論文 96-21-2-24

An Adaptive Time-Delay Recurrent Neural Network for Temporal Learning and Prediction

Sung Suk Kim* *Regular Member*

시계열패턴의 학습과 예측을 위한 적응 시간지연 회귀 신경회로망

正會員 김 성 석*

ABSTRACT

This paper presents an Adaptive Time-Delay Recurrent Neural Network (ATRN) for learning and recognition of temporal correlations of temporal patterns. The ATRN employs adaptive time-delays and recurrent connections, which are inspired from neurobiology. In the ATRN, the adaptive time-delays make the ATRN choose the optimal values of time-delays for the temporal location of the important information in the input patterns, and the recurrent connections enable the network to encode and integrate temporal information of sequences which have arbitrary interval time and arbitrary length of temporal context. The ATRN described in this paper, ATNN proposed by Lin, and TDNN introduced by Waibel were simulated and applied to the chaotic time series preditcion of Mackey-Glass delay-differential equation. The simulation results show that the normalized mean square error (NMSE) of ATRN is 0.0026, while the NMSE values of ATNN and TDNN are 0.0114, 0.0117, respectively, and in temporal learning, employing recurrent links in the network is more effective than putting multiple time-delays into the neurons. The best performance is attained by the ATRN. This ATRN will be well applicable for temporally continuous domains, such as speech recognition, moving object recognition, motor control, and time-series prediction.

요 약

본 논문에서는 시계열패턴의 학습과 예측을 위한 적응 시간지연 회귀 신경회로망(Adaptive Time-Delay Recurrent Neural Network : ATRN)을 제안한다. ATRN은 적응 시간지연 요소와 회귀연결 요소를 가지며, 신경생물학에서 밝혀진 증거를 바탕으로 모델링되었다. ATRN은 두가지의 구조적인 특징, 즉 가변시간지연과 회귀연결로 구분된다. 가변시간지연은 입력패턴의 중요한 정보인 시계열 위치에 대하여 회로망의 시간지연 값이 최적화되도

*용인대학교 전산통계학과
論文番號 : 95161-0427
接受日字 : 1995년 4월 27일

록 함으로써 시계열적 정보를 효과적으로 기록하도록 하고, 회귀연결은 시계열패턴의 문맥정보가 임의의 길이로 넓게 퍼져 있더라도 이를 잘 표현할 수 있도록 한다. 시계열패턴의 예측 실험은 본 논문의 ATRN, 동적 신경회로망 가운데 가장 뛰어난 것으로 알려진 ATNN(Adaptive Time-Delay Neural Network)과 TDNN(Time-Delay Neural Network)을 카오스 시계열패턴에 대하여 각각 수행하였다. 실험결과에서 ATRN이 ATNN과 TDNN에 비하여 회로망의 규모가 적음에도 불구하고 정규화 분산 예측 오차가 각각 0.0026, 0.0114, 0.0117로 나타났으며, 또한 시계열패턴의 학습과 예측에는 회귀연결이 다중 시간지연보다 더욱더 효과적임을 알 수 있었다. 이러한 ATRN은 음성, 동영상 등과 같은 시간적으로 변화하는 신호의 인식과 예측에 잘 적용될 수 있을 것으로 예상된다.

I. Introduction

Spatiotemporal pattern recognition and temporal sequence learning are challenging tasks in neural network researches. Learning and recognition of temporal correlations are crucial in hearing and vision, motor control, prediction, and other time-dependency applications. The key issue in learning temporal sequences is that there needs to be some means of recognizing and storing temporal natures of sequences. There are multiple ways representing temporal information in the neural network [1]-[6]. These include; (1) creating a spatial representation of temporal pattern, (2) putting time-delays into the neurons or their connections, (3) employing recurrent connections, (4) using neurons with activations summing inputs over time, and (5) using combination of the above. The simplest, and the most common, is to spatially encode the temporal information, as can be done by using Fourier transform preprocessing. Time-delay networks have shown great promise in such applications as speech recognition, and are highly competitive with existing technologies. However, both spatial coding and time-delay approaches have some inherent problems with scaling, in terms of how much(or how long) of a temporal pattern can be encoded. Recurrent networks are a useful and interesting possibility, and have been applied to speech recognition and robotic control tasks [7]. This approach also has problems with stability, and requires a great deal of training time.

The time-delay neural network (TDNN) proposed by Waibel has been successfully applied to phoneme recognition and trajectory recognition [8][9]. However, a limitation of the TDNN is its inability to learn or adapt the values of the time-delays. Time-delays are fixed initially and remain the same throughout training. As a result, the system may have poor performance due to the inflexibility of time-delays, and a mis-match between the choice of time-delay values and the temporal location of the important information in the input patterns. To overcome this limitation, Lin proposed the adaptive time-delay neural network (ATNN), which adapts time-delays as well as synaptic weights during training, to better accommodate to temporally changing patterns and to provide more flexibility for optimization tasks [10]. The time-delay networks such as TDNN and ATNN can not recognize temporal patterns which have arbitrary interval times and arbitrary lengths of temporal context, and these networks can not integrate temporal information explicitly.

The recurrent network supposes that all temporal patterns are to be recognized and/or stored, even if the patterns include redundancy. We present an Adaptive Time-Delay Recurrent Neural Network (ATRN) for learning and recognition of temporal correlations of temporal patterns, which meets the above-mentioned issues. The ATRN is inspired from neurobiology: time-delays do occur along axons due to different conduction time and different lengths of axonal fibers, and temporal properties such as temporal decay and integration occur at synapses, and in addition, within areas of the cortex there is a great deal of feedback between the cortical strata. The proposed ATRN has been applied to the chaotic time-series prediction of Mackey-Glass delay-differential equation. And the comparision of the performance of ATRN, ATNN and TDNN is made.

Ⅱ. Adaptive Time-Delay Recurrent Neural Network

The network employs adaptive time-delays and recurrent connections for processing temporal information. The architecture of the ATRN is shown in Figure 1. The network mainly consists of three-layer perceptron, and also has an internal state layer. The activations of the hidden units at time τ-1 are copied into the internal state units, which integrate the internal state of the system and act as the additional inputs at time $\tau+D$. The ATRN employs modifiable weights and time-delays along interconnections between two processing units, and both time-delays and weights are adjusted. However, the feedback connections from the hidden unit to the internal state unit are not subject to training. The configuration of n interconnections, each with its own delay, from the input unit to the hidden unit is called a Delay Box which is depicted in Figure 2. The interconnections between the internal state unit and the hidden unit, and from the hidden

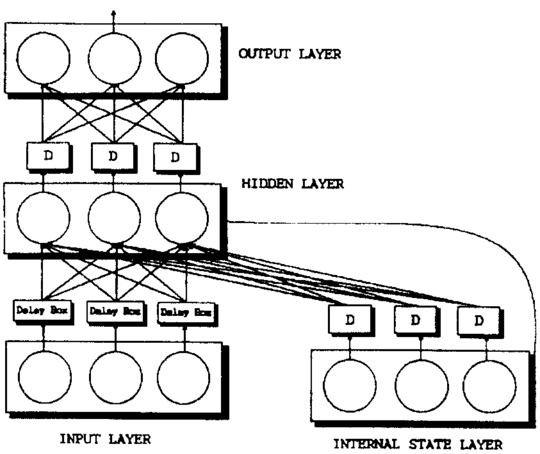


Fig 1. The architecture of the ATRN.

unit to the output unit have modifiable weights and only one time-delay which is depicted as D in Figure 1. Node i of layer h-1 is connected to node j of the next layer h, with the connection line having an independent time-delay $\tau_{jik,h-1}$ and synaptic weights $w_{jik,h-1}$.

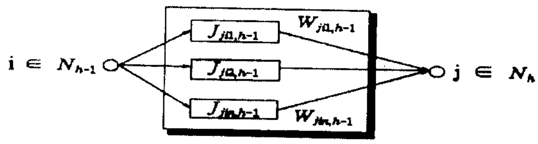


Fig. 2. Delay Box.

Each node sums up the net inputs from the activation values of the previous neurons, through the corresponding time-delays on each connection line, i. e., at time t_n unit j on the layer h receives a weighted sum:

$$S_{j,h}(t_n)=\sum_{i\in N_{h-1}}\sum_{k=1}^{K_{ji,h-1}} w_{jik,h-1}\, a_{i,h-1}(t_n-\tau_{jik,h-1}) \qquad (1)$$

where $a_{i,h-1}(t_n-\tau_{jik,h-1})$ is the activation level of unit i on the layer h-1 at time $t_n-\tau_{jik,h-1}$, N_{h-1} denotes the set of nodes of layer h-1, and $K_{ji,h-1}$ represents the total number of connections to node j(layer h) from node i of layer h-1. Then the output of node j is governed by a nondecreasing differential function f of the net input (sigmoidal function is selected in this paper):

$$a_{j,h}(t_n)=\begin{cases} f_{j,h}(S_{j,h}(t_n)) & \text{if } h\ge 2 \\ a_{j,0}(t_n) & \text{if } h=1 \end{cases} \qquad (2)$$

where

$$f_{j,h}(x)=\frac{\beta_{j,h}}{1+e^{-\alpha_{j,h}x}}-\gamma_{j,h} \qquad (3)$$

and $a_{j,0}(t_n)$ denotes the jth channel of the input signal at time t_n and $\alpha_{j,h}$, $\beta_{j,h}$ and $\gamma_{j,h}$ are real numbers which define the upper bound of sigmoidal funtcion, $-\gamma_{j,h}$ and the lower bound, $\beta_{j,h}-\gamma_{j,h}$, and the steepness of $f_{j,h}(x)$, $f'_{j,h}(0)=(\alpha_{j,h}\beta_{j,h})/4$.

An instantaneous error measure is defined as MSE:

$$E(t_n)=\frac{1}{2}\sum_{j\in N_L}(d_j(t_n)-a_{j,L}(t_n))^2 \quad (4)$$

where L denotes the output layer and $d_j(t_n)$ indicates the desired(target) value of output node j at time t_n.

The weights and time-delays are updated by an amount proportional to the opposite direction of the error gradient respectively:

$$\Delta w_{jik,h}=-\eta_1\frac{\partial E(t_n)}{\partial w_{jik,h}} \quad (5)$$

$$\Delta \tau_{jik,h}=-\eta_2\frac{\partial E(t_n)}{\partial \tau_{jik,h}} \quad (6)$$

where η_1 and η_2 are the learning rate. The derivation of this learning algorithm was addressed explictly in [10]. The learning rules are summerized as follows:

$$\Delta w_{jik,h-1}=\eta_1\delta_{j,h}(t_n)a_{i,h-1}(t_n-\tau_{jik,h-1}) \quad (7)$$

$$\Delta \tau_{jik,h-1}=\eta_2\rho_{j,h}(t_n)w_{jik,h-1}a'_{i,h-1}(t_n-\tau_{jik,h-1}) \quad (8)$$

where

$$\delta_{j,h}(t_n)=\begin{cases}(d_j(t_n)-a_{j,h}(t_n))f'(S_{j,h}(t_n)), \text{ if } j \text{ is an output unit.}\\ (\sum_{p\in N_{h+1}}\sum_{q=1}^{K_{pj,h}}\delta_{p,h+1}(t_n)w_{pjq,h}(t_n))f'(S_{j,h}(t_n)), \quad (9)\\ \text{if } j \text{ is a hidden unit.}\end{cases}$$

and

$$\rho_{i,h}(t_n)=\begin{cases}-(d_j(t_n)-a_{j,h}(t_n))f'(S_{j,h}(t_n)), \text{ if } j \text{ is an output unit.}\\ (-\sum_{p\in N_{h+1}}\sum_{q=1}^{K_{pj,h}}\rho_{p,h+1}(t_n)w_{pjq,h}(t_n))f'(S_{j,h}(t_n)), \quad (10)\\ \text{if } j \text{ is a hidden unit.}\end{cases}$$

Ⅲ. Relation to ATNN and TDNN

The ATRN is indeed a generalization of the Adaptive Time-Delay Neural Network (ATNN) and Time-Delay Neural Network(TDNN) if the values of certain parameters of the ATRN are adjusted.

Case 1: If we set $K_{js,h-1}=0$(where $K_{js,h-1}$ is the number of connections to the hidden unit j from the internal state unit s) and $K_{kj,h}=n>1$ (where $K_{kj,h}$ is the number of connections to the output unit k from the hidden unit j), the ATRN becomes a typical Adaptive Time-Delay Neural Network.

Case 2: In addition to Case 1, if we fixed the time-delay $\tau_{jik,h-1}$ in Equation (1) and applied weight learning without updating time-delay variables, it becomes a model of Time-Delay Neural Network.

Ⅳ. Chaotic Time Series Prediction

We have carried out the chaotic time-series prediction of the Mackey-Glass delay-differential equation:

$$\frac{dx(t)}{dt}=-bx(t)+a\frac{x(t-\tau)}{1+x(t-\tau)^{10}} \quad (11)$$

This differential equation possesses many dynamic properties such as nonlinearity, limit cycle oscillations, aperiodic wave forms and other dynamic behaviors [11], and provides a useful benchmark for temporal learning. We chose $\tau=17$, $a=0.2$, $b=0.1$, and integrated (11) using a four-point Runge-Kutta method with a step size of $t=1$ up to $t=1000$.

To evaluate performances, the ATRN described in this paper, ATNN and TDNN were simulated. These networks were trained to predict values 6 time steps ahead by using the same training samples and the same set of initial weights. The training set consisted of the first three hundred samples (from $t=1$ to $t=300$), and the testing set included a thousand samples of the differential equation (from $t=1$ to $t=1000$). The ATRN, ATNN and TDNN were configured with 1 input unit, 3 hidden units and 1 output unit, and the connection parameters of the simulated networks

Table 1. The connection parameters of ATRN, ATNN, and TDNN

Network Topology	$K_{ji,h-1}$	$K_{js,h-1}$	$K_{kj,h}$	Total number of weights
ATRN	4	1	1	24
ATNN	4	X	8	36
TDNN	4	X	8	36

※X: None

were summerized in Table 1, where $K_{ji,h-1}$, $K_{js,h-1}$ represent the total number of connections (or time-delays) to the hidden node j from the input node i and from the internal state node s, respectively, and $K_{kj,h}$ is the number of connections to the output node k from the hidden node j.

We used the normalized mean square error (NMSE) to monitor prediction performance. The NMSE is defined as the following:

$$NMSE = \frac{E[(x(t)-\hat{x}(t))^2]}{E[x^2(t)]} \tag{12}$$

where $x(t)$ is the original signal and $\hat{x}(t)$ is the network prediction value. Figure 3 shows the NMSE values of the networks during learning. The NMSE values of the ATRN, ATNN and TDNN were 0.0026, 0.0114, and 0.0117, respectively.

The prediction outputs of the ATRN, ATNN, and TDNN are shown in Figure 4, Figure 5, and Figure 6, respectively. As we can observe from these figures, the ATRN catches the variations of the bumping peaks better than the ATNN and TDNN do, and the best performance is attained by the ATRN. These results indicate that employing recurrent links in the network is more effective than putting multiple time-delays into the neurons for temporal learning and prediction.

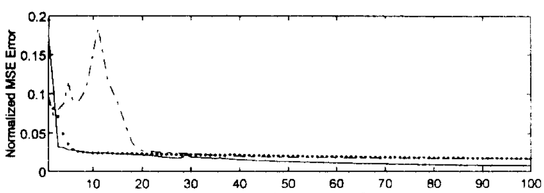


Fig. 3. Learning curves showing the normalized mean square error (NMSE) of networks trained to predict the Mackey-Glass signal an interval $t=6$, into future. The solid curve, point curve, and dashdot curve are for the ATRN, ATNN and TDNN, respectively.

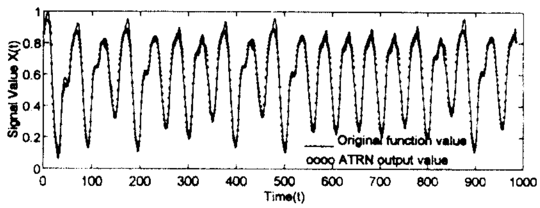


Fig. 4. The simulation results of ATRN (NMSE = 0.0026).

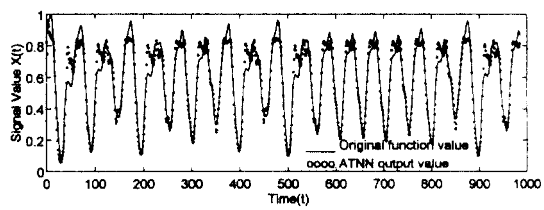


Fig. 5. The simulation results of ATNN (NMSE = 0.0114).

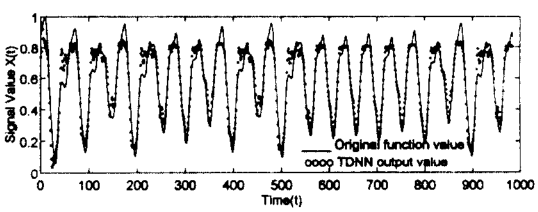


Fig. 6. The simulation results of TDNN (NMSE = 0.0117).

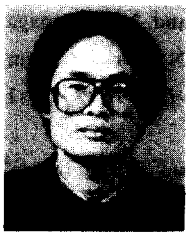
V. Conclusions

We have presented an Adaptive Time-Delay Recurrent Neueal Network(ATRN) for learning and recognition of temporal correlations of temporally changing signals. The ATRN employs adaptive time-delays and recurrent connections. The adaptive time-delays make the ATRN choose the optimal values of time-delays for the temporal location of the important information in the temporal patterns, and the recurrent connections enable the network to encode and integrate temporal information of sequences which have arbitrary interval time and arbitrary length of temporal context.

The ATRN described in this paper, ATNN proposed by Lin, and TDNN introduced by Waibel were simulated and applied to the chaotic time-series preditcion of Mackey-Glass delay-differential equation. The simulation results show that the NMSE values of ATRN, ATNN and TDNN were 0.0026, 0.0114, and 0.0117, respectively, and employing recurrent links in the network is more effective than putting multiple time-delays into the neurons for temporal learning and prediction. The best performance was attained by the ATRN. This ATRN will be well applicable for speech recognition, moving object recognition, motor control, and time-series prediction.

References

1. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme Recognition Using Time-Delay Neural Networks," *IEEE trans. on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 328-339, March 1989.
2. R. J. Williams and D. Zipser, "A Learning Algorithm for Continually Running Fully Recurrent Neural Networks," *Neural Computation*, vol. 1, no. 2, pp. 270-280, 1989.
3. T. Uchiyama, K. Shimohara, and Y. Tokunaga, "A Modified Leaky Integrator Network for Temporal Pattern Processing," *Proc. IJCNN'89*, vol. 1, pp. 469-475, June 1989.
4. U. Bodenhausen and A. Waibel, "The tempo2 algorithm: Adjusting time-delays by supervised learning," *in Advances in Neural Information Processing Systems 3*, R. P. Lippmann, J. E. Moody, and D. S. Touretzky, Eds. Denver 1991, vol. 3, pp. 155-161, Morgan Kaufmann, San Mateo, 1991.
5. S. P. Day and M. R. Davenport, "Continuous-Time Temporal Back-Propagation with Adaptive Time Delays," *IEEE trans. on Neural Networks*, vol. 4, no. 2, pp. 348-354, March 1993.
6. Sung Suk Kim et al., "Learning Sequential Structures in Partially Connected Recurrent Neural Network," *KITE Journal of Electronics Engineering*. vol. 2, no. 1, pp. 109-116, June 1991.
7. Sung Suk Kim, "Temporal Processing of Robot Inverse Kinematics Using Recurrent Neural Network," *Proc. of IJCNN'92*, vol. 1, pp. 172-179, 1992.
8. A. Waibel "Modular Construction of Time-Delay Neural Networks for Speech Recognition" *Neural Computation*, vol. 1, no. 2, pp. 39-46, 1989.
9. D. T. Lin, J. E. Dayhoff, and P. A. Ligonenides, "Trajectory Recognition with a Time-Delay Neural Network," *Proc. of IJCNN'92*, vol. 3, pp. 197-202, 1992.
10. D. T. Lin, J. E. Dayhoff, and P. A. Ligonenides, "Adaptive Time-Delay Neural Network for Temporal Correlations and Prediction," *in SPIE Intelligent Robots and Computer Vision XI: Biological, Neural Net, and 3-D Method*, vol. 1826, pp. 170-181, Boston, 1992.
11. D. Farmer, "Predicting Chaotic Time Series," *Physical Review Letters*, 59:845-848, 1987.



Sung-Suk Kim

Sung-Suk Kim was born in Kyungju, Korea, on October 12, 1959. He received the B. S. degree in electrical engineering from Yeungnam University, Kyungsan, Korea, in 1985, and the M. S. and Ph. D. degree in electronics and computer engineering from the University of Ulsan, Ulsan, Korea, in 1987 and 1990, respectively. From 1985 to 1990, he joined the Korea Electric Power Corporation (KEPCO). From 1991 to 1994, he was with the department of computer information processing, Sangji Junior College, Andong, Korea, where he was an assistant professor and a director of computing center. Since 1995, he has been with the department of computer science and statistics, Yongin University, Yongin, Korea. His research interests include neural networks, speech recognition and artifitial intelligence.