Interaction between Recurrent Multilayer Perceptrons and Sparsely Interconnected Neural Networks

Takeshi KAMIO and Mititada MORISUE

Hiroshima City University,
3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima, 731-3194, JAPAN
Phone: +81-82-830-1720, Fax: +81-82-83-1792
E-mail: kamio@im.hiroshima-cu.ac.jp

1. Introduction

Useful neural hardware has been developed. However, it is still difficult to implement large-scale neural hardware without any conditions to reduce hardware cost. This fact has promoted the studies on quantized neural networks (QNNs) and sparsely interconnected neural networks (SINNs). The QNNs are fabricated at a low cost by quantizing weights. The digital multilayer perceptrons (MLPs) are useful QNNs [1]. As the number of quantization levels becomes smaller, the generalization capability gets lower. On the other hand, SINNs are suitable for analog hardware implementation because the sparse interconnected structure can decrease the number of synaptic circuits. The cellular neural networks (CNNs) [2] are well known as SINNs. Although CNNs are often used as image filters, CNNs can work as associative memories for 2-valued images [3]. However, sparse interconnections decrease the capability of SINN associative memories [4].

As mentioned above, the conditions to reduce hardware cost exercise a bad influence on the capability of neural hardware. Although loosening such conditions is the easiest way to improve the capability, hardware cost goes up rapidly. In this paper, we propose associative memories using interaction between recurrent MLPs and SINNs as one of solutions for the above problem.

2. Sparsely Interconnected Neural Networks

The SINN associative memory is based on CNN with space-varying cloning templates as shown in Fig.1. A variety of synthesis procedures for SINNs have been proposed. They are founded on the idea that memory patterns are stored at the asymptotically stable equilibrium points of the dynamics of SINNs. The dynamics of the continuous-time SINN with n-neurons can be expressed by,

$$\dot{\mathbf{x}} = -\mathbf{x} + \mathbf{T}\mathbf{y} + \mathbf{I}, \tag{1}$$

where $\mathbf{x} \in \Re^n$ is the state vector, $\dot{\mathbf{x}}$ is the derivative of $\mathbf{x}$ with respect to time t, $\mathbf{y} \in \Re^n$ is the output vector, $\mathbf{T} \in \Re^{n \times n}$ is the interconnection matrix, and $\mathbf{I} \in \Re^n$ is the threshold vector. The initial state vector $\mathbf{x}(0)$ is used as the input. Furthermore, the output $y(t)$ is expressed as a piecewise linear function of the state $x(t)$,

$$y = f_{\mathrm{PWL}}(x) = \tfrac{1}{2}(|x+1| - |x-1|). \tag{2}$$

Here we define the set of n-dimensional bipolar vectors as $\mathrm{B}^n \equiv \{\mathbf{x} \in \Re^n : x_i = +1 \text{ or } -1,\ i = 1, \cdots, n\}$. When a memory vector and the equilibrium point are given as $\boldsymbol{\alpha} \in \mathrm{B}^n$ and $\boldsymbol{\beta} \in \Re^n$ respectively, the condition of the asymptotic stability for $\boldsymbol{\beta}$ can be obtained as follows [3],

$$\begin{cases} \boldsymbol{\beta} = \mathbf{T}\boldsymbol{\alpha} + \mathbf{I} \in \mathrm{C}(\boldsymbol{\alpha}) \\ \mathrm{C}(\boldsymbol{\alpha}) \equiv \{\mathbf{x} \in \Re^n : x_i \alpha_i > 1,\ i = 1, \cdots, n\} \end{cases}. \tag{3}$$

For SINNs with a neighborhood radius r, each neuron has $(2r+1)\times(2r+1)$ synapse weights at most. Therefore, SINNs with $r=1$ (r=1-SINNs) can be more effectively implemented on the analog hardware than fully interconnected neural networks (FINNs). Since each neuron must remember a correct value from the local information of an input pattern, the capability of r=1-SINN is much worse than that of FINN. However, it is confirmed by simulations that the local association of r=1-SINN might work well for the cases where FINN fails to associate memory patterns.

3. Multilayer Perceptrons with Quantized Weights

Three-valued weights ($-w$, 0, $+w$) essentially eliminate

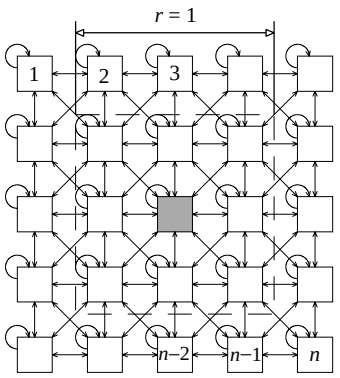


Fig.1 SINN with r=1.

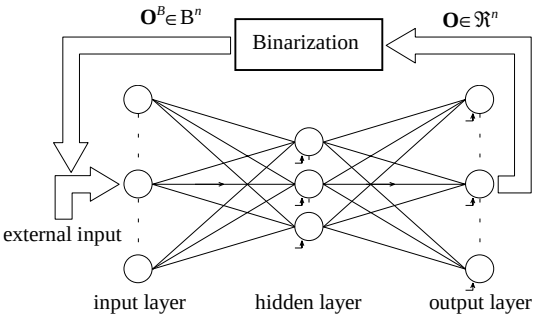


Fig.2 RMLP with 3-valued weights.

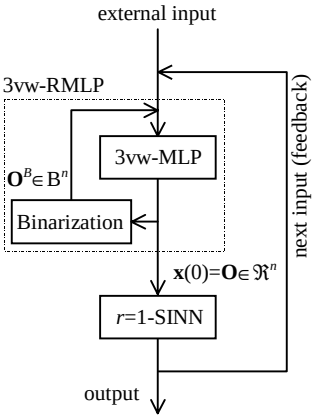


Fig.3 RMLP-SINN.

multipliers from synaptic circuits and they are stored in 2-bit digital memories. Therefore, the digital MLPs with 3-valued weights (3vw-MLPs) are implemented at a low cost. Moreover, 3vw-MLPs can execute high-speed parallel processing, because many neurons are effectively implemented on the hardware. However, 3vw-MLPs do not have good performance as associative memories. This is because 3-valued weights decrease the generalization capability (GC) and MLPs are feed-forward networks. Especially, the latter damages associative memories composed of 3vw-MLPs more seriously than the former.

From the above facts, we propose recurrent MLPs (RMLPs) shown in Fig.2 to make good use of 3vw-MLPs. The activation function is f_{PWL} in Eq.(2). The RMLP is constructed by adding feedback to 3vw-MLP designed by modified backpropagation algorithm (see, e.g., [4]). The feedback signal ($\mathbf{O}^{\mathrm{B}}\in \mathrm{B}^n$) is given by binarizing the output of 3vw-MLP ($\mathbf{O}\in \Re^n$):

$$O_i^B = \begin{cases} +1, \text{ if } O_i \geq 0 \\ -1, \text{ otherwise} \end{cases}. \tag{4}$$

As a result, the data bit length for states and outputs of neurons is fixed. The feedback signal $\mathbf{O}^{\mathrm{B}}$ is easily obtained from the sign bit of the output $\mathbf{O}$. Therefore, RMLPs with 3-valued weights (3vw-RMLPs) are also suitable for hardware implementation. Furthermore, it is clear that the capability of 3vw-RMLP is better than that of 3vw-MLP.

Since the GC of MLP affects the capability of RMLP, we consider how to set w in order to improve the GC of 3vw-MLP. According as w becomes large, the state of a neuron is easily influenced by noise included in the input patterns. Therefore, very large w makes GC low. On the other hand, if w becomes small to some extent, the above problem does not occur. As a result, GC should be improved. However, when w is too small, the states of hidden neurons always stay in the linear region of f_{PWL}. In this case, it is clear that GC becomes bad. From the above facts, we have proposed the following equation as one of conditions to decide appropriate w [4],

$$w(n+1) >> 1, \tag{5}$$

where n is the number of input neurons. Since $w(n+1)$ is the maximum value of states of hidden neurons, Eq.(5) indicates that the states may be in the saturation region of f_{PWL} when memory patterns are inputted. Therefore, we suggest that w should be set to smaller value within Eq.(5) to keep high GC of 3vw-MLP. However, one should notice that small w demands a lot of hidden neurons.

4. RMLP-SINN Associative Memories

Since the sparse interconnections make SINNs suitable for analog hardware implementation, the r=1-SINN has the local association and the continuous associative process. Although the local association might work well for the cases where FINN fails to associate memory patterns, it causes serious damage to the capability of r=1-SINN. To improve r=1-SINN, each neuron has to be supplied with the global information of an input pattern. Increasing the interconnections is the easiest way to satisfy the above

requirement. Ideally FINN should be exploited instead of r=1-SINN. However, a marked increase in interconnections means that hardware cost goes up rapidly.

The 3vw-RMLP has the global association and the discrete associative process, because each neuron is connected to all neurons in the previous layer and the feedback signal is binarized. Although the quantized weights and the binarized feedback signal make 3vw-RMLP suitable for digital hardware implementation, they restrict the capability of 3vw-RMLP. The reasons are as follows. The quantized weights decrease the generalization capability of 3vw-MLP included in 3vw-RMLP. The binarized feedback signal ignores the small correction given by 3vw-MLP. These problems can be solved by increasing the number of quantization levels of weights and feedback signal. Although it is ideal that these parameters are close to continuous values, RMLPs need large hardware cost. Furthermore, if the number of neurons included in such RMLPs becomes large, it is almost impossible to implement all the neurons on the digital hardware.

Comparing r=1-SINN with 3vw-RMLP, we can find followings. First, loosening conditions to reduce hardware cost is not a good way to improve the capabilities of r=1-SINN and 3vw-RMLP, because hardware cost goes up rapidly. Second, r=1-SINN and 3vw-RMLP can cooperate each other. Therefore, we suggest using interaction between 3vw-RMLP and r=1-SINN to get high-performance associative memories implemented at a low cost. Figure 3 shows our proposed system named RMLP-SINN associative memories. In this system, an external input is given to RMLP. The output of RMLP ($\mathbf{O}\in\Re^n$) is set to the initial state of SINN ($\mathbf{x}(0)\in\Re^n$). Moreover, the stable output of SINN is fed back to RMLP as the next input. When the stable output of SINN stops changing, it is the final output of this system.

5. Simulation Results

In this section, we show the capability of RMLP-SINN. Our proposed system is compared with FINN, r=1-SINN, 3vw-MLP, and 3vw-RMLP. Each associative memory stores N_p-memory patterns, which are randomly selected from 26 images shown in Fig.4. Each pattern is illustrated by a (7×5)-pixel image. A noisy pattern is constructed by inverting some pixels in a memory pattern. The number of the inverted pixels depends on a given Hamming distance (*HD*). Moreover, considering the processing speed of the hardware, feedback iterations in RMLP and from SINN to

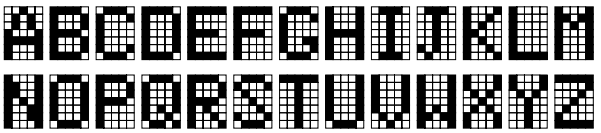
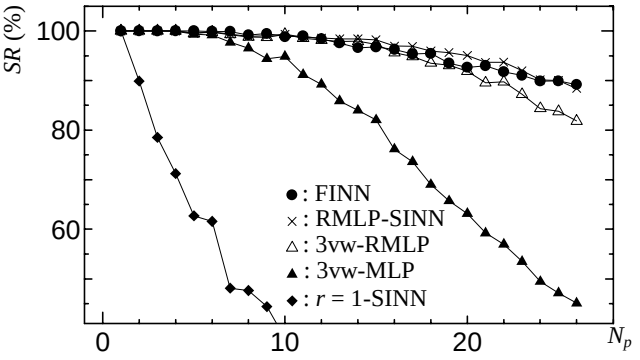
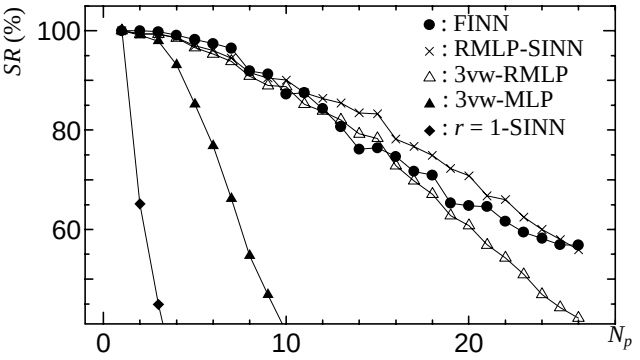


Fig.4 Memory patterns.



(a) *HD* = 4



(b) *HD* = 8

Fig.5 Success rate (*SR*) of each associative memory.

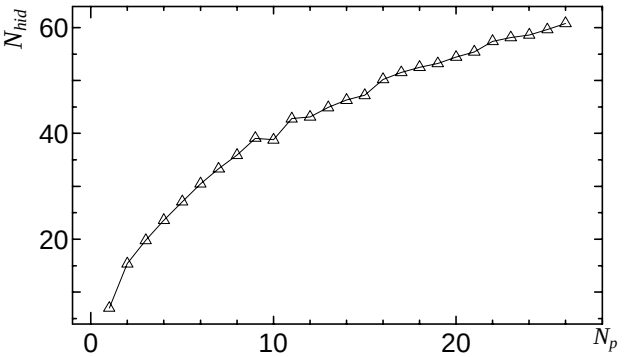
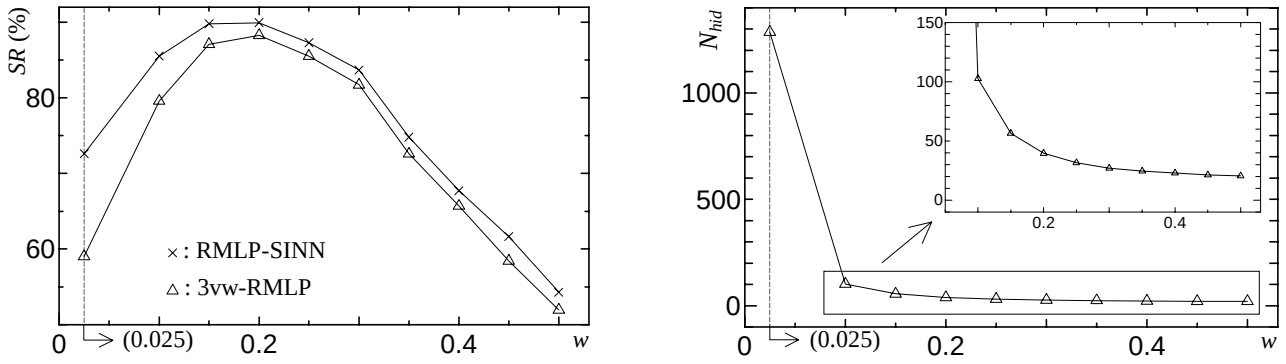


Fig.6 The number of hidden neurons (N_{hid}) in 3vw-RMLP.



(a) *SR* (*HD* = 8) (b) N_{hid}

Fig.7 Influence of *w* on the capability of RMLP-SINN, when N_p=10.

RMLP are restricted to 20 times at the maximum. The other parameters are as follows. The SINN and FINN consist *n*=35 neurons. The MLP and RMLP have 35 input and 35 output neurons. For a given parameter *w*, the number of hidden neurons (N_{hid}) is set as small a value as possible.

Figure 5 presents the success rate (*SR*) for each associative memory. In this simulation, the parameter *w* is 0.2. Here, let us define the *SR* as the ratio of the number of noisy patterns which converge to the desired memory patterns to the number of total trials (i.e., 1000× N_p trials) for each N_p and *HD*. Figure 6 shows N_{hid} used in the above simulation. From these results, we can find followings. The *SR* of *r*=1-SINN decreases rapidly. Although *SR* of 3vw-MLP is also low, the 3vw-RMLP can keep much higher *SR* than 3vw-MLP. Therefore, it is found that we can achieve to make good use of 3vw-MLP. However, when N_p and *HD* are large, the *SR* of 3vw-RMLP is lower than that of FINN. On the other hand, RMLP-SINN is better than FINN in almost all the cases. This result shows that *r*=1-SINN and 3vw-RMLP can cooperate each other.

Next, we investigate the influence of the parameter *w* in our proposed system. If *w* is 0.027, $w(n+1)$ is about 1. Figure 7 presents the *SR* and N_{hid} for various *w*, when N_p= 10 and *HD*=8. The results show that it is almost right that smaller *w* within Eq.(5) makes for good performance in terms of *SR* of 3vw-RMLP. Moreover, our system also satisfies this characteristic. In other words, Eq.(5) can be used to keep high performance in RMLP-SINN. On the other hand, as *w* becomes smaller, N_{hid} increases rapidly. Therefore, we suggest that *w* should be set to a small value within both Eq.(5) and the permissible hardware resources.

From the above results, it is confirmed that our proposed system is one of useful and practical associative memories realized at a low cost.

6. Conclusions

In this paper, we have proposed associative memories using interaction between 3vw-RMLPs and *r*=1-SINNs without loosening their conditions to reduce hardware cost. Simulation results have shown that the success rate of RMLP-SINN is better than that of FINN. Therefore, we conclude that our proposed system is useful and practical in terms of the capability and hardware cost.

References

[1] T. Kamio, H. Fujisaka, and M. Morisue, "Back-propagation algorithm for logic oriented neural networks with quantized weights and multilevel threshold neurons," IEICE Trans. Fundamentals, vol.E84-A, no.3, pp.705-712, Mar. 2001.

[2] L. O. Chua and L. Yang, "Cellular neural networks: Theory," IEEE Trans. Circuits & Syst., vol.35, no.10, pp.1257-1272, Oct. 1988.

[3] D. Liu and Z. Lu, "A new approach for feedback neural networks based on the perceptron training algorithm," IEEE Trans. Neural Networks, vol.8, pp.1468-1482, Nov. 1997.

[4] T. Kamio, H. Fujisaka, and M. Morisue, "Associative memories using interaction between multilayer perceptrons and sparsely interconnected neural networks," IEICE Trans. Fundamentals, vol.E85-A, no.6, pp.1220-1228, Mar. 2001.