

具有双储层结构的动态误差补偿回声状态网络

张昭昭¹, 朱应钦^{1,2†}, 余 文²

(1. 西安科技大学 计算机科学与计算学院, 陕西 西安 710054;

2. 墨西哥国立理工大学高级研究中心(CINVESTAV)自动化研究所, 墨西哥城 07360)

摘要: 针对传统回声状态网络难以有效应对高阶非线性复杂模型问题, 本文在理论分析的基础上提出了一种双储层结构的误差补偿回声状态网络, 并设计了该网络的学习算法. 该网络由计算层和补偿层构成, 计算层主要承担拟合任务, 补偿层则作为状态跟随器, 实时补偿由于计算层对期望方差估计不足而导致的幅值偏差. 对多阶振荡器和真实高阶非线性数据集的实验结果表明, 本文所提网络结构较常规网络具有更高的稳定性和泛化性能, 尤其对高阶非线性复杂模型的预测精度大幅度提升.

关键词: 回声状态网络; 高阶非线性复杂模型; 补偿回声状态网络; 多阶振荡器

引用格式: 张昭昭, 朱应钦, 余文. 具有双储层结构的动态误差补偿回声状态网络. 控制理论与应用, 2024, 41(3): 385 – 395

DOI: 10.7641/CTA.2023.20539

A new echo state network with a double reservoir compensates for dynamic error

ZHANG Zhao-zhao¹, ZHU Ying-qin^{1,2†}, YU Wen²

(1. College of computer Science and Technology, Xi'an University of Science and Technology, Xi'an Shaanxi 710054, China;

2. Department of the Control Automatic, CINVESTAV-IPN (National Polytechnic Institute), Mexico city 07360, Mexico)

Abstract: The traditional echo state network is challenging to deal with the high-order nonlinear complex model effectively. We proposed an error trace reservoir computing and designed the optimal network algorithm. This new reservoir computing structure consists of a computing layer and a compensation layer. The computing layer mainly undertakes the fitting task, and the compensation layer acts as an error trace function. Because the computing layer always has an insufficient variance estimation, it will lead to unstable neural network prediction. Thus, we proposed the compensation layer to trace neural network error in real-time. The numerical experiments on modeling the multiple superimposed oscillators and nonlinear data sets demonstrate that error trace reservoir structure has higher stability and generalization performance than the conventional network, especially in the high order nonlinear complex models.

Key words: echo state network; high order nonlinear complex model; compensating echo state network; multiple superimposed oscillator

Citation: ZHANG Zhaozhao, ZHU Yingqin, YU Wen. A new echo state network with a double reservoir compensates for dynamic error. *Control Theory & Applications*, 2024, 41(3): 385 – 395

1 引言

回声状态网络(echo state network, ESN)^[1]通过观测时间序列数据来对动态系统建模, 其主要特点是所需训练数据集和计算资源小, 目前已在时间序列预测控制等领域得到了广泛应用^[2–5]. 一般情况下, ESN通过最小二乘法或岭回归的方法最小化均方误差(mean

squared error, MSE)来实现建模的过程^[6]. 最小二乘法会导致输出权重估计剧烈波动, 影响ESN的稳定性与建模精度, 而岭回归方法虽然降低了估计方差, 但该估计是一个有偏估计^[7], 不能保证ESN的建模精度. 尤其在高阶非线性复杂系统中, 由于误差自相关的原因, 导致方差无法准确估计, 这使得传统 ESN 很难实

收稿日期: 2022–06–17; 录用日期: 2023–04–20.

†通信作者. E-mail: yzhu@ctrl.cinvestav.mx; Tel.: +52 5566374377.

本文责任编辑: 周平.

陕西省自然科学基金计划陕煤联合基金资助项目(2019JLZ–08), 陕西省自然科学基金基础研究计划资助项目(2020JM–522, 2021JM–396)资助.

Supported by the Basic Research Plan of Natural Science Foundation of Shaanxi Coal Joint Fund (2019JLZ–08) and the Basic Research Plan of Nature Science in Shaanxi Province of China (2020JM–522, 2021JM–396).

现对高阶非线性复杂系统的精确建模. 然而, 大多数的工业过程是高阶非线性复杂系统^[8], 因此, 构建能够对高阶非线性复杂系统精确建模的ESN模型是该领域研究的难点与热点^[9].

近年来, 国内外众多学者通过优化参数或结构的方法来提高ESN的性能^[10-12]. 在参数优化方面, Chen等人^[13]在灰狼优化算法的基础上, 引入选择对抗策略来增加搜索能力. 但由于储层结构生成的随机性, 导致优化策略可重复性较差. 文献[14]针对参数优化策略可重复性较差及算法效率低的问题, 提出构建行为空间来确定网络结构是否具备优化条件以及后续优化方向, 并结合新颖搜索遗传算法最终克服上述问题. 然而, 参数优化受到网络结构本身的局限, 仅能逼近理论最优, 也难以拟合高阶非线性复杂模型. 文献[15]认为, 由于常规储层计算难以产生多样化复杂动力学行为, 仅仅依靠参数优化无法从根本上解决传统ESN不能应对高阶非线性复杂系统的问题. 在结构优化方面, 薄迎春等人从提高网络记忆容量的角度出发, 提出时间分割池计算网络(temporally segregated reservoir computing, TSRC)^[16]、稀疏连接的异步池计算网络(sparsely connected asynchronous reservoir computing network, SARC)^[17]、异步深度池网络(asynchronously deep reservoir computing, ADRC)^[18], 通过延迟使网络具有不同时刻信息的记忆能力, 最终提升复杂时间序列数据建模的能力. 王磊等人^[19]提出增量式学习回声状态网络模型(incremental regularized ESN, IRESN), 根据问题复杂度或残差来设计合适大小的网络结构. 然而, 上述方法, 无论是延迟深度网络还是增量网络, 均是通过增加网络的深度和网络的个数来提升网络的性能^[20], 而网络深度和规模的增大无疑会导致网络学习成本的增加和网络稳定性的下降, 且上述方法也未能针对常规ESN无法有效应对高阶非线性复杂模型从理论上进行分析.

本文针对常规ESN难以有效应对高阶非线性复杂系统的问题, 通过理论与实验分析常规ESN预测高阶系统误差产生的原因, 借鉴控制领域中采用动态补偿解决动态误差的思想^[21-22], 提出一种带有误差补偿机制的回声状态网络结构(error trace reservoir computing, ETRC). 网络主要由计算层与补偿层构成, 其中计算层初步构建(multiple superimposed oscillator, MSO)问题的数学模型, 而补偿层则对由于计算层对期望方差估计不足而导致的幅值偏差进行补偿. 最后基于人工数据和真实数据的实验结果表明, ETRC能有效降低误差之间的相关性, 提升对高阶非线性复杂问题的建模能力. 本文贡献主要在于: 1) 理论分析回声状态网络误差来源以及当存在自相关误差时对网络稳定性的影响; 2) 提出一种新的储层计算结构, 用于消除由自相关误差带来的影响, 并证明其收敛性.

2 回声状态网络

2.1 ESN误差分析

经典ESN由输入层、储层和输出层构成, 假设神经元节点数分别为 K, N, L . 拓扑结构如图1所示, 其状态矩阵与输出矩阵更新如下:

$$X(n) = f(W^{\text{in}}u(n) + W^{\text{res}}X(n-1)), \quad (1)$$

$$y(n) = W^{\text{out}}X(n) + \varepsilon, \quad (2)$$

记特征矩阵 $X = [x(1) \cdots x(l)]^T$, 特征向量 X 会形成一个 N 维特征超平面, 对应的输出矩阵为 $Y = [y(1) \cdots y(l)]^T$, ε 为误差项.

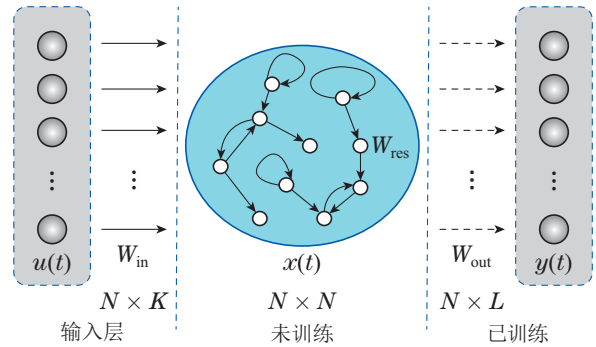


图1 ESN结构图

Fig. 1 Scheme diagram of ESN

回声状态网络的本质是训练向量 W^{out} 使特征矩阵 X 到期望输出矩阵 Y 之间距离最小, 即估计最佳参数 $\hat{W}^{\text{out}}$ 来最小化均方误差(mean squared error, MSE)误差, 如式(3)所示:

$$\hat{W}^{\text{out}} = M\hat{y}, \quad (3)$$

其中: $M \in \mathbb{R}^{N \times L}$, $\hat{y}$ 为真实值. 常规最小二乘法(ordinary least squares, OLS)求解可得 $M_{\text{ols}} = (X^T X)^{-1} X^T$; 岭回归方法求解可得 $M_{\text{ols}} = (X^T X + \lambda I_P)^{-1} X^T$; 代入式(2)可得

$$\hat{W}^{\text{out}} - W^{\text{out}} = (MX - I_P)W^{\text{out}} + M\varepsilon. \quad (4)$$

根据式(4)可得偏差与方差的关系为

$$\begin{aligned} \text{Bias}(\hat{W}^{\text{out}}) &= E(\hat{W}^{\text{out}}) - W^{\text{out}} = \\ &= (MX - I_P)W^{\text{out}}, \end{aligned} \quad (5)$$

$$\text{Var}(\hat{W}^{\text{out}}) = \text{Var}(M\varepsilon) = \sigma^2 MM^T, \quad (6)$$

其均方误差如下所示:

$$\begin{aligned} \text{MSE} &= \sum_t \|e_t\|_2^2 = \sum_t \|X_t - \hat{X}_t\|_2^2 = \\ &= E(\hat{W}^{\text{out}} - W^{\text{out}})^T (\hat{W}^{\text{out}} - W^{\text{out}}) = \\ &= [\text{Bias}(\hat{W}^{\text{out}})]^T [\text{Bias}(\hat{W}^{\text{out}})] + \sigma^2 \text{Tr}(MM^T) = \\ &= (W^{\text{out}})^T (MX - I_P)^T (MX - I_P) W^{\text{out}} + \\ &= \sigma^2 \text{Tr}(MM^T). \end{aligned} \quad (7)$$

优化均方误差面临的难题是现实中很难确定参数

W^{out} 和 σ^2 , 错误估计方差 σ^2 与输出权重 W^{out} 均会增加MSE误差, 从而使得网络预测性能不理想. 目前针对ESN参数的确定大都采用元启发等优化算法, 其实质是优化模型特征提取的能力, 以此来保证输出权重 W^{out} 的准确性, 但针对方差 σ^2 却鲜有方法. 因此, 本文提出一种带有误差补偿的回声状态网络, 采用补偿的方式尽可能的降低 σ^2 估计误差, 进而提高网络的性能.

2.2 回声状态网络自相关误差的影响

回声状态网络中常采用最小二乘法和岭回归的方法得到模型参数 W^{out} . 当满足基本假设时最小二乘法能够得到无偏估计量, 但由于真实系统中误差 $e_t \in \mathbb{R}$ 存在自相关特性, 难以满足正态分布 $e_t \sim N(0, \sigma^2)$ 和 $\text{Cov}(e_t, e_{t-\Delta t}) = 0, \forall \Delta t \neq 0$. 但真实误差常具有自相关特性^[23], 如式(8)所示:

$$e_t = \rho_1 e_{t-1} + \dots + \rho_p e_{t-p} + \tau_t, \quad |\rho_i| < 1, \quad \forall i, \quad (8)$$

其中: ρ 为不同时刻的相关系数, e 为不同时刻的误差, τ_t 为高斯噪声. 由于误差与前一时刻常表现为高度相关, 所以, 本文将对一阶自相关误差 $e_t = \rho e_{t-1} + \tau_t$ 的协方差进行推导.

$$\begin{aligned} e_t &= \rho(\rho e_{t-2} + \tau_{t-1}) + \tau_t = \\ &\rho^2 e_{t-2} + \rho \tau_{t-1} + \tau_t = \dots = \\ &\rho^h e_{t-h} + \sum_{i=0}^{h-1} \rho^i \tau_{t-i}, \end{aligned} \quad (9)$$

假设 $h \rightarrow \infty$,

$$e_t = \lim_{h \rightarrow \infty} (\rho^h e_{t-h} + \sum_{i=0}^{h-1} \rho^i \tau_{t-i}), \quad (10)$$

若 $\rho \geq 1$, 则 $\lim_{h \rightarrow \infty} \rho^h = \infty$, 假设 $|\rho| < 1$,

$$e_t = \lim_{h \rightarrow \infty} \sum_{i=0}^{h-1} \rho^i \tau_{t-i}, \quad (11)$$

e_t 的方差为

$$\begin{aligned} \text{Var}(e_t) &= \lim_{h \rightarrow \infty} \sum_{i=0}^{h-1} \text{Var}(\rho^i \tau_{t-i}) = \\ \lim_{h \rightarrow \infty} \sum_{i=0}^{h-1} \rho^{2i} \sigma^2 &= \frac{1}{1 - \rho^2} \sigma^2, \end{aligned} \quad (12)$$

根据式(12)可以得出, 当存在一阶自相关误差时, 若 $\rho = 0.2$, 则 $\text{Var}(e_t) = 1.042\sigma^2$, 将会导致方差增大4%左右. 最后结合式(10)–(13)求出协方差矩阵 $\text{Cov}(e_t, e_{t-h})$ 为

$$\begin{aligned} \text{Cov}(e_t, e_{t-h}) &= \\ \text{Cov}(\rho^h e_{t-h} + \sum_{i=0}^{h-1} \rho^i \tau_{t-i}, e_{t-h}) &= \\ \rho^h \text{Cov}(e_{t-h}, e_{t-h}) &= \frac{\rho^h}{1 - \rho^2} \sigma^2 \neq 0. \end{aligned} \quad (13)$$

由上述证明可以得知, 当存在一阶自相关误差时, Gauss-Markov假设不成立, 即系统方差被低估, 导致ESN预测的MSE增大.

3 误差补偿回声状态网络

3.1 ETRC网络模型

本文所提ETRC结构如图2所示, 其动力学模型可表示为

$$x^{\text{RC}}(t) = f(W_{\text{in}}^{\text{RC}} u(t) + W_{\text{res}}^{\text{RC}} x^{\text{RC}}(t-1)), \quad (14)$$

$$e(t) = \hat{y}(t) - y(t), \quad (15)$$

$$x^{\text{ET}}(t) = f(W_{\text{in}}^{\text{ET}} e(t) + W_{\text{res}}^{\text{ET}} x^{\text{ET}}(t-1)), \quad (16)$$

$$y(t) = W_{\text{out}}^{\text{RC}} x^{\text{RC}}(t) + W_{\text{out}}^{\text{ET}} x^{\text{ET}}(t), \quad (17)$$

其中: t 为时间步; f 为神经元激活函数; $x^{\text{RC}}(t)$ 和 $x^{\text{ET}}(t)$ 分别代表计算层及补偿层的内部状态; $e(t)$ 为 t 时刻误差; $W_{\text{in}}^{\text{RC}}, W_{\text{res}}^{\text{RC}}$ 分别为计算层的输入及神经元的连接权重矩阵; $W_{\text{out}}^{\text{RC}}$ 与 $W_{\text{out}}^{\text{ET}}$ 分别为计算层及补偿层的输出权重.

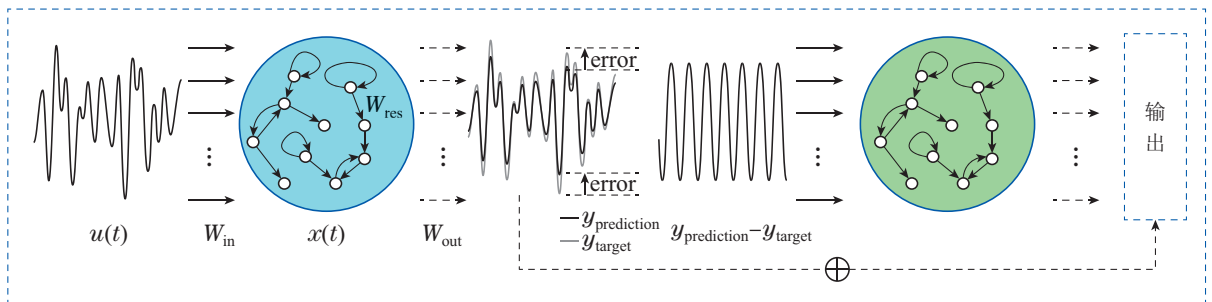


图2 ETRC结构图

Fig. 2 The scheme diagram of ETRC

在ETRC中, 输入与储层权重在初始化时随机给定并保持不变, 仅 $W_{\text{out}}^{\text{RC}}$ 和 $W_{\text{out}}^{\text{ET}}$ 需要通过学习确定, 通常采用最小二乘法. 本文所提ETRC网络由计算层和补偿层组成, 构建ESN补偿层的理由在于: 1) 由本文证明可知, 储层计算过程中由于自相关误差导致Gauss-

Markov假设不成立, 使得预测误差增大. 补偿层作为状态跟随器, 可以实时补偿计算层输出, 消除部分自相关误差, 进而保证网络预测的稳定性. 2) 降低网络构建的复杂性. 研究表明, 针对复杂问题ESN短时记忆容量受神经元规模限制^[24]. 对长时依赖问题, 需

满足其最低记忆容量要求^[14], 通常采用构建大规模神经元池来提高网络记忆容量, 但神经元池规模的增大会导致学习成本增加和网络稳定性的下降^[25]. ETRC采用一种分而治之的思想, 计算层可认为是对预测信号的一次分解, 补偿层用于拟合分解后的信号. 将复杂的问题分解为难度较小的问题, 最终降低网络构建的复杂性.

3.2 ETRC收敛性分析

引理 1 设ESN储层具有一个固定的内部权值矩阵 W_{res} 和 $f(x) = \tanh(x)$. 若内部权值矩阵的最大奇异值 $\bar{\sigma}(W) < 1$, 则ESN具有回声状态特性, 使得 $\lim_{k \rightarrow \infty} \|x_k - \tilde{x}_k\| = 0$, 其中 x_k 和 $\tilde{x}_k$ 是两个不同的状态向量.

定理 1 ETRC由两个ESN独立构成, 若 $\bar{\sigma}(W) < 1$, 则具备回声状态特性. 当误差存在自相关性且补偿层具备回声状态特性, 存在 $\rho - \hat{\rho} \rightarrow 0$, 使得 $e_{\text{new}} \leq e_{\text{regular}}$, 即 $\lim_{k \rightarrow \infty} |e_t - \hat{e}_t| = 0$.

证 由式(8)代入可得

$$\begin{aligned} e_{\text{regular}} &= |e_t|^2 = \langle e_t, e_t \rangle = \\ &\langle \rho_1 e_{t-1} + \cdots + \rho_p e_{t-p}, \rho_1 e_{t-1} + \cdots + \rho_p e_{t-p} \rangle = \\ &\rho_1^2 |e_{t-1}|^2 + \cdots + \rho_p^2 |e_{t-p}|^2 = \\ &\rho_1^2 |e_{t-1}|^2 + \sum_{j=2}^p \rho_j^2 |e_{t-j}|^2, \quad (18) \\ e_{\text{new}} &= |e_t - \hat{e}_t|^2 = \langle e_t - \hat{e}_t, e_t - \hat{e}_t \rangle = \\ &\langle (\rho_1 - \hat{\rho}_1) e_{t-1} + \cdots + (\rho_p - \hat{\rho}_p) e_{t-p}, \\ &(\rho_1 - \hat{\rho}_1) e_{t-1} + \cdots + (\rho_p - \hat{\rho}_p) e_{t-p} \rangle = \\ &(\rho_1 - \hat{\rho}_1)^2 |e_{t-1}|^2 + \cdots + (\rho_p - \hat{\rho}_p)^2 |e_{t-p}|^2 = \\ &(\rho_1 - \hat{\rho}_1)^2 |e_{t-1}|^2 + \sum_{j=2}^p (\rho_j - \hat{\rho}_j)^2 |e_{t-j}|^2 \leq \\ &e_{\text{regular}}, \quad (19) \end{aligned}$$

其中: e_{regular} 和 e_{new} 分别为经典ESN和ETRC的误差, e_t 为 t 时刻的误差, $\hat{e}_t$ 为 t 时刻补偿的误差, ρ_p 为 $t-p$ 时刻的相关系数, $\hat{\rho}_p$ 为补偿层学习得到的相关系数. 所以当ESN具备回声状态特性时, 存在 $(\rho_1 - \hat{\rho}_1)^2 \leq \rho_1^2$, 可推出 $e_{\text{new}} \leq e_{\text{regular}}$. 显然, 序列 e_{new} 是递减的且有界为零, 故ETRC收敛. 证毕.

3.3 ETRC参数优化算法设计

由于计算层对ETRC的性能至关重要, 且计算层的性能影响着补偿层的预测精度, 所以, 对计算层进行优化尤为必要. 从优化成本角度考虑, 本文采用一种行为空间算法优化计算层, 补偿层使用热重启的优化方法^[26-27]. 行为空间是根据ETRC不同行为来评估网络质量的方法, 其优势在于相对传统评判网络质量采用训练或测试误差, 而行为空间则仅通过状态矩阵或少量训练即可评判性能好坏. 热重启则是根据优化算

法继续所消耗的期望时间 $E(T_{\text{continue}})$ 与重启算法 $E(T_{\text{restart}})$ 的期望运行时间关系来制定相应的策略. 由于计算层已采用行为空间确定大致范围, 所以补偿层仅需通过较少次数的随机初始化来寻找最优参数. 其原因在于: 重启随机优化的期望时间远小于优化算法, 即 $E(T_{\text{restart}}) < E(T_{\text{continue}})$.

行为空间采用核心等级(kernel rank, KR)、泛化等级(generalization rank, GR)、记忆容量(memory capacity, MC)3个参数构建. 核心等级与泛化等级用于评估储层特征提取以及泛化能力. 记忆容量则评估整个网络的重现能力. 具体构建方法如下:

$$\text{KR} = \text{rank} \begin{pmatrix} \mathbf{x}(u_1) \\ \vdots \\ \mathbf{x}(u_m) \end{pmatrix}, \quad (20)$$

$$\text{GR} = \text{rank} \begin{pmatrix} \mathbf{x}(y_1) \\ \vdots \\ \mathbf{x}(y_m) \end{pmatrix}, \quad (21)$$

其中: u_m 为输入区间 $[-1, 1]$ 上均匀分布的随机信号; $y_m(t)$ 为添加干扰信号后的输入, $y_m(t) = u_m(t) + z(t)$, $z(t) \in [-0.15, 0.05]$; $x(u_m)$ 和 $x(y_m)$ 为搜集不同输入所对应的状态矩阵.

ETRC的记忆容量定义为

$$\text{MC} = \sum_{k=1}^N \text{MC}_k = \sum_{k=1}^N \frac{\text{cov}^2(u(t-k), y_k(t))}{\sigma^2(u(t))\sigma^2(y_k(t))}, \quad (22)$$

其中: $y_k(t)$ 为ESN的第 k 个输出值, 其对应的期望值为 $u(t-k)$; cov^2 和 sigma^2 分别为协方差及方差算子. MC反映了ESN输出对输入信号的复现能力, MC_k 是 $y_k(t)$ 和 $u(t-k)$ 的相关系数, 为神经元个数. Jaeger等人^[1]表示网络的记忆容量 $\text{MC} \leq N$, 所以定义 k 的最大值为 N .

由于行为空间能够表征储层性能的好坏, 所以, 采用该方法进行合理的筛选, 可以有效的提升算法的性能. 筛选策略分为初始化筛选与优化过程筛选两个部分.

$$S_i = \begin{cases} 0, & \text{KR} - \text{GR} \leq 10, \\ 0, & \text{GR} \leq 0, \\ 0, & \text{MC} \leq 3, \\ 1, & \text{其他}. \end{cases} \quad (23)$$

初始化筛选是网络需要达到的基础条件, 研究表明, 当满足式(23)条件时^[14], 往往不具备预测的能力. 依据网络行为空间的3个指标: KR, GR和MC, 判断个体是否满足最低行为配置, 其中: $S_i = 0$ 表示淘汰, $S_i = 1$ 表示保留. 优化过程筛选则是根据训练误差判定具体的筛选标准. 具体优化方法如下.

步骤 1 随机初始化 P 个不同参数的回声状态网络, 计算泛化等级GR、核心等级KR、记忆容量MC这

3个指标, 以此构建行为空间. 本文取 $P = 75$, 输入权重和储层权重初始化范围为 $[-1, 1]$ 且服从均匀分布.

步骤2 采用新颖搜索遗传算法对回声状态网络参数进行寻优, 将式(24)作为遗传算法的适应度函数. 并结合行为空间筛选机制(23)进行初次淘汰, 最终采用最优参数进行预测.

$$\rho(x) = \frac{1}{k} \sum_{i=1}^k \text{dist}(x, \xi_i), \quad (24)$$

其中: ξ_i 表示与节点 x 相邻节点, $\text{dist}(\cdot)$ 表示Euclidean距离. 由式(24)可以看出, 新颖性实质上是行为个体与 k 个最近邻个体之间的平均距离, 所以, 个体稠密区域则平均距离小, 即新颖性低, 反之, 个体稀疏区域则平均距离大, 则新颖性高.

步骤3 将训练集样本分为训练集和校验集两部分, 由步骤2可以得到初步的行为空间分布, 继而采用误差搜索遗传算法, 用于得到误差最低的解. 并采用热重启生成20个符合条件当前行为分布的储层, 作为补偿层的初始化参数.

步骤4 将步骤2中最优模型根据式(16)计算训练误差 $E(t) = (e_1, \dots, e_t)$, 并作为补偿层训练集.

步骤5 根据最终误差, 从步骤3中20个个体中选择最优的一组作为补偿层参数.

步骤6 由式(17)集成计算层与补偿层, 并输入相应的测试数据, 根据式(26)计算测试误差.

4 实验及分析

4.1 实验对象设置

选择MSO问题对ETRC的动力学特性进行研究, 主要考察神经元池能否存储多种频率的信息, MSO模型为

$$y(t) = \sum_{i=1}^Q \sin(\alpha_i t), \quad (25)$$

其中: Q 为正弦波数量, $\alpha_1 = 0.2 + 0.11(i - 1)$. 选择MSO问题主要有以下几个原因: 1) 针对常规储层计算难以解决MSO问题, 缺乏相关文献分析其内在原因; 2) MSO问题具有确定的动力学, 有利于本文进行动力学分析. 实验中采用规范化的均方根误差(normalized root MSE, NRMSE)作为测试误差指标, 其定义为

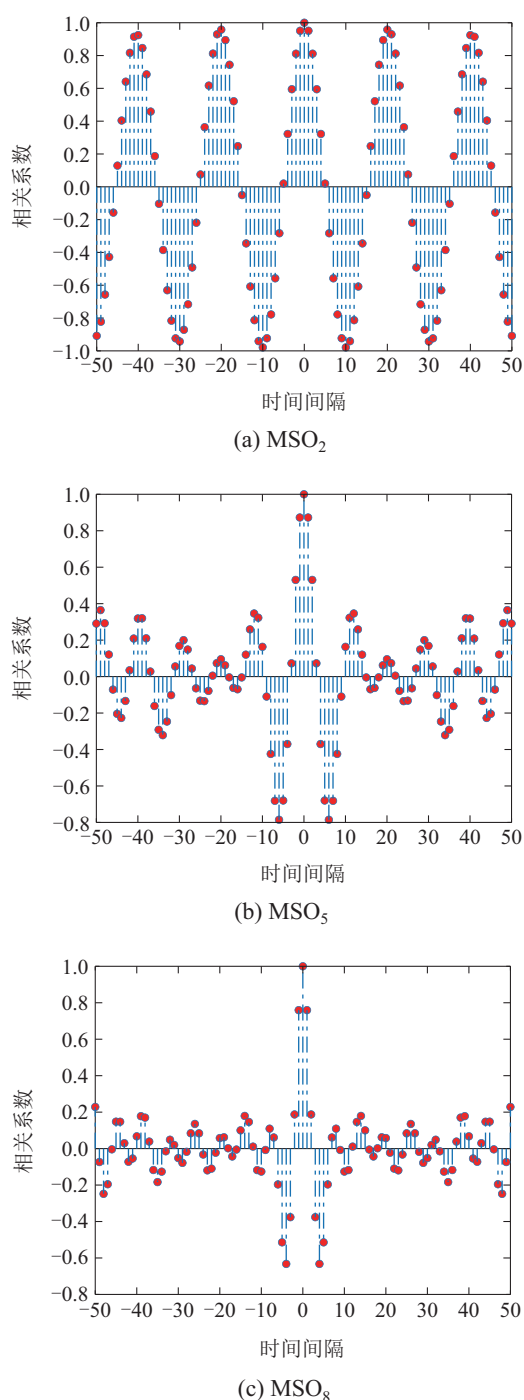
$$\text{NRMSE} = \sqrt{\frac{1}{P\sigma_y^2} \sum_{k=1}^P (\hat{y}(k) - y(k))^2}, \quad (26)$$

其中: P 为样本数量, σ_y^2 为测试样本的方差, y 和 $\hat{y}$ 分别为预测输出以及目标输出.

4.2 预测MSO问题的自相关误差分析

图3所示为常规RC在针对MSO₂, MSO₅, MSO₈, MSO₁₂任务的误差自相关系数. 从图3(a)–(b)可看出,

当MSO维数较低时, 网络与邻近时刻存在强相关性且呈现一定周期性. 从图3(c)–(d)可看出, 随着MSO的维数增加, 误差自相关周期性规律逐渐消失, 但误差仍然与前一时刻存在强相关. ESN学习过程中忽略了误差自相关性, 导致预测误差增加. 为解决以上问题, 本文所提ETRC计算层用于尽可能逼近真实值, 补偿层用于解决预测过程中可能出现的波幅预测不足与误差自相关等问题, 其实质等同于式(8)数学模型. 由于补偿层只对计算层的学习误差进行补偿, 其复杂程度远低于计算层学习的原始任务, 且网络的最终误差来源于补偿层.



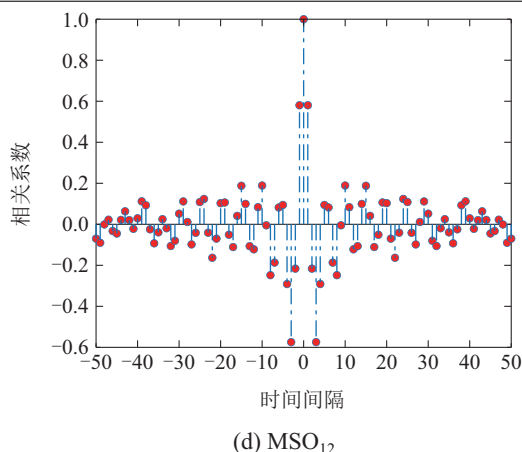


图3 常规RC预测MSO问题的自相关误差

Fig. 3 Autocorrelation error of conventional RC prediction for MSO problems

4.3 探究ESN无法预测高维MSO问题的原因

本实验使用 MSO₂, MSO₅, MSO₈, MSO₁₂ 作为测试模型, 随着维数的上升, 模型所包含的频率也随之增加. 从图 4(a)–(d) 可以看出, 针对相对简单的 MSO₂

问题, ESN拥有较好的预测效果, 但随着维数的上升, ESN的预测误差整体上升. 观察误差走势可以看出, 误差高的区域往往出现在数据波峰和波谷的区域. 原因在于传统时间序列预测时, 网络无法完全获得学习任务的全面知识, 而学习任务新信息的引入导致神经网络预测产生偏差. 所以, 一个可靠的预测模型不仅应该包括数据点的预测, 还应降低预测过程中不确定性对稳定性的影响.

从误差曲线形态上分析, 图4(a)误差曲线呈现出近似正弦形状, 图4(b)–(d)也均近似于目标函数, 但振幅远低于目标函数. 上述说明, ESN特征超平面虽能够较准确的提取出相关信号频率特征, 但由于各频率幅度较弱导致无法准确预测; 另一方面, 由于问题维数的上升, 多种不同频率信号之间的叠加导致网络稳定性下降, 而ESN本身的学习算法难以处理此类问题. 本文引入补偿层的目的是降低由于原模型对新信息学习不足而影响系统的稳定性, 其次, 可以弥补计算层忽略的模型特征, 从而提升预测精度和稳定性.

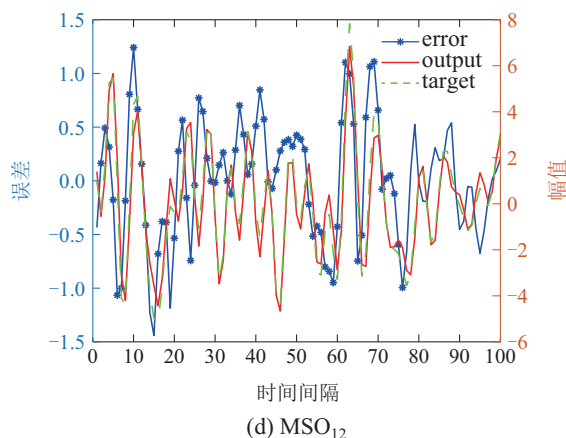
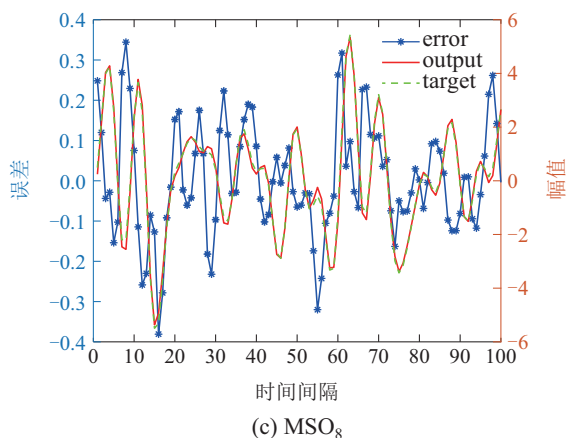
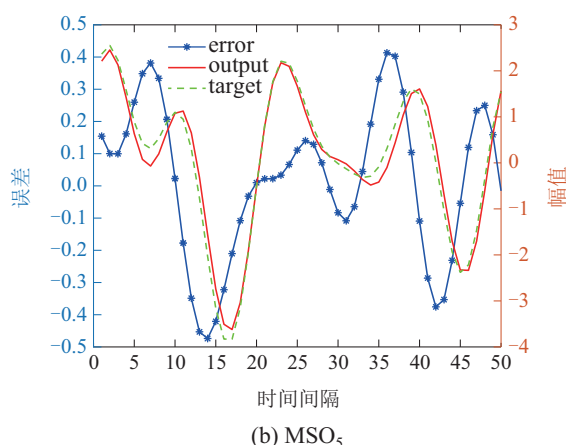
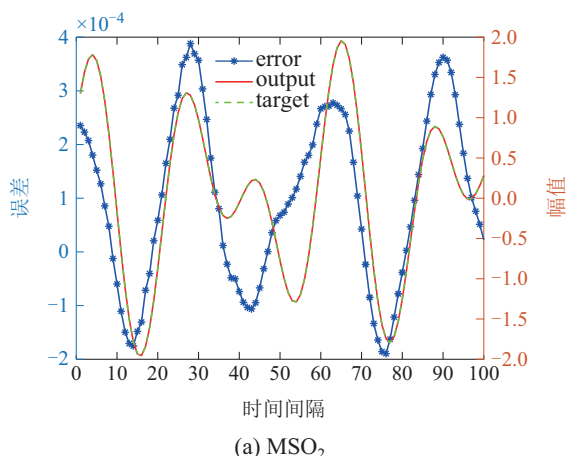


图4 常规RC预测MSO问题的真实输出和误差曲线

Fig. 4 The real output and error curves of conventional RC prediction MSO problems

4.4 实验结果与分析

为进一步验证本文所提方法的有效性, 分别将ET-RC与现有方法进行对比实验. 包括常规ESN网络、

D&S RC (delay and sum readout reservoir computing)^[28]、平衡回声状态网络 (balanced ESN)^[29] 及 TSRC^[13]. ETRC储层大小设置为200, 经30次不同初

始化的独立运行后, 测试NRMSE的平均值如表1所示.

由表1可见, ETRC相较于ESN能够用于解决更复杂的MSO问题. 随着MSO维数上升, 误差补偿的优势逐渐体现, 由于计算层预测误差增加, 从而使得补偿层的效果更加明显. 补偿层的添加弥补了计算层本身的不足, 而对于复杂度较小的问题, 由于计算层本身能够保证预测精度, 且补偿层本身也具有误差, 从而使得提升效果不明显. 图5为ETRC针对MSO₂及MSO₁₆问题的测试效果.

ETRC的鲁棒性定义如式(27)–(28)所示, 其中: e_i

为测试误差; θ 为阈值; $\rho(\theta)$ 表示在 M 次测试中误差小于 θ 的概率; $v(e_i - \theta)$ 作为判断条件, 0表示误差大于 θ 并剔除, 1表示误差保留并计算其概率. 本文设定 θ 为计算层测试误差的平均值 $\text{mean}(e_{rc})$ 与最小值 $\min(e_{rc})$, 用于研究添加补偿层后性能改善的效果.

$$\rho(\theta) = \frac{1}{M} \sum_{i=1}^M v(e_i - \theta), \quad (27)$$

$$v(e_i - \theta) = \begin{cases} 0, & e_i > \theta, \\ 1, & \text{其他}. \end{cases} \quad (28)$$

表1 不同ESN模型性能比较

Table 1 Performance comparison of different models.

数据集	ESN	D&S RC	Balanced RC	TSRC	ETRC
MSO ₂	1.48×10^{-4}	3.02×10^{-9}	2.51×10^{-12}	3.73×10^{-8}	5.70×10^{-7}
MSO ₅	—	8.21×10^{-5}	1.06×10^{-6}	6.84×10^{-7}	3.98×10^{-6}
MSO ₈	—	—	2.73×10^{-4}	6.89×10^{-6}	1.05×10^{-5}
MSO ₁₂	—	—	—	1.50×10^{-5}	9.62×10^{-5}
MSO ₁₆	—	—	—	5.50×10^{-3}	1.44×10^{-4}

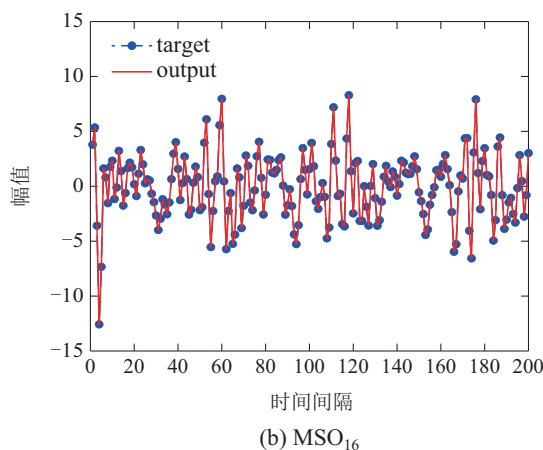
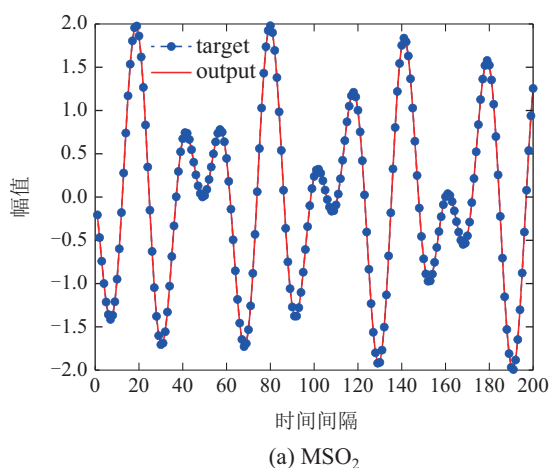


图5 ETRC对MSO₂及MSO₁₆的预测效果

Fig. 5 The predictive performance of ETRC for MSO₂ and MSO₁₆

图6为常规RC及ETRC训练时神经元池输出信号的差值. 从图6(a)可见, 针对MSO₂和MSO₁₆问题, 当 $\theta = \text{mean}(e_{rc})$ 的情况下, ETRC较常规RC优化率为100%. 说明较常规RC而言, 每一个预测点的精度均得到提升. 当 $\theta = \min(e_{rc})$ 时, 对MSO₂问题的优化率仍为100%, 但针对MSO₁₆, 问题优化率 $\rho(\theta) = 82.5\%$. 图7表示ETRC预测误差的自相关系数分布图. 从图7中可看出, 误差与上一时刻不再存在强相关.

误差补偿能够提升网络整体性能, 首先是因为MSO₂和MSO₁₆的误差自相关问题均得到改善, 其次是因为误差与原始信号频率特性相似, 使得误差补偿的实质在于增强原始网络所捕获特征的强度. 最后补偿层根据实际情况不断调整网络的误差情况, 一定程度上提升了网络的稳定性. 图6(b)可以看出, 误差整

体呈现下降趋势, 且保证误差方差的稳定性. 需要指出的是, 虽然计算层和补偿层都具有误差, 但其所需满足预测精度不同. 相对于计算层需要精确预测, 补偿层仅需满足趋势即可降低整体误差. 如图8(b)所示, 虽然预测精度降低, 但仍具备不错的补偿能力. 补偿层无法完全精确预测误差的原因在于, 误差信号中包含了一部分白噪音信号, 导致部分信息不可预测.

4.5 瓦斯浓度预测实验

实验数据采用石家庄某矿工作面4组不同区域的瓦斯浓度采样点数据, 分别为进风口、上隅角、回风口、混合回风检测点, 每间隔2 min收集一次数据. 数据共记录2021-10-30至2021-11-18时间段数据, 本实验采用18日12:12至18:12共6 h数据作为测试数据, 其余数据作为训练数据集, 测试误差为平均绝对误差

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)|, \quad (29)$$

式中: m 为预测样本总数, $\hat{y}_i$ 为真实值, y_i 为预测值。

从表2中可以看出,在4个监测点的预测任务中

ETRC模型均表现出良好的预测性能。值得注意的是,随着预测步数的增加,网络在保证预测精度的前提下,预测误差仅在小范围波动。其原因在于随着计算层误差的增加,补偿层能够保持网络整体的稳定性。

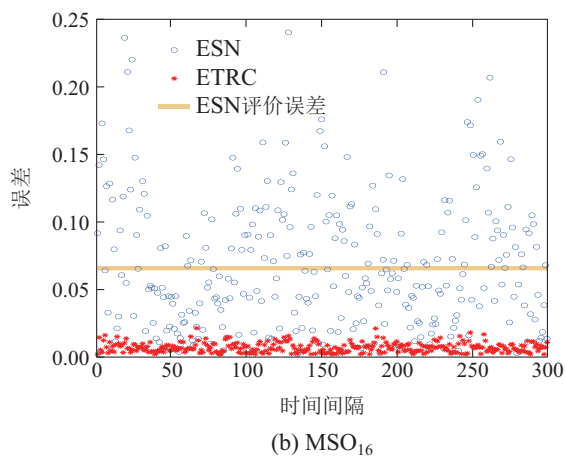
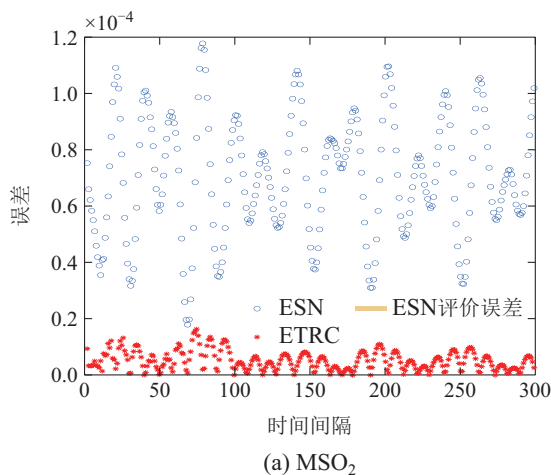


图6 ETRC与RC预测误差分布

Fig. 6 The error distributions of ETRC and RC

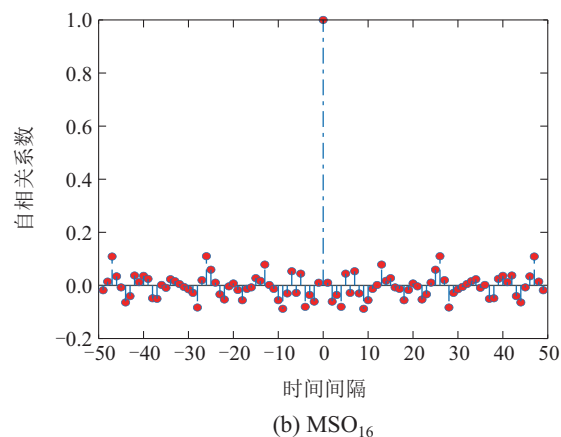
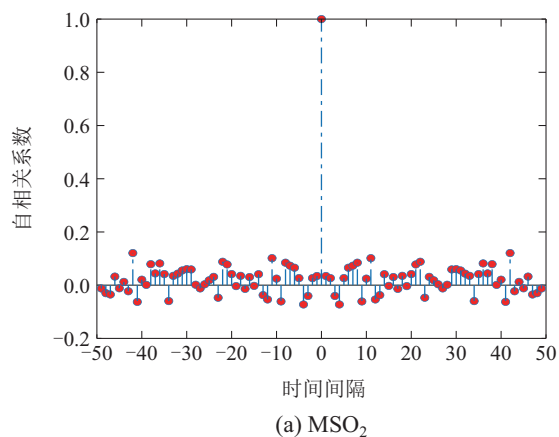


图7 ETRC误差自相关系数分布

Fig. 7 The error autocorrelation distributions of ETRC

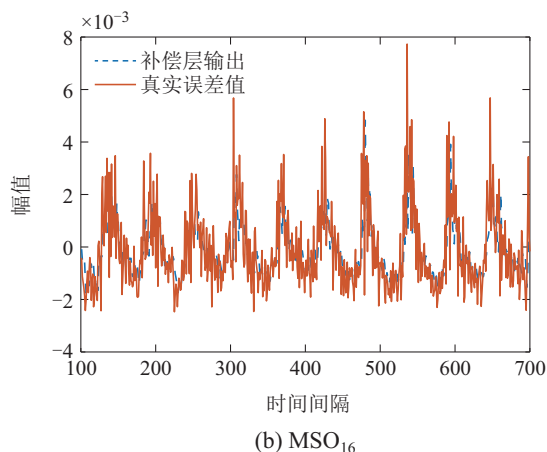
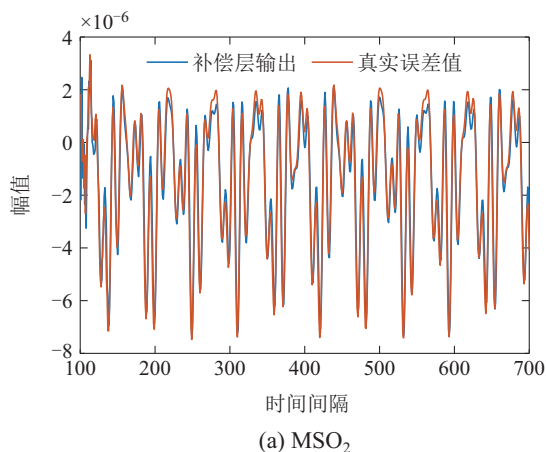


图8 ETRC补偿信号输出

Fig. 8 The compensation signal output of ETRC

表2 各监测点预测结果的平均绝对误差

Table 2 The mean absolute error of the predicted results at each monitoring point

MAE	2 min	4 min	6 min	8 min	10 min
监测1	0.0007	0.0008	0.0008	0.0009	0.0009
监测2	0.0029	0.0036	0.0036	0.0035	0.0034
监测3	0.0023	0.0035	0.0037	0.0038	0.0038
监测4	0.0034	0.0042	0.0074	0.0072	0.0073

表3为将ETRC模型与自回归滑动平均模型(auto regressive moving average, ARMA), 混沌模型Chaos和长短期记忆网络(long short-term memory, LSTM)的编码-解码(encoder-decoder, ED)模型^[30]进行对比实验. 实验结果表明, 本文所提ETRC模型对真实时间序列数据多步预测的有效性.

为了更直观的验证ETRC模型的预测效果, 图9-

10给出了ETRC模型单步与多步的预测效果图. 可以看出, ETRC对瓦斯浓度拟合较好. 从误差分布可以看出, 优化后低误差预测点个数增多, 说明点预测处于高误差的概率明显低于优化前. 但单步预测补偿效果略差于5步预测, 原因在于预测步数增加使得计算层难以捕获全部有效信息, 使得补偿层的作用愈加明显, 进一步验证了本文所提误差补偿网络结构的有效性.

表3 3号监测点瓦斯浓度数据预测结果的比较

Table 3 The comparison of prediction results of gas concentration data in No.3 monitoring site

MAE	2 min	4 min	6 min	8 min	10 min
ARMA	0.0115	0.0214	0.0287	0.0358	0.0409
CHAOS	0.0106	0.0188	0.0264	0.0331	0.0370
ED	0.007	0.0123	0.0141	0.0150	0.0165
ETRC	0.0023	0.0035	0.0037	0.0038	0.0038

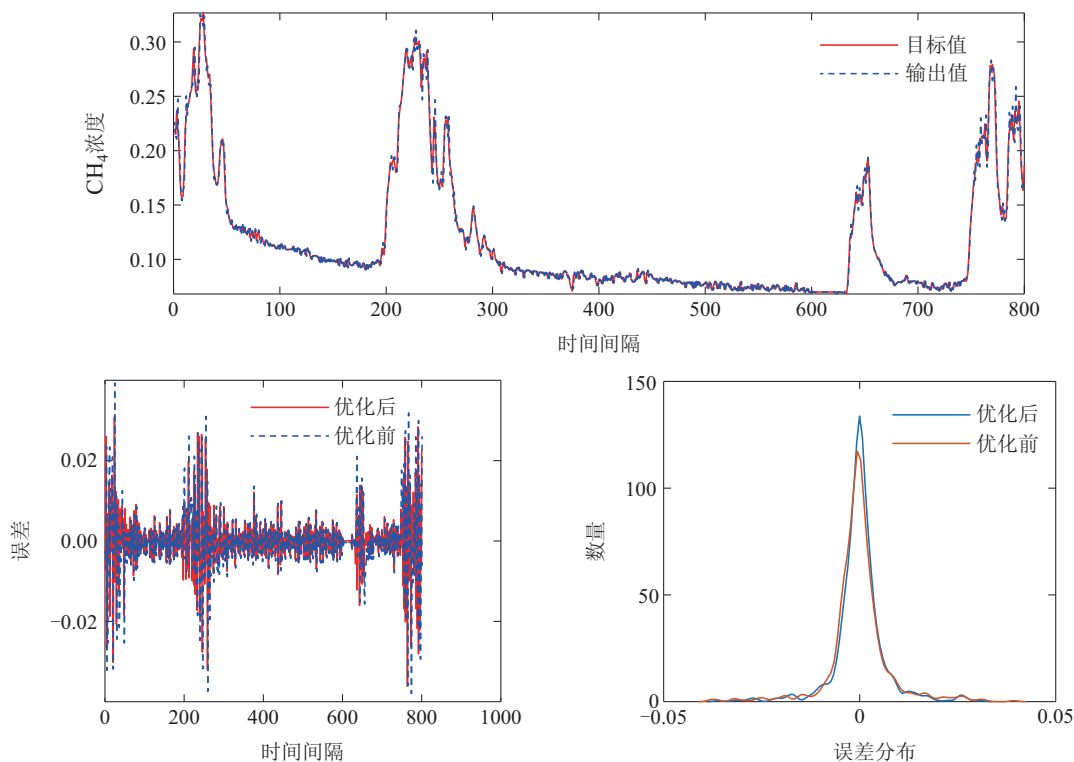
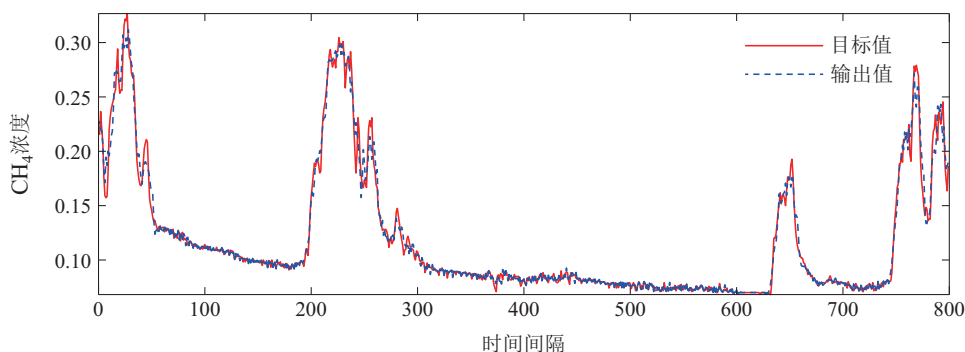


图9 监测点4瓦斯浓度单步预测效果图

Fig. 9 The performance of single step prediction gas concentration at monitoring point 4



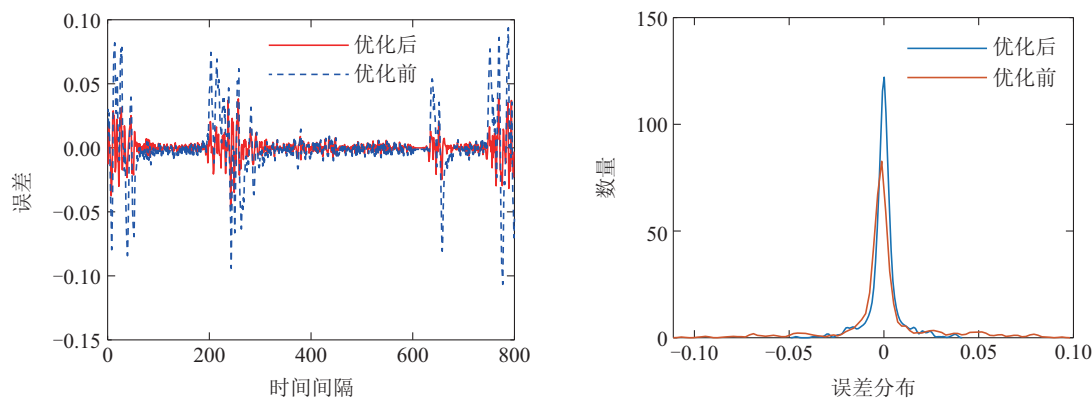


图10 监测点4瓦斯浓度5步预测效果图

Fig. 10 The performance of five step prediction gas concentration at monitoring point 4

5 结论

针对传统回声状态网络难以有效应对高阶非线性复杂模型问题, 本文从自相关误差以及高阶模型不确定性两个影响预测性能的方向出发, 在理论与实验分析的基础上, 提出一种带有补偿的回声状态网络. 在网络学习算法方面, 采用行为空间与热重启的方法改善传统优化算法存在可重复性差以及优化效率低的问题. 从MSO实验可见, ETRC框架能够有效降低由自相关误差与新信息所导致的不确定性, 尤其对于高阶MSO模型的鲁棒性和准确性大幅提升. 最后, 采用石家庄某矿工作面4组不同的真实数据进行验证, 结果表明, 本文所提ETRC网络预测准确性高于其他网络, 尤其是当预测步数增加时, ETRC网络预测精度相较于其他网络变化很小, 证明了本文所提框架ETRC的有效性和稳定性. ETRC的模型本质是将误差反馈至计算层来提升性能, 未来的工作是否可以估计输入数据和预测数据的概率密度分布, 将其KL (Kullback-Leibler)散度作为预测误差反馈至计算层. 改善计算层性能的同时, 增加概率决策判定, 仍值得进一步的研究.

参考文献:

- [1] JAEGER H, HAAS H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 2004, 304(5667): 78 – 80.
- [2] YAO X, FAN S, ZHAO B, et al. Controller design based on echo state network with delay output for nonlinear system. *Complexity*, 2020, 2020: 8643029.
- [3] SUN L, BO J, YANG H, et al. Unsupervised EEG feature extraction based on echo state network. *Information Sciences*, 2019, 475: 1 – 17.
- [4] BO Y, WANG P, ZHANG X, et al. Modeling data-driven sensor with a novel deep echo state network. *Chemometrics and Intelligent Laboratory Systems*, 2020, 206: 104062.
- [5] WANG Z, YAO X, LI T, et al. Design of PID controller based on echo state network with time-varying reservoir parameter. *IEEE Transactions on Cybernetics*, 2022, 52(7): 6615 – 6626.
- [6] HAN Min, REN Weijie, XU Meiling. An improved echo state network via L_1 -norm regularization. *Acta Automatica Sinica*, 2014, 40(11): 2428 – 2435.
- (韩敏, 任伟杰, 许美玲. 一种基于 L_1 范数正则化的回声状态网络. *自动化学报*, 2014, 40(11): 2428 – 2435.)
- [7] ZHANG Liyan, XIANG Kui, LONG Rong, et al. Unmodeled dynamics compensation and control of nonlinear system based on ESN. *Acta Electronica Sinica*, 2016, 44(1): 60 – 66.
- (张立炎, 向隳, 龙容, 等. 基于ESN的非线性系统未建模动态补偿及控制. *电子学报*, 2016, 44(1): 60 – 66.)
- [8] BO Yingchun, ZHANG Xin, LIU Bao. Delayed deep echo state network and its application on time series prediction. *Acta Automatica Sinica*, 2020, 46(8): 1644 – 1653.
- (薄迎春, 张欣, 刘宝. 延迟深度回声状态网络及其在时间序列预测中的应用. *自动化学报*, 2020, 46(8): 1644 – 1653.)
- [9] LEE Jie, LI Xiang, XU Yuanming, et al. Recent advances and prospects in industrial AI and applications. *Acta Automatica Sinica*, 2020, 46(10): 2031 – 2044.
- (李杰, 李响, 许元铭, 等. 工业人工智能及应用研究现状及展望. *自动化学报*, 2020, 46(10): 2031 – 2044.)
- [10] QIAO J, WANG L, YANG C. Adaptive lasso echo state network based on modified bayesian information criterion for nonlinear system modeling. *Neural Computing and Applications*, 2019, 31(10): 6163 – 6177.
- [11] WU Z, LI Q, ZHANG H. Chain-structure echo state network with stochastic optimization: Methodology and application. *IEEE Transactions Neural Network Learn System*, 2022, 33(5): 1974 – 1985.
- [12] HU R, TANG R, SONG X, et al. Ensemble echo network with deep architecture for time-series modeling. *Neural Computing and Applications*, 2021, 33(10): 4997 – 5010.
- [13] CHEN H, WEI D. Chaotic time series prediction using echo state network based on selective opposition grey wolf optimizer. *Nonlinear Dynamics*, 2021, 104(4): 3925 – 3935.
- [14] ZHANG Zhaozhao, ZHU Yingqin, QIAO Junfei, et al. An echo state network parameter optimization method based on behavior space. *Information and Control*, 2021, 50(5): 556 – 565.
- (张昭昭, 朱应钦, 乔俊飞, 等. 一种基于行为空间的回声状态网络参数优化方法. *信息与控制*, 2021, 50(5): 556 – 565.)
- [15] GALLICCHIO C, MICHELI A, PEDRELLI L. Deep reservoir computing: A critical experimental analysis. *Neurocomputing*, 2017, 268: 87 – 99.
- [16] BO Yingchun, ZHANG Xin, LIU Bao. Temporally segregated reservoir computing and its dynamics. *Control Theory & Applications*, 2019, 36(8): 1315 – 1321.
- (薄迎春, 张欣, 刘宝. 时间分割的池计算网络及其动力学. *控制理论与应用*, 2019, 36(8): 1315 – 1321.)
- [17] BO Yingchun, ZHANG Xin, LIU Bao, et al. Sparsely connected asynchronous reservoir computing network (in Chinese). *Scientia*

- Sinica Informationis*, 2021, 51(5): 764 – 778.
(薄迎春, 张欣, 刘宝等. 稀疏连接的异步池计算网络. 中国科学: 信息科学, 2021, 51(5): 764 – 778.)
- [18] BO Y, WANG P, ZHANG X. An asynchronously deep reservoir computing for predicting chaotic time series. *Applied Soft Computing*, 2020, 95: 106530.
- [19] WANG Lei, SU Zhong, QIAO Junfei, et al. Design of incremental regularized echo state network. *Control and Decision*, 2022, 37(3): 661 – 668.
(王磊, 苏中, 乔俊飞, 等. 基于增量式学习的正则化回声状态网络. 控制与决策, 2022, 37(3): 661 – 668.)
- [20] KIM T, KING B. Time series prediction using deep echo state networks. *Neural Computing and Applications*, 2020, 32: 17769 – 17787.
- [21] HAN Tailin, ZHANG Yanxue, WANG Xiao, et al. Application of improved grey wolf algorithm in dynamic compensation of thermocouple. *Control and Decision*, 2021, 36(1): 61 – 67.
(韩太林, 张延雪, 王啸, 等. 改进型灰狼算法在热电偶动态补偿中的应用. 控制与决策, 2021, 36(1): 61 – 67.)
- [22] LIN Di, WU Yiming, ZHU Fan. Implementation of IMM Kalman filtering feed-forward compensation techonlogy in search and track ststems. *Control and Decision*, 2020, 35(5): 1253 – 1258.
(林笛, 吴易明, 朱帆. IMM Kalman滤波前馈补偿技术在搜索跟踪系统中的应用. 控制与决策, 2020, 35(5): 1253 – 1258.)
- [23] SUN F, CHRISTOPHER I L, DUANE S B. Adjusting for autocorrelated errors in neural networks for time series regression and forecasting. *Advances in Neural Information Processing Systems*, 2021, 34: 29806 – 29819.
- [24] JAEGER H. Short term memory in echo state networks. *GMD Report 152*. Munich, Gemany: Fraunhofer Institute for Autonomous Intelligent Systems, 2002.
- [25] QIAO J, WANG L, YANG C, et al. Adaptive levenberg-marquardt algorithm based echo state network for chaotic time series prediction. *IEEE Access*, 2018, 6(99): 10720 – 10732.
- [26] BERNARDO A H, RAIAN M L, TAD H. An economics approach to hard computational problems. *Science*, 1997, 275(5296): 51 – 54.
- [27] ZHANG H, WANG Z, CHEN W, et al. Ensemble mutation-driven salp swarm algorithm with restart mechanism: Framework and fundamental analysis. *Expert Systems with Applications*, 2021, 165: 113897.
- [28] HOLZMANN G, HAUSER H. Echo state networks with filter neurons and a delay & sum readout. *Neural Networks*, 2010, 23(2): 244 – 256.
- [29] KORYAKIN D, LOHMANN J, BUTZ M. Balanced echo state networks. *Neural Networks*, 2012, 36(8): 35 – 45.
- [30] LYU P, NING C, MAO S, et al. LSTM based encoder-decoder for short-term predictions of gas concentration using multi-sensor fusion. *Process Safety and Environmental Protection*, 2020, 137: 93 – 105.

作者简介:

张昭昭 副教授, 目前研究方向为智能信息处理及神经网络结构优化设计, E-mail: zzzhao123@126.com;

朱应钦 博士研究生, 目前研究方向为智能信息处理、池网络结构优化设计, E-mail: yzhu@ctrl.cinvestav.mx;

余文 教授, 目前研究方向为神经网络、自适应控制、模糊系统, E-mail: yuw@ctrl.cinvestav.mx.