A Recurrent Neural Network with Serial Delay Elements for Memorizing Limit Cycles

Seiji MIYOSHI* and Kenji NAKAYAMA†
*Graduate School of Natural Science and Technology, Kanazawa Univ.
†Department of Electrical and Computer Eng. Faculty of Eng., Kanazawa Univ.
*†2-40-20, Kodatsuno, Kanazawa 921, Japan
*E-mail miyoshi@ddec1.kobe-kosen.ac.jp

ABSTRACT

A recurrent neural network(RNN), in which each unit has serial delay elements, is proposed for memorizing limit cycles(LCs). This network is called DRNN in this paper. An LC consists of several basic patterns. The hysteresis information of LCs, realized on the connections from the delay elements to the units, is very efficient in the following reasons. First, the same basic patterns can be shared by different LCs. This make it possible to drastically increase the number of LCs, even though using a small number of the basic patterns. Second, noise performance, that is, probability of recalling the exact LC starting from the noisy LC, can be improved. The hysteresis information consists of two components, the order of the basic patterns included in an LC, and the cross-correlation among all the basic patterns. The former is highly dependent on the number of LCs, and the latter the number of all the basic patterns. In order to achieve good noise performance, a small number of the basic patterns is preferred. These properties of the DRNN are theoretically analized and confirmed through computer simulations. It is also confirmed that the DRNN is superior to the RNN without delay elements for memorizing LCs.

1. Introduction

An associative memory is one of hopeful applications of artificial neural networks(NNs). Memorizing patterns to equilibrium points of a recurrent neural network (RNN) is one approach. Researches on algorithms of learning and recall which improve memory capacity and noise performance have been carried out[1],[2],[3]. Limit cycle(LC) memories have been also studied [4]–[8]. The RNNs have been mainly used, in which patterns, components of an LC, are stored at equilibrium points. Furthermore, dynamical operation of such RNNs has been investigated, taking delay effects into account[9]. However, in existing approaches, hysteresis information of LCs have not been well used in a recall process.

In this paper, an RNN, in which units have serial delay elements in order to take the hysteresis information of an LC, is proposed. This network is called DRNN in this paper. The DRNN operates as a discrete time system. Patterns included in LCs can be used as hysteresises information through feedback weights from the delay elements.

First, we show capability of the DRNN to memorize many LCs which include common basic patterns. Next, basic properties of the DRNN such as memory capacity and recall performance from noisy patterns are investigated through theoretical analysis and computer simulation. The effect of the hysteresis information is investigated. Comparisons between the RNN memory with equilibrium memory points and the DRNN are discussed. Finally, design methodology for good noise performance is discussed.

2. RNN with Serial Delay Elements (DRNN)

2.1. Structure of DRNN

Structure of the DRNN is shown in Fig.1. The number of the delay elements is three. $\{u_{i0}\}$, $\{u_{i1}\}$, $\{u_{i2}\}$, $\{u_{i3}\}$, $i = 1 \sim N$, form four basic patterns. A black square indicates a connection weight w_{ijk} from the kth delay element of the ith unit to the jth unit, where k=0 means the ith unit itself. It is assumed $w_{iik} = 0$.

2.2. Network Equations

Input and output relations of the units and the delay elements are given by

$$v_j(n) = \sum_{i=1}^{N}\sum_{k=0}^{L-1} w_{ijk}u_{ik}(n) \tag{1}$$

$$u_{j0}(n+1) = \begin{cases} 1, & v_j(n) \geq T \\ u_{j,L-1}(n), & |v_j(n)| < T \\ 0, & v_j(n) \leq -T \end{cases} \tag{2}$$

$$u_{j,k+1}(n+1) = u_{jk}(n) \quad k = 0, \cdots, L-2 \tag{3}$$

N is the number of the units. L-1 is the number of the delay elements. T is a hysteresis threshold level.

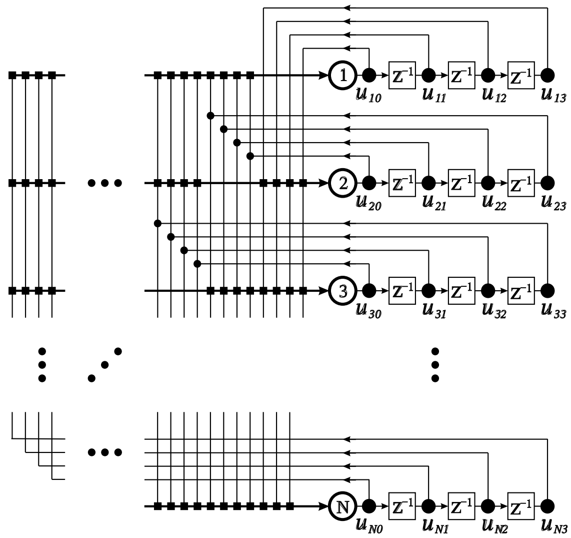

Fig. 1: Structure of the DRNN

3. Limit Cycles

In this paper, we impose the following conditions on the basic patterns which are elements of LCs: Fifty percent of the units, randomly selected, take the state "1", and the other units "0". An example of an LC, whose period is 4, is shown in Fig.2.

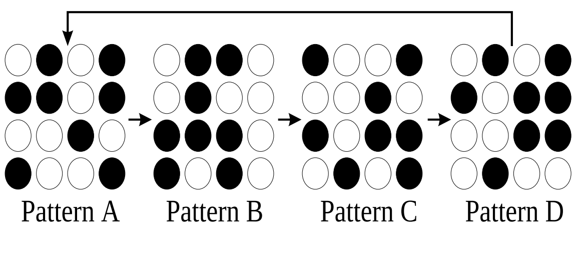

Fig. 2: Example of LC. (period is 4)

4. Improved Error Correction Learning

The improved error correction learning algorithm is employed for adjusting connection weights of the DRNN. Hysteresis threshold $\pm T$ and hysteresis margin dT were introduced in order to improve recall performance from noisy patterns and stabilize a learning process, respectively[1],[3]. This learning algorithm is modified for the DRNN as shown in the following.

I. Initialize connection weights.

$$w_{ijk}(0)=0 \ (1 \leq i,j \leq N, 0 \leq k \leq L-1) \tag{4}$$

II. Calculate network state.
The weighted sum input of all units are calculated after setting the network state to a certain phase of one of LCs to be memorized. Letting $p_{ik}(n)$ be the state, that is the output, of the kth delay element of the ith unit for the given basic pattern, the weighted sum input $v_j(n)$is given by

$$u_{ik}(n) = p_{ik}(n), \ k = 0, \cdots, L-1 \tag{5}$$

$$v_j(n) = \sum_{i=1}^{N}\sum_{k=0}^{L-1} w_{ijk}u_{ik}(n) \tag{6}$$

III. Calculate correction of weights.
When $p_{j0}(n) = 1$,
if $v_j(n) \geq T$, then

$$\Delta w_{ijk}(n) = 0 \tag{7}$$

if $v_j(n) < T$, then

$$\Delta w_{ijk}(n) = \frac{T + dT - v_j(n)}{\sum_{k=0}^{L-1} N_k - 1} p_{ik}(n) \tag{8}$$

When $p_{j0}(n) = 0$,
if $v_j(n) \leq -T$, then

$$\Delta w_{ijk}(n) = 0 \tag{9}$$

if $v_j(n) > -T$, then

$$\Delta w_{ijk}(n) = \frac{-T - dT - v_j(n)}{\sum_{k=0}^{L-1} N_k} p_{ik}(n) \tag{10}$$

dT is a margin for the threshold level, which can avoid unstable behavior in an error correction learning process and can accelerate the learning speed[1],[3].

IV. Update connection weights.
Using $\Delta w_{ijk}(n)$ calculated in III, all connection weights are adjusted at the same time by

$$w_{ijk}(n+1) = w_{ijk}(n) + \Delta w_{ijk}(n) \tag{11}$$

V. Shift LC phase.
Phase of LC in II is shifted by one step.

VI. II-V are repeated until $\Delta w_{ijk}(n) = 0$ for all phases of the LC.

VII. II-VI are executed for the next LC.

VIII. II-VII are repeated until $\Delta w_{ijk}(n) = 0$ for all LCs to be memorized.

After convergence, the network always satisfies the followings.
if $p_{j0} = 1$ then

$$v_j \geq T \tag{12}$$

if $p_{j0} = 0$ then

$$v_j \leq -T \tag{13}$$

5. Recalling from Incomplete Patterns

In order to recall the memorized LCs, the method proposed in [1],[3], which uses variable hysteresis threshold, is applied. The original one is modified to recall the LCs instead of the patterns stored at the equilibrium point. The algorithm for recalling the LCs from incomplete initial patterns is shown in the following.

I. The initial LC patterns are set to all units and their delay elements.

II. The weighted sum input $v_j(n)$ of all units are calculated by Eq.(1).

III. A variable hysteresis threshold $T(n)$ is chosen to be the maximum value of $|v_j(n)|$ of the unit which satisfies either ($v_j(n) > 0$ and $u_{j,L-1}(n) = 0$) or ($v_j(n) < 0$ and $u_{j,L-1}(n) = 1$). Hence, the output $u_{j0}(n+1)$ of the unit, which satisfies the above and its $|v_j(n)|$ is maximum, is replaced by the reverse of the last delay element output of the previous time step, denoted by $\overline{u_{j,L-1}(n)}$, as follows:

$$u_{j0}(n+1) = \overline{u_{j,L-1}(n)} \tag{14}$$

The bar indicates reverse operation from 1 to 0, and vice versa. The other units follow the following state transition. Namely, the output of the jth unit is changed to the output of the last delay element at one time step before.

$$u_{j0}(n+1) = u_{j,L-1}(n) \tag{15}$$

IV. State of delay elements are shifted by one step.

$$u_{j,k+1}(n+1) = u_{jk}(n) \ , \ k = 0, 1, \cdots, L-2 \tag{16}$$

V. II-IV are repeated until the network settles down to some stable state, which may include the memorized LCs and spurious LCs.

6. Computer Simulation and Discussions

6.1. Generation of LCs

In the following simulation, the number of units is 16, the number of delay elements connected to a unit is 3, and $\eta = dT/T = 0.2$. When the number of all basic patterns is m, and a period of LCs is l, the full number of LCs is given by

$$_{m}C_{l} \times (l-1)! = \frac{m!}{(m-l)! \times l} \tag{17}$$

A set of LCs with m=8 and l=4 include the following LCs, which have the common basic pattern B.

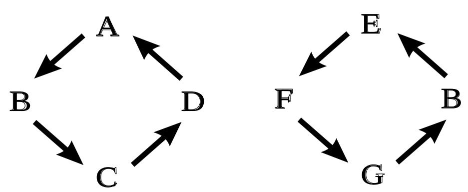

Fig. 3: Examples of LC, which include a common pattern B.

6.2. Capability of Learning a Set of LCs

Capability of learning a set of LCs is investigated. Especially, a combination of periods of the LCs in the same set is taken into account. Table 1 shows simulation results. In this table, "l=4,3,2,1" means a set of LCs includes period of 4,3,2,1. "X" means the learning didn't converge within 1000 iterations. The number in () means the number of LCs used in this learning.

Table 1: The number of iterations in learning a set of LCs, in which several periods are combined

m	4		6	8	10
l=4,3,2,1	25(4)	X(24)	X(151)	X(568)	X(1555)
l=4,3	40(4)	X(14)	X(130)	X(532)	X(1500)
l=4,2,1	-	7(16)	4(111)	3(456)	2(1315)
l=4	-	2(6)	28(90)	13(429)	7(1260)

From Table 1, there are two kinds of convergence properties. One of them includes l=4,3,2,1 and l=4,3, and the other l=4,2,1 and l=4. In the former case, called "Case 1" here, the learning didn't converge except for the number of LCs is 4. On the other hand, in the latter case, called "Case 2", the learning always converged with a small number of iterations.

A main difference between them is relations among the periods. In Case 1, l=3 is not a factor of the maximum period l=4. In this case, directions of adjusting connection weights for LCs with periods of 4 and 3 conflict with each other. Thus, by increasing the number of LCs, this confliction is also increased, and finally the learning becomes very difficult. This phenomenon will be further explained in the following.

In Case 2, directions of learning for all LCs are the same. Because, the smaller periods than the maximum one can be equivalently the same as the maximum. For instance, an LC with a period of 2, $\rightarrow(A\rightarrow B)\rightarrow(A\rightarrow B)\rightarrow$, can be regarded as an LC with a period of 4, $\rightarrow(A \rightarrow B \rightarrow A \rightarrow B)\rightarrow$. For this reason, the learning can converge toward the same direction for all LCs, resulting in stable and fast convergence.

In Case 2, the following relation is always valid.

$$u_{i0}(n+1) = u_{i3}(n) \tag{18}$$

Therefore, the role of connection weights w_{ijk} is divided into the following two categories.

w_{ij3} : Connection weights concerned in auto-association.
$w_{ij0}, w_{ij1}, w_{ij2}$: Connection weights concerned in hysteresis of an LC and cross-correlation among the basic patterns.

In Case 1, however, Eq.(18) cannot be always held. Hence, all w_{ijk} are concerned in the hysteresis and the cross-correlation.

In order to confirm the above analysis, variances and averages of connection weights are investigated, and are shown in Table 2.

In Case 2, it is possible to distinguish a role of w_{ij3} from that of the others as described above. The LC association follows the auto-association by w_{ij3}. The other weights, however, hold important information, that is the hysteresis and the cross-correlation information. Their significant effects on the LC association will be shown in the later section.

In Case 1, however, the above distinctive role cannot be recognized. Especially, the auto-association by w_{ij3} may be very weak. For example, let consider LCs with periods of 4 and 3, such as $\rightarrow(A\rightarrow B\rightarrow C\rightarrow D)\rightarrow$ and $\rightarrow(A\rightarrow B\rightarrow C)\rightarrow$, respectively. The following associations through w_{ij3} occur: Period of 4;$(A\rightarrow A)$, $(B\rightarrow B)$, $(C\rightarrow C)$, $(D\rightarrow D)$, and period of 3; $(A\rightarrow B)$, $(B\rightarrow C)$, $(C\rightarrow A)$. Thus, different patterns should be recalled from the same pattern. This is a reason why w_{ij3} cannot be grown, at the same time, the other weights cannot extract the necessary information, as shown in Case 2.

6.3. Recall Performance from Noisy Patterns

6.3.1. Uniformly Distributed Noise

Association rates are evaluated through computer simulation when the initial 4 patterns $u_{i0}(0)$, $u_{i1}(0)$, $u_{i2}(0)$, $u_{i3}(0)$ are changed from the memorized patterns by adding noises. Adding noise to a unit reverses its state. The combination of the periods of the LCs follows Case 1. The relation between the number of LCs, the number of noises and association rates are shown in Fig.4. The number of all basic patterns is fixed to 12. In this figure, $k/16$ means a single pattern includes k noises.

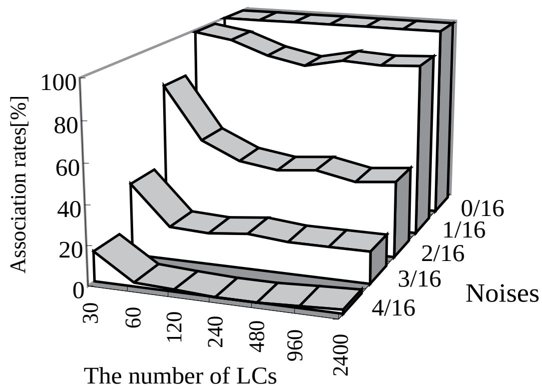

Fig. 4: Relation between the number of LCs, the number of noises and association rates. Noises are uniformly distributed.

Figure 4 shows the following typical properties.
(1)When the number of LCs is fixed, the association rate is inversely proportional to the amount of noises.
(2)When the amount of noises is fixed, the association rate is inversely proportional to the number of LCs until around 100. After that it is almost saturated.

The latter property (2) can be explained as follows: In the first interval, the hysteresis information represented by $w_{ijk}, k=0\sim L-2$, except for $w_{ij,L-1}$, may be weakened by increasing the number of LCs. For example, let consider two LCs, $\rightarrow(A\rightarrow B\rightarrow C\rightarrow D)\rightarrow$ and $\rightarrow(A\rightarrow B\rightarrow C\rightarrow E)\rightarrow$. In this case, the hysteresis information help to recall both D and E from the same pattern sequence $\rightarrow(A\rightarrow B\rightarrow C)\rightarrow$. The association of different patterns from the same patterns cannot grow the connection weights $w_{ij0}\sim w_{ij,L-2}$. However, this property does not continue until the maximum number of LCs. The association rate saturates after around 100 LCs. This means another information may be extracted by these weights. That is, deterministic information come from a small number of the basic patterns compared with that of LCs. Probability of taking the state "1" or "0" is not al-

Table 2: Variances and averages of connection weights

periods	4,3,2,1		4,3		4,2,1		4	
m	4		4		10		10	
the number of LCs	4		4		1315		1260	
	var.	ave.	var.	ave.	var.	ave.	var.	ave.
w_{ij3}	369	-1.19	697	-2.32	4290	-11.7	4733	-13.7
w_{ij2}	234	0.06	151	-0.80	17	3.92	54	4.40
w_{ij1}	203	0.16	387	1.98	14	3.32	51	4.39
w_{ij0}	449	1.83	1077	2.87	21	4.10	57	4.51

ways fifty percent. Some units may have high probability for either "1" or "0". This deviation of the unit states from the uniform distribution can help the LC association. This effect can be regarded as cross-correlation among the basic patterns. Effects of the hysteresis information will be further investigated in Sec.6.4.

6.3.2. Non-uniformly Distributed Noise

The number of noises is fixed to 4 in all cases. The noises are distributed non-uniformly. Simulation results are shown in Fig.5. For example, the noise pattern (0,0,1,3) means that $u_{i0}(0)$ and $u_{i1}(0)$ do not include any noise, while $u_{i2}(0)$ and $u_{i3}(0)$ contain one noise and three noises, respectively.

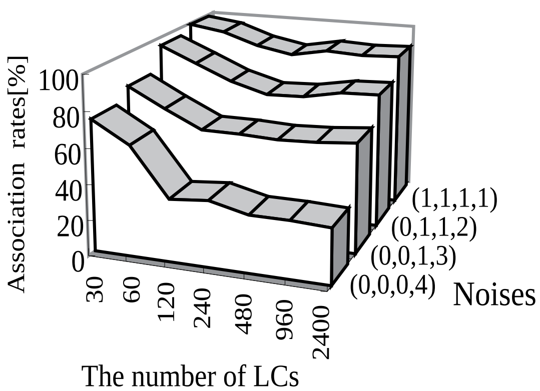

Fig. 5: Relation between the number of LCs, the number of noises and association rates. Noises are non-uniformly distributed.

Figure 5 leads to the followings:
(1) When the total amount of noises included in initial patterns is fixed, the larger the deviation to a specific pattern is, the lower association rates is.
(2) Even when the noise distribution is concentrated to a specific pattern, association rates become almost constant after about 100 LCs.
(3)When noises included in initial patterns are 0,0,0,4 and the number of LCs is greater than about 100, the association rates are 32%.

6.4. Effects of Hysteresis Information

Association rates are evaluated through computer simulation when the hysteresis information represented by $w_{ijk}, k = 0 \sim L-2$, except for $w_{ij,L-1}$ are reduced after convergence. The relation between the reduction rates, the number of noises and association rates when the number of LCs is 2400 are shown in Fig.6.

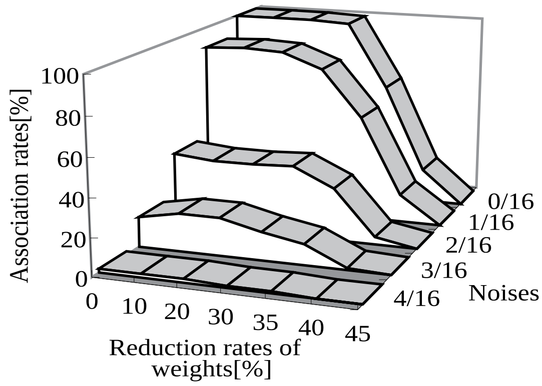

Fig. 6: Relation between the reduction rates, the number of noises and association rates.

As discribed in Sec.6.3, the hysteresis information represented by $w_{ijk}, k = 0 \sim L-2$, except for $w_{ij,L-1}$, is weakened by increasing the number of LCs. In fact, the variances of $w_{ijk}, k = 0 \sim L-2$ are about 1/100 of the variance of $w_{ij,L-1}$ when the number of LCs is large. But Fig.6 shows that the hysteresis information represented by $w_{ijk}, k = 0 \sim L-2$ plays an important role.

Next, effect of the cross-correlation among all the basic patterns is investigated. The relation between the number of basic patterns, the number of noises and association rates when the number of LCs is 420, which is the maximum when the number of basic patterns is 8 as shown in Eq.(17), are shown in Table 3.

This table shows that when the numbers of LCs are the same, the smaller the number of basic patterns is, the higher the association rates is. Because the cross-correlation information may be weakened by increasing the number of the basic patterns, which

Table 3: The relation between the number of basic patterns, the number of noises and association rates

The number of basic patterns	12	8
0/16 noise	100	100
1/16 noise	85	100
2/16 noise	43	84
3/16 noise	9	44
4/16 noise	3	13

are generated randomly. So generally speaking, when DRNN is used as an associative memory, the number of basic patterns should be selected as small as possible.

6.5. Comparison between the RNN and the DRNN

Efficiency of $w_{ijk}(k = 0 \sim 2)$ is evaluated in this section. If $w_{ijk}(k = 0 \sim 2)$ don't effect at all, properties of the DRNN must be equivalent to those of independent 4 RNNs, in which no delay element is included, and patterns are stored at equilibrium points. Simulation results concerning the noise performance under the following conditions are shown in Table 4.

Table 4: Recall performance of RNN associative memory with hysteresis threshold

Noises	0/16	1/16	2/16	3/16	4/16
Association rates[%]	100	95	79	34	20

In the case of association by the DRNN, when all basic patterns of an LC of interest are associated correctly, this LC is considered as exactly recalled. Therefore, to compare the DRNN in Fig.4 and the RNN in Table 4, association rates of the RNN must be raised to the 4th power. When 4 initial patterns include 1/16 noises, the association rate is 89% by the DRNN. On the other hand, the association rate of the RNN is 95%, and the 4th power of that is 81%, which is lower than that of the DRNN.

When noises are included in only one pattern, it is not necessary to use the 4th power of that. In the case of the DRNN, when 4/16 noise are included in only one pattern, the association rate is 34% as shown in Fig.5 , which is higher than 20% by the RNN with 4/16 noise as shown in Table 4.

From these results, it can be concluded that $w_{ijk}(k = 0 \sim 2)$ are effective. Even when the number of LCs is large enough and w_{ijk} do not include information of the order of basic patterns enough, they can hold information of cross-correlation among a set of the basic patterns, which can increase the association rates.

7. Conclusion

In this paper, we have proposed an associative memory of LCs by a recurrent neural network whose units have serial delay elements(DRNN). It can be concluded that, first, the DRNN can memorize many LCs which have common basic patterns. Next, the DRNN memory has higher association rates for the noisy LCs. Finally, efficiency of the feedback weights from the delay elements has been also confirmed.

References

[1] K. Nakayama and N. Mitsutani, "An adaptive hysteresis threshold method for an associative memory using mutually connected neural network," *Proc. IJCNN'91, Seattle,* Vol.II, p.A-914, June 1991.

[2] K. Nakayama, K. Nishimura and H. Katayama, "Memory capacity bound and threshold optimization in recurrent neural network with variable hysteresis threshold," *Proc. IJCNN'93, Nagoya,* pp. 2603-2606, Oct. 1993.

[3] K. Nakayama and K. Nishimura, "A delta rule algorithm using double hysteresis thresholds for recurrent associative memory," *Proc. ICNN'94, Florida,* pp. 1163-1168, June 1994.

[4] T. Kojima, H. Kita and Y. Nishikawa, "Learning of a limit cycle by a recurrent neural network (in Japanese)," *Proc.of the 36th Annual Conference of the Systems,Control and Information Engineers, ISCIE Japan,* pp. 149-150, May 1992.

[5] T. Kumagai,R. Hashimoto and M. Wada, "Learning of limit cycles in discrete-time neural network with time delay (in Japanese)," *Proc. FAN symposium of The Japan Society of Mechanical Engineers,* Vol.3 pp. 121-124, 1993.

[6] K. Nowara and T. Saito, "Storing limit cycles into a discrete asynchronous neural network," *Technical Report of IEICE Japan,* NLP91-77, 1992.

[7] E. A. Wan, "Temporal backpropagation for FIR neural networks," *Proc. Int. Joint Conf. Neural Networks, San Diego,* pp. **I**575–580, June 1990.

[8] Simon Haykin, "Neural Networks –A Comprehensive Foundation–," *Macmillan College Publishing Company,* 1994.

[9] P. Baldi and A.F. Atiya, "How delay affect neural dynamics and learning," *IEEE Trans.Neural Networks,* Vol.5, no.4, pp. 612-621, July 1994.