMI*, vol. 28, no. 9, 2006, pp. 1385-1392.
- [13] R. Li, W. Liu, and J. C. Principe, "A Unifying Criterion for Instantaneous Blind Source Separation Based on Correntropy," *Signal Processing*, vol. 87, no. 8, 2007, pp. 1872-1881.
- [14] S. Ekici, S. Yildirim, and M. Poyraz, "Energy and Entropy-Based Feature Extraction for Locating Fault on Transmission Lines by Using Neural Network and Wavelet packet Decomposition," *Expert Systems with Applications*, vol. 34, 2008, pp. 2937-2944.
- [15] D. Erdogmus and J. C. Principe, "Information Transfer Through Classifiers and Its Relation to Probability of Error," *Proc. IJCNN'01*, vol. 1, 2001, pp. 50-54.
- [16] S. J. Lee, M. T. Jone, and H. L. Tsai, "Constructing Neural Networks for Multiclass-Discretization Based on Information Theory," *IEEE Trans. Sys., Man, and Cyb.-Part B*, vol. 29, 1999, pp. 445-453.
- [17] R. Kamimura and S. Nakanishi, "Hidden Information maximization for Feature Detection and Rule Discovery," *Network: Computation in Neural Systems*, vol. 6, 1995, pp. 577-602.
- [18] K. Torkkola, "Nonlinear Feature Transforms Using Maximum Mutual Information," *Proc. IJCNN'01*, vol. 4, 2001, pp. 2756-2761.
- [19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, second ed., New York: McGraw-Hill, 1984.
- [20] Y. Lee, S. H. Oh, and M. W. Kim, "An Analysis of Premature Saturation in Back-Propagation Learning," *Neural Networks*, vol. 6, 1993, pp. 719-728.
- [21] Y. Lee and S. H. Oh, "Input Noise Immunity of Multilayer Perceptrons," *ETRI Journal*, vol. 16, 1994, pp. 35-43.
- [22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, INC. 1991.
- [23] S. H. Oh, "Decreasing of Correlations Among Hidden Neurons of Multilayer Perceptrons," *Journal of the Korea Contents Association*, vol. 3, no. 3, 2003, pp. 98-102.

**Sang-Hoon Oh**

received his B.S. and M.S. degrees in Electronics Engineering from Pusan National University in 1986 and 1988, respectively. He received his Ph.D. degree in Electrical Engineering from Korea Advanced Institute of Science and Technology in 1999. From 1988 to 1989,

he worked for LG semiconductor, Ltd., Korea. From 1990 to 1998, he was a senior research staff in Electronics and Telecommunication Research Institute (ETRI), Korea. From 1999 to 2000, he was with Brain Science Research Center, KAIST. In 2000, he was with Brain Science Institute, RIKEN, Japan, as a research scientist. In 2001, he was an R&D manager of Extell Technology Corporation, Korea. Since 2002, he has been with the Department of Information Communication Engineering, Mokwon University, Daejeon, Korea, and is now a full professor. Also, he was with the Division of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, USA, as a visiting scholar from August 2008 to August 2009. His research interests are machine learning, speech signal processing, pattern recognition, and bioinformatics.