文章编号:1672-6553-2023-21(8)-031-007

DOI:10.6052/1672-6553-2023-072

基于注意力机制回声状态神经网络的混沌系统预测*

刘建明[1]　徐一宸[2†]

(1.吉林大学 数学学院,长春　130012)

(2.中国人民大学 信息学院,北京　100872)

摘要　混沌系统在电路、保密通讯、加密解密等方面具有重要的研究意义.由于其对初值非常敏感,传统的统计学时间序列预测方法在处理混沌时间序列预测问题是具有挑战的.回声状态网络是一种特殊的循环神经网络,在复杂动态系统动力学与控制方面具有优势.经典的回声状态网络将每个样本置于同一地位,然而实际问题中不同的样本的重要性往往是有差异的.本文提出注意力机制回声状态神经网络模型,将回声状态网络与注意力机制相结合体现样本之间的差异性以及样本之间的相互作用对预测的影响.对混沌系统的预测结果表明注意力机制回声状态神经网络具有更好的预测性能.

关键词　混沌系统,　回声状态网络,　储备池计算,　注意力机制,　机器学习

中图分类号:O322　　**文献标志码**:A

Chaotic Systems Prediction Using the Echo State Network with Attention Mechanism*

Liu Jianming[1]　Xu Yichen[2†]

(1. College of Mathematics, Jilin University, Changchun　130012, China)

(2. School of Information, Renmin University of China, Beijing　100872, China)

Abstract　Chaotic systems have important research significance in circuit, secure communication, encryption and decryption. The traditional statistical time series prediction methods are challenging in dealing with the chaotic systems because the chaotic systems are very sensitive to the initial values. Echo state network is a special cyclic neural network, and has advantages in the dynamics and control of complex dynamic systems. The classical echo state network places all samples in the same importance, however in practical problems, the importance of the samples are often different. This paper proposes the attention mechanism echo state network by combining the echo state network and the attention machine to reflect the differences and interactions between samples. The prediction results on chaotic systems show that the prediction performance of the echo state network with attention mechanism is better than that of the classical methods.

Key words　chaotic systems,　echo state network,　reservoir computing,　attention mechanism,　machine learning

2023-06-26 收到第 1 稿,2023-07-30 收到修改稿.

* 国家自然科学基金资助项目(12072128),吉林大学研究生教改项目(2021JGZ26,2021JGZ15), National Natural Science Foundation of China (12072128), Graduate Teaching project in Jilin University (2021JGZ26,2021JGZ15).

† 通信作者 E-mail:yichenruc@126.com

引言

混沌动力系统广泛存在于物理学、生物学、航空气象学等学科中，混沌系统的预测是一个重要的实际问题.混沌系统对初值非常敏感，传统的时间序列预测方法(例如自回归模型)对混沌等复杂非线性数据的处理效果不佳.

近年来，机器学习的快速发展给很多领域带来了革命性的突破.卷积神经网络[1-2]进行图像识别的精度远远高于其它方法；循环神经网络[3-4]在自然语言处理等时序问题的建模上效果显著；对抗生成网络[5-6]能生成逼真的图像，几乎能以假乱真.这些突破有多方面的原因，而注意力机制[7]的引入起着重要作用.

回声状态网络(Echo State Network，ESN)[8-10]是一类特殊的循环神经网络，在时间序列预测上表现极为优异.ESN 的隐藏层是大量神经元构成的储备池，输入数据经由储备池后被映射到高维空间，然后由读出权值映射为输出.读出权值是 ESN 中唯一需要学习的参数，可由岭回归算法计算，避免了传统神经网络模型训练过程中存在的梯度爆炸和梯度消失问题[11-12].韩敏等将 ESN 与主成分分析、Kalman 滤波及 Adaboost 算法相结合，研究多元时间序列预测问题[13-15]；薄迎春等对具有小世界(small-world)特性的 ESN 进行结构分析与设计[16]；伦淑娴等提出基于小世界网络的时间序列预测方法[17]；宋青松等给出连接权重的一个稳定训练方法[18]；沈力华等提出多稀疏 ESN 的预测模型[19]；而王磊等将 ESN 用于 PM2.5 的预测研究[20].

经典 ESN 模型假设每个样本的重要程度都是相同的，然而现实世界中的很多问题中，不同样本的重要程度往往是不同的.本文提出注意力机制回声状态网络(Attention-Mechanism-Based Echo State Network，AM-ESN)，将回声状态网络与注意力机制相结合来体现不同样本之间的差异，并应用于混沌动力系统的预测.

1 回声状态网络

回声状态网络由输入层，隐藏层和输出层组成，如图 1 所示.隐藏层是由大量神经元组成的储备池，其作用是将输入数据映射到高维空间.ESN 的核心思想是将输入序列映射到高维空间，在高维空间中这些状态表现出线性的性质，从而实现序列的有效预测.将 t 时刻的输入、储备池状态、输出状态分别记作 $u(t)$、$x(t)$ 和 $y(t)$，则有[21]：

$$\begin{cases} x(t+1) = \tanh[W_{in}u(t+1)+Wx(t)] \\ y(t) = W_{out}x(t) \end{cases} \tag{1}$$

$u(t) \in R^{d_1}$，$x(t) \in R^{N}$，$y(t) \in R^{d_2}$，$W_{in} \in R^{N\times d_1}$，$W \in R^{N\times N}$，$W_{out} \in R^{N\times d_2}$；$d_1$，$d_2$ 是输入和输出维度，N 是储备池中神经元的数量.权值矩阵 W_{in} 和 W 是随机生成的，一经生成在整个过程中保持不变.

设参与计算读出权重的样本数目为 n，记：

$$X = [x(1), x(2), \cdots, x(n)]$$

$$Y = [y(1), y(2), \cdots, y(n)]$$

于是 W_{out} 是如下问题的最优解：

$$\min_{W_{out}}(\|W_{out}X - Y\|^2 + \gamma\|W_{out}\|^2) \tag{2}$$

这里 γ 是正则化参数.用岭回归方法，的解为：

$$W_{out} = YX^{\mathrm{T}}(XX^{\mathrm{T}} + \gamma I)^{-1} \tag{3}$$

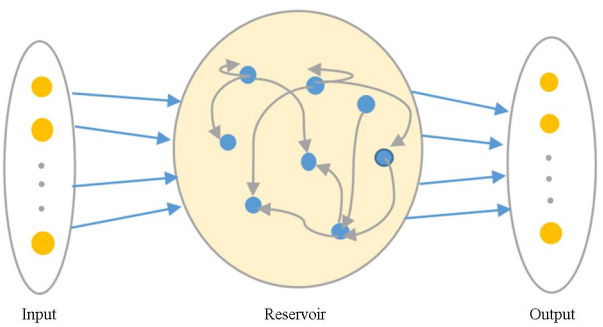

图 1 回声状态神经网络示意图
Fig.1 Sketch of the echo state neural network

2 注意力机制回声状态网络(AM-ESN)

注意力机制起源于对人类视觉的研究，在认知科学中，人类会有选择性地关注一部分特定的信息，同时忽略其他可见的信息.注意力机制是解决信息超载问题的重要手段之一，将计算资源分配给更重要的任务或者某一任务的更重要的部分[22-24].

现实世界中的很多问题中，不同样本的重要程度往往是不同的.如果将所有样本置于同一重要性，读出权值将会损失很多对预测有用的信息.本文认为在 ESN 的训练过程中，不同样本对读出权值的贡献是不同的，提出注意力机制回声状态网络，图 2 给出了结构示意图.

图 2 注意力机制回声状态网络示意图

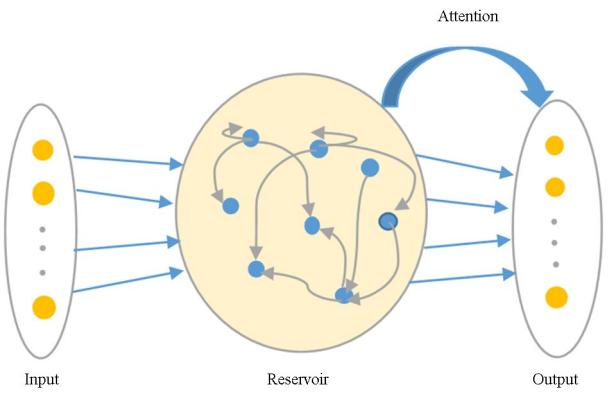

Fig.2　Sketch of the AM-ESN

我们在目标函数中添加权系数 $p_i \geqslant 0$ 来体现不同样本之间的差异,则目标函数可以写成:

$$\min_{W_{out}} = \sum_{i=1}^{n} \| p_i [W_{out} x(i) - y(i)] \|^2 + \gamma \| W_{out} \|^2 \tag{4}$$

这里 $p_i \geqslant 0$, $i = 1,2,\cdots,n$ 是样本的权重系数.

记 $P = \mathrm{diag}(p_1, p_2, \ldots, p_n)$,则公式可以写成如下矩阵形式:

$$\min_{W_{out}} f(W_{out}) \tag{5}$$

这里: $f(W_{out}) = \| (W_{out} X - Y) P \|^2 + \gamma \| W_{out} \|^2$.

简单计算可以得到:

$$\frac{\partial f(W_{out})}{\partial W_{out}} = 2(W_{out} X - Y) P P^{\mathrm{T}} X^{\mathrm{T}} + 2\gamma W_{out}$$

令公式两端为零,则得到最优的权值 W_{out}

$$W_{out} = YA X^{\mathrm{T}} (XA X^{\mathrm{T}} + \gamma I)^{-1} \tag{7}$$

这里注意力矩阵 $A = \mathrm{diag}(a_1, a_2, \cdots, a_n) = P P^{\mathrm{T}}$.

公式给出 AM−ESN 的读出权值 W_{out} 的计算方法,有如下的性质成立.

定理 1: 对于输入−输出 $\{[x(1), y(1)], [x(2), y(2)], \cdots\}$ 以及注意力矩阵 $A = \mathrm{diag}(a_1, a_2, \cdots, a_n)$.考虑第 i 个样本,假设其相应的注意力系数为 $a_i = 1$,其余的注意力系数为 $a_k = a \geqslant 0 (i \neq k)$.记由公式计算出来的权矩阵为 W'_{out} ,公式计算出来的权矩阵为 W''_{out} ,于是有:

(1)如果 a>1, 那么

$$\| W''_{out} x(k) - y(k) \|^2 \leqslant \| W'_{out} x(k) - y(k) \|^2;$$

(2)如果 $0 \leqslant a < 1$,那么

$$\| W''_{out} x(k) - y(k) \|^2 \geqslant \| W'_{out} x(k) - y(k) \|^2$$

证明: 我们首先来证明(1).记

$$\Phi(W_{out}) = \sum_{i=1}^{n} \| W_{out} x(i) - y(i) \|^2 + \gamma \| W_{out} \|^2,$$

$\min_{W_{out}} [\Phi(W_{out})]$ 是一个凸优化问题,由公式计算的 W'_{out} 是全局最优解.同样地 W''_{out} 也是全局最优解并且有

$$W''_{out} = \mathrm{argmin}_{W_{out}} [\Phi(W_{out}) + (a-1) \| W_{out} x(k) - y(k) \|^2] \tag{8}$$

由于 $\min[\Phi(W_{out})] = \Phi(W'_{out})$,且 $a > 1$,因此

$$\min[\Phi(W_{out}) + (a-1) \| W_{out} x(k) - y(k) \|^2] \leqslant \Phi(W'_{out}) + (a-1) \| W'_{out} x(k) - y(k) \|^2 \tag{9}$$

等号成立当且仅当 $W''_{out} = W'_{out}$,在这种情形下有

$$\| W''_{out} x(k) - y(k) \|^2 = \| W'_{out} x(k) - y(k) \|^2 \tag{10}$$

如果 $W''_{out} \neq W'_{out}$,式取.注意到

$$\Phi(W''_{out}) > \Phi(W'_{out}) \tag{11}$$

因此,

$$(a-1) \| W''_{out} x(k) - y(k) \|^2 < (a-1) \| W'_{out} x(k) - y(k) \|^2 \tag{12}$$

所以有

$$\| W''_{out} x(k) - y(k) \|^2 < \| W'_{out} x(k) - y(k) \|^2 \tag{13}$$

由式和式可知结论成立.

与(1)的证明类似,可以得到(2)的结论.

注意力矩阵 A 使得不同的样本在计算读出权值的过程中具有不同的重要性.通常来说,第 i 个样本 a_i 的越大,拟合误差 $\| W_{out} x(i) - y(i) \|$ 就越小.为了计算简单,注意力机制矩阵 A 可以为(14)式中的对角矩阵 A_1,也可以为(15)式中所示的非对角矩阵 A_2,以体现不同样本之间的差异性以及样本之间的相互作用.

$$A_1 = \begin{pmatrix} a_1 & 0 & \cdots & 0 & 0 \\ 0 & a_2 & \cdots & \vdots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \vdots & \ddots & a_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & a_n \end{pmatrix} \tag{14}$$

$$A_2 = \begin{pmatrix} 1 & v & 0 & \cdots & 0 \\ u & 1 & v & \cdots & 0 \\ \vdots & u & \ddots & \ddots & \vdots \\ 0 & \vdots & \cdots & 1 & v \\ 0 & 0 & \cdots & u & 1 \end{pmatrix} \tag{15}$$

注意力系数的选取有如下几种方法:

(1) 如果对系统或者样本有更好的了解,则更容易地给出注意力参数的值, 对重要的样本赋予更大的注意力系数.对于混沌预测问题,可以根据

混沌的具体特性来选择注意力系数,例如可以对转捩点和突变点等特殊样本赋予大的注意力系数.

(2) 用 softmax 函数来确定注意力系数.首先令所有的注意力系数都为 1,按照经典的 ESN 方法计算 W_{out} ;然后按照 $p_i=\mathrm{softmax}(\|W_{out}x(i)-y(i)\|)$,计算注意力系数,并根据公式(7)重新计算 W_{out} ,则训练误差大的样本点将获得更大的注意力系数.

(3) 用优化的方法确定注意力系数.例如对于公式(15)的注意力矩阵,只有两个参数需要确定,则可以将参数(u,v)在$(0,0)$附近做网格划分,通过搜索的方式获得最优的注意力矩阵参数.

1 数值模拟

本节考虑两个经典的混沌系统的预测问题,使用如下的均方根误差(RMSE)来度量预测的精度:

$$\mathrm{RMSE}=\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\tilde{y}(t)-y(t))^2} \tag{16}$$

这里 $y(t)$是目标信号(Ground truth), $\tilde{y}(t)$ 是预测信号.RMSE 的值越小,预测精度越高.

3.1 Mackey-Glass 混沌系统

Mackey-Glass 系统[25]是描述白细胞繁殖的模型,其方程为:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=\alpha x(t)+\frac{\beta x(t-\tau)}{1+x^n(t-\tau)} \tag{17}$$

这里 $x(t)$表示血液循环中成熟细胞的质量分数, τ 是在骨髓中产生未成熟细胞和在血液中释放成熟细胞的时滞参数,α 为系统的反馈率,β 是过去的状态对当前状态的影响率,n 是正常数. 根据文献[8], 当参数设 $\alpha=-0.1,\beta=0.2,n=10,\tau=17$ 时系统展现出混沌特性.我们用经典的龙格库塔方法对 进行积分,然后选取 3000 个样本,其中前 1000 个样本用来预热,后 2000 个样本直接参与读出权值的计算.在计算过程中,储备池的大小 $N=1000$, $\gamma=10^{-8}$,谱半径 $\rho=1.1$ 和稀疏度 $sp=0.01$. 训练过程中采用如下的注意力矩阵:

$$\begin{pmatrix} 1 & 2.9 & & & 0 \\ 0.2 & 1 & 2.9 & & \\ & 0.2 & \ddots & \ddots & \\ & & \ddots & 1 & 2.9 \\ 0 & & & 0.2 & 1 \end{pmatrix} \tag{18}$$

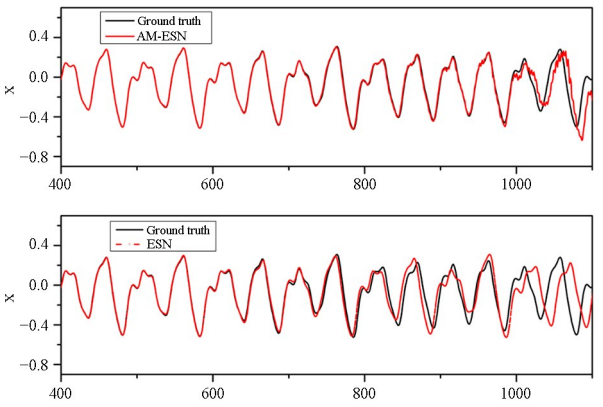

图 3 Mackey-Glass 系统预测时间历史图

Fig.3 Time history of the prediction of the Mackey-Glass system

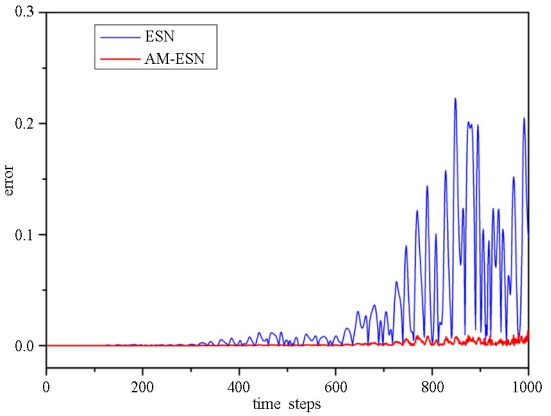

图 4 Mackey-Glass 混沌系统的预测误差

Fig.4 Prediction error of the Mackey-Glass system

图 3 画出了应用 ESN 方法与 AM−ESN 方法预测的 Mackey-Glass 混沌系统的时间历史,图 4 给出了预测误差的比较图.可以看出,应用 ESN 方法在小于 600 个时间步时的预测精度$<1\%$;而 AM-ESN 可以在 900 个时间步之内其预测误差$<1\%$.这说明注意力机制的引入可以在使得模型在更长的时间内实现混沌系统的精确预测,有效地提升模型的预测性能.

3.2 Rössler 混沌系统

Rössler 系统是典型的混沌系统,其运动方程为:

$$\begin{cases}\dot{x}=-(y+z)\\ \dot{y}=x+\alpha y\\ \dot{z}=\beta+z(x-\theta)\end{cases} \tag{19}$$

取 $\alpha=0.15,\beta=0.2,\theta=10$ 时系统展现为混沌.以(1,1,1)为初值,用龙格库塔方法以 0.05 的步长生成数据.训练样本的数量为 2000,前 100 个样本用于预热.由于 Rossler 混沌系统的数据在 z 方向具

有脉冲性质,用 ESN 不能对其进行有效的预测,所以这里我们只考虑在 x-和 y-方向的系统预测.我们比较了 ESN 以及其注意力机制版 AM-ESN 的预测性能,具体的参数如下:储备池的大小 $N=1000$,$\gamma=10^{-8}$,谱半径 $\rho=1.25$,稀疏度 $sp=0.01$.我们采用公式所示的注意力矩阵,这里主对角线前 1000 元素取值 1,后 1000 元素取值 50,其他元素为 0.

$$\begin{pmatrix} 1 & 0 & & & 0 \\ 0 & 1 & 0 & & \\ & 0 & \ddots & \ddots & \\ & & \ddots & 50 & 0 \\ 0 & & & 0 & 50 \end{pmatrix} \tag{20}$$

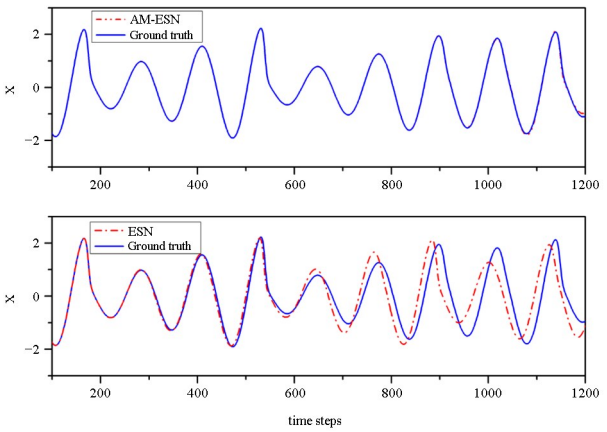

图 5　Rössler 混沌系统在 x-方向的预测结果
Fig.5　Prediction of Rössler system in the x-direction

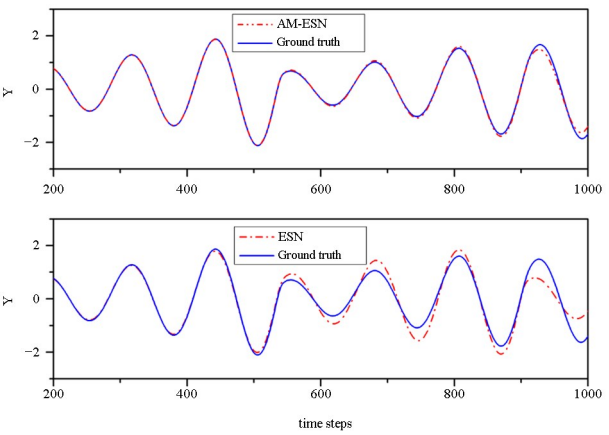

图 6　Rössler 混沌系统在 y-方向的预测结果
Fig.6　Prediction of Rössler system in y-direction

图 5 和图 6 分别画出了应用 ESN 和 AM-LESN 方法在 x-和 y-方向的预测结果,图 7 给出了在 x-和 y-方向的预测误差比较图.可以看出 ESN 方法在小于 200 个时间步时的预测精度<1%;而 AM-LESN 可以在 800 个时间步之内其预测误差<1%.这说明注意力机制的引入可以使得模型在更长的时间内实现混沌系统的精确预测,有效地提升模型的预测性能.

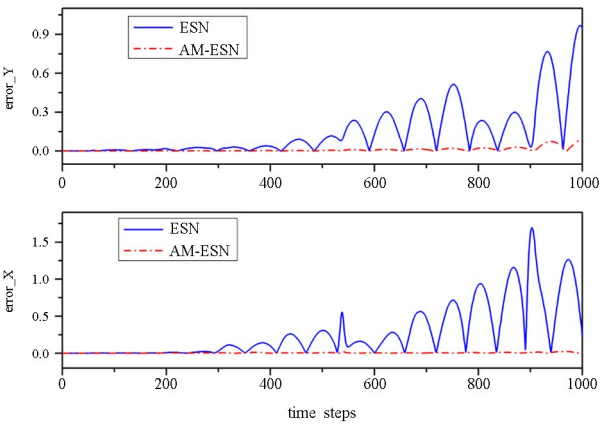

图 7　Rössler 混沌系统的预测误差
Fig.7　Prediction Error of the Rössler system

注意力机制参数的选择是获得更好预测性能的重要因素,这往往是与问题相关的,目前还没有统一的有效的方法给出理论上的指导.文中给出了两种注意力机制,其中一种是通过在注意力机制矩阵的两个次对角线增加非零元素,另外一种是使得注意力机制矩阵的主对角线元素选取不同的值来实现更好地性能.第一种方式是增加了样本之间的相互作用,也就是一个样本的临近样本会对神经网络的训练产生影响;第二种方式是考虑了样本的差异性,考虑了重要样本对训练和预测的影响.文中的算例表明,注意力矩阵的参数可以采用简单的方式来选取就可以获得很好的性能.此外如果对系统或者样本有充分的先验知识,则更容易地给出注意力参数的值.实验结果表明,通过引入注意力机制可以获得更好的混沌预测精度,可以有效地跟踪混沌系统的运动.

4　结论

目前现有的基于神经网络的复杂系统预测方法中,通常将所有样本置于相同的地位,并不区分样本之间的差异性.然而实际生活中的许多问题中不同样本的重要性往往是有差异的,因此考虑不同样本之间的差异性,对于复杂系统的机器学习方法至关重要.本文提出了注意力机制回声状态神经网络用于混沌系统预测,考虑训练集中不同样本的重要性不同,通过在目标函数中引入注意力机制矩阵来体现样本的差异性.文中讨论了两种简单的注意

力机制矩阵的构成方式,给出了确定注意力矩阵中元素的几种方法.提出的方法对于 Mackey-Glass 和 Rössler 混沌系统具有很好的预测结果,这表明与经典的回声状态神经网络相比,注意力机制的引入能使得回声状态网络具有更好的预测能力.

参考文献

[1] LECUN Y, BOSER B E, DENKER J S, et al. Handwritten digit recognition with a back-propagation network [C]. In Proceedings Advances in Neural Information Processing Systems, 1990, 396－404.

[2] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [C]. Proceedings IEEE, 1998, 86, 2278－2324.

[3] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735－1780.

[4] CHUNG J, GULCEHRE C, CHO K, et al. Gated feedback recurrent neural networks[C]. In Proceedings of the 32th International Conference on Machine Learning, 2015, 37, 067－2075.

[5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]. Advances in Neural Information Processing Systems (NIPS), 2014, 27, 2672－2680.

[6] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [C]. In Proceedings of the 36th International Conference on Machine Learning, 2019, 97, Long Beach,CA.

[7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]. Advances in Neural Information Processing Systems (NIPS), 2017, 30, 5998－6008.

[8] JAEGER H. The 'echo state' approach to analysing and training recurrent neural networks [R]. German National Research Center for Information Technology, GMD Report 148, 2001.

[9] MAASS W, NATSCHLAGERER T, MARKRAM H. Real-time computing without stable states: a new framework for neural computation based on perturbations [J]. Neural Computation, 2002, 14, 2531－2560.

[10] JAEGER H, HASS H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communications [J]. Science, 2004, 304 (5667): 78－80.

[11] PEARLMUTTER B A. Gradient calculations for dynamic recurrent neural networks: a survey [J]. IEEE Transactions on Neural Networks, 1995, 6 (5): 1212－1228.

[12] PASCANU R, MIKOLOV T, BENGIO Y. On the difficulty of training recurrent neural networks [C]. In Proceedings of the 30th International conference on machine learning, 2013, 1310－1318.

[13] 韩敏,王亚楠. 基于储备池主成分分析的多元时间序列预测研究 [J]. 控制与决策,2009,24(10):1526－1530.
HAN M, WAND YN. Prediction of multivariate time series based on reservoir principal component analysis [J]. Control and Decision, 2009,24(10): 1526－1530. (in Chinese)

[14] 韩敏,王亚楠. 基于 Kalman 滤波的储备池多元时间序列在线预报器 [J]. 自动化学报, 2010,36(1): 169－173.
HAN M, WANG YN. Multivariate time series online predictor with Kalman Filter trained reservoir [J]. Acta Automatic Sinica, 2010,36(1):169－173. (in Chinese)

[15] 韩敏, 穆大云. 基于 Adaboost 算法的回声状态网络预报器 [J]. 控制理论与应用, 2011, 28(4): 601－604.
HAN M, MU DY. Improvement of echo state network accuracy with adaboost [J]. Control Theory & Applications, 2011, 28(4): 601－604 (in Chinese)

[16] 薄迎春,乔俊飞,张昭昭. 一种具有 small world 特性的 ESN 结构分析与设计 [J]. 控制与决策,2012,27 (3):383－388.
BO YC, QIAO JF, ZHANG ZZ. Analysis and design on structure of small world property ESN [J]. Control and Decision, 2012,27(3): 383－388. (in Chinese)

[17] 伦淑娴,林健,姚显双. 基于小世界网络的时间序列预测 [J]. 自动化学报, 2015, 41(9): 1669－1679.
LUN SX, LIN J, YAO XS. Time series prediction with an improved echo state network using small world network [J]. Acta Automatic Sinica, 2015, 41(9): 1669－1679. (in Chinese)

[18] 宋青松,冯祖仁,李人厚. 回响状态网络输出连接权重的一个稳定训练方法 [J]. 控制与决策,2011,26

(1):22－26.

SONG Q S, FENG Z R, LI R H. Stable training method for output connection weights of echo state networks [J]. Control and Decision, 2011,26(1): 22－26. (in Chinese)

[19] 沈力华,陈吉红,曾志刚,等. 多稀疏回声状态网络预测模型 [J]. 控制理论与应用,2018,35(4):421－428.

SHEN L H, CHEN J H, ZENG Z G, et al. Prediction model with multiple sparse echo state network [J]. Control Theory & Applications, 2018,35(4): 421－428. (in Chinese)

[20] 王磊,杨翠丽,乔俊飞. 基于回声状态网络的 PM2.5 预测研究 [J]. 控制工程, 2019, 26(1): 1－5.

WANG L, YANG C L, QIAOA J F. Study on prediction of atmospheric PM2.5 based on echo state network [J]. Control Engineering of China, 2019, 26(1): 1－5. (in Chinese)

[21] LUKOSEVICIUS M, JAEGER H, SCHRAUWEN B. Reservoir computing trends [J].Kunstliche Intelligenz,2012, 26(4): 365－371.

[22] VOLODYMYR M, NICOLAS H, ALEX G, et al. Recurrent models of visual attention [C]. 28th Conference on Neural Information Processing Systems (NIPS), 2014.

[23] BA J, MNIH V, KAVUKCUOGLU K. Multiple object recognition with visual attention [DB]. The ArXiv:1412.7755, 2014.

[24] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]. In International Conference on Learning Representations (ICLR), 2014.

[25] MACKEY MC, GLASS L. Oscillation and chaos in physiological control system [J]. Science, 1977, 197: 287－289.