Deep Learning Improving Learning Efficiency by Using Noise for Back Propagation

Kazuki NAGAO, Shinsaburo KITTAKA, Yoko UWATE and Yoshifumi NISHIO
Dept. of Electrical and Electronic Engineering of Tokushima University
2-1 Minami-Josanjima, Tokushima-shi, Tokushima-ken 770-8506, Japan
Email: {k-nagao, kittaka, uwate, nishio}@ee.tokushima-u.ac.jp

***Abstract*—Deep learning began to be rapidly attention such as image recognition from the 2010. Now, the third time of artificial intelligence is boom. Deep learning is made to imitate the human nervous system. It automatically learns valid features of discrimination. That is why it is popular. In this study, we study about deep learning using noise for back propagation. Then, we improve the learning accuracy and reduce learning loops.**

I. Introduction

Human nervous system has neurons which are connected. It is controlled by the strength of neurons. Then, human can think things. Deep learning is made to imitate the human nervous system[1]. There are concepts and methods of deep learning from before and after 1980. It began to be rapidly attention such as image recognition from the 2010. Now, the third time of artificial intelligence is boom.

Deep learning is superposition of neural networks and has deep structure. Deep learning has more three layers include input layer and output layer. It is repeated learning by the time, and the input data is gradually being transferred to the deeper and deeper from the first layer. Previously, researchers and engineers had set the parameters manually. Now, that deep learning automatically learns valid features of discrimination. It has been improved accuracy of pattern recognition. It is attracting attention.

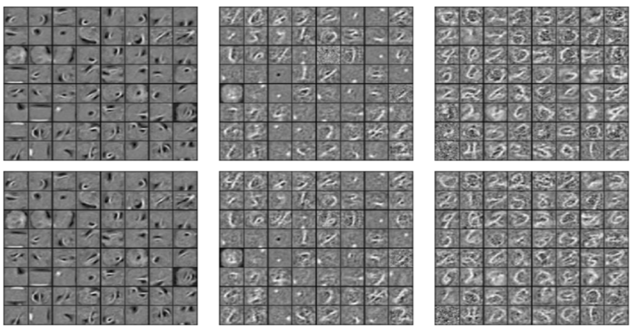

Fig. 1. Recognition of numbers by deep learning.

In June 2012, it was able to recognize the cat on computer by research of Google. After that, it came to be used in various fields. For example, image field, emotional awareness, module of self-driving, medical field, financial field and so on... Google released a software library of machine learning techniques including deep learning in 2015, and we raised interest in deep learning. Figure 1 shows recognition of numbers by deep learning with Google's software library[2].

In deep learning, there is a variety of techniques in order to improve the learning accuracy. Technique of back propagation compares input data and output data to find deviations of feature values and optimize weights. This technique can reduce learning loops. If the learning period is too long or the training data is not typical, it adapts to features unrelated to the features that should be originally learned. This phenomenon is called over learning. In such a case, it is used the method of dropout to correspond to input data. Deep learning requires a large amount of learning data. When there is not too much input data, input data is shifted slightly due to use noise to increase features. There are various techniques like these.

In this study, we propose new system of deep learning. This system is noisy feed back. Weights of neurons are optimized using a noise in the back propagation. Purposes of this study are to improve learning accuracy and to reduce learning loops. There are two patterns which are a small deep learning model (A) and a deep learning model (B). We compared conventional system and two patterns of proposed system. [1]

II. Proposed method

Figs. 2 and 4 show conventional system. Figs. 3 and 5 show proposed system. Arrows of left to right indicate propagation of input signal and right to left indicate learning of back propagation.

(A) Iris data set
We use a model consisting of two layers of intermediate layers and Gaussian distribution (0, 0.01) for noise and add it to equations (2). In addition, the putting place of Gaussian distribution (0, 0.01) is changed. We experiment 8 patterns about conventional system without noise, (h_{1back}), (h_{2back}), (y_{back}), (h_{1back}, h_{2back}), (h_{1back}, y_{back}), (h_{2back}, y_{back}) and (h_{1back}, h_{2back}, y_{back}). These combinations indicate where noise is added.

[1]This research is that we add one dataset to what was announced at SJCIEE.

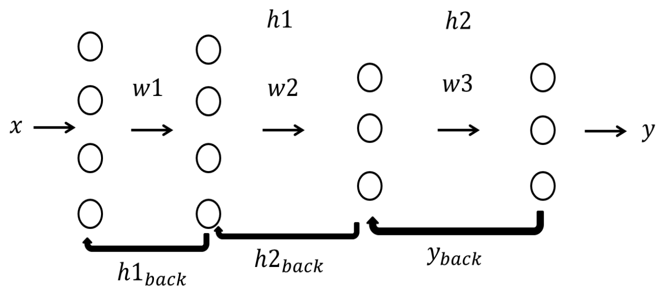

Fig. 2. Schematic drawing of deep learning consisting of two intermediate layers.

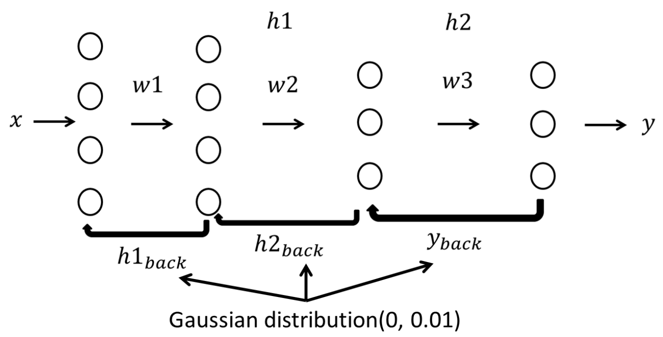

Fig. 3. Schematic drawing of proposed deep learning consisting of two intermediate layers.

$$
\begin{cases}
h_1 = \dfrac{1}{1+e^{(-\sum w_1 x)}} \\
h_2 = \dfrac{1}{1+e^{(-\sum w_2 h_1)}} \\
y = \dfrac{1}{1+e^{(-\sum (w_3 h_2))}}
\end{cases}
\tag{1}
$$

$$
\begin{cases}
y_{back} = (y-ty)(1-y)y \\
h_{2back} = \sum w_3 y_{back}(1-h_2)h_2 \\
h_{1back} = \sum w_2 h_{2back}(1-h_1)h_1
\end{cases}
\tag{2}
$$

$$
\begin{cases}
{w_1}^{(l+1)} = {w_1}^{l} - \varepsilon x h_{1back} \\
{w_2}^{(l+1)} = {w_2}^{l} - \varepsilon h_1 h_{2back} \\
{w_3}^{(l+1)} = {w_3}^{l} - \varepsilon h_2 y_{back}
\end{cases}
\tag{3}
$$

ε is the rate of decay which is 0.01. h means activation function. w means weights between neurons. The initial value of the weights is chosen by random from 0.04 or less.

(B) Cars data set
We use Gaussian distribution (0, 0.01) for noise and add it to equations (5). In addition, the putting place of Gaussian distribution (0, 0.01) is changed. We experiment 15 patterns about conventional system without noise, (h_{1back}), (h_{2back}), (h_{3back}), (y_{back}), (h_{1back}, h_{2back}), (h_{1back}, h_{3back}), (h_{1back}, y_{back}), (h_{1back}, h_{2back}, h_{3back}), (h_{1back}, h_{2back}, y_{back}), (h_{1back}, h_{3back}, y_{back}), (h_{1back}, h_{2back}, h_{3back}, y_{back}), (h_{2back}, h_{3back}), (h_{2back}, y_{back}), (h_{2back}, h_{3back}, y_{back}) and (h_{3back}, y_{back}). These combinations indicate where noise is added.

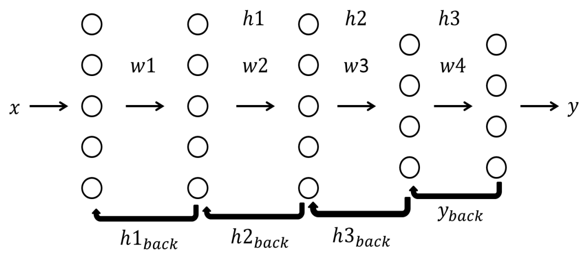

Fig. 4. Schematic drawing of deep learning consisting of three intermediate layers.

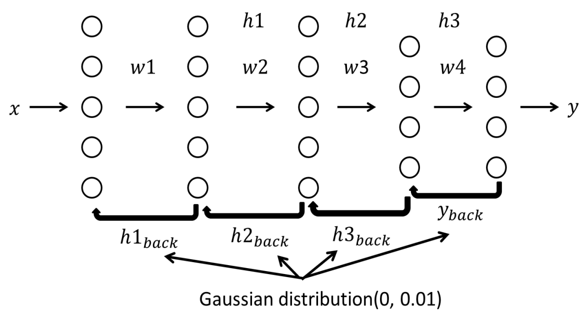

Fig. 5. Schematic drawing of proposed deep learning consisting of three intermediate layers.

$$
\begin{cases}
h_1 = \dfrac{1}{1+e^{(-\sum w_1 x)}} \\
h_2 = \dfrac{1}{1+e^{(-\sum w_2 h_1)}} \\
h_3 = \dfrac{1}{1+e^{(-\sum w_3 h_2)}} \\
y = \dfrac{1}{1+e^{(-\sum (w_4 h_3))}}
\end{cases}
\tag{4}
$$

$$
\begin{cases}
y_{back} = (y-ty)(1-y)y \\
h_{3back} = \sum w_4 y_{back}(1-h_3)h_3 \\
h_{2back} = \sum w_3 h_{3back}(1-h_2)h_2 \\
h_{1back} = \sum w_2 h_{2back}(1-h_1)h_1
\end{cases}
\tag{5}
$$

$$
\begin{cases}
{w_1}^{(l+1)} = {w_1}^{l} - \varepsilon x h_{1back} \\
{w_2}^{(l+1)} = {w_2}^{l} - \varepsilon h_1 h_{2back} \\
{w_3}^{(l+1)} = {w_3}^{l} - \varepsilon h_2 h_{3back} \\
{w_4}^{(l+1)} = {w_4}^{l} - \varepsilon h_3 y_{back}
\end{cases}
\tag{6}
$$

ε is the rate of decay which is 0.4. h means activation function. w means weights between neurons. The initial value of the weights is chosen by random from 0.3 or less.

III. Simulation Result

(A) Iris data set

Learning accuracy and learning loops are better to be smaller. We define as the learning loops = 25000 and the number of learning data sets = 150.

Figure 6 shows change of the learning accuracy and learning loops of conventional system and proposed system. Table I shows learning accuracy for all kinds of patterns. Table II shows learning loops when learning accuracy is 0.15 or less.

From Figure 6, the learning accuracy and learning loops of using Gaussian distribution (0, 0.01) are better than not using it. From Table I, most results by using the noise are found to be better.

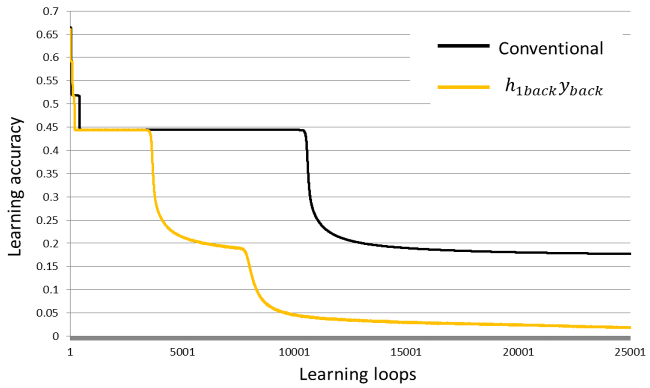

Fig. 6. Simulation results of schematic drawing of deep learning consisting of two intermediate layers.

TABLE I
LEARNING ACCURACY OF MODEL (A).

	accuracy
Conventional	0.1771
h_{1back}	0.1787
h_{2back}	0.1727
y_{back}	0.0269
$h_{1back}h_{2back}$	0.1750
$h_{1back}y_{back}$	0.0178
$h_{2back}y_{back}$	0.1708
$h_{1back}h_{2back}y_{back}$	0.0261

TABLE II
LEARNING LOOPS OF MODEL (A).

	loops
y_{back}	12612
$h_{1back}y_{back}$	8002
$h_{1back}h_{2back}y_{back}$	11103

(B) Car data set

Learning accuracy and learning loops are better to be smaller. We define as the learning loops = 1000 and the number of learning data sets = 1728.

Figure 7 shows the learning accuracy and learning loops of conventional system and proposed systems. Table III shows learning accuracy of minimum value and average for all kinds of patterns. Table IV shows learning loops for all kinds of patterns when learning accuracy is 0.07 or less.

From Figure 7, the learning accuracy and learning loops of using Gaussian distribution (0, 0.01) are better than not using it. From Table IV, $h_{2back}y_{back}$ and y_{back} are better than conventional about minimum value and average value.

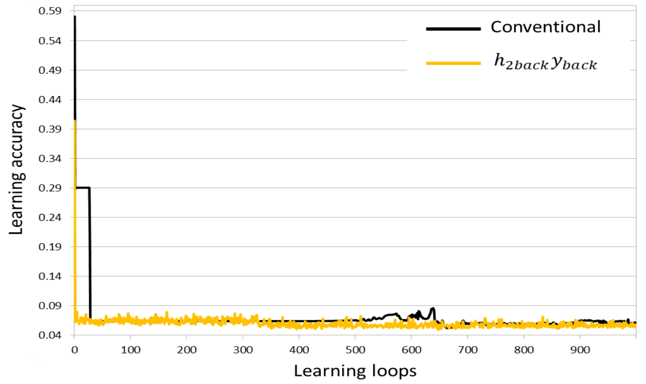

Fig. 7. Simulation results of schematic drawing of deep learning consisting of three intermediate layers.

TABLE III
LEARNING ACCURACY OF MODEL (B).

	minimum	average
Conventional	0.052	0.0634
h_{1back}	0.060	0.0734
h_{2back}	0.057	0.0697
h_{3back}	0.064	0.0639
y_{back}	0.050	0.0591
$h_{1back}h_{2back}$	0.060	0.0667
$h_{1back}h_{3back}$	0.058	0.0702
$h_{2back}h_{3back}$	0.056	0.0723
$h_{1back}h_{2back}h_{3back}$	0.057	0.0650
$h_{1back}y_{back}$	0.054	0.0660
$h_{2back}y_{back}$	0.049	0.0595
$h_{3back}y_{back}$	0.057	0.0718
$h_{1back}h_{2back}y_{back}$	0.052	0.0707
$h_{1back}h_{3back}y_{back}$	0.053	0.0678
$h_{2back}h_{3back}y_{back}$	0.054	0.0738
$h_{1back}h_{2back}h_{3back}y_{back}$	0.054	0.0724

TABLE IV
LEARNING LOOPS OF MODEL (B).

	loops
Conventional	29
h_{1back}	11
h_{2back}	84
h_{3back}	6
y_{back}	2
$h_{1back}h_{2back}$	6
$h_{1back}h_{3back}$	76
$h_{2back}h_{3back}$	46
$h_{1back}h_{2back}h_{3back}$	10
$h_{1back}y_{back}$	45
$h_{2back}y_{back}$	2
$h_{3back}y_{back}$	39
$h_{1back}h_{2back}y_{back}$	221
$h_{1back}h_{3back}y_{back}$	8
$h_{2back}h_{3back}y_{back}$	5
$h_{1back}h_{2back}h_{3back}y_{back}$	13

IV. CONCLUSION

In this study, we proposed new technique of deep learning. It is using noise at back propagation. In the proposed technique, we choose gaussian distribution (0, 0.01) for noise. Then, we examined whether this technique is effective in simulation that is using two data sets.

In simulations, we were able to improve learning accuracy and reduce learning loops to use noise at back propagation. From the tendency of research results of two data sets, we found that results are improved by using noise at back propagaion from the output layer to the final intermediate layer. Because back propagation of the output layer is compared with input data and output data, so it is properly carried out weight adjustment.

In the future work, we use other data sets to demonstrate usability and a variety of noise. Then, we obtain the good performance of the proposed technique.

ACKNOWLEDGMENT

This work was partly supported by JSPS Grant-in-Aid for Challenging Exploratory Research 26540127.

REFERENCES

[1] Chihiro Ikuta, Yoko Uwate and Yoshihumi Nishio, "Investigation of Behavior of Deth of Neuron and Neurogenesis in Multi-Layer Perceptron," 2013 IEEE Workshop on Nonlinear Circuit Networks, pp. 65-68, 2013

[2] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, Samy Bengio, "Why Does Unsupervised Pre-training Help Deep Learning?," Journal of Machine Learning Research 11, pp. 625-660, 2010

[3] Shohei Gotoda, Yoko Uwate and Yoshihumi Nishio, "Cellular Neural Networks with Changing Templates for Image processing," 2015 IEEE Workshop on Nonlinear Circuit Networks, pp.21-23, 2015

[4] Shinsaburo Kittaka, Ryota Oshima, Yoko Uwate and Yoshifumi Nishio, "Feed-forward Neural Netwotks with Changing Sigmoid Functions," Journal of Shikoku-Section Joint Convention of the Institutes of Electrical and Related Engineers, No.1-27, p.27, 2015