Approach with Reducing Time-Delays to Rise Prediction Ability for Neural Networks

Jian Zheng\textsuperscript{1,*}, Zhaoni Li\textsuperscript{2}, and Jia He\textsuperscript{1}

\textsuperscript{1}Chongqing Aerospace Polytechnic, 401331 Chongqing, China
\textsuperscript{2}Qinghai Normal University, 811300 Qinghai, China

Abstract. The time-delays induce instability, then the instability influences the prediction accuracy of neural networks, hence, time-delays should be considered as a key indicator for constructing neural networks. Based on above mentioned problem, we reconstructed time-dependent Lyapunov function. Next following, we found the sufficient condition which ensured the convergence of neural networks and prevented high-frequency oscillation, which effectively increased the prediction accuracy of neural networks. This results show that the prediction accuracy about raise 2.5\% for neural networks after reducing time-delays, and the mean square error decreased about 0.02 for training samples using neural networks.

1 Introduction

Neural networks regard as a famous bionic systems, due to between neurons own transfer time, time-delays are easy to be induced. If neural network occurs seriously time-delays, it will lead to poor performance, even damage the stability of neural construction. Recently, this problem that time-delays affect the stability of neural networks have been extensively studied, and several significant results have been also derived. H.W et al.[1] implement system stability via feedback control or other technologies. Li et al.[2] studied the problem of event-triggered control, which applied nonlinear continuous-time systems under strict-feed- back condition. Tao et al.[3] analyzed the stability conditions for stochastic neural networks with Markovian jumping parameters, but the designed Markovian process is complex, furthermore, it improved the complexity of solution. Because the time-dependent Lyapunov function owns the ability of capturing the information that presents the feature of time-varying delays, the conditions derived of Lyapunov functional were less open. Above works did not take into account the stability of neural networks.

2 Theory analysis

The neural networks system are given by

\[ \Delta(x(t)) = -\delta x(t) + W(f(x(t)) + \theta(x(t))) \]

(1)

where, \( x(t) = [x_1(t), x_2(t), ..., x_n(t)]^T \) is the neurons state vector with \( n \) neurons. \( \delta = \text{diag}(\delta, \delta, ..., \delta) \) is a diagonal matrix, which has positive entries \( \delta > 0(i=1,2,\ldots,n) \).
\( f(x(t)) = [f_1(x_1(t))^T, f_2(x_2(t))^T, \ldots, f_n(x_n(t))^T] \) denotes the neuron activation function.

\[ w_0 = [w_{ij}(x(t))]_{n \times m} \] is the interconnection matrices representing the weight coefficients of the neurons. The \( \theta(x(t)) \), namely extra controller, is constant matrix with appropriate dimensions.

Assumption 1. The neuron activation function \( f(*) \) (i=1, 2, ..., n) is boundary and continuous in system (1), then satisfies:

\[
\begin{align*}
\text{real}
\end{align*}
\]

where, \( a, b \in R \), and \( a \neq b \), \( \text{real}_{\eta}^l \) and \( \text{real}_{\eta}^r \) are real constants. Next following, we introduce several Lemmas that will be used for analyzing the stability of system (1).

Theorem 1. For a scalar \( c_i > 0 \), if there exists matrices \( P \in R^{n \times n} > 0 \), \( Q \in R^{n \times n} > 0 \), symmetrical matrices \( M_{11}, M_{22}, M_{33} \in R^{n \times n} \), matrices \( M_{12}, M_{21}, M_{13}, M_{31} \in R^{n \times n} \), \( N_1, N_2, N_3 \in R^{n \times n} \), then we have:

\[
\Xi_1 + \Xi_2 < 0 \quad \text{and} \quad \begin{pmatrix}
M_{11} & M_{12} & N_1 \\
M_{21} & M_{22} & N_2 \\
N_1 & N_2 & N_3
\end{pmatrix} \geq 0
\]

where, \( e_i = [0_{n \times (i-1)}, I_n, 0_{n \times (4-n)}] \), \( \varepsilon_i \in e \), \( \Pi_1 = \varepsilon_i - \varepsilon_2 \), \( \Pi_2 = \varepsilon_i - 2\varepsilon_1 \), \( \Pi_3 = \varepsilon_i + \varepsilon_2 - 2\varepsilon_1 \).

Proof. We reconstructed a Lyapunov function, following that:

\[
L(x(s)) = \theta^T(s)P\theta(s) + \int_{t-k}^{t}\theta^T(s)Q\theta(s)ds + \int_{t-k}^{t}\int g^T(\tau)d\tau d\tau
\]

where, \( \theta(s) = [g^T(s), \bar{g}^T(s)]^T \),

\[
\theta(s) = [g^T(s), g^T(s+k)]^T = \int_{t-k}^{t}g(\tau)d\tau
\]

Let \( L(x(s)) \) be derivative, having

\[
L(x(s)) = \theta^T(s)[P\theta^T(s) \text{Sym} \Omega^T \Omega^T \theta^T(s) - \int_{t-k}^{t}g^T(\tau)d\tau \bar{g}^T(\tau)ds - \int_{t-k}^{t}g^T(\tau)ds \bar{g}^T(s)]\theta(s)
\]

According to the Theorems in [4] and [5], the following inequality holds

\[
-\int_{t-k}^{t}g^T(\tau)d\tau \leq \bar{g}^T(s)\int_{t-k}^{t}g^T(\tau)d\tau
\]

We introduced (5) into (4), having \( L(x(s)) \leq \theta^T(s)[\Xi_1 + \Xi_2] \theta(s) \). If \( \Xi_1 + \Xi_2 < 0 \), then \( L(x(s)) < 0 \). The system (1) exhibit asymptotically stability. This completed above proof.

For matrices \( P, Q \) of the Theorem 1, \( C \) is a matrices, and \( X \) is a unknown matrices. \( \phi_1, \phi_2 \) are eigenvalue of matrices \( P, Q \), respectively. \( A, B \) are the similar transformation matrix of \( P, Q \), respectively. \( \phi_{ij} \) is the element of \( D = A^{-1}CB \). \( \Lambda_{ij} = \text{diag}(\phi_1, \phi_2, \ldots, \phi_{ij}) \).

\[ y = A^T \begin{pmatrix}
d_{ij} \\
\phi_1 + \phi_2
\end{pmatrix}
\]

we can get the unknown matrix \( X \), having that:

\[
X = A \begin{pmatrix}
d_{ij} \\
\phi_1 + \phi_2
\end{pmatrix} B^{-1}
\]

311
Matrix $X$ is an extra controller of system (1). $\theta(x(t)) = X$, i.e., $\theta(x(t))$ of system (1) is an appropriate constant matrix. Because the $PX + XQ = C$ is Lyapunov equation, the matrix $X$ presents stability. Hence, extra controller $\theta(x(t))$ enhances the stability of system (1).

### 3 Experimental results

We generated the initial states $x(0)$ of data vectors. $\theta(x(0))$ was calculated by matrices $P$, $Q$, $x(0)$. We took the nonlinear squashing function, namely Sigmoid function, as activation function $1(x)$, $1(x) = \frac{1}{1+ \exp(-x)}$, $-\infty < x < +\infty$. Sigmoid function is treated as a nonlinear gain value.

Gain value is determined via the slope of point $x$, i.e., $\text{real}_1^{\text{\wedge}} = 0$ and $\text{real}_1^{\text{\wedge}} = 1$. We expect that mean square error (MSE) can rapidly converge $E^{-5}$, and the process of converging does not exhibit high-frequency oscillation.

$$
\begin{bmatrix}
0.7348 & -9.1895 & -11.2498 \\
-5.1588 & 7.2146 & -10.0277 \\
3.8553 & 4.4885 & 0.025
\end{bmatrix}
$$

Gain value is determined via the slope of point $x$.

We apply MNIST data to verify our theory. We train on 10000 samples, then validate on 2000 samples, which address the prediction accuracy and MSE.

![Figure 1](image1.png)

**Figure 1.** MSE results. The error results that time-delays influence the neural networks system (1).

![Figure 2](image2.png)

**Figure 2.** Prediction accuracy. We apply MNIST data to verify our theory. We train on 10000 samples, then validate on 2000 samples, which address the prediction accuracy and MSE.

Fig. 1a shows that it applies our approach into neural networks with 20 hidden units, which implies that the neural networks system (1) converged steadily. For 20 hidden units, time-delays were superimposed, so system (1) occurred low-frequency oscillation in earlier
stage. In later period, it did not take place oscillation. Fig. 1b shows that the system (1) without our approach hardly converged, which presented continuous oscillation too. The results of MSE mean that our approach are better than without our approach.

Fig. 2a shows that the time-delays impacts the prediction accuracy of neural networks, which means that the prediction accuracy is improved with our approach. Fig. 2b presents that the MSE for training samples is also affected, due to the time-delays induced oscillation. Controlled the time-delays, the prediction accuracy approximately raised 2.5% for neural networks, and the mean square error decreased about 0.02.

The research funding provided by the Chun-hui Planning Project of Ministry of Education (Z2016084).

References