

An Evaluation of the Dynamics of Diluted Neural Network

Lijuan Wang¹, Jun Shen², Qingguo Zhou^{3,*}, Zhihao Shang³, Huaming Chen², Hong Zhao⁴

¹ School of Cyber Engineering, Xidian University, Xi'an, China

² School of Computing and Information Technology, University of Wollongong, Wollongong, Australia

³ School of Information Science and Engineering, Lanzhou University, Lanzhou, China

⁴ Department of Physics, Xiamen University, Xiamen, China

Received 5 January 2016

Accepted 2 August 2016

Abstract

The Monte Carlo adaptation rule has been proposed to design asymmetric neural network. By adjusting the degree of the symmetry of the networks designed by this rule, the spurious memories or unwanted attractors of the networks can be suppressed completely. We have extended this rule to design full-connected neural networks and diluted neural networks. Comparing the dynamics of these two neural networks, the simulation results indicated that the performance of diluted neural network was poorer than the performance of full-connected neural network. As to this point, further research is needed. In this paper, we use the annealed dilution method to design a diluted neural network with fixed degree of dilution. By analyzing the dynamics of the diluted neural network, it is verified that asymmetric full-connected neural network do have significant advantages over the asymmetric diluted neural network.

Keywords: diluted neural network; annealed dilution; dynamics; spurious memory.

1. Introduction

Spurious memory is a classical problem in neural networks.^{1,2,3,4,5,6,7,8} A variety of algorithms have been proposed to suppress the spurious memories. These algorithms could decrease the number of spurious memories to some extent, but it is unknown whether the spurious memories would completely disappear by using one of these algorithms. With the development of biology knowledge, we become more and more acquainted with biological neural systems. It has been indicated that asymmetric neural networks can suppress spurious memories.^{9,10,11,12,13}

One of the authors developed a strategy for designing asymmetric neural networks of associative

memory with controllable degree of symmetry and controllable basins of attraction.¹⁴ The strategy is called Monte Carlo Adaptation (MCA) rule. It is shown that the performance of the networks designed by MCA rule depends on the degree of the symmetry. By adjusting the degree of the symmetry, the spurious memories or unwanted attractors can be suppressed completely. The values of the coupling matrix are set to be two states: the binary numbers ± 1 . By extending the original MCA rule to design feedback neural networks with discrete weights, the performance of feedback neural networks with single-layer full connection as well as diluted connection synapses have been studied.¹⁵ And their storage capacity and global dynamics behaviors were evaluated. The simulated results indi-

* Corresponding author, E-mail: zhouqg@lzu.edu.cn

cated that the full-connected neural network had better performance than the diluted one.¹⁵ This is different from the traditional views, which thought the diluted neural network had better performance than the full-connected neural network.^{16,17,18}

The authors of Ref. 19 designed a new learning algorithm towards self-generating fuzzy neural work. The proposed algorithm was efficient and was able to generate a fuzzy neural network with high accuracy and compact structure. Ref. 20 presented a survey about the development of neural associative memories, and pointed out the importance of sparse coding of associative memory patterns. A series of pipelined decision feedback recurrent neural networks were proposed by Ref. 21–25. The simulations demonstrated these feedback recurrent neural networks with less computational complexity and better convergence rate. The author of Ref. 26 suggested the asymmetric neural network could store more patterns when the dilution and learning processes were performed together. The time order of the learning and dilution processes affects the storage capacity of diluted neural networks. There are three different types of diluted neural networks: dilution after learning, dilution before learning, and dilution during learning. The last one is different from the former two. In the last type of diluted neural network, the dilution process and the learning process are carried out at the same time. When the dilution process occurs, it will consider the stored patterns in the network and select the best broken bonds to have more patterns to be stored. This method is called annealed dilution and will be adopted in this paper.

In our previous work¹⁵, we had compared the dynamics of the full-connected neural network and the diluted neural network, but we did not fix the degree of dilution. In other words, the size of the full-connected neural network and the size of the diluted neural network were the same. In order to study the performance of the diluted neural network more deeply, this paper designs a diluted neural network with fixed degree of dilution and compares its dynamics with the full-connected neural network which has the same real links with the diluted one. Meanwhile, this paper analyzes the valid storage ra-

tio, the training time, and the retrieval time of the two types of neural networks, as well as the three phases of diluted neural networks with different degrees of dilution. This paper also applies the two types of neural networks for solving the handwritten digital image problem.

The outline of this paper is the following. Section 2 describes the rule to design asymmetric neural network with binary weighted values, and the rule to design diluted neural network with fixed degree of dilution. Section 3 presents the dynamic behaviors of the designed diluted neural network and the full-connected neural network. Then section 4 introduces the application of the two types of neural networks for solving the handwritten digital image problem. Finally, section 5 concludes this paper.

2. The Design Rule

2.1. The rule to design asymmetric neural network with binary weighted values

The goal of MCA rule is to control attraction basins of the stored patterns, by means of distributing the local fields of the stored patterns. We employ the single-layer feedback neural network, which is given by (1), to introduce this idea.

$$s_i(t+1) = \text{sgn}(h_i(t))$$

$$h_i(t) = \sum_{j=1}^N J_{ij}s_j(t), i = 1, 2, \dots, N \quad (1)$$

Here, N is the number of neurons, $s_i(t) \in \{\pm 1\}$ is the state of i_{th} neuron at time t , $h_i(t)$ is the instantaneous local field of the i_{th} neuron, and $\text{sgn}(h)$ is the sign function. The goal of the design is to find the coupling matrix J for which P patterns $\{\varepsilon^\mu, \mu = 1, 2, \dots, P\}$ are stored as fixed points of (1). Every pattern $\varepsilon^\mu = \{\varepsilon_1^\mu, \varepsilon_2^\mu, \dots, \varepsilon_N^\mu\}$ should satisfy, $\varepsilon_i^\mu = \text{sgn}(h_i^\mu)$, or, equivalently,

$$b_i^\mu = \varepsilon_i^\mu h_i^\mu > 0$$

$$h_i^\mu = \sum_{j=1}^N J_{ij}\varepsilon_j^\mu, i = 1, 2, \dots, N \quad (2)$$

Being stored as a fixed point is just the primary requirement for a memory pattern. An efficiently

stored pattern should be practically retrievable. The quality of the retrieval is related to the attraction basin of the memory pattern. The larger the attraction basin of the pattern, the more easily it can be retrieved. To illustrate this, we may analyze the following example.

Suppose ε^μ is a memory pattern and there are a set of mutations, which will take place on ε^μ , namely, $\varepsilon_k^\mu \rightarrow -\varepsilon_k^\mu$, for $k \in \{r\}$, where $\{r\}$ is the index set of the mutations, then we have

$$\widetilde{b}_i^\mu = b_i^\mu - 2\varepsilon_i^\mu \sum_{k=1}^r J_{ik}\varepsilon_k^\mu \quad (3)$$

If $\widetilde{b}_i^\mu > 0$ for $i = 1, 2, \dots, N$, the mutated pattern should be attracted to the original one via one step of evolution under (2). For convenience, we define the set of patterns that can be attracted to the memory pattern ε^μ by only one step of evolution, as the core of the attraction basin of the memory pattern. To keep \widetilde{b}_i^μ positive under a number of simultaneous mutations, b_i^μ should be as large as possible, while $|\varepsilon_i^\mu \sum_{k=1}^r J_{ik}\varepsilon_k^\mu|$ in Eq. (3) being as small as possible. Thus, designing a feedback neural network is a typical problem of constrained optimization.

2.2. The rule to design diluted neural network with fixed degree of dilution

We extend the original MCA rule to design the diluted neural network with fixed degree of dilution. Note that b_i^μ is determined only by the i_{th} row of synapse matrix J , in this way the synapse matrix can be designed row by row independently. We start with a random matrix with random value ± 1 , suppose a set of P patterns to be stored, and limit the diagonal element $J_{ii} = 0$. The design process consists of the following steps.

- (i) The first step, for the i_{th} row of J , we find the minimum value of b_i^μ . Usually there are a set of elements having the same minimum. We denote the minimum value as b_i^{min} and record the indexes of these elements as $\{\mu^1, \mu^2, \dots, \mu^m\}$, namely, $b_i^\mu = b_i^{min}$ if $\mu \in \{\mu^1, \mu^2, \dots, \mu^m\}$.

- (ii) The second step, for every element in i_{th} row of matrix J , namely, $\{J_{ij}, j = 1, 2, \dots, N, j \neq i\}$, we compute $\{s_i^\mu J_{ij}^\mu s_j^\mu, \mu = \mu^1, \mu^2, \dots, \mu^m\}$ and find the items that are lower than or equal to zero, denoted as m_{ij} . Using m_{ij}^{max} to represent the biggest item among them, we find a set $\{j^1, j^2, \dots, j^m\}$ that satisfies $m_{ij} = m_{ij}^{max}$, where $j \in \{j^1, j^2, \dots, j^m\}$.
- (iii) The third step, we randomly select one item j from the set $\{j^1, j^2, \dots, j^m\}$, if $J_{ij} < 0$, then $J_{ij} = 0$; if $J_{ij} > 0$, then $J_{ij} = -1$; otherwise $J_{ij} = 1$. Before doing this, we have fixed the number of nil elements of the coupling matrix J , and make every row have the same number of nil elements. But the location of nil elements is adjusted to maximize the total number of patterns to be stored. Here we adopt the annealed dilution. We suppose every row has $(1 - f) * N$ nil elements. The parameter f ($0 \leq f \leq 1$) represents the fraction of non-zero synapses afferent to a neuron. It is used to denote the degree of dilution: $f = 1$ represents the network is full-connected (no dilution) and $f = 0$ stands for the disconnected network (extreme dilution).

3. Dynamic Behaviors of the Two Neural Networks

The two neural networks in this paper refer to the full-connected neural network with binary weighted values and the diluted neural network with fixed degree of dilution, respectively. In this section, the results of numerical simulations are discussed and the performances of the two neural networks are illustrated.

Generally, the MCA rule indicates that there are two transformation points ($C1, C2$) and three phases in the relational graph of absorptivity P_{ac} and control parameter C in the full-connected neural network. C is a positive constant, being used to control the absolute value of local field. Meanwhile, C determines the size of attraction basin of the stored patterns. Larger values of C imply larger basins of attraction, thus this parameter is called the stability parameter.¹⁵ In the region $C < C1$, $P_{ac} = 0$, namely,

almost no initial state is attracted to the memory patterns. This region is referred as chaos phase. In the region $C1 < C < C2$, $P_{ac} = 1$, namely, all the initial states are attracted to memory patterns. It is referred as pure memory phase. In other words, there are no spurious memories at all in this region. In the region $C > C2$, P_{ac} decreases as C increases, which implies extra attractors besides the memory patterns coexist in this region.

Meanwhile, the storage capability is always the key point when designing neural network. For a neural network with N neurons and P stored patterns, suppose R denotes the storage capability, then $R = P/N$. The storage capability of Hopfield networks has been studied for a long time, and the maximum storage capability of Hopfield neural networks is $0.138N$. Above this value the network will not learn and cannot store patterns, then the patterns will not be the stable points of the network.

There are two conditions when a stored pattern is considered as the attractor of the network. One is that the pattern is the fixed point of the network, and the other is that it must have the domain of attraction. Our proposed rules indicate that when the controlled parameter $C > 0$, the stored patterns satisfy the above two conditions.

Fig. 1 shows the phase diagrams of the full-connected neural network, and Fig. 2 gives an example of the diluted neural network with $f = 0.98$. $C1$ and $C2$ are the two transformation points. The diluted neural network is set with 500 neurons, and the storage ratio is 0.02 (10 stored patterns). As mentioned earlier, the two neural networks are designed to have the same real neuron links in this paper. Therefore, the full-connected neural network has 495 neurons. The three phases a , b , r in Fig. 1 and Fig. 2 represent chaos phase, pure memory phase, and mixed phase. It is clear that the dynamics behaviors of the two neural networks are similar. However, the pure memory phase b and the first transformation point $C1$ of the diluted neural network are smaller than that of the full-connected one.

In the chaos phase and pure memory phase, there are no spurious memories and all the stored patterns are the fixed points of the networks. Thus, the first transformation point $C1$ can denote the maximum

storage capability of the network without spurious memories. It is also called maximum effective storage capacity, denoted by α_{max} .

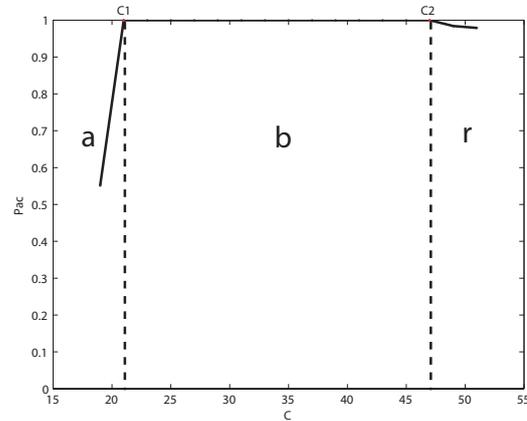


Fig. 1. The transformation points and three phases of full-connected neural network

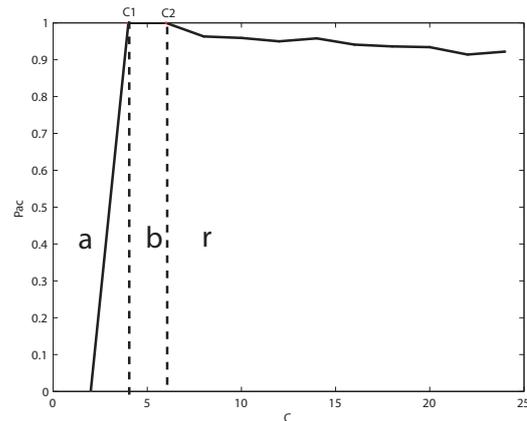


Fig. 2. The transformation points and three phases of diluted neural network with $f = 0.98$

Fig. 3 and Fig. 4 on the next page show the storage capability of the two neural networks. According to these two figures, $\alpha_{max} = 0.12$ in full-connected neural network and $\alpha_{max} = 0.08$ in the diluted neural network. The results indicates that the storage capability of the diluted neural network with fixed degree of dilution is lower than the full-connected one. Meanwhile, the first transformation point of the full-connected neural network does not change as the storage ratio changes. In other words, the first

transformation point is independent of the storage ratio. However, in our designed diluted neural network, the first transformation point is stable at first but later increases when the storage ratio increases. The second turning point increases first and later decreases. Thus, in this diluted neural network, the first transformation point is not stable.

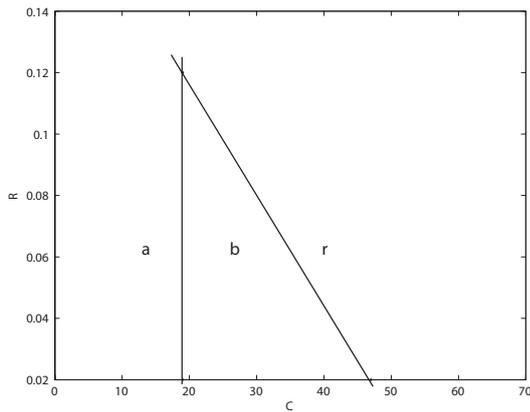


Fig. 3. The value of α_{max} in full-connected neural network

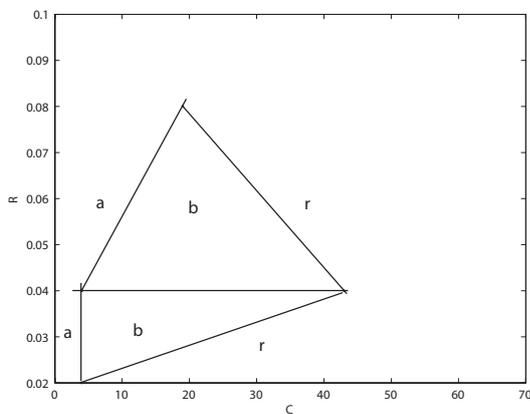


Fig. 4. The value of α_{max} in diluted neural network

Fig. 5 and Fig. 6 show the storage capacity of the discrete multi-state neural network, which was designed in our previous work.¹⁵ According to Fig. 5 and Fig. 6, $\alpha_{max} = 0.11$ in full-connected neural network and $\alpha_{max} = 0.04$ in diluted neural network. Hence, this paper has similar conclusion, that

the storage capability of the diluted neural network with fixed degree of dilution is lower than the full-connected one.

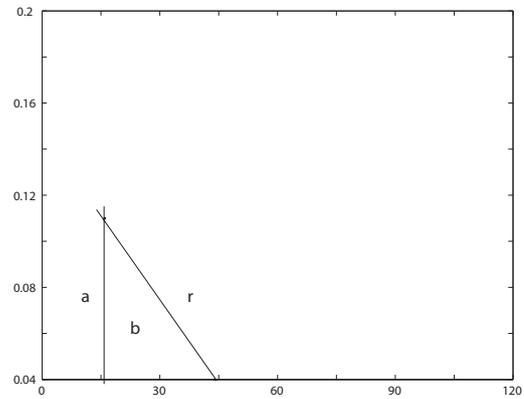


Fig. 5. The value of α_{max} in full-connected neural network

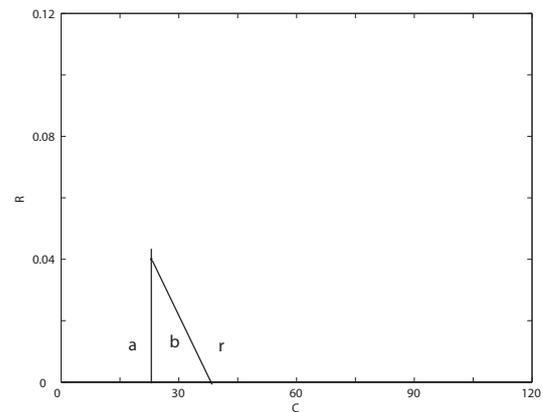


Fig. 6. The value of α_{max} in diluted neural network

Fig. 7, Fig. 8 and Fig. 9 on the next page show the phase diagram of the diluted neural network with $f = 0.96$, $f = 0.90$ and $f = 0.80$, respectively. The diluted neural network is set with 500 neurons, and the number of stored patterns is 10. Compared with the phase diagram of Fig. 2, the pure memory phase becomes larger when $f = 0.96$, then it becomes smaller with the increasing of the degree of dilution. When $f = 0.80$, there is no pure memory phase in the diluted neural network, hence it is not necessary to continue exploring greater dilutions.

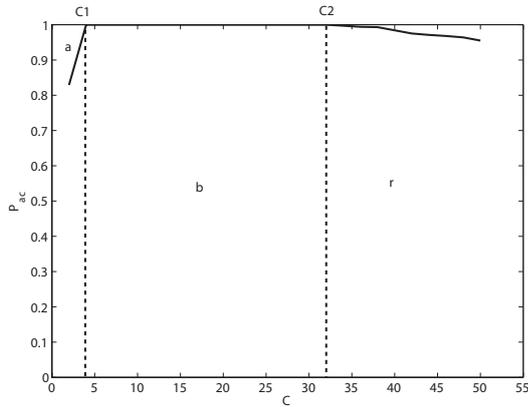


Fig. 7. The transformation points and three phases of the diluted neural network with $f = 0.96$

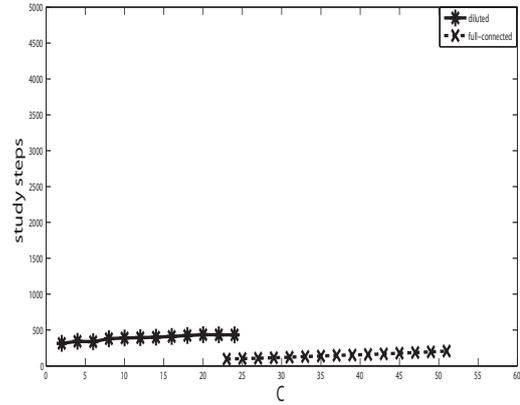


Fig. 10. The training time of the two neural networks with 10 stored patterns

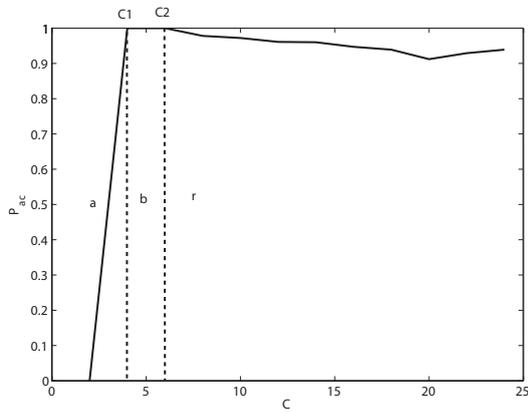


Fig. 8. The transformation points and three phases of the diluted neural network with $f = 0.90$

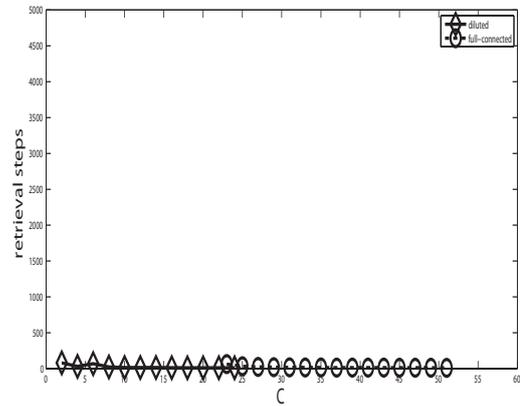


Fig. 11. The retrieval time of the two neural networks with 10 stored patterns

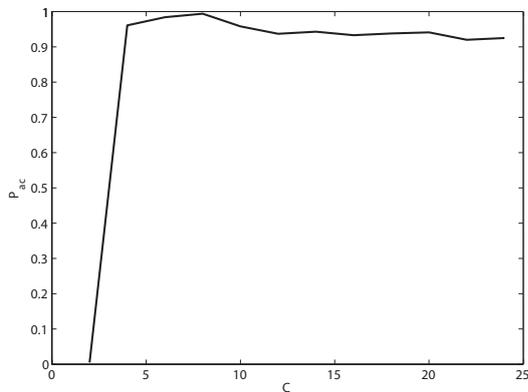


Fig. 9. The transformation points and three phases of the diluted neural network with $f = 0.80$

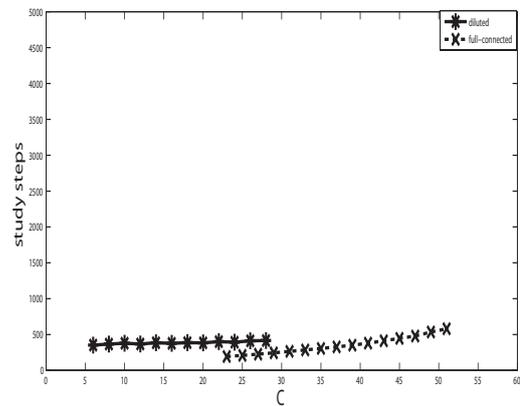


Fig. 12. The training time of the two neural networks with 20 stored patterns

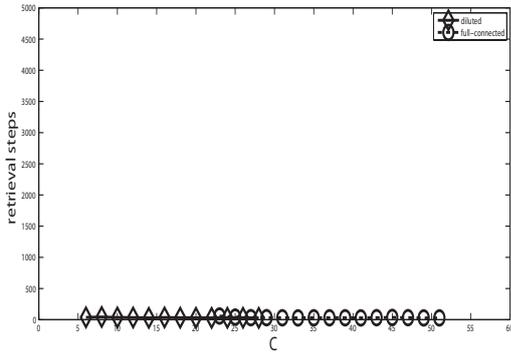


Fig. 13. The retrieval time of the two neural networks with 20 stored patterns

Fig. 10 and Fig. 11 show the training time and retrieval time of the full-connected neural network and diluted neural network with 10 stored patterns. Fig. 12 and Fig. 13 show the training time and retrieval time of the full-connected neural network and diluted neural network with 20 stored patterns. The training time and retrieval time represent the training step length and retrieval step length, respectively. The size of the diluted neural network is 500 and $f = 0.98$. The full-connected neural network has 495 neurons. The results indicated that the diluted neural network needs more training time (study step) than that of the full-connected one but it needs less retrieval time. Due to space limit, herein we only show the results in the pure memory phase, the chaos phase and mixed phase have the similar results.

4. A Case Study

After presenting the design rule and studying the dynamic behaviors of the full-connected neural network and the diluted neural network, the two neural networks will also be applied for solving the handwritten digital image problem.

In our evaluation, we experimented with the U.S. Postal (USPS) handwritten digits dataset.²⁷ The dataset consists of a training set with 7291 images and a test set with 2007 images. The images are 16*16 eight-bit grayscale maps, with each pixel ranging in intensity from 0 to 255. The task is to predict, from the 16*16 matrix of pixel intensities, the identity of each image (0, 1, ..., 9) quickly and

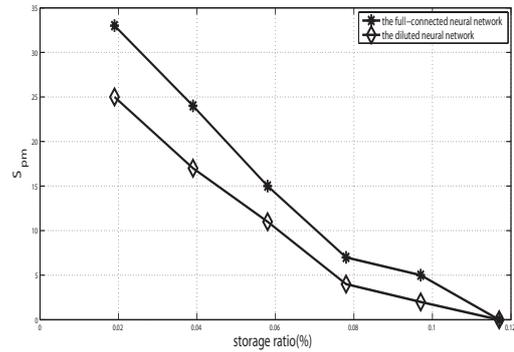


Fig. 14. The size of the pure memory phase of the two neural networks

accurately.²⁷ In order to get binary data, we converted each pixel into two bits, and each image has 512 bits. Then we set the number of neurons in the diluted neural network to 512.

In the evaluation, we chose ns ($ns \in \{1, 2, \dots, 6\}$) samples for each image (0, 1, ..., 9), and we got $10 * ns$ training cases. Then the storage capability is $R = 10 * ns / 512$. The size of the pure memory phase is denoted as $S_{pm} = C2 - C1$. Fig. 14 shows the value of S_{pm} of the full-connected neural network and the diluted neural network. The degree of dilution in the diluted neural network is set to $f = 0.98$. The results indicated that the value of S_{pm} of the full-connected neural network is greater than that of the diluted neural network when they had the same storage capability.

5. Conclusion and Future Work

This paper presented the designed diluted neural network with fixed degree of dilution. The dynamics behaviors of this network are consistent with the discrete multi-state diluted neural network in previous works. In our previous work, the degree of dilution was not fixed and the size of the full-connected neural network was same with that of diluted neural network. In this paper, the degree of dilution is fixed and only the real neuron links of the two neural networks are the same. Thus, the sizes of the two neural networks are different. The simulation results demonstrated that the maximum storage capability of the designed diluted neural network is

smaller than that of the full-connected neural network designed by the MCA rule. Meanwhile, in our experiment, the pure memory phase of the diluted neural network exists when the degree of dilution $f \in \{0.8, 1\}$. And in the beginning, the pure memory phase of the diluted neural network becomes larger with the increasing of the degree of dilution ($f \in \{0.9, 1\}$), then it becomes smaller with the increasing of the degree of dilution ($f \in \{0.8, 0.9\}$). The results also indicated that the second transformation point of the diluted neural network does not decrease with the increase of the number of stored patterns, but fluctuates.

We will conduct further research on other unsolved issues. For example, the diluted elements in this paper are asymmetry, and every row of the coupling matrix has the same number of nil elements. The performance of the symmetry diluted elements is also worth to investigate for future work.

Acknowledgments

This work has been partly supported by National Natural Science Foundation of China under Grant No. 60973137, Gansu Sci. & Tech. Program under Grant No. 1104GKCA049 and No. 1204GKCA061, the Fundamental Research Funds for the Central Universities under Grant No. lzujbky-2012-44 and No. 8002/20103156225, and UOW's UIC Project 2015. We would like to thank Dr. Tao Jin for suggestions and comments during this work.

References

1. J.J.Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proceedings of the National Academy of Sciences of the United States of America* 79 (8) (1982) 2554–2558.
2. S.Hu, X.Liao, X.Mao, Stochastic hopfield neural networks, *Journal of Physics A: Mathematical and General* 36 (9) (2003) 2235–2249.
3. G.Huang, J.Cao, Multistability in bidirectional associative memory neural networks, *Physics Letters A* 372 (16) (2008) 2842–2854.
4. T.Isokawa, H.Nishimura, N.Kamiura, N.Matsui, Associative memory in quaternionic hopfield neural network, *International Journal of Neural Systems* 18 (2) (2008) 135–145.
5. A.Singh, S.Kumar, T.P.Singh, Analysis of hopfield associative memory with combination of mc adaptation rule and an evolutionary algorithm, *International Journal of Computer Applications* 78 (11) (2013) 37–42.
6. T.P.Vogels, K.Rajan, L.F.Abbott, Neural network dynamics, *Annual Review of Neuroscience* 28 (2005) 357–376.
7. F.Yasar, The investigation of the spin glass properties of the hopfield neural network model, *International Journal of Modern Physics C* 16 (3) (2005) 427–437.
8. H.Zhao, General vector machine, <http://arxiv.org/abs/1602.03950> (2016).
9. M.Gao, B.Cui, Adaptive synchronization in an array of asymmetric coupled neural networks, *Chinese Physics B* 18 (1) (2009) 76–83.
10. T.Jin, H.Zhao, Pattern recognition using asymmetric attractor neural networks, *Physical Review E* 72 (6) (2005) 06611:1–7.
11. J.Lu, D.W.C.Ho, M.Liu, Globally exponential synchronization in an array of asymmetric coupled neural networks, *Physics Letters A* 369 (5-6) (2007) 444–451.
12. H.Soula, G.Beslon, O.Mazet, Spontaneous dynamics of symmetric random recurrent spiking neural networks, *Neural Computation* 18 (1) (2006) 60–79.
13. P.Zheng, W.Tang, J.Zhang, Efficient continuous-time asymmetric hopfield networks for memory retrieval, *Neural Computation* 22 (6) (2010) 1597–1614.
14. H.Zhao, Designing asymmetric neural networks with associative memory, *Physical Review E* 70 (6) (2004) 066137:1–4.
15. L.Wang, Q.Zhou, T.Jin, H.Zhao, Feed-back neural networks with discrete weights, *Neural Computing and Application* 22 (6) (2013) 1063–1069.
16. J.W.Bohland, A. Minai, Efficient associative memory using small-world architecture, *Neurocomputing* 38-40 (2001) 489–496.
17. N. Davey, R. Adams, High capacity associative memories and connection constraints, *Connection Science* 16 (1) (2004) 47–65.
18. A. Sankar, R. Mammone, Optimal pruning of neural tree networks for improved generalization, in: *IEEE International Joint Conference on Neural Networks*, 1991, pp. 219–224.
19. F.Liu, M.Er, A novel efficient learning algorithm for self-generating fuzzy neural network with applications, *International Journal of Neural Systems* 22 (1) (2012) 21–35.
20. C.Palm, Neural associative memories and sparse coding, *Neural Networks* 37 (1) (2013) 165–171.
21. H. Q. Zhao, S. B. Gao, Z. Y. He, X. P. Zeng, W. D. Jin, T. R. Li, Identification of nonlinear dynamic system using a novel recurrent wavelet neural network based on the pipelined architecture, *IEEE Transactions on Industrial Electronics* 61 (8) (2014) 4171–4182.

22. H. Q. Zhao, J. S. Zhang, Adaptively combined fir and functional link artificial neural network equalizer for nonlinear communication channel, *IEEE Transactions on Neural Networks* 20 (4) (2009) 665–674.
23. H. Q. Zhao, X. P. Zeng, Z. Y. He, Low-complexity nonlinear adaptive filter based on a pipelined bilinear recurrent neural network, *IEEE Transactions on Neural Networks* 22 (9) (2011) 1494–1507.
24. H. Q. Zhao, X. P. Zeng, Z. Y. He, W. D. Jin, T. R. Li, Complex-valued pipelined decision feedback recurrent neural network for non-linear channel equalization, *IET Communications* 6 (9) (2012) 1082–1096.
25. H. Q. Zhao, X. P. Zeng, J. S. Zhang, R. Tian, Y. G. Liu, D. Ruan, Pipelined functional link artificial recurrent neural network with the decision feedback structure for nonlinear channel equalization, *Information Sciences* 181 (17) (2011) 3677–3692.
26. M. Bouten, Storage capacity of diluted neural networks, *Lecture Notes in Physics* 368 (1990) 225–236.
27. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2009.