A New Elman Neural Network and Its Dynamic Properties

Yuan-Chu Cheng , Wei-Min Qi , Jie Zhao
College of Power and Mechanical engineering
Wuhan University
Wuhan, China
yccheng@whu.edu.cn

***Abstract*—Elman neural network was one of the dynamic recurrent neural networks. In this paper, a new modified Elman network was proposed first. Then the dynamic characteristics of Elman series neural network are fully discussed. The theory study proved that the new Elman network inherently has proportional (P), integral (I), derivative (D) properties and traditional Elman networks only have PI or I characteristics. Simulation results based on ideal mathematical model show that the PID Elman network is prior to the modified Elman network in identifying nonlinear dynamic system.**

***Keywords*—dynamic recurrent network, Elman neural network, PID properties, dynamic system identification**

I. INTRODUCTION

Elman neural network is a partial recurrent network model first proposed by Elman in 1990 [1]. It lies somewhere between a classic feed-forward perception and a pure recurrent network. The feed-forward loop consists of input layer, hidden layer and output layer in which the weights connecting two neighboring layers are variable. In contrast to the feed-forward loop, the back-forward loop employs context layer which is sensitive to the history of input data so the connections between context layer and hidden layer are fixed. Further more, because the dynamic characteristics of Elman network are provided only by internal connections, so it doesn't need to use the state as input or training signal, which makes Elman neural network prior to static feed-forward network and be used in dynamic system identification widely.

However, although the Elman neural network has found various applications in speech recognition and time series prediction, its training and converge speed are usually very slow and not suitable for some critical applications. Correlative study shows that the basic Elman network can only identify one order linear dynamic system with standard back-propagation learning algorithm. In order to improve the ability to identify the high-order system, some modified Elman neural network models have been proposed recently which proved to have some advantages to the basic Elman neural network [2,3,4].

In this paper, we aimed at the study of the dynamic characteristics of partial recursive neural networks, and proposed a novel modified Elman neural network based on the discussion of the dynamic properties, which inherently has proportional (P), integral (I), derivative (D) properties. In section 2, the original Elman network and some modified Elman network is reviewed, and the novel PID Elman neural network is illustrated in section 3. Then, the dynamic characteristics of Elman series neural network are fully discussed in section 4. In section 5, a case study about the simulation results of an ideal mathematical model and some comparisons about the dynamical characteristics are presented respectively. Finally, we give some comments on the Elman network and the modified Elman network in section 6.

II. THE TRADITIONAL MODIFIED DYNAMICAL RECURRENT ELMAN NEURAL NETWORK

The original Elman network and a traditional modified dynamical recurrent Elman neural network [4,5] are illustrated in Fig.1 and Fig.2 respectively. We can see clearly that the original Elman network is the simplification of the modified Elman network. It's obvious that the output of the context layer is used as an input of both the output layer and the hidden layer. It is easy to find that an auto feedback connection with fixed gain exists in the context unit and the output of context layer at k time equals to the output of hidden layer at k-1 time.

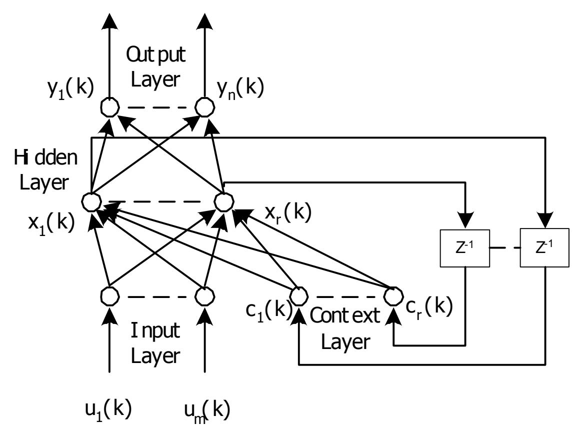

Figure 1. The structure of the Elman neural network.

The notations used in this section are as follows:

$wl_{i,j}$: The weights that connect node i in the context layer to node j in the hidden layer.

$w2_{i,j}$: The weights that connect node i in the input layer to node j in the hidden layer.

$w3_{i,j}$: The weights that connect node i in the hidden layer to node j in the output layer.

$w4_{i,j}$: The weights that connect node i in the context layer to node j in the output layer.

m, n, r: The number of nodes in the input, output, and hidden layers respectively.

$u_i(k)$, $y_j(k)$: Inputs and outputs of the Elman neural network, where i=1,2,⋯⋯m, j=1,2,⋯⋯n.

$x_i(k)$: Output of the node i in hidden layer, where i=1,2,⋯⋯r.

$c_i(k)$: Output of the node i in context layer which equals to the output of he hidden node i of last time. where i=1,2,⋯⋯r.

z^{-1} : A unit delay.

$f(\cdot)$, $g(\cdot)$: The linear or nonlinear output function of hidden and output layers respectively.

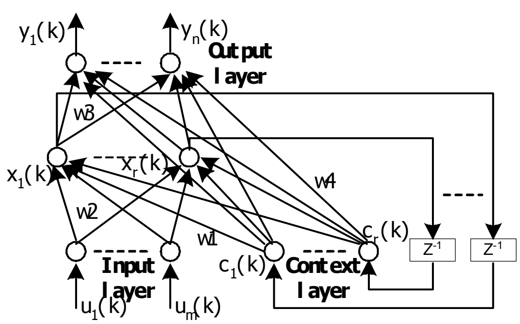

Figure 2. The structure of traditional modified Elman neural network.

The inputs and output of the network are: $u(k)\in R^m$, $y(k)\in R^n$, $x(k)\in R^r$, then the outputs in each layer can be given by (1)(2)(3).

$$x_j(k) = f(\sum_{i=1}^{m} w2_{i,j}u_i(k) + \sum_{i=1}^{r} w1_{i,j}c_i(k)) \tag{1}$$

$$c_i(k) = x_i(k-1) \tag{2}$$

$$y_j(k) = g(\sum_{i=1}^{r}(w3_{i,j}x_i(k) + w4_{i,j}c_i(k))) \tag{3}$$

III. A NOVEL DYNAMIC RECURRENT NEURAL ELMAN NETWORK

The traditional modified Elman neural network has been applied in modeling nonlinear system since its dynamic propertie is improved to a certain degree. To improve the characteristics of the Elman neural network, a new partial recurrent network model is proposed as in Fig.3.

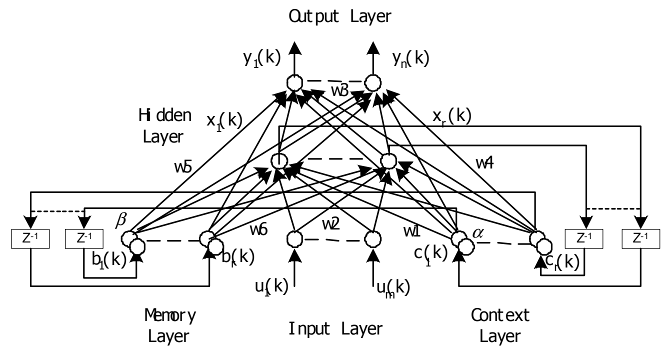

Figure 3. The structure of PID Elman neural network.

From the figures above we can find the difference between the two modified Elman networks is that an additional context layer named memory layer is added to the traditional modified Elman neural network. The output of the former context layer is used as the input of the memory layer through a unit delay and the output of it as the input of the hidden and output layers respectively. Before we discuss the novel Elman neural network, additional definition must be done as following:

α , β : The self-feedback coefficient of the context and memory layers respectively.

$w5_{i,j}$: The weights that connect node i in the memory layer to node j in the output layer.

$w6_{i,j}$: The weights that connect node i in the memory layer to node j in the hidden layer.

$b_i(k)$: Output of the node i in memory layer which equals to the output of he context node i of last time. where i=1,2,⋯⋯r.

The expression of nonlinear state space described by traditional modified Elman neural network is:

$$x_j(k) = f(\sum_{i=1}^{m} w2_{i,j}u_i(k) + \sum_{i=1}^{r} w1_{i,j}c_i(k) + \sum_{i=1}^{r} w1_{i,j}b_i(k)) \tag{4}$$

$$c_i(k) = x_i(k-1) + \alpha \times c_i(k-1) \tag{5}$$

$$b_i(k) = c_i(k-1) + \beta \times b_i(k-1) \tag{6}$$

$$y_j(k) = g(\sum_{i=1}^{r}(w3_{i,j}x_i(k) + w4_{i,j}c_i(k) + w5_{i,j}b_i(k)) \tag{7}$$

From systemic view, the new Elman network can utilize both the system static information and the dynamic information owing to its internal connection. It will further enforce the dynamic performance through the two context layers and dual-loop feedback control.

IV. DYNAMIC PROPERTIES OF ELMAN NETWORK AND MODIFIED ELMAN NETWORK

In order to analyze conveniently, we consider only a simplest modified Elman network with one hidden layer node, as shown in Fig. 4. Especially, to one input and one output system, when hidden unit and output unit use the linear function, and the threshold values of hidden layer and output layer are zero.

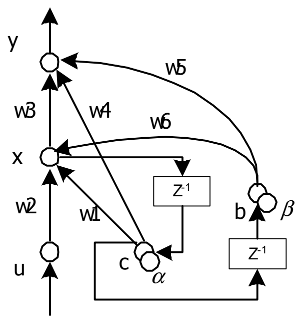

Figure 4. Simple modified Elman network with 1 output and 1 input

The equation of input and output relation is:

$$y(k) = w3\times x(k) + w4\times c(k) + w5\times b(k) \quad (8)$$

$$x(k) = w2\times u(k) + w1\times c(k) + w6\times b(k) \quad (9)$$

$$c(k) = x(k-1) + \alpha\times c(k-1) \quad (10)$$

$$b(k) = c(k-1) + \beta\times b(k-1) \quad (11)$$

Using z transform, we get:

$$y(z) = w3\times x(z) + w4\times c(z) + w5\times b(z) \quad (12)$$

$$x(z) = w2\times u(z) + w1\times c(z) + w6\times b(z) \quad (13)$$

$$zc(z) = x(z) + \alpha\times c(z) \quad (14)$$

$$zb(z) = c(z) + \beta\times b(z) \quad (15)$$

Then, we have：

$$c(z) = \frac{1}{z-\alpha}x(z) \quad (16)$$

$$b(z) = \frac{1}{z-\beta}c(z) = \frac{1}{(z-\beta)(z-\alpha)}x(z) \quad (17)$$

$$x(z) = \frac{w2(z-\alpha)(z-\beta)}{(z-\alpha)(z-\beta) - w1(z-\beta) - w6}u(z) \quad (18)$$

$$G(z) = \frac{y(z)}{u(z)} = \frac{w2\times w3(z-\alpha)(z-\beta) + w2\times w4(z-\beta) + w2\times w5}{(z-\alpha)(z-\beta) - w1(z-\beta) - w6} \quad (19)$$

Using reverse z transform to (19), we have：

$$\begin{aligned} y(k) &= (\alpha+\beta+w1)y(k-1) - (\alpha\beta+\beta w1 - w6)y(k-2) \\ &+ w2\times w3\times u(k) - w2[(\alpha+\beta)\times w3 - w4]u(k-1) \\ &+ w2(\alpha\beta w3 + w5 - w4\beta)u(k-2) \end{aligned} \quad (20)$$

Nowadays, the well-known dynamic characteristics are proportional (P), integral (I) and derivative (D) properties. The PID controllers have been used in industrial control.

The general PID control algorithm with practical derivative is described by (21).

$$y(s) = (K_P + K_I\frac{1}{s} + \frac{K_D s}{1+Ts})u(s) \quad (21)$$

Where, K_P is proportional gain, K_I is integral coefficient, K_D is derivative factor, and T is an inertia time constant.

The discrete formula of (21) is written as (22).

$$\begin{aligned} y(k) &= \frac{2T+\Delta T}{T+\Delta T}y(k-1) - \frac{T}{T+\Delta T}y(k-2) \\ &+ \frac{(K_P+K_I T)\Delta T + K_D + K_P T + K_I\Delta T^2}{T+\Delta T}u(k) \\ &- \frac{(K_P+K_I T)\Delta T + 2K_D + 2K_P T}{T+\Delta T}u(k-1) \\ &+ \frac{K_D + K_P T}{T+\Delta T}u(k-2) \end{aligned} \quad (22)$$

Here, ΔT is sampling time.

Comparing (20) and (22), we get:

$$\alpha+\beta+w1 = \frac{2T+\Delta T}{T+\Delta T} \quad (23)$$

$$\alpha\beta + \beta w1 - w6 = \frac{T}{T+\Delta T} \quad (24)$$

$$w2\times w3 = \frac{(K_P+K_I T)\Delta T + K_D + K_P T + K_I\Delta T^2}{T+\Delta T} \quad (25)$$

$$w2[(\alpha+\beta)w3 - w4] = \frac{(K_P+K_I T)\Delta T + 2K_D + 2K_P T}{T+\Delta T} \quad (26)$$

$$w2(\alpha\beta w3 + w5 - w4\beta) = \frac{K_D + K_P T}{T+\Delta T} \quad (27)$$

From above analysis, we can find that new modified Elman network has the proportional, integral and derivative properties, which can be called PID Elman neural network. Its proportional gain and integral coefficient as well derivative action were changed as the weights were adjusted. Different from general PID algorithm, the controlled increment of new modified network is not only changed along with the change of input unit, but also it reinforces or restrains the output at last time with $\alpha+w1+\beta$ weight function. When $\alpha+w1+\beta$>1, it will enlarge the control output at last time; when $\alpha+w1+\beta$<1, it will reduce the control output at last time; and when $\alpha+w1+\beta$ =1, it is the standard variable-parameters PID control algorithm.

When r is not 1, the output of the network is the sum of the output of the r variable-parameters PID controllers. Meanwhile, the value of w3 means different selection or different weight of r PID control algorithm.

For traditional modified Elman networks, $\beta = 0$, $w5 = 0, w6 = 0$, so, we get:

$$T = 0$$

$$K_D = 0$$

$$\alpha + w1 = 1$$

$$w2 \times w3 = K_P + K_I \Delta T$$

$$w2(w4 - \alpha \times w3) = K_P$$

Equation (20) becomes (28).

$$y(k) = y(k-1) + K_P(u(k) - u(k-1)) + K_I \Delta T u(k) \tag{28}$$

It's clear that the traditional modified Elman neural network has the property of proportionality (P) and integral (I).

Furthermore, If α=0, and $w4 \neq 0$ (corresponding to second modified Elman network), or if $\alpha \neq 0$ and $w4$=0(corresponding to first modified Elman network), the network keeps proportional-integral characteristics. And if α=0 and $w4$=0, the modified Elman network degenerated to basic Elman one, we have:

$$y(k) = w1 \times y(k-1) + w2 \times w3 \times u(k) \tag{29}$$

It's the standard integral equation. So its dynamic response will slow down that will result in poor convergence speed. From above theoretical analysis, we can find out clearly that new modified Elman network has better dynamic properties than traditional modified Elman network and basic Elman network.

V. The Identification of the Dynamic System Using Modified Elman Network and the PID Elman Neural Network

The structure adopted for the dynamic system identification [6,7,8] is shown in the Fig. 5. And the simulation was carried on using the modified Elman neural network and the PID Elman network respectively.

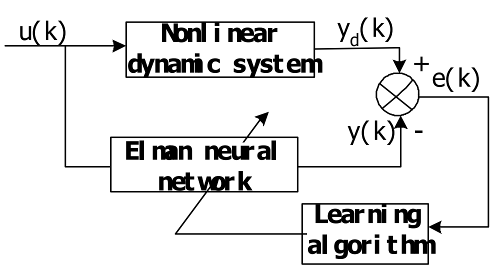

Figure 5. The structure of dynamic system identification.

The identification model is designed to minimizing the error between expected and actual outputs ($e(k) = y_d(k) - y(k)$). The mathematical model used as the nonlinear dynamic system is:

$$y(k+1) = 0.8y(k) + \frac{1}{y(k)} - 0.5u(k) \tag{30}$$

And the input signal u(k) is sinusoidal parameter:

$$u(k) = 0.4\sin(\frac{3\pi k}{14}) + 0.6\sin(\frac{2\pi k}{11}) \tag{31}$$

Where, k=1,2······200, thus we get 200 training data. We defined 15 nodes in the hidden, context, memory layers respectively. We use the sigmoid function as the transfer function and nonlinear function as the output unit of the neural network. The learning cycle sets 600. The converge speed of the two kinds of modified Elman neural network is given in Fig. 6 and the simulation results in Fig. 7 and Fig. 8.

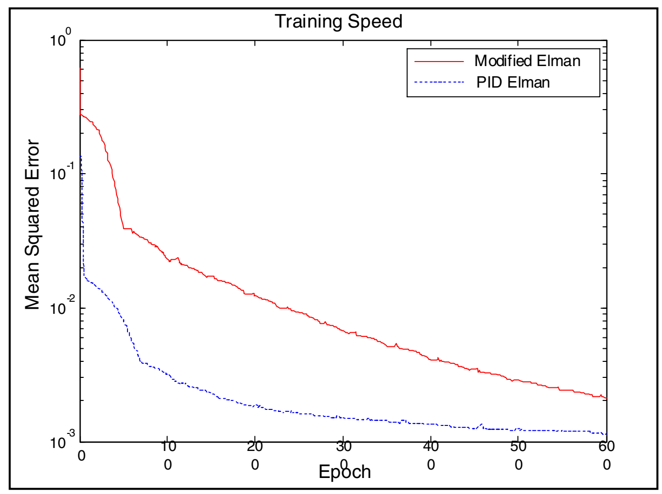

Figure 6. The converge speed of the traditional modified Elman neural network and the PID Elman neural network.

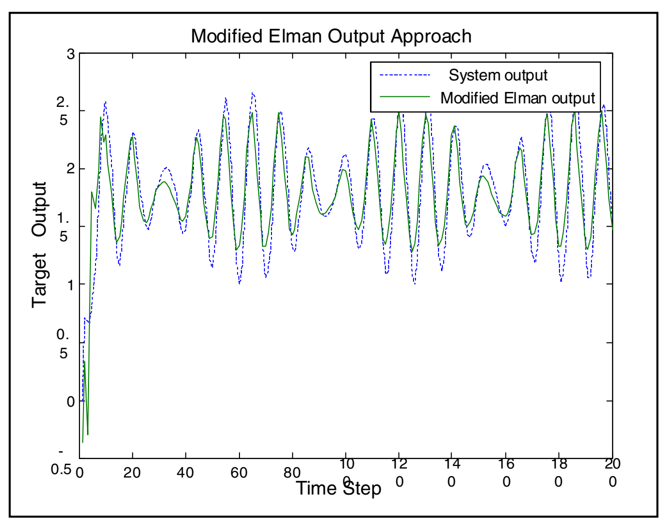

Figure 7. Identification result of the traditional modified Elman neural network.

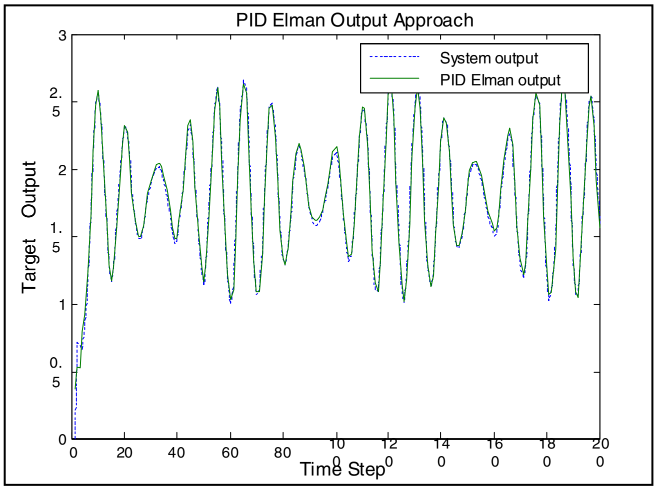

Figure 8. The simulation results of PID dynamic recurrent Elman neural network.

After the training process finished, the output mean square errors of the traditional modified and the PID Elman neural networks are 0.006133 and 0.001049 respectively. Comparing the Fig. 5 and Fig.6 it can be concluded that the latter has lower output MSE and better nonlinear dynamic system identification effect than the traditional modified Elman neural network.

VI. Conclusion

A new modified Elman neural network is proposed and the property of the new dynamic recurrent Elman neural network is discussed in detail by comparing to the traditional modified Elman neural network. The new Elman network inherently has proportional (P), integral (I), derivative (D) properties, the traditional modified Elamn network have proportional and integral function, and basic Elman networks have only integral function. Based on the above discuss and the comparison of the computer simulation results, the conclusion can be drawn that the PID Elman neural network is prior to the traditional modified Elman neural network in identification of nonlinear dynamic system. The model established may be referenced in the design of on-line identification and model predictive control.

References

[1] J.Elman, Finding structure in time, Cognitive Science, 1990, 14(2):179-211.

[2] Yuan-chu Cheng; Wei-min Qi; Wei-you Cai; Dynamic properties of Elman and modified Elman neural network, Proceedings of the first International Conference on Machine Learning and Cybernetics. Beijing,2002,2(2):637-640

[3] D.R.Hush and B.G.Horne, Progress in Supervised Neural Networks, IEEE Signal Processing Magazine, 1993, 10(1):8-39.

[4] Ren Xue-mei, Chen Jie, Gong Zhi-hao, Approximation Property of the Modified Elman network, Journal of Beijing Institute of Technology, 2002, 11(1):19-23

[5] P.S.Sastry, G. Santharam and K.P.Unnikrishnan, Memory Neuron Networks for Identification and Control of Dynamic Systems, IEEE Transactions on Neural Networks, 1994, 5(2): 306-319.

[6] X.Z.Gao, X.M.Gao,and S.J.Ovaska, A modified Elman neural network model with application to dynamical systems identification, IEEE International Conference on Systems, Man, and Cybernetics, Beijing, 1996.1376-1381

[7] Yonghong Tan, Xuanju Dang, Feng Liang, et al. Dynamic wavelet neural network for nonlinear dynamic system identification, Proceedings of the 2000 IEEE International Conference on Control Applications, 2000.214 - 219

[8] Calderon G., Draye J.-P., Pavisic D., et al. Nonlinear dynamic system identification with dynamic recurrent neural networks, Proceedings of International Workshop on Neural Networks for Identification, Control, Robotics, and Signal/Image Processing, 1996.49-54