15-7

Learning Method Adding Noise to Inertia Term of Feedforward Neural Network

Yusuke NOMURA Kazuki NAGAO Yoko UWATE Yoshifumi NISHIO
(Tokushima University)

1. Introduction

Neural Network (NN) is imitated the working of the human nervous system. NN attracts attention because it is specialized in pattern recognition and image recognition and so on. However, it takes time to process pattern recognition and image recognition and so on.

To solve this problem we experiment to improve learning accuracy and to reduce learning loops. In this study, we propose Feed Forward NN(FFNN) with inertia term having Gaussian distribution noise. Conventional system and proposed system are compared.

2. Proposed method

Conventional and proposed systems that is used in this study are shown in Fig. 1. Arrows of left to right indicate propagation of input signal and right to left indicate learning of back propagation. We propose Feed Forward NN (FFNN) with inertia term having Gaussian distribution noise. In this study, we use FFNN about 4 neurons input layer, 4 neurons hidden layer and 3 neurons output layer. We use Gaussian distribution (0, 0.01) for noise and add it to Equation. (4). In addtion, the putting place of noise is changed. We experiment 4 patterns about *conventonal* system, (input layer-hidden layer) = IH, (hidden layer-output layer) = HO and (input layer-hidden layer, hidden layer - output layer) = $IHHO$.

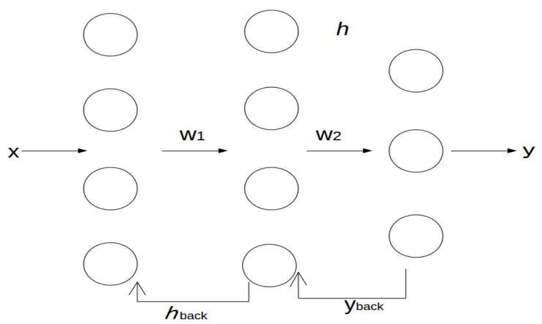

Figure 1: Schematic diagram of FFNN.

Equations (1) to (4) are general FFNN's propagation equations, back propagation equations and weight correction equations.

$$\begin{cases} h = \dfrac{1}{1+e^{(-\sum(w_1 x))}} \\ y = \dfrac{1}{1+e^{(-\sum(w_2 h))}} \end{cases} \tag{1}$$

$$\begin{cases} y_{back} = (y-ty)(1-y)y \\ h_{back} = \sum w_2 y_{back}(1-h)h \end{cases} \tag{2}$$

$$\begin{cases} {w_1}^{(l+1)} = {w_1}^{l} - \varepsilon\ x h_{back} \\ {w_2}^{(l+1)} = {w_2}^{l} - \varepsilon\ h y_{back} \end{cases} \tag{3}$$

$$\begin{cases} {w_1}^{(l)} = \alpha w_1 \\ {w_2}^{(l)} = \alpha w_2 \end{cases} \tag{4}$$

ε is the rate of decay which is 0.01. h means propagation. w means weights between neurons. The initial value of the weights is chosen by random from 0.01 to 0.04. α is inertia coefficient.

Equation (5) is proposed systems equation in this study.

$$\begin{cases} {w_1}^{(l)} = \alpha w_1 + (\alpha w_1+\beta) \\ {w_2}^{(l)} = \alpha w_2 + (\alpha w_2+\beta) \end{cases} \tag{5}$$

β is Gaussian distribution noise.

3. Simulation result

We define as the number of input layer = 4, the number of hidden layer = 4, the number of output layer = 3, the learning loops = 50000 and the number of learning data sets = 150. We used Iris data set as learning data. We simulate proposed method five times and calculate the average error rate and minimum error rate. Table 1 shows average error rate and minimum error rate for all kinds of patterns.

Table 1: Average and minimum error rate.

	average error rate	minimum error rate
Conventional	0.0945	0.0853
IH	0.0355	0.0234
HO	0.0534	0.0261
IHHO	unmeasurable	unmeasurable

We comfirm that average error rate and minimum error rate of the proposed method are better than the conventional method from Table 1. Moreover, we understand that average error rate and minimum error rate of IH is better than HO from Table 1. In case of $IHHO$, we could not measure the average and minimum error rate.

4. Conclusion

We comfirm that it is good to add noise to the inertia term. However, when we use the proposed method at IHHO, its measurement became impossible. The reason for this is considered to be caused by over training. As a cause of over training, there are lack of learning data and a lot of the propotion of biased data in learning data. In this study, we consider to make a lot of the propotion of biased data in learning data cause over trainig.

In the future,we developed a model to suppress over training and we would like to combine to proposed model. We use a variety of noise and want to obtain better results.