2014 22nd International Conference on Pattern Recognition

To be Bernoulli or to be Gaussian, for a Restricted Boltzmann Machine

Takayoshi Yamashita*, Masayuki Tanaka†, Eiji Yoshida‡, Yuji Yamauchi‡ and Hironobu Fujiyoshi‡
*Chubu University
†Tokyo Institute of Technology
‡Tome R&D
Email: yamashita@cs.chubu.ac.jp

***Abstract*—We introduce a method that automatically selects appropriate RBM types according to the visible unit distribution. The distribution of a visible unit strongly depends on a dataset. For example, binary data can be considered as pseudo binary distribution with high peaks at 0 and 1. For real-value data, the distribution can be modeled by single Gaussian model or Gaussian mixture model. Our proposed method selects appropriate RBM according to the distribution of each unit. We employ the Gaussian mixture model to determine whether the visible unit distribution is the pseudo binary or the Gaussian mixture. According to this distribution, we can select a Bernoulli-Bernoulli RBM(BBRBM) or a Gaussian-Bernoulli RBM(GBRBM). Furthermore, we employ normalization process to obtain a smoothed Gaussian mixture distribution. This allowed us to reduce variations such as illumination changes in the input data. After experimentation with MNIST, CBCL and our own dataset, our proposed method obtained the best recognition performance and further shortened the convergence time of the learning process.**

I. Introduction

A restricted Boltzmann machine (RBM)[1] is becoming more and more popular stochastic graphical model in recent years. iFollowing the RBM proposed, various RBM and the stacked RBM methods have proposed as new machine learning methods [3][4][5][7]. These methods are applied to image recognition, speech recognition, etc., and show state-of-the-art performance in benchmark tests [6][8][10]. A conventional RBM assume that the visible unit is of binary data, it becomes a critical limitation to the various applications. The method such as softmax which approximate the binary data from real-value, is employed to address this issue. Alternatively, Hinton also proposed a method to apply real-value data through a formulation of the RBM by assuming the input data to be of Gaussian distribution[2]. These existing approaches used to handle real-value data are quite effective when the distribution of the input data is known. However, it is difficult to determine the data distribution for general purpose, i.e. whether the input is from a binary image or a gray scale image. In this paper, a Bernoulli-Bernoulli RBM (BBRBM) refers to the RBM assuming that the distribution of both visible and hidden units is binary. A Gaussian-Bernoulli RBM (GBRBM) refers to the RBM assuming that the distribution of the visible unit is the Gaussian distribution and that the distribution of the hidden unit is binary.

The Stacked RBM can achieve higher recognition performance compared with the single RBM. In order to apply to gray scale image, Hinton proposed the stacked RBM which contain the GBRBM in the first layer and the BBRBM in other layers.[2]. On the other hand, Cho used the GBRBM for all stacked layers[9][11]. In those approaches, one needs to heuristically pre-determine which RBM is appropriate for each layer. We propose a method to automatically select the appropriate RBM type according to the distribution of visible unit for each layer. In the stacked RBM, the conditional probability or the inference of the hidden unit is used as the input of the next layer of the RBM, even though the distribution of hidden unit is assumed as binary in both the BBRBM and the GBRBM. In this sense, we need to select RBM for each layer according to the distribution of actual input data. For the first layer, we can select RBM according to the distribution of the given data. For example, the GBRBM should be selected for the real-valued data such as image. However, it is unclear which type of the RBM is suitable for the second layer. Therefore, we need to evaluate the distribution of the inferred hidden state of the first layer. When the GBRBM is selected, we employ the normalization process where it is possible to suppress variations in the distribution of the input data to obtain the Gaussian distribution. We refer this GBRBM with normalization process to nGBRBM.

The key contributions of our proposed method are, 1) the proposed automatic RBM selection improves the recognition performance. In particular, it is no longer necessary to heuristically pre-determine the RBM type of each layer; 2) the normalization for the GBRBM improves the performance of the stacked RBM which includes the GBRBM. This normalization is not only successful in the first layer, but also obtains better results in later layers.

The rest of paper, we review the BBRBM and the GBRBM in Section 2. Then, we propose the automatic RBM type selection and the normalization for the GBRBM in Section 3. The comparison results for predefined multiple layers of the RBM and the GBRBM in various dataset, such as binary data and real-value data, are shown in Section 4. Finally, we will discuss about the effectiveness of our proposed method with analysis of the distribution of each unit in Section 5, and present our conclusions in Section 6.

II. Restricted Boltzmann Machine

A. Bernoulli-Bernoulli RBM

The restricted Boltzmann machine (RBM) consists of m visible units $\boldsymbol{v} = (v_1, ..., v_m)$ and n hidden units $\boldsymbol{h} = (h_1, ..., h_n)$, with fully connecting between them. But there are no visible to visible and hidden to hidden connections. The

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.270

IEEE computer society

visible units in the first layer correspond to the measurements. In computer vision tasks, one visible unit often corresponds to one pixel value. The hidden units $\boldsymbol{h}$ are independent conditionally to the measurements. Because of this, each hidden unit becomes an independent specific feature. The RBM has been mainly developed to model binary variables $(\boldsymbol{v}, \boldsymbol{h})$ which take the binary values of $(\boldsymbol{v}, \boldsymbol{h}) \in \{0, 1\}^{m+n}$. A joint probability of $(\boldsymbol{v}, \boldsymbol{h})$ can be expressed with the BBRBM as:

$$p(\boldsymbol{v}, \boldsymbol{h}) = \frac{1}{Z} e^{-E(v,h)}, \tag{1}$$

where Z is the normalizing constant and the energy function E,

$$E(\boldsymbol{v}, \boldsymbol{h}) = -\sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} v_i h_j - \sum_{i=1}^{m} b_i v_i - \sum_{j=1}^{n} c_j h_j, \tag{2}$$

for all $i \in \{1, ..., n\}$ and $j \in \{1, ..., m\}$, w_{ij} is a real valued weight associated with the edge between units v_i and h_j and b_i and c_j are real valued bias terms associated with the i-th visible and the j-th hidden variable, respectively. From Eqs. (1) and (2), $E(\boldsymbol{v}, \boldsymbol{h})$ with a low energy are given a high probability.

In terms of probability, the hidden units are independent from the visible units and vice versa, as shown in Eqs. (3) and (4). When binary data are given in visible units, the conditional probability is estimated from the neural network propagation rule by Eqs. (5) and (6).

$$p(\boldsymbol{v}|\boldsymbol{h}) = \prod_{i=1} p(v_i|\boldsymbol{h}), \tag{3}$$

$$p(\boldsymbol{h}|\boldsymbol{v}) = \prod_{i=1} p(h_j|\boldsymbol{v}), \tag{4}$$

$$p(v_i = 1|\boldsymbol{h}) = f(b_i + \sum_{j=1} h_j w_{ij}), \tag{5}$$

$$p(h_j = 1|\boldsymbol{v}) = f(c_j + \sum_{i=1} v_i w_{ij}), \tag{6}$$

where $f(\cdot)$ is the sigmoid activation function.

The model with the energy function has been developed to model the random binary variables. Therefore, this model is not suitable to model a continuous value data. To address this issue, the GBRBM has proposed.

B. Gaussian-Bernoulli RBM

The Gaussian-Bernoulli RBM (GBRBM) has visible units with real-value v_m and binary hidden units h_n. Based on the same idea as the BBRBM, the energy function of the GBRBM is defined as

$$E(\boldsymbol{v}, \boldsymbol{h}) = -\sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} h_j \frac{v_i}{\sigma_i} - \sum_{i=1}^{m} \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^{n} c_j h_j, \tag{7}$$

where b_i and c_j are biases corresponding respectively to visible and hidden units, w_{ij} are the connecting weights between the visible and hidden units and σ_i is the standard deviation associated with Gaussian visible units v_i. Conditional probabilities for visible and hidden units are

$$p(v_i = v|\boldsymbol{h}) = N(\boldsymbol{v}|b_i + \sum_j h_j w_{ij}, \sigma_i^2), \tag{8}$$

$$p(h_j = 1|\boldsymbol{v}) = f(c_j + \sum_i w_{ij} \frac{v_i}{\sigma_i^2}), \tag{9}$$

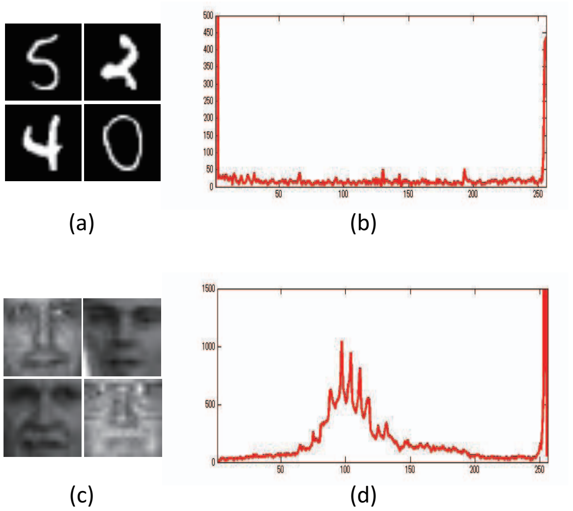

Fig. 1. Distribution of visible unit. (a) is MNIST, (b) is the intensity distribution of certain unit for MNIST images, (c) is CBCL, and (d) is the intensity distribution of certain unit for CBCL images.

where $N(\cdot|\mu, \sigma^2)$ denotes the Gaussian probability density function with mean μ and standard deviation σ.

In the parameter updating process, a contrastive divergence (CD) learning is highly successful and is becoming the standard learning method to train the RBM parameters[1]. The CD learning only samples for k steps to approximate the second term in the log-likelihood gradient from a sample from the RBM distribution. In the CD learning, k is usually set to 1. The update rules for the parameters are the following,

$$\nabla w_{ij} =< \frac{1}{\sigma_i^2} v_i h_j >_d - < \frac{1}{\sigma_i^2} v_i h_j >_m, \tag{10}$$

$$\nabla b_i =< \frac{1}{\sigma_i^2} v_i >_d - < \frac{1}{\sigma_i^2} v_i >_m, \tag{11}$$

$$\nabla c_j =< h_j >_d - < h_j >_m, \tag{12}$$

where shorthand notations $< \cdot >_d$ and $< \cdot >_m$ denote the expectation computed over the data and model distributions accordingly.

III. Proposed Method

As shown in Eqs.(5) and (6), the random binary variables are assumed for the inputs of the BBRBM. This assumption becomes a critical issue in considering real applications such as image processing. Hand writing digit data as shown in Fig.1 (a) can be considered as pseudo binary variables whose density functions condense at two values as shown in Fig.1 (b). In this case, the BBRBM can obtain an efficient classification feature in the hidden units because most of the inputs are of binary value. On the other hand, if the input data is real-value like grayscale images in Fig.1(c), the density function of each input is distributed in wide range as shown in Fig.1(d). The GBRBM may obtain efficient features for these distribution input data. In this sense, the performance of the RBM can be improved by selecting the GBRBM or the BBRBM depending on the distribution of the inputs. First, we propose the RBM selection method that determines whether to apply either the

BBRBM or the GBRBM according to the distribution of the visible unit.

Training of the GBRBM is a challenging task. One of the reasons for that the data distribution of each visible unit has many variations. The distribution of each visible unit is assumed to be single Gaussian distribution, but it becomes an obscure and vague Gaussian distribution. We therefore propose a normalized Gaussian-Bernoulli RBM (nGBRBM) where each visible unit is normalized in order to obtain a distinct Gaussian distribution from the Gaussian mixture distribution.

A. *Proposed RBM type selection*

We assume that the distribution of each input unit belongs to a pseudo binary distribution, single Gaussian distribution, or Gaussian mixture distribution. The pseudo binary distribution has two sharp peaks at two values like 0 and 1. The pseudo binary distribution can be considered the special case of the Gaussian mixture distribution. We model distribution of each inputs by the Gaussian mixture model (GMM). Then, the distribution is classified by parameters of the GMM. The distribution of visible unit $\boldsymbol{v_i}$ is modeled by

$$p(v_i) = \sum_{k=1}^{K} \pi_k N(v_i|\mu_k, \Sigma_k). \tag{13}$$

Note that K is the number of Gaussian components, π_k is the mixture weight of k-th Gaussian, μ_k and Σ_k are respectively the mean and variance. We estimated the parameters π, μ and Σ via a maximum likelihood estimation, which matches the distribution of the training data. In the maximum likelihood estimation, the parameters can be obtained iteratively by using an EM algorithm. On the EM iteration, we compute a posteriori probability for component k as E-step, given by

$$pr(k|v_i, \mu_k, \Sigma_k) = \frac{\pi_k N(v_i|\mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l N(v_i|\mu_k, \Sigma_l)}. \tag{14}$$

In M-step, the parameters are updated as follows,

$$\mu_k = \frac{\sum_{t=1}^{T} pr(k|v_i^{(t)}, \mu_k, \Sigma_k) v_i^{(t)}}{\sum_{t=1}^{T} pr(k|v_i^{(t)}, \mu_k, \Sigma_k)}, \tag{15}$$

$$\Sigma_k = \frac{\sum_{t=1}^{T} pr(k|v_i^{(t)}, \mu_k, \Sigma_k) v_i^{(t)2}}{\sum_{t=1}^{T} pr(k|v_i^{(t)}, \mu_k, \Sigma_k)} - \mu_k^2, \tag{16}$$

$$\pi_k = \frac{1}{T} \sum_{t=1}^{T} pr(k|v_i, \mu_k, \Sigma_k). \tag{17}$$

We estimate the parameters for each k. Then, we find best K with minimum fitting error between the distribution of visual unit and the distribution with estimated parameters. If K is equal or greater than 3, the distribution of visible unit is identified as the Gaussian mixture, and if K equals 1, the distribution is identified as the single Gaussian. On the other hand, if K equals 2, it must be identified as either the pseudo binary distribution or the Gaussian mixture. The pseudo distribution has two Gaussians with small standard variations around 0 and 1. This mean that two μ_k are separated by significant distance Th_m. Moreover, both standard deviations Σ_k are smaller than the threshold Th_d. If these two conditions are satisfied, we are able to identify the distribution as the pseudo binary. After identification of the distribution for all visible unit using the above method, we determine the RBM types by taking a vote of the distribution of them. For example, if pseudo binary distribution is majority, the BBRBM is selected. On the other hand, if major distribution is single or the Gaussian mixture, then the nGBRBM is selected.

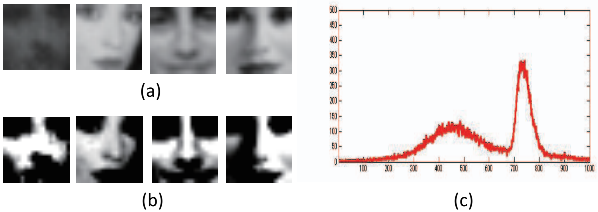

Fig. 2. Normalized image and distribution of visible unit after input data normalization. (a) is the original image in CBCL, (b) is normalized images, (c) is the distribution of visible unit using normalized images (original distribution is Fig.1(d)).

B. *Normalization of Visible Unit*

We normalize each training data to obtain a distinct Gaussian mixture distribution for visible units. Each pixel of the training data is normalized with the zero-mean and the unit variance of each training data. Figure 2 shows the normalized images and distribution of visible unit using them. Advantages to this normalization are that 1) the training data becomes uniform with no condition variations, such as illumination change, 2) the distribution of each visible unit becomes smoother.

C. *Stacked selected RBMs*

The stacked RBM is sometimes called a deep Boltzmann machine (DBM) or a deep belief network (DBN). It has simple layer-wise structure where the RBM is stacked on top repeatedly multiple times. The hidden units of the learned RBM can be used as visible unit for learning a following RBM. It can be characterized as follows: 1) high representation can be built from a large amount of unlabeled input data, and just a few minimal labeled data can be used to slightly fine-tune the model. 2) better propagation of ambiguous inputs can be obtained with approximate inference procedure of fine-tuning feedback. In our proposed method, the RBM type is identified according to the distribution of the visible unit in the learning of the RBM in each layer. Accordingly, the appropriate RBM type for each layer can be decided without the pre-definitions required in the conventional methods.

IV. Experiments

A. *Implementation*

We evaluate the reconstruction error of the a single layer RBM and the recognition performance with stacked layers in various datasets. For the comparison, we use the standard handwritten dataset MNIST[13] and face image dataset CBCL[12]. MNIST and CBCL are famous databases but they include strictly controlled data for their background. In order to evaluate under various real-scene situations, we use our original hand shape dataset including 6 hand shape classes in clutter background. We define the following learning parameters of the BBRBM, the GBRBM and our proposed method. The size

TABLE I. STRUCTURE OF PROPOSED METHOD IN EACH DATASET.

Layer	RBM type		
	MNIST	CBCL	Hand shape
1	BBRBM	nGBRBM	nGBRBM
2	nGBRBM	BBRBM	nGBRBM
3	BBRBM	BBRBM	BBRBM
4	BBRBM	BBRBM	BBRBM

of minibach is 100, while the initial learning rate and its upper bound are set to 0.001 for pre-training. The weight-updating ratio is set to 0.1. As parameters of the RBM selection, Th_m is 0.99 and Th_d is 0.01.

We learn for all methods with 4-layer structure for each dataset. We use the 60000 handwritten images for learning and 10000 for testing in MNIST. The image size is 28 × 28 and the number of visible unit is 784. All methods are of 4-layer structure with respectively 400, 255, 100 and 10 units in each layer. The 10 units in the final layer are the number of labels that are connected with the previous layer. In CBCL, we learn for all methods with 2000 face images, and we use 1000 images for testing with a 19 ×19 image size. The number of visible units is 361, and later layers respectively has 400, 255, 100 and 2 units. In this network, the output layer discriminates between faces and non-faces for given image. As a third dataset for comparison, we use a hand shape dataset with a 40 × 40-sized images. The number for learning data is 28000 and for test data is 10000. The number of visible units is 1600, while hidden units are 1024, 400, 100 and 7.

We also employ fine tuning to find optimized parameters for whole of network with dropout. Fine tuning is based on gradient decent on a supervised training criterion[3][4]. Dropout ratio is set to 0.5, iteration of fine tuning is 1000. All methods apply same parameters for all dataset.

B. RBM Selection

The selected RBM type in each layer of the proposed method is shown in Table I. For MNIST, since the distribution of visible unit is close to the pseudo binary distribution, the BBRBM is selected for the first layer. The nGBRBM is selected for the second layer because the distribution is near the Gaussian mixture distribution. For the later layers, the BBRBM is selected since most of units have the pseudo binary distribution. As shown in Fig.1, CBCL data has Gaussian distribution. This Gaussian distribution data is given as visible unit and therefore, the nGBRBM is selected for the first layer. For the second and following layers, the BBRBM is selected because many units have pseudo binary distribution. Like CBCL, the nGBRBM is selected in the first layer for hand shape database. The nGBRBM is also selected since the Gaussian distribution still remained in the second layer. For the later layers, the BBRBM is selected since most of units have a pseudo binary distribution. As mentioned above, the appropriate RBM type depends on the dataset and the layer. In MNIST, most of region is black, and hand written pixels are white like the binary image. It is possible to learn with high discrimination even when we apply the BBRBM in first layer. However, CBCL and hand shape dataset are grayscale images and that data represents Gaussian distribution. When the nGBRBM is selected first, the best suitable RBM type depending on the difficulty of the datasets will be selected for the following layers. In addition, the normalization process of the nGBRBM is efficient in order to obtain the pseudo binary distribution in earlier layers. This normalization reduce condition variation such as illumination change and distribution of unit becomes smoother. Therefore, the nGBRBM can learn more easily even from datasets including difficult data. We describe more details the distribution in each layer in the following section.

TABLE II. RECOGNITION PERFORMANCE IN EACH DATASET.

Dataset	BBRBM	GBRBM	Ours
MNIST	96.02	96.39	97.10
CBCL	93.59	96.39	97.60
Hand Shape	86.21	87.77	89.99

C. Reconstruction Result

We reconstruct of images to compare each RBM type. The reconstructed images in MNIST are shown in Fig.3. While most regions are correctly reconstructed with all methods in MNIST, the details in handwritten images can not be reproduced with the BBRBM or the GBRBM. For example, for the '7' in the third column, the BBRBM can not reconstruct the upper region with enough thickness. The '3' in the 5th column also does not have enough thickness. The nGBRBM can reconstruct much more correctly not only shapes but also thickness. In CBCL, while each pixel is real-value data, their images are uniform since the face size and background are not contained in the image. Therefore, even if it is grayscale image, the reconstruction is not such a difficult task for all methods, as shown in Fig.4. All methods can reconstruct correctly various faces. However, as shown in Fig.5, the nGBRBM had significantly better reconstruction performance compared with the BBRBM and the GBRBM. The BBRBM and the GBRBM can not correctly reconstruct both the hand shape and background region. While most of reconstructed images with them are blurry, the reconstruction with the nGBRBM is clearer than that of others.. Moreover, the nGBRBM reconstructs the hand region correctly even with a cluttered background. It had the surprising ability to reconstruct not only hand shape but also the background.

D. Recognition Performance

The comparison results of the recognition performance on each dataset are shown in Table II. Our proposed method can improve the recognition performance in all dataset. In MNIST, the recognition rate of our proposed method reaches 97.10%, an increase of more than 1% compared to the BBRBM. Even for binary images in MNIST, our method performs the best accuracy. Our proposed method also has the highest recognition rate among comparison methods in CBCL and the hand shape dataset. With these results, we can say that our proposed method offers significant improvement for real-value data. In other words, our proposed method is able to obtain effective initial parameters throughout the whole network. Furthermore, as shown in Fig.6, convergence comes much earlier in pre-training with our proposed method than with the other methods.

V. DISCUSSION

In Fig.7, we show the distributions of typical units in each dataset. The visible units of the first layer in MNIST receives

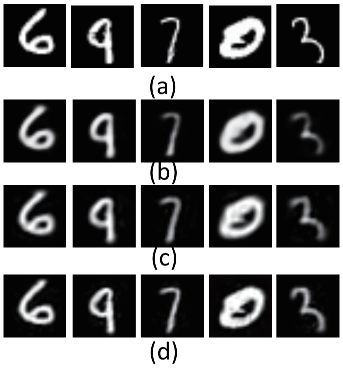

Fig. 3. Reconstructed images in MNIST. (a)original image, (b)BBRBM result, (c)GBRBM result and (d)our proposed method.

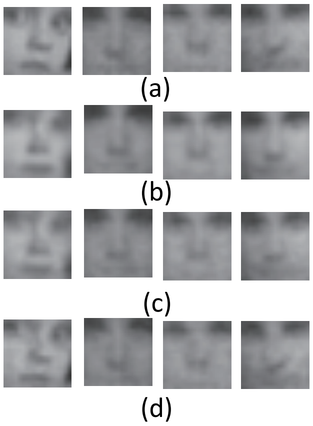

Fig. 4. Reconstructed images in CBCL. (a)original image, (b)BBRBM result, (c)GBRBM result and (d)our proposed method.

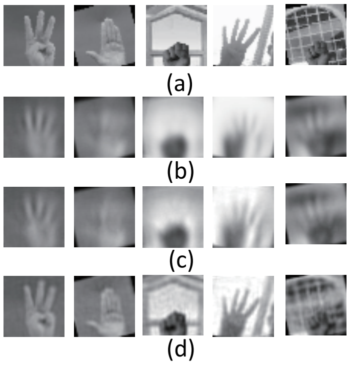

Fig. 5. Reconstructed images in hand shape data. (a)original image, (b)BBRBM result, (c)GBRBM result and (d)our proposed method.

the data that have the distribution as shown in Fig.1(b). Those distributions have two peaks at 0 and 1 because most of the pixels in MNIST consist of black and white. These distribution can be seen as the pseudo binary, and our method selects the BBRBM in the first layer. The output distribution of the typical units in the first layer has two peaks, as shown with the yellow line in Fig.7(a). Although the distribution has high peak at

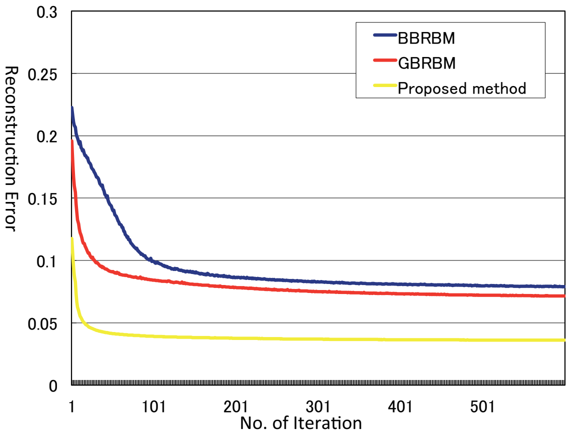

Fig. 6. Reconstruction error of each iteration in CBCL.

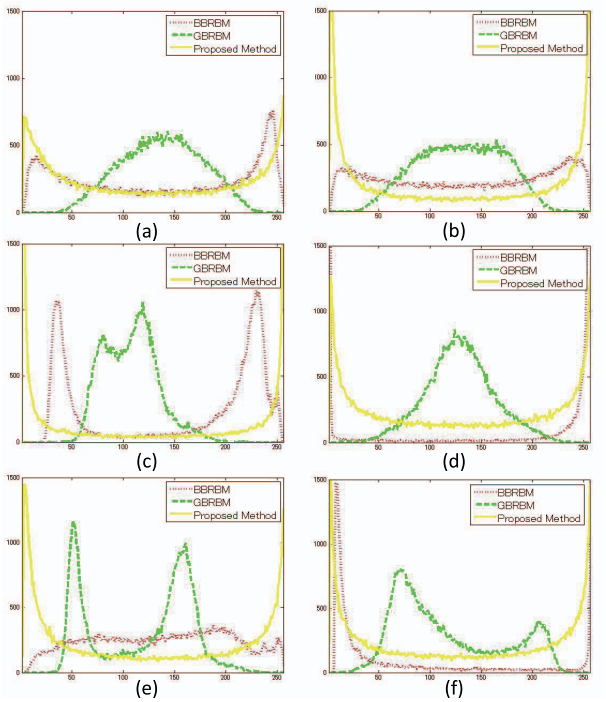

Fig. 7. Distribution of hidden units. (a) is the 80th unit in the first layer, and (b) is the 205th unit in the second layer, in MNIST. (c) is the 15th unit in the first layer, and (d) is the 57th unit in the second layer, in CBCL. (e) is the 71th unit in the first layer, and (f) is the 95th unit in the second layer, in the hand shape data set.

0 and 1, it is identified the Gaussian mixture instead of the pseudo binary. Because the most of distributions have large variance than threshold TH_m . Therefore, the nGBRBM is selected as the 2nd layer. As a result, the proposed method outputs the distribution with high peak at 0 and 1, while the BBRBM and the GBRBM output like Gaussian distribution as shown in Fig.7(b). The BBRBM is selected in later layers, because the most of distributions are identified the pseudo binary like yellow line in Fig.7(b).

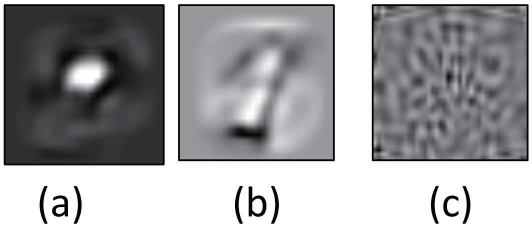

Fig. 8. Weight Visualization. (a) the distribution has a peak in 1, it is common feature to discriminate several classes, (b) the distribution has two peaks near 0 and 1, it is specific feature to discriminate specific class, in this case it seem is specified to '1' in MNIST, (c) the distribution is uniform, it does not focus on specific classes.

In CBCL, our proposed method selects the nGBRBM by reason that many distribution of visible units are Gaussian as shown in Fig.1(c). The distribution has high peaks at 0 and 1 and it is identified the pseudo binary distribution. Therefore, the BBRBM is selected as the 2nd layer. While the BBRBM has two peaks near 0 or 1, the distance between them is not far enough than threshold TH_d like red line in Fig.1(c). In the later layer, our proposed method selects the BBRBM because the distribution is the pseudo binary distribution like yellow line in Fig.1(d).

The hand shape dataset has wide variation due to pose and illumination changes. Moreover, since the background is complex, the distribution of visible unit is a Gaussian mixture. The proposed method selects the nGBRBM in the first layer as same as CBCL. As shown in Fig.7(e)(f), the distributions have two peaks. However, the proposed method selected the nGBRBM as the second and third layers by reason that it has not enough distance between them and variance.

We show the visualization of typical units of MNIST in Fig.8. We find the characteristics of the unit through this observation of the distribution and Fig.8 as follows. 1)if the unit has a high peak near 1, it seems common features for the training data. As shown in Fig.8, white region that has high weights, commons to several class such as '2', '6', '8', etc. It means that some weights connected with this unit are high for several classes and other weights are low to discard rest of classes. 2)if the unit has two high peak near 0 and 1, it has a high discrimination for the specific class like Fig.8(b). In this case it attention to specific class like '1'. 3)if the unit has uniform distribution, it has not enough discrimination for training data. It means that the weights are not high to specific class like Fig.8(c).

We show the distribution of the units for the BBRBM, the GBRBM and the proposed methods in each layer. We illustrated the effect of RBM type selection through the distribution of the visible units. We could obtain effective parameters with high discrimination with our proposed RBM type selection.

VI. Conclusion

In this paper, we have proposed the method to select the RBM types according to the distribution of the visible unit. By doing so, it is no longer necessary to pre-determine the type of RBM in advance. Furthermore, the normalization process for the GBRBM to suppress variations of the input data is incorporated. With the proposed method, one can select the appropriate RBM types automatically. As a results, the proposed method improve the performance compared to conventional methods.

References

[1] G. E. Hinton, "Training products of experts by minimizing contrastive divergence", Neural Computation, Vol.14, No.8, pp.1771–1800, 2002.

[2] G. E. Hinton and R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", Science, Vol.313, No.5786, pp.504–507, 2006.

[3] G. E. Hinton and S. Osindero, "A fast learning algorithm for deep belief nets", Neural Computation,Vol.18, No.7, pp.1527–1554, 2006.

[4] Y. Bengio, P. Lamblin, V. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks", In Advances in Neural Information Processing Systems 19 (NIPS'06), pp.153–160, 2007.

[5] R. Salakhutdinov and G. E. Hinton, "Deep Boltzmann machines", In Proceedings of the International Conference on Artificial Intelligence and Statistics, pp.448–455, 2009.

[6] A. Krizhevsky, "Learning multiple layers of features from tiny images", Technical report, Computer Science Department, University of Toronto, 2009.

[7] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines", In Proceedings of the 27th International Conference on Machine Learning (ICML10), pp. 807-814, 2010.

[8] B. Lake, R. Salakhudinov, J. Gross, and J. Tenenbaum, V. Ramesh, P. Meer, "One shot learning of simple visual concepts", Proceedings of the 33rd Annual Conference of the Cognitive Science Society, 2011.

[9] K. Cho, A. Ilin, and T. Raiko, "Improved Learning of Gaussian-Bernoulli Restricted Boltzmann Machines", Artificial Neural Networks and Machine Learning(ICANN 2011), Springer Berlin Heidelberg, pp.10–17, 2011.

[10] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks", In Advances in Neural Information Processing Systems 25 (NIPS'12). pp.1106-1114, 2012.

[11] K. Cho, T. Raiko, and A. Ilin, "Gaussian-Bernoulli deep Boltzmann machine", In NIPS 2011 Workshop on Deep Learning and Unsupervised Feature Learning, 2011.

[12] CBCL Face Database MIT Center For Biological and Computation Learning, http://www.ai.mit.edu/projects/cbcl

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, Vol.86, No.11, pp.2278–2324, 1998.