Analysis variation value momentum algorithm backpropagation method in the recognizing process of temperature pattern in Medan

Purwa Hasan Putra, Muhammad Zarlis

Fakultas Ilmu Komputer dan Teknologi Informasi, University of Sumatera Utara, Medan, Indonesia

purwahasanputra21@gmail.com m.zarlis@usu.ac.id

Abstract. In this paper, From the results of the study it can be concluded. A different target error will produce a different number of iterations. The smaller the target error, the greater the number of iterations. The smaller the target error, the better the accuracy value (the greater). The number of iterations in the training process with variations in momentum values is smaller than without variations in momentum values. The highest level of accuracy is achieved at the target error of 0.001 with a momentum value of 0.4 which is 96.62%.

1. Introduction

In general the pattern of seasons in Indonesia is known as the Monsoon pattern. This monsoon pattern is strongly influenced by the monsoon wind that produces two seasons namely the rainy season and the dry season. The peak of the rainy season occurs in December, January and February while the peak of the dry season occurs in June, July and August. Weather conditions are very influential in everyday life, such as in agriculture, transportation and industry. Therefore, observation of weather conditions, especially temperature conditions is very important [1].

Temperatures that are too hot or too cold can be categorized as one extreme weather phenomenon. Based on the observation data from the Center of Meteorology, Climatology, and Geophysics Region I period of 1998 to 2017, the normal average temperature for the city of Medan is 27.5° C, and the normal average temperature is 32.9° C. The low or high temperature is cannot be determined with certainty, but it can be predicted or estimated. There are many ways can be done to recognize the pattern of high or low temperature in one place, one of them is using artificial neural network technique (Artificial Neural Network).

2. Research Background

Gupta Akashap, Gautam Anjali, et al. [2] were applying Neural Network to predict the rainfall with backpropagation method. The result is more accurate, the prediction value was closer to the true value and designed to be used to predict the rainfall in India.

Priya, et al. [3] is using backpropagation method to predict the rainfall in India. The testing with backpropagation method gave an accurate result and could be used to predict the rainfall. The conclusion of this research is that by the increasing number of the hidden layer then the target error would decrease to a certain limit.

Sukmawati N Endah, et al. [4] research compared the convergence of gradient descent, gradient descent + momentum, gradient descent + adaptive learning rate and gradient descent + momentum + adaptive learning rate for diabetic detection. Based on the experimental results, it is known that gradient descent has better performance than others in terms of convergence velocity and result of TPR and FPR value. Training algorithm with combining of gradient descent + momentum + adaptive learning rate faster convergent compared to gradient + momentum or gradient + adaptive learning rate.

3. Proposed Method

3.1. Neural Network

Artificial neural networks basically deal with the nodes and connectors. Nodes are analogous to natural neurons those receive signals through synopses situated on dendrites or membraneof the neurons which are connected through Axons as connectors. Synopses are analogous to weights of the inputs. When the strong enough signals, more than a certain threshold, are passed through axons, neurons would be activated. In the same way, other signals may be passed through other synapses, those may instigate other neurons [5].

Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the biological system of the nerves, such as the process of information in the human brain. A key element of this paradigm is the structure of an information processing system consisting of a large number of interconnected processing elements (neuron), working simultaneously to solve a particular problem.

Way ANN works like the way humans work, learning by example. The JST layers are divided into 3, the input layer, the hidden layer, and the output layer (output layer) [6].

A. Backpropagation

Backpropagation neural network (BPNN) is a conceptof re-framing the obtained output towards the desired output. The output neurons are again connected back to the input neurons in the BPNN system to train the network by modifying the weights according to the obtained error in such a way that desired output relationship is eventually realized. In the a schematic sketch of BPNN (represented in a darker shade of line) has been drawn comprising n input neurons, r output neurons, and a single hidden layer with m neurons. In a model of ANN, a number of hidden layers play a crucial role for determining the rate of convergence during the period of parameters training for a given number of neurons at input and output layers. In this paper, the constructed ANN model, input nodes have a transfer function of identity, and the hidden and outputlayers have the activation functions as sigmoid S(.) and linear respectively [5].

Backpropagation is a supervised learning algorithm and is commonly used by perceptrons with multiple layers to change the weights associated with neurons in the hidden layer. The backpropagation training algorithm basically consists of 3 stages, namely [7] :

a. Input the value of training data to obtain the output value.
b. Back propagation of the error value obtained.
c. Adjust the connection weight to minimize the error value.

All three stages are repeated continuously to get the desired error value. After the training is done.

3.1. Momentum Backpropagation

Backpropagation momentum algorithm is the development of standard backpropagation algorithm which in its learning uses momentum where the Momentum constant value has range 0 to 1.

Backpropagation momentum algorithm has the same step with standard backpropagation algorithm but different when the backward propagation. Here is the backpropagation algorithm [8].

a. Initialize weights (take the initial weight with fairly small random value), Epoh = 1 and MSE = 1.
b. Determine the Epoh Maximum, Learning Rate (α), and Target Error.
c. Perform the following 4 to 12 steps during (Epoh < Epoh maximum) and (MSE > Target Error).
d. Epoh = Epoh + 1.
e. Feed forward
   1) Each of input unit *(X_1, i=1,2,3,…,n)* receives the signal xi and forwads the signal to all unit on the layer above it (hidden layer).
   2) Each of hidden unit *(Z_1, j=1,2,3,…,p)* sums the weight of the input signal, shown by the equation (1).
   $$z\_in_j = b_{1j} + \sum_{i=1}^{n} x_i\, v_{ij} \tag{1}$$
   Then use the activation function to calculate the output signal shown by the equation (2).
   $$z_j = f(z_{in_j}) \tag{2}$$
   And send those signals to all unit in the top layer (output layer)
   3) Each of output unit (YK, K=1,2,3,..,m) sums weighted input signals, shown by the equation (3).
   $$y\_in_k = w_{0k} + \sum_{i=1}^{p} z_i\, w_{jk} \tag{3}$$
   Then use the activation function to calculate the output signal, shown by the equation (4).
   $$y_k = f(y\_in_k) \tag{4}$$
f. Feedback (Momentum Backpropagation)
   1) Each of output unit *(Y_k, k=1,2,3,…,m)* receives the target pattern as the training input pattern, then calculate the error, shown by the equation (5).
   $$\delta_k = (t_k - y_k)f'(y_{in_k})$$
   $$\varphi 2_{jk} = \delta_k \text{ - } z_j \tag{5}$$
   $$\beta 2_k = 8_k$$
   Then calculate the weight correction (which will later be used to correct Wjk value), shown by the equation (6).
   $$\Delta w_{jk} = a\varphi 2_{jk} + \mu \Delta w_{jk} \tag{6}$$
   Performed as many as the number of hidden layers to the previous hidden layer.
   2) Each of hidden unit (Zj, j=1,2,3,..,p) sums the input delta (from units located on the top layer), shown by the equation (7).
   $$\delta\_in_j = \sum_{k=1}^{m} \delta 2_k\, w_{jk}$$
   Multiply this by a derivative of the activation error function, shown by the equation (7).
   $$\delta_j = \delta_{in_j} f'\left(z_{in_j}\right) \tag{7}$$
   $$\varphi 1_{ij} = \delta 1_j \text{ - } x_j$$
   $$\beta 1_j \text{ - } \delta 1_j$$
   Then calculate the weight correlation (which will later be used to fix Vij value), shown by the equation (8).

$$\Delta v_{ij} = \alpha\varphi_{ij} + \mu\Delta v_{ij} \quad (8)$$

Calculate the bias correction also (which will later be used to improve the value of bij), shown by the equation (9).

$$\Delta b1_j = \alpha\varphi 1_j + \mu\Delta b1_j \quad (9)$$

μ is a constant of momentum with a range [0.1].

g. Improvement weight

1) Each of output unit (Yk, k=1,2,3,..,m) fixes its bias and weights (j=0,1,2,..p), shown by the equation (10).

$$w_{jk}(new) = w_{jk}(long) + \Delta w_{jk} \quad (10)$$

$$b2_k(new) = b2_k(long) + \Delta b2_k$$

2) Each of hidden unit (Zj, j=1,2,3,..p) fixes its bias and weights (i=0,1,2,..n), shown by the equation (11).

$$v_{ij}(new) = v_{ij}(long) + \Delta v_{ij} \quad (11)$$

$$b1_j(new) = b1_j(long) + \Delta b1_j$$

3) Calculate the value of MSE, shown by the equation (12).

$$\text{MSE} = \frac{\sum \text{tk} - \text{Y k}}{\text{n}} \quad (12)$$

4. Results and Analysis

In this research, the pattern recognition of Medan city temperature data would be done using artificial neural network Backpropagation method that is done by dividing into 2 parts, namely: data for training, data for testing. Data used in Medan city temperature is the data from 1998-2017.

To know whether variation value Momentum algorithm with Backpropagation method, then some tests were done. The first test the researcher did the training with the number of hidden 6, alpha 1, target error differed between 0.001, 0.009 max epoh 50, and variation value momentum.

Before it is processed the data was normalized first. The normalization of the data was done in order that the network output was in accordance with the activation function used. The data was normalized in the interval [0, 1] because the data used is positive or 0. It is also related to the activation function given, that is the binary sigmoid.

Sigmoid function is asymptotic function (never reach 0 or 1), then data transformation was done at smaller interval [0,1; 0.8], shown by the equation (13).

$$x' = \frac{0.8(x-a)}{b-a} + 0.1 \quad (13)$$

a is the minimum data, *b* is the maximum, *x* is the data to be normalized, and x' is the data that has been transformed.

The next stage is training. The training process is carried out until the resulting error is appropriate or smaller than the target error. Figure 1 is a graph of the decrease in error squares of training results in random weights that have been generated without adding momentum (standard backpropagation) where the target error is 0.001 and the number of hidden layers 6.

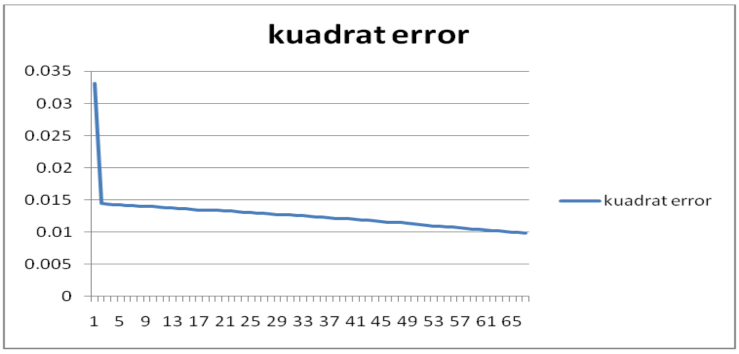

Figure 1. Decreasing backpropagation error squares without momentum value variations

Figure 1 indicates that the target error is achieved in the 50th iteration. The next test is done by varying the value of momentum in the process of changing the weight. Decreasing the squared error of training results with variations in momentum values is shown in figure 2.

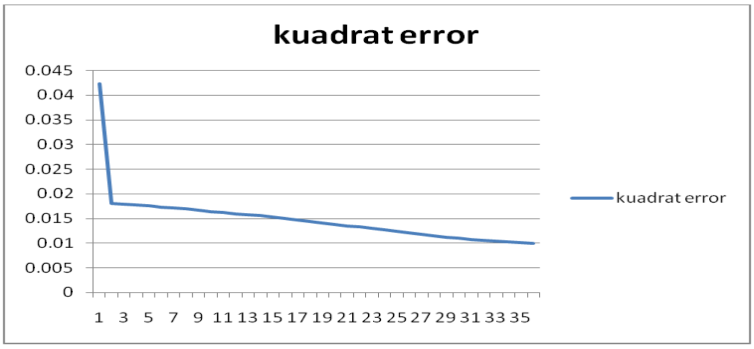

Figure 2. Decreasing backpropagation error squares with momentum value variations

After the data was normalized, then the training was done to recognize the given data pattern. The training was done with several different target errors. Initialization of weights was also done in starting the training process. The authors propose Momentum algorithms to add the weight at the backpropagation method to recognize patterns of data to be trained. The results of the training can be seen at table 1.

Table 1. Test results target error 0.001 with variations in momentum values

Target error	Without Momentum		Value Momentum	Variation of Values Momentum	
	Iteration	Accuracy		Iteration	Accuracy
0.001	50	96.22%	0.1	50	96.45%
			0.2	50	96.47%
			0.3	50	96.4%
			0.4	50	96.62%
			0.5	50	96.57%
			0.6	50	96.47%
			0.7	50	96.52%

5. Conclusion

From the results of the study it can be conclusion. A different target error will produce a different number of iterations. The smaller the target error, the greater the number of iterations. The smaller the target error, the better the accuracy value (the greater). The number of iterations and accuracy in the training process with variations in momentum values are better than without variations in momentum values are shown in Table 1.

6. References

[1] Ihwan, Andi, 2013, Artificial Neural Network Method Backpropagation For Monthly Rainfall Estimation In Ketapang West, *Proceed Semirata FMIPA University Of Lampung.*

[2] Akashdeep, et al., 2013, Time Series Analysis Of Forecasting Indian Rainfall, *International Journal Of Inventive Engineering And Sciences (IJIES): ISSN 2319-9598, Volume-1, Issue-6, May 2013.*

[3] Priya, et al., 2014, Time Series Analysis Of Forecasting Indian Rainfall, *International Journal Of Innovatios & Advancement In Computer Science (IJIACS), Volume 3, Issue 1, April.*

[4] Sukamawati N Endah, et al. Betond Back-Propagation Learning For Diabetic Detection: Converge Comparison Of Gradient Descent, Momentum And Adaptive Learning Rate. International Conference On Informatics And Computational Sciences (ICICOS) 2017.

[5] Jaiswal J K, Das R 2017, Application Of Artificial Neural Networks With Backpropagation Technique In The Financial Data, *IOP Conf. Series: Materials Science and Engineering*

[6] Sutojo, T., et al, 2010, *Artificial Intelligence,* Yogyakarta: Andi Offset.

[7] Izhari F, Dhany H W, Zarlis M, et al, 2017, Analysis Backpropagation Methods With Neural Network For Prediction Of Children’s Ability In Psychomotoric, *2nd International Conference on Computing and Applied Informatics*

[8] Nurcahaya, Pradana, T.P., et al, 2013, Utilization Of Speed Region Growing Segmentation And Backpropagation Neural Network Momentum For Classification Of White Blooed Cell Types.