https://doi.org/10.5392/IJoC.2017.13.4.023

Improving the Water Level Prediction of Multi-Layer Perceptron with a Modified Error Function

Sang-Hoon Oh
Department Division of Information Communication Convergence Engineering,
Mokwon University, Daejeon, 302-729, Korea

ABSTRACT

Of the total economic loss caused by disasters, 40% are due to floods and floods have a severe impact on human health and life. So, it is important to monitor the water level of a river and to issue a flood warning during unfavorable circumstances. In this paper, we propose a modified error function to improve a hydrological modeling using a multi-layer perceptron (MLP) neural network. When MLP's are trained to minimize the conventional mean-squared error function, the prediction performance is poor because MLP's are highly tunned to training data. Our goal is achieved by preventing overspecialization to training data, which is the main reason for performance degradation for rare or test data. Based on the modified error function, an MLP is trained to predict the water level with rainfall data at upper reaches. Through simulations to predict the water level of Nakdong River near a UNESCO World Heritage Site "Hahoe Village," we verified that the prediction performance of MLP with the modified error function is superior to that with the conventional mean-squared error function, especially maximum error of 40.85cm vs. 55.51cm.

Key words: *Flooding, Water Level Prediction, Hahoe Village, Error-Back Propagation, Modified Error Function, Muti-Layer Perceptron, Deep Learning.*

1. INTRODUCTION

Flood protection and forecasting have been focused because of its severe impact on the lives and properties in a wide area [1]-[4]. Accordingly, there have been many efforts to forecast flood or river flow which can be categorized into deterministic, conceptual, and parametric models [5]. Deterministic models use physical laws of mass and energy transfer to describe the relationship between rainfall and runoff or flood discharge. Conceptual models describe simplified representations of key hydrological process using a perceived system. Parametric models try to find mathematical transfer functions to relate several variables to runoff. In this point of view, neural network models belong to the parametric models [5].

Based on the mathematical proofs that multi-layer perceptron (MLP) neural networks can approximate any function with enough number of hidden nodes [6]-[8], neural networks have been widely applied in various fields such as fraud detection, pattern recognition, speech recognition, telecommunications, time series prediction, hydrology, etc [9]-[12]. Furthermore, deep learning enlarges the application area to image understanding and language processing [13]. Especially, it is not necessary to elucidate complex mechanisms of phenomena to be modeled by MLP's, and this property expands hydrological modeling using MLP neural networks [5].

Feng and Lu developed neural networks to forecast the peak stage in the lower reaches [2]. Atiya and Shaheen applied neural networks to the problem of forecasting the flow of the River Nile in Egypt with multi-step ahead predictions [14]. Dawson et al. predicted T-year flood events and the index flood for 850 catchments across the UK [15]. Wei et al. applied neural networks to predict the flood disaster area in China [16]. Rajurkar et al. modeled daily flows during flood events using neural networks [17]. Riad et al. showed that neural networks could model the rainfall-runoff relationship in a semiarid region in Morocco although there were extreme events such as floods and droughts with irregularity [18], [19]. Furthermore, Chua et al. combined neural networks with a kinematic wave approach for improving event-based rainfall-runoff modeling [20]. However, all these neural networks were trained based on the error back-propagation (EBP) algorithm to minimize mean-squared error (MSE) function for training data [21]. Also, there have been efforts of applying recurrent neural networks to water level prediction [22], [23]. Still these methods try to minimize MSE functions for training data.

In hydrological modeling, it is essential to predict the peak of hydrograph or water level [1]. On the contrary, given the nature of hydrological data, there is an imbalance of data in which low or medium level data are very much dominant than high-level data [1], [15]. When training neural networks to

* *Corresponding author, Email: ohsanghoon16@gmail.com*
Manuscript received Jul. 12, 2017; revised Dec. 04, 2017; accepted Dec. 15, 2017

minimize MSE, neural networks become highly tuned to the dominant training data with low or medium level and this overspecialization causes poor performance for the peak or test data [15]. In pattern recognition applications of neural networks, there have been reports that overspecializations degrade performance for test data and this can be prevented through modified error or objective functions [24], [25]. Also, algorithmic level approaches were attempted to improve the classification performance for heavily imbalanced data [26], [27]. Although these methods were successful in classifications, they cannot be adopted in hydrological modeling since target values in classification problems are not continuous but discrete.

In this paper, we propose a modified error function to improve the performance of water level prediction through preventing overspecialization for the dominant training data. This purpose is achieved by weak updating of weights when MLP outputs are near their desired values. UNESCO concluded the "Convention Concerning the Protection of the World Cultural and Natural Heritage" in 1972 for national and international protection activities of world heritages. Among many UNESCO world heritage sites in Korea, "Hahoe Village" in Andong region is adjacent to Nakdong River and, therefore, the water level near the village should be carefully monitored for protection from flooding. So, we simulate the proposed method to predict the water level of Nakdong River near "Hahoe Village" with rainfalls at the upper reaches. In section 2, we briefly introduce MLP neural networks and its EBP training algorithm. Also, we propose a modified error function for preventing overspecialization. In section 3, the hydrological modeling of rainfalls and water level near "Hahoe Village" is described and verified through simulations of the proposed mehod with real data. Finally, section 4 concludes this paper.

2. ERROR BACK-PROPAGATION ALGORITHM WITH A MODIFIED ERROR FUNCTION

2.1 Error Back-Propagation Algorithm

Consider an MLP consisting of *N* inputs, *H* hidden nodes, and *M* output nodes, as shown in Fig. 1. When a *p*-th training data $\mathbf{x}^{(p)} = [x_1^{(p)}, x_2^{(p)}, \ldots, x_N^{(p)}] \quad (p = 1, 2, \ldots, P)$ is presented to the input layer of MLP, by the forward propagation, the *j*-th hidden node is given by

$$h_j^{(p)} = h_j(\mathbf{x}^{(p)}) = \tanh((w_{j0} + \sum_{i=1}^{N} w_{ji} x_i^{(p)})/2), \; j = 1, 2, \ldots, H. \tag{1}$$

Here, w_{ji} denotes the weight connecting x_i to h_j, w_{j0} is a bias, and tanh(.) is the sigmoidal activation function of hidden node. We usually adopt the sigmoidal activation function of output node for classification problems, in which desired values of output nodes are in two extremely saturated regions of the sigmoid function [21]. If we want to generate "warning-no warning signal" for flood forecasting, it belongs to classification problems [1]. However, our goal is to predict the water level near "Hahoe Village" which is a real number above zero, and we adopt a linear function as an activation function of output node. Consequently, the *k*-th output node with a linear activation function is given by

$$y_k^{(p)} = y_k(\mathbf{x}^{(p)}) = v_{k0} + \sum_{j=1}^{H} v_{kj} h_j^{(p)}, \; k = 1, 2, \ldots, M. \tag{2}$$

Also, v_{k0} is a bias and v_{kj} denotes the weight connecting h_j to y_k.

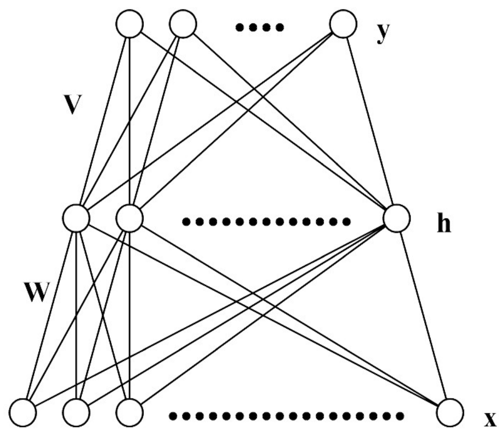


Fig. 1. The architecture of a multilayer perceptron.

Let the desired output vector corresponding to a training sample $\mathbf{x}^{(p)}$ be $\mathbf{t}^{(p)} = [t_1^{(p)}, t_2^{(p)}, \ldots, t_M^{(p)}]$. As a distance measure between the actual and desired outputs, the mean-squared error (MSE) function for *P* training data is defined by

$$E = \frac{1}{2} \sum_{p=1}^{P} \sum_{k=1}^{M} \left(t_k^{(p)} - y_k^{(p)}\right)^2. \tag{3}$$

To minimize *E,* according to the negative gradient of MSE, output weights v_{kj} 's are iteratively updated by

$$\Delta v_{kj} = -\eta \frac{\partial E}{\partial v_{kj}} = \eta \delta_k^{(p)} h_j^{(p)}, \tag{4}$$

where

$$\delta_k^{(p)} = -\frac{\partial E}{\partial y_k^{(p)}} = \left(t_k^{(p)} - y_k^{(p)}\right) \tag{5}$$

is the error signal of output node and η is the learning rate. Also, by the backward propagation of the error signal, hidden weights w_{ji} 's are updated by

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}} = \eta x_i^{(p)} \frac{(1 - h_j^{(p)})(1 + h_j^{(p)})}{2} \sum_{k=1}^{M} v_{kj} \delta_k^{(p)}. \tag{6}$$

In Eq. (6), the error signal of output node is back-propagated through output weight v_{kj}. The above weight-updating procedure for training of MLP is the EBP (error back-propagation) algorithm [21].

2.2 Modified Error Function for Preventing Over-specialization

During the learning process, the direction of weight update for reducing an error associated with a specific input will assist or compete with that for reducing total error [28]. For instance, some output nodes are pushed away from desired values by competition in the network. In this case, a strong error is necessary for output nodes far from desired values [29]. For output nodes near desired values, a weak error signal needs to be generated so that the weight update associated with a training data scarcely perturb the weights trained for whole training data [24]. Above all, training a network too much to minimize error may mean the network has become highly tuned to the training data leading to an inability to generalize [15]. Furthermore, this overspecialization for training data degrades performance for rare data such as peak values. For preventing overspecialization of learning, the weak error signal also must be generated for output nodes near desired values, like the classification figure of merit method [25].

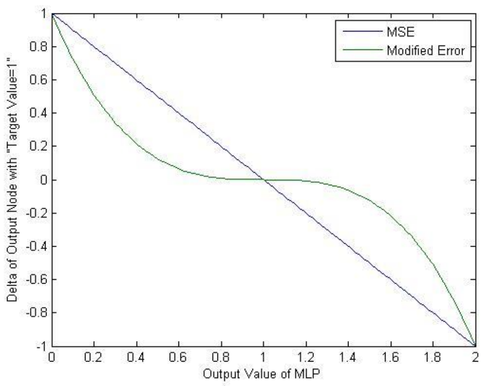


Fig. 2. Error signals ($\delta_k^{(p)}$'s) using MSE and the modified error function with n=4 when the desired value is '1'

In this sense, we propose the modified error function

$$E_m = \frac{1}{n}\sum_{p=1}^{P}\sum_{k=1}^{M}\left(t_k^{(p)} - y_k^{(p)}\right)^n, \qquad (7)$$

where n is an even number. Using the above error function, the error signal of output node is

$$\delta_k^{(p)} = -\frac{\partial E_m}{\partial y_k^{(p)}} = \left(t_k^{(p)} - y_k^{(p)}\right)^{n-1} \qquad (8)$$

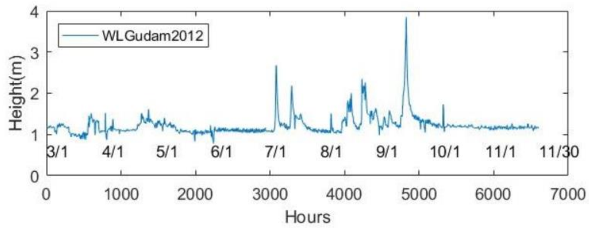
The other equations in the EBP algorithm are the same. As shown in Fig. 2, the proposed error signal can satisfy the above criterion, which requests a weak error signal for output node near desired values and a strong error signal for output node far from desired values.

3. HYDROLOGICAL MODELING NEAR "HAHOE VILLAGE"

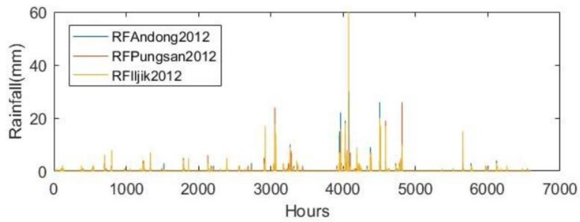
3.1 Hahoe Village

Fig. 3. The map of Andong region. Red circles indicate the locations of water level gauge at "Gudam" and rainfall gauges at "Pungsan," "Iljik," and "Andong"

Fig. 3 is a map of Andong region, in which Nakdong River flows from east of Andong to "Hahoe Village." Red circles in Fig. 3 indicate the locations of water level gauge at "Gudam" and three rainfall gauges at "Pungsan," "Iljik," and "Andong." Since there is not a water level gauge at "Hahoe Village" and the nearest one is the gauge at "Gudam," we select the gauge at "Gudam" to monitor the water level near "Hahoe Village." Also, the rainfalls at "Pungsan" and "Iljik" should be considered for the hydrological modeling, since there are tributary rivers from the two locations to Nakdong River [30].



(a)



(b)

Fig. 4. Collected data in 2012. (a) Water level at "Gudam" (WL Gudam2012), (b) Rainfalls at "Andong" (RF Andong2012), "Pungsan" (RF Pungsan2012) and "Iljik" (RF Iljik2012)

Avoiding the winter season of icing and snowing, the data at each gauge is collected from March 1st to November 30th in the year of 2012, 2013, and 2014 with the interval of one hour.

So, there are 6,600 records in each year. The total data was provided by "Nakdong River Flood Control Office" and "Korea Water Resource Corp" [30]. We use the data in 2012 and 2013 to train MLP's for the hydrological modeling and the other to test the performance of water level prediction. Fig. 4 shows the water level and rainfalls in the year of 2012, in which we can find many irregularities between the water level and the rainfalls. For easy readability of the horizontal axis, we include the index "month/day" just above the data index on the horizontal axis in Fig. 4(a).

3.2 Predicting the Water Level with the Modified Error Function

In order to predict the water level at "Gudam" after *D* hours, we construct an MLP whose input layer consists of the water level at "Gudam" and the rainfalls at "Pungsan", "Iljik", and "Andong" from current (denoted by "*c*") to previous *c*-*L* hours [17], [30]. When deciding the parameter values in MLP's, we usually adopt the method of trial and error since there is not a concrete theoretical guidance. Accordingly, we determine that *L* is two and the number of hidden nodes is forty through many trials and errors. Also, we adopt a single output node with linear activation function to predict the water level at "Gudam" after *D* hours. Therefore, the MLP architecture is 12 inputs, 40 hidden nodes with tanh(.) activation function, and one linear output node [30].

After initializing the MLP with random weights uniformly distributed on $[-1\times10^{-4}, 1\times10^{-4}]$, the EBP algorithm updates the weights v_{kj} 's and w_{ji} 's to minimize the modified error function with *n*=4 for training data in 2012 and 2013. Also, we simulate the conventional EBP algorithm with MSE function for comparison [30]. Since no fair comparison is possible if the learning rate is kept the same for all methods [29], we derive the learning rate so that $\int_0^1 \eta\delta_k^{(p)} dy = 0.005$ for each method under the assumption that *y* is uniform on [0,1]. As a result, the learning rates of 0.01 and 0.02 are used for the conventional EBP and proposed methods, respectively. The training procedure of water level prediction is described as follows:

1. Initialization of MLP with random weights uniformly distributed on $[-1\times10^{-4}, 1\times10^{-4}]$
2. Presentation of training data to the input layer of MLP
3. Calculation of hidden and output node values according to Eqs. (1) or (2)
4. Calculation of $\delta_k^{(p)}$'s according to Eqs. (5) or (8)
5. Weight updates according to Eqs. (4) and (6)
6. Estimation of error for training and test data
7. Termination of training or go to step 2.

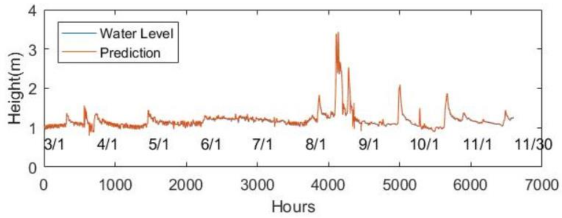
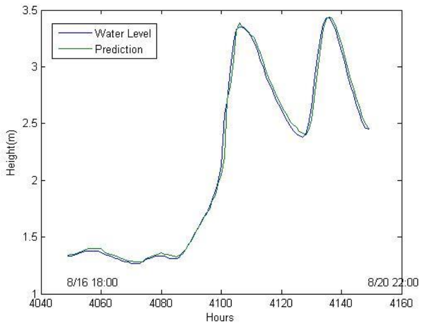
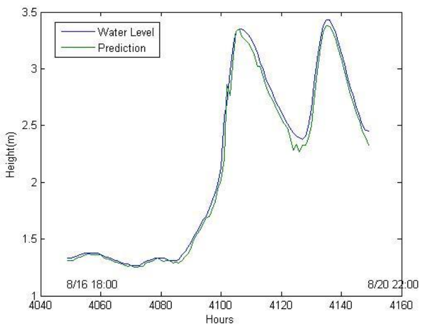


Fig. 5. The water level at "Gudam" and its predicted value by the MLP with *D*=1 after 10,000 iterations of the proposed learning (March 1st ~ November 30th, 2014)



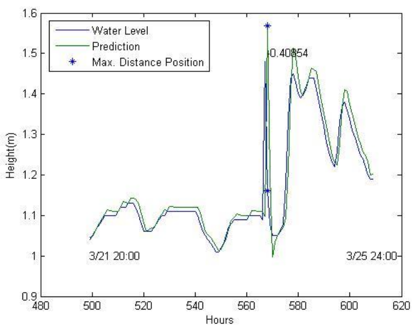
(a)



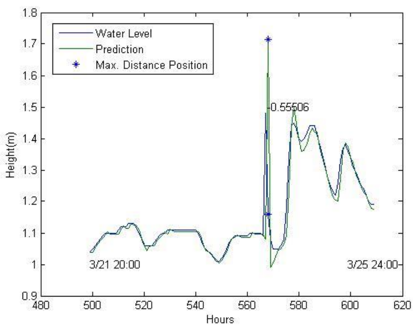
(b)

Fig. 6. The water level at "Gudam" and its predicted value in August 16th 18:00 ~ 20th 22:00, 2014. (a) the proposed method, (b) the conventional EBP method

Firstly, we train the MLP with *D*=1 for 10,000 iterations. For verifying the prediction performance for test data, we plot the real and predicted values of the water level at "Gudam" in 2014. Fig. 5 shows the predicted results using the proposed method. We can find that the predicted values are globally very close to real values. Also, we plot detail curves of the real and predicted ones in the period of August 16th to 20th – the period with the highest peak- in Fig. 6. Fig. 6(a) shows that, in the proposed method, the prediction for the high-level values is very close to the real values. Contrary, as shown in Fig. 6(b), the predicted value using the conventional EBP method is rough and worse than that using the proposed method [30].



(a)



(b)

Fig. 7. The water level at "Gudam" and its predicted value in March 21st 20:00~25th 24:00, 2014. (a) the proposed method, (b) the conventional EBP method. 0.40854 and 0.55506 are the maximum differences between real and predicted values in (a) and (b), respectively

Also, we estimate the period of the maximum difference between real and predicted values and the detail curves in the period of March 21st to 25th are drawn in Fig. 7. The maximum difference in the proposed method is 40.85cm (Fig. 7(a)), which is much smaller than 55.50cm in the conventional EBP method (Fig. 7(b)). From Figs. 6 and 7, we can argue that the proposed method prevents overspecialization of learning to dominant data with low or medium level and closely approximates rare data including peaks.

For further comparison, we measure the correlation coefficient between real data **t** and its predicted one **y** defined by

$$Corr[\mathbf{t},\mathbf{y}] = \frac{E[\mathbf{ty}] - E[\mathbf{t}]E[\mathbf{y}]}{\sigma_{\mathbf{t}}\sigma_{\mathbf{y}}}, \quad (9)$$

where E[.] is the expectation operator and $\sigma_{\mathbf{t}}$ and $\sigma_{\mathbf{y}}$ are standard deviations of **t** and **y**, respectively. The correlation coefficient is independent of the scale of data used and ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation) [5]. In the proposed method, the correlation coefficient of water level prediction for test data is 0.99714, which is greater than 0.99692 in the conventional EBP method. Therefore, the proposed method globally predicts the water level better.

4. CONCLUSIONS

Although it is essential to predict the peak of hydrograph or water level, low or medium level data are very much dominant than high-level data. When training MLP neural networks by the conventional EBP method, MLP's become highly tuned to the dominant data and this overspecialization degrades prediction performance of rare data including peaks. In this paper, we proposed a modified error function to improve the water level prediction in hydrological modeling by preventing overspecializations. The proposed method improved the prediction performance through weakening weight update for output nodes near desired values and intensifying weight update for output nodes far from desired values. This strategy had effects that learning of rare data is encouraged but learning of dominant data is discouraged, contrary to the conventional EBP method.

Among many UNESCO world heritages in Korea, "Hahoe Village" in Andong region is adjacent to Nakdong River and it is important to monitor the water level near the village for protection from flooding. We developed the hydrological modeling to predict the water level near "Hahoe Village" using the proposed learning method. Since the proposed method prevents overspecialization for the dominant data with low or medium levels, the prediction for peak values is more precise than that using the conventional EBP method. Through estimating the correlation coefficient between real and predicted values, we also verified that the global prediction of proposed method is better despite weak updating of weights for output nodes near desired values.

REFERENCES

[1] H. L. Cloke and F. Pappenberger, "Ensemble flood forecasting: A review," Journal of Hydrology, vol. 375, 2009, pp. 613-626.

[2] L. H. Feng and J. Lu, "The practical research on flood forecasting based on artificial neural networks," Expert Systems with Applications, vol. 37, 2010, pp. 2974-2977.

[3] D. Demeritt, H. Cloke, F. Pappenberger, J. Thielen, J. Bartholmes, and M.-H. Ramos, "Ensemble predictions and perceptions of risks, uncertainty, and error in flood forecasting," Environmental Hazards, vol. 7, no. 2, 2007, pp.115-127.

[4] A. de Roo, B. Gouweleeuw, J. Thielen, J. Bartholmes, P. Bongioannini-Cerlini, E. Todini, P. D. Bates, M. Horritt, N. Hunter, K. Beven, F. Pappenberger, E. Heise, G. Rivin, M. Hils, A. Hollingsworth, B. Holst, J. Kwadijk, P. Reggiani, M. van Dijk, K. Sattler, and E. Sprokkereef, "Development of a European flood forecasting system," International Journal of River Basin Management, vol. 1, 2003, pp. 49-59.

[5] C. W. Dawson and R. L. Wilby, "Hydrological modeling using artificial neural networks," Progress in Physical Geography, vol. 25, 2001, pp. 80-108.

[6] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feed-forward Networks are Universal Approximators," Neural Networks, vol. 2, 1989, pp. 359-366.

[7] K. Hornik, "Approximation Capabilities of Multilayer Feedforward Networks," Neural Networks, vol. 4, 1991, pp. 251-257.
[8] S. Suzuki, "Constructive Function Approximation by Three-Layer Artificial Neural Networks," Neural Networks, vol. 11, 1998, pp. 1049-1058.
[9] C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection: classification of skewed data," ACM SIGKDD Explorations Newsletter, vol. 6, issue. 1, 2004, pp. 50-59.
[10] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Science+Business Media LLC, 2006.
[11] F. Altiparmak, B. Dengiz, and A. E. Smith, "A general neural network model for estimating telecommunications network reliability," IEEE Trans. Reliability, vol. 58, issue. 1, 2009, pp. 2-9.
[12] I. Kaaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," Neurocomputing, vol. 10, issue. 3, 1996, pp. 215-236.
[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, 2015, pp. 436-444.
[14] A. Atiya and S. Shaheen, "A comparison between neural-network forecasting techniques-case study: river flow forecasting," IEEE Trans. Neural Networks, vol. 10, no. 2, 1999, pp. 402-409.
[15] C. W. Dawson, R. J. Abrahart, A. Y. Shamseldin, and R. L. Wilby, "Flood estimation at ungauged sites using artificial neural networks," Journal of Hydrology, vol. 319, 2006, pp. 391-409.
[16] Y. Wei, W. Xu, Y. Fan, and H.-T. Tasi, "Artificial neural network based predictive method for flood disaster," Computers & Industrial Engineering, vol. 42, 2002, pp. 383-390.
[17] M. P. Rajurkar, U. C. Kothyari, and U. C. Chaube, "Modelling of the daily rainfall-runoff relationship with artificial neural network," Journal of Hydrology, vol. 285, 2004, pp. 96-113.
[18] S. Raid, J. Mania, L. Bouchaou, and Y. Najjar, "Rainfall-runoff model using an artificial neural network approach," Mathematical and Computer Modelling, vol. 40, 2004, pp. 839-846.
[19] S. Raid, J. Mania, L. Bouchaou, and Y. Najjar, "Predicting catchment flow in a semi-arid region via an artificial neural network," Hydrological Processes, vol. 18, 2004, pp. 2387-2393.
[20] L. H. C. Chua and T. S. W. Wong, "Improving event-based rainfall-runoff modeling using a combined artificial neural network-kinematic wave approach," Journal of Hydrology, vol. 390, 2010, pp. 92-107.
[21] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, Cambridge, MA, 1986.
[22] Q. K. Tran and S. K. Song, "Water level forecasting based on deep learning: a use case of Trinity River-Texas-The United Sattes," Journal of KIISE, vol. 44, 2017, pp. 607-612.
[23] F. J. Chang, P. A. Chen, Y. R. Lu, E. Huang, and K. Y. Chang, "Real-time multi-step-ahead water level forecasting by recurrent neural neyworks for urban flood control," Journal of Hydrology, vol. 517, 2014, pp. 836-846.
[24] S. H. Oh, "Improving the error back-propagation algorithm with a modified error function," IEEE Trans. Neural Networks, vol. 8, 1997, pp. 799-803.
[25] J. B. Hampshire II and A. H. Waibel, "A novel objective function for improved phoneme recognition using time-delay neural networks," IEEE Trans. Neural Networks, vol. 1, 1990, pp. 216-228.
[26] S. H. Oh, "Error back-propagation algorithm for classification of imbalanced data," Neurocomputing, vol. 74, 2011, pp. 1058-1061.
[27] L. Bruzzone and S. B. Serpico, "Classification of imbalanced remote-sensing data by neural networks," Pattern Recognition Letters, vol. 18, 1997, pp. 1323-1328.
[28] R. K. Cheing, I. Lustig, and A. L. Kornhauser, "Relative effectiveness of training set patterns for backpropagation," Proc. IJCNN, Washington, D.C., Jan 15-19, 1990, vol. I, pp. 673-678.
[29] A. van Ooyen and B. Nienhuis, "Improving the convergence of the backpropagation algorithm," Neural Networks, vol. 5, 1992, pp. 465-471.
[30] S. H. Oh and H. Wakuya, "Hydrological modeling of water level near "Hahoe Village" based on multi-layer perceptron," Int. Journal of Contents, vol. 12, no. 1, 2016, pp. 49-53.

Sang-Hoon Oh

He 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, Daejon, Korea, and is now an associate 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.