

基于改进差分进化和回声状态网络的时间序列预测研究

许美玲^{1,2} 王依雯¹

摘要 针对回声状态网络无法根据不同的时间序列有效地选择储备池参数的问题, 本文提出一种新型预测模型, 利用改进的差分进化算法来优化回声状态网络. 其中差分进化算法的缩放因子 F 、交叉概率 CR 和变异策略自适应调整, 以提高算法的寻优性能. 为验证本文方法的有效性, 对 Lorenz 时间序列、大连月平均气温 - 降雨量数据集进行仿真实验. 由实验结果可知, 本文提出的模型可以提高时间序列的预测精度, 且具有良好的泛化能力及实际应用价值.

关键词 时间序列, 预测模型, 差分进化, 回声状态网络

引用格式 许美玲, 王依雯. 基于改进差分进化和回声状态网络的时间序列预测研究. 自动化学报, 2021, 47(7): 1589-1597

DOI 10.16383/j.aas.c180549

Time Series Prediction Based on Improved Differential Evolution and Echo State Network

XU Mei-Ling^{1,2} WANG Yi-Wen¹

Abstract For the echo state network, it is difficult to select the suitable reservoir parameters for different time series. In this paper, we propose a new prediction model which uses an improved differential evolution algorithm to optimize the parameters of the echo state network. The scale factor and crossover probability of differential evolution algorithms are adaptively adjusted. In addition, offspring generation strategy is also adaptively adjusted. To verify the effectiveness of the proposed method, experiments were conducted on Lorenz time series, and monthly average temperature - rainfall time series of Dalian. Experimental results show that the model can improve forecast accuracy and has good generalization ability and practicability.

Key words Time series, predictive model, differential evolution, echo state network

Citation Xu Mei-Ling, Wang Yi-Wen. Time series prediction based on improved differential evolution and echo state network. *Acta Automatica Sinica*, 2021, 47(7): 1589-1597

时间序列广泛存在于人类社会生活的方方面面, 比如经济领域的生产总值、商品的销售量、股票的增长幅度等^[1]; 天气领域中城市的降雨量、平均气温等^[2-3]; 社会领域中城市交通流量、外来人口迁徙量及汽车能量消耗值等^[4]; 工业领域中高炉炼铁的中心温度^[5]、钢铁企业能源预测^[6]和原油物性预测^[7]等.

随着人工智能研究的不断发展, 人工神经网络、

支持向量机^[8]等机器学习方法逐渐成为非线性时间序列建模和预测的主要工具. 但是传统的人工神经网络如 BP (Back propagation) 神经网络^[9]存在训练算法复杂、网络结构难以确定、易陷入局部最优和收敛速度慢等问题. 支持向量机模型若训练样本过大, 存在训练时间过长且效果不佳的问题. 极限学习机 (Extreme learning machine, ELM)^[10]是针对单层前馈神经网络设计的新型学习算法, 学习速度快, 泛化性能好, 但其静态前馈结构不适用于动态时间序列的建模. 随着学者们的不断研究, 发现回声状态网络 (Echo state network, ESN)^[11]在时间序列预测方面存在一定的优势, 其训练算法简单, 计算快速, 且能保证解的全局最优性. 虽然回声状态网络具有以上优点, 但是也存在一些问题, 例如储备池的适应性问题、稳定性问题及病态解的出现会影响网络的泛化能力, 容易产生过拟合等.

近年来, 众多学者们针对回声状态网络存在的问题进行一些改进. 文献 [12] 提出一种基于 L1 范数正则化的改进回声状态网络, 将 L1 范数作为惩

收稿日期 2018-08-15 录用日期 2018-12-05

Manuscript received August 15, 2018; accepted December 5, 2018

国家自然科学基金 (61702077), 中央高校基本科研业务费 (DUT16RC(3)123) 资助

Supported by National Natural Science Foundation of China (61702077), Fundamental Research Funds for the Central Universities (DUT16RC(3)123)

本文责任编辑 刘艳军

Recommended by Associate Editor LIU Yan-Jun

1. 大连理工大学电子信息与电气工程学部 大连 116023 2. 东北大学智能工业数据解析与优化教育部重点实验室 沈阳 110004

1. Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023 2. Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Ministry of Education, Shenyang 110004

罚项,防止模型出现过拟合现象,提高模型的稳定性.文献[13]利用改进的小世界网络优化泄露积分型回声状态网络,有效地改进了储备池神经元的连接方式,减少了稀疏连接的随机性,提高了储备池的适应性.文献[14]提出拉普拉斯回声状态网络解决了由于实际训练样本的数量小于隐藏层节点数存在的不适应问题并获得低维的输出权重.文献[15]提出一种鲁棒回声状态网络预测模型,采用更具鲁棒性的拉普拉斯分布代替高斯分布对数据的噪声进行描述,消除了异常点对模型泛化能力的影响.文献[16]结合回声状态网络和 Kalman 滤波,利用 Kalman 滤波直接对网络的输出权值进行在线更新,扩展了算法的适用范围,提高预测精度.文献[17]利用遗传算法直接优化 ESN,实现自适应非线性动力系统的控制.文献[18]采用人工鱼群算法优化回声状态网络,满足了聚氯乙烯聚合过程的实时控制要求.文献[19]利用粒子群优化算法对 ESN 中未经训练的权重进行优化,提高网络的预测性能.

虽然已有学者采用智能优化方法优化回声状态网络参数,但由于回声状态网络的储备池包含节点较多、搜索空间较大,优化效果还有待改善.遗传算法 (Genetic algorithm, GA)^[17, 20] 通过模拟自然进化过程搜索最优解,但容易过早收敛,存在早熟现象.粒子群算法 (Particle swarm optimization, PSO)^[21] 的鲁棒性较差而且容易陷入局部最优解,参数的初始化对粒子群算法的性能影响较大.蚁群算法 (Ant colony optimization)^[22] 尽管具有较强的鲁棒性,但是有明显的经验性,收敛慢,对实际问题适用性不强.人工鱼群算法 (Artificial fish, AF)^[23] 虽然不易陷入局部最优,但不适合同时优化不同范围或者范围相差较大的参数,而且人工鱼群算法本身较为复杂,效率低.研究中发现差分进化算法^[24] 简单,具有较强的全局搜索能力,寻优速度快,且能平衡局部及全局的信息来进行搜索,更具实用性,鲁棒性.教与学优化算法 (Teaching-learning-based optimization, TLBO)^[25] 原理简单、易实现,需要调优的参数极少,且计算效率比传统的方法计算效率高.该算法自提出以来,已被广泛用于函数优化、神经网络优化、工程优化等领域.然而,每种算法都不是万能的,教与学优化算法也存在早熟收敛现象.综上,本文选用全局性较强的差分进化算法.

为使差分算法可以更加准确地找到最优解及适用性更强.本文提出一种改进的差分算法,一方面对缩放因子 (Scale factor, F) 和交叉概率 (Cross rate, CR) 采用自适应,另一方面对变异策略采用自适应.综上所述,本文利用改进的差分进化算法来

优化回声状态网络中储备池的 4 个重要参数,对每个个体都分配一个适合的 F 、 CR 和变异策略,使个体变异更加趋向于适应度好的个体,从而加快该算法的收敛速度,更准确地找到最优解,提高预测精度.

1 回声状态网络

Jaeger 等^[11, 26] 提出的回声状态网络 (Echo state network, ESN) 是一种简化的递归神经网络,用稀疏连接的储备池代替全连接的隐含层,增强了对动态系统的建模能力,避免一般神经网络基于梯度下降原理的收敛速度慢、易陷入局部最优的问题.

1.1 ESN 基本原理

如图 1 所示,回声状态网络是一种三层递归神经网络,由输入层、隐含层和输出层三部分组成,隐含层又称储备池,含有成百上千个稀疏递归连接的神经元,神经元之间的连接权值随机生成且固定不变.

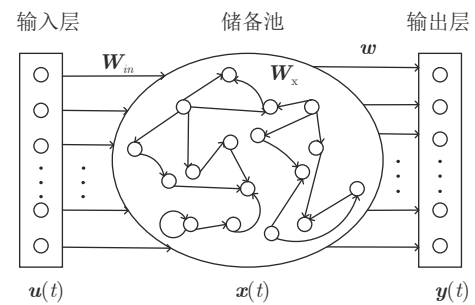


图 1 ESN 结构示意图

Fig.1 Structure of ESN

ESN 的状态方程和输出方程如下:

$$\mathbf{x}(t) = \varphi\{\mathbf{W}_{in}\mathbf{u}(t) + \mathbf{W}_x\mathbf{x}(t-1) + \mathbf{b}_x\} \quad (1)$$

$$\mathbf{y}(t) = \mathbf{w}^T\mathbf{x}(t) + \mathbf{b} \quad (2)$$

式中, $\mathbf{u}(t) \in \mathbf{R}^{M \times 1}$ 是 M 元输入向量, $\mathbf{y}(t) \in \mathbf{R}^{M \times 1}$ 是 M 元输出向量, $\mathbf{b}_x \in \mathbf{R}^{N \times 1}$ 为输入偏置, $\mathbf{b} \in \mathbf{R}^{M \times 1}$ 为输出偏置. 由当前时刻 t 的输入 $\mathbf{u}(t)$ 和前一时刻 $t-1$ 的储备池内部状态 $\mathbf{x}(t-1)$ 通过激活函数的映射得到当前时刻的状态 $\mathbf{x}(t) \in \mathbf{R}^{N \times 1}$. $\varphi(\cdot)$ 为神经元的激活函数,可以选取 Sigmoid 函数或者 tanh 函数. 输入连接矩阵 $\mathbf{W}_{in} \in \mathbf{R}^{N \times M}$ 的元素在区间 $[-1, 1]$ 取值. $\mathbf{W}_x \in \mathbf{R}^{N \times N}$ 为储备池内部连接矩阵,稀疏连接. 通过伪逆求得输出连接矩阵 $\mathbf{w} \in \mathbf{R}^{N \times M}$. 输入连接矩阵 \mathbf{W}_{in} 和内部的连接矩阵 \mathbf{W}_x 随机生成,且在回声状态网络的训练阶段始终保持不变. 网络只需求输出连接矩阵 \mathbf{w} , 因此降低

了计算复杂度.

1.2 网络关键参数

回声状态网络的核心是储备池, 储备池性能的好坏取决于 4 个重要的参数: 储备池规模 N 、储备池的谱半径 R 、稀疏度 D 以及输入变换因子 S , 如何选取这 4 个参数至关重要. 下面介绍储备池的参数选择对模型性能的影响.

1) 储备池规模

储备池的规模 N 即为储备池中神经元数目, 是影响 ESN 预测性能最重要的参数. N 的值过大, 造成过拟合; N 的值过小, 造成欠拟合.

2) 储备池的谱半径

谱半径 R 为内部连接矩阵 \mathbf{W}_x 的最大特征值的绝对值. R 的取值范围一般为 $[0, 1]$ 之间, 但对于不同的时间序列其取值将视情况而定.

3) 稀疏度

储备池内部神经元连接的稀疏程度称为稀疏度 D . 储备池的神经元之间不是全连接, 而是少部分连接. 具体实现方法是使储备池的连接权值 \mathbf{W}_x 中的大多数元素等于零. Jaeger 等认为稀疏度 $D \in [0, 0.1]$ 即可保证储备池具有足够的动力特性.

4) 输入变换因子

输入变换因子 S 是指在信号输入储备池前缩放的比例因子, 表征输入连接权值的取值范围. 根据式 (1) 可知, 其大小决定激活函数的工作区间, 也决定了输入对储备池的状态变量作用的强度. 其取值范围通常在 $[0, 1]$ 之间.

回声状态网络参数的选择往往会针对数据的不同特性而变化, 因此如何选择适合不同数据的储备池参数是本文研究的重点. 本文采用差分进化算法来自动选择适合当前数据的储备池的参数, 提高网络的预测性能.

2 改进差分进化算法

差分进化算法 (Differential evolution, DE) 是基于群体智能的优化算法^[24]. 由于算法简单易于实现、鲁棒性好、搜索能力强等优点被广泛应用于电力、炼油、工业及科学研究等领域^[27-28].

2.1 差分进化算法

一般来说, 常常选用差分进化算法解决模型参数优化问题. 标准的差分进化算法包括 4 个操作: 初始化、变异、交叉和选择. 下面介绍这 4 个操作的具体实现.

1) 随机初始化个体数量为 NP , 迭代次数为 G

的一个种群, 记 $\mathbf{X} = (\mathbf{X}_{1,G}, \mathbf{X}_{2,G}, \dots, \mathbf{X}_{i,G}, \dots, \mathbf{X}_{NP,G})$, 其中 $\mathbf{X}_{i,G} = (x_{i,G}^1, x_{i,G}^2, \dots, x_{i,G}^j, \dots, x_{i,G}^Q)$, Q 为待优化问题的维数, 初始化公式:

$$\mathbf{X} = (\mathbf{X}_{\max} - \mathbf{X}_{\min}) \times \text{rand}(NP, Q) + \mathbf{X}_{\min} \quad (3)$$

\mathbf{X}_{\max} 和 \mathbf{X}_{\min} 分别为种群个体范围的上限和下限, $\text{rand}(\cdot)$ 函数表示生成范围在 $(0, 1)$ 之间的随机数.

2) 变异操作则是利用变异策略来产生新的个体, 有利于后代进行搜索以便找到最好的解. 变异策略有:

$$\mathbf{V}_{i,G} = \mathbf{X}_{r_1,G} + F \times (\mathbf{X}_{r_2,G} - \mathbf{X}_{r_3,G}) \quad (4)$$

$$\mathbf{V}_{i,G} = \mathbf{X}_{r_1,G} + F \times (\mathbf{X}_{r_2,G} - \mathbf{X}_{r_3,G}) + F \times (\mathbf{X}_{r_4,G} - \mathbf{X}_{r_5,G}) \quad (5)$$

$$\mathbf{V}_{i,G} = \mathbf{X}_{\text{best},G} + F \times (\mathbf{X}_{r_1,G} - \mathbf{X}_{r_2,G}) \quad (6)$$

$$\mathbf{V}_{i,G} = \mathbf{X}_{\text{best},G} + F \times (\mathbf{X}_{r_1,G} - \mathbf{X}_{r_2,G}) + F \times (\mathbf{X}_{r_3,G} - \mathbf{X}_{r_4,G}) \quad (7)$$

$$\mathbf{V}_{i,G} = \mathbf{X}_{i,G} + F \times (\mathbf{X}_{\text{best},G} - \mathbf{X}_{i,G}) + F \times (\mathbf{X}_{r_1,G} - \mathbf{X}_{r_2,G}) \quad (8)$$

$$\mathbf{V}_{i,G} = \mathbf{X}_{r_1,G} + F \times (\mathbf{X}_{\text{pbest},G} - \mathbf{X}_{r_2,G}) + F \times (\mathbf{X}_{r_3,G} - \mathbf{X}'_{r_4,G}) \quad (9)$$

得到变异种群 $\mathbf{V}_G = (\mathbf{V}_{1,G}, \mathbf{V}_{2,G}, \dots, \mathbf{V}_{i,G}, \dots, \mathbf{V}_{NP,G})$, 其中 $\mathbf{V}_{i,G} = (v_{i,G}^1, v_{i,G}^2, \dots, v_{i,G}^j, \dots, v_{i,G}^Q)$, 代表变异的新个体, $i = 1, 2, \dots, NP$, $j = 1, 2, \dots, Q$. r_1, r_2, r_3, r_4, r_5 为随机生成的数, 代表不同于 i 的个体, $\mathbf{X}_{\text{best},G}$ 代表当前种群适应度最好的个体, F 为缩放因子. 变异策略 “DE/current-to-pbest/1” 由 Zhang 等提出^[29], 如式 (9), 将当前种群按适应度排序, 选择出适应度较好的前 $p\%$ 个个体, $\mathbf{X}_{\text{pbest},G}$ 从中随机选择. 该变异策略带有小范围的外部存档 arch 来存储迭代时竞争失败的父代个体, 大小为 NP . 若外部存档 arch 的个体数大于 NP , 则随机删除多余的个体. $\mathbf{X}'_{r_4,G}$ 从当前种群和外部 arch 中随机选择, 提高种群多样性. F 为缩放因子, 其决定着每个个体变异的尺度来得到新的个体, 因此它对差分算法的性能有着决定性影响.

3) 交叉操作的交叉公式为:

$$\mathbf{u}_{i,G}^j = \begin{cases} v_{i,G}^j, & \text{rand} \leq CR \text{ 或 } j = j_{\text{rand}} \\ x_{i,G}^j, & \text{其他} \end{cases} \quad (10)$$

其中, $i = 1, 2, \dots, NP$, $j = 1, 2, \dots, Q$, 生成的试验向量为 $\mathbf{U}_G = (\mathbf{U}_{1,G}, \mathbf{U}_{2,G}, \dots, \mathbf{U}_{i,G}, \dots, \mathbf{U}_{NP,G})$, 每个试验个体为 $\mathbf{U}_{i,G} = (u_{i,G}^1, u_{i,G}^2, \dots, u_{i,G}^Q)$, CR 为交叉概率因子.

4) 选择操作就是对新旧个体进行淘汰制操作, 规则就是比较新旧个体的适应度, 适应度好的个体顺利进入下一代, 适应度差的就被淘汰, 以此来获得适应度较好的种群. 选择公式:

$$\mathbf{X}_{i,G+1} = \begin{cases} \mathbf{X}_{i,G}, & f(\mathbf{X}_{i,G}) \leq f(\mathbf{U}_{i,G}) \\ \mathbf{U}_{i,G}, & f(\mathbf{X}_{i,G}) > f(\mathbf{U}_{i,G}) \end{cases} \quad (11)$$

式中, $\mathbf{X}_{i,G}$ 为第 G 代的第 i 个个体, 如果目标矢量 $\mathbf{X}_{i,G}$ 的适应度比试验矢量 $\mathbf{U}_{i,G}$ 适应度好, 则 $\mathbf{X}_{i,G}$ 进入下一代成为 $\mathbf{X}_{i,G+1}$; 反之, $\mathbf{U}_{i,G}$ 进入下一代成为 $\mathbf{X}_{i,G+1}$.

2.2 参数的自适应

在标准的差分进化算法中, 缩放因子 F 和交叉概率 CR 是一个确定值, 意味着种群中的每个个体都是用同样的 F 和 CR 来进行后面的变异和交叉操作, 但是合适的参数选择通常与个体成员相关, 不同的个体对应着不同的控制参数. 为了使参数 F 和 CR 的选择适合每一个个体, 本文采用自适应方法, 控制参数在进化的过程中根据个体适应度之间的关系自适应变化.

1) 缩放因子 F 的自适应

从变异策略方程式可知, 差分矢量实际上是对基向量各维的扰动, F 则控制扰动量的大小. 如果生成差分矢量的两个个体 $\mathbf{X}_{p,G}$ 和 $\mathbf{X}_{q,G}$ 在搜索空间中离得很远, 则生成的差分矢量比较大, 此时 F 应减小, 否则对基向量 $\mathbf{X}_{b,G}$ 的各维的扰动量太大, 将会使变异个体超出整个搜索空间, 不利于局部搜索. 如果 $\mathbf{X}_{p,G}$ 和 $\mathbf{X}_{q,G}$ 在搜索空间中的位置相近, 则生成的差分矢量的值 ($\mathbf{X}_{p,G} - \mathbf{X}_{q,G}$) 就会很小, 此时 F 应增大, 否则基向量 $\mathbf{X}_{b,G}$ 的各维的扰动量太小, 就起不到变异的作用, 不利于在进化初期的全局搜索. 因此, 本文采用自适应策略为每个个体选择合适的缩放因子 F_i , 在全局搜索和局部搜索之间取得平衡.

$$F_i = F_l + (F_u - F_l) \frac{f_p - f_b}{f_q - f_b} \quad (12)$$

式中, f_b, f_p, f_q 分别为个体 $\mathbf{X}_{b,G}, \mathbf{X}_{p,G}, \mathbf{X}_{q,G}$ 的适应度, F_l 和 F_u 分别为 F_i 给定的最小值和最大值. 上式表明如果 $\mathbf{X}_{p,G}$ 和 $\mathbf{X}_{q,G}$ 的适应度相差很小, 说明这两个个体在搜索空间里离得很近, 则 F_i 取值大, 以防止对基向量的各维扰动过小; 如果 $\mathbf{X}_{p,G}$ 和 $\mathbf{X}_{q,G}$ 的适应度相差很大, 说明这两个个体在搜索空间里位置离得很远, 则 F_i 取值小, 以限制扰动量过大.

2) 交叉概率 CR 的自适应

从交叉策略方程式可知, CR 是用来控制变异矢

量 $\mathbf{V}_{i,G}$ 对试验矢量 $\mathbf{U}_{i,G}$ 的贡献的. 如果 CR 过大, 则控制变异矢量 $\mathbf{V}_{i,G}$ 对试验矢量 $\mathbf{U}_{i,G}$ 的贡献越多, 对目标矢量的破坏也就越大, 从而使原本适应度很好的个体结构遭到破坏, 不利于种群的进化. 如果 CR 过小, 则不易产生新的个体, 会使整个搜索过程缓慢甚至停止, 不利于种群后代的搜索. 因此, CR 的取值应该根据个体的适应度的变化而变化, 对适应度很好的个体, 此时 CR 应减小, 避免对该个体造成破坏, 使其进入下一代的机会更大; 对适应度差的个体, 此时 CR 应增大, 有利于改变该个体的结构则使产生适应度好的个体的可能更大. 每个个体的交叉概率 CR_i 计算如下:

$$CR_i = \begin{cases} CR_l + (CR_u - CR_l) \frac{f_i - f_{\min}}{f_{\max} - f_{\min}}, & f_i \geq \bar{f} \\ CR_l, & f_i < \bar{f} \end{cases} \quad (13)$$

式中, CR_u, CR_l 是 CR_i 给定的最大值和最小值, f_i, \bar{f}, f_{\min} 和 f_{\max} 分别代表第 i 个个体的适应度、种群适应度的平均值、最优个体的适应度和最差个体的适应度. 上式表明如果当前第 i 个个体的适应度大于当前种群适应度的平均值, 说明该个体不好, 此时 CR 应增大; 如果该个体的适应度小于当前种群适应度的平均值, 说明该个体性能好, 此时 CR 应减小, 使该个体更可能进入下一代.

2.3 变异策略的自适应

在标准的差分进化算法中, 变异策略只使用一种, 意味着种群中的每个个体都使用同样的变异策略进行变异, 但是合适的变异策略选择通常与问题和初始的种群息息相关, 不同的优化问题对应着不同的变异策略, 且初始化具有随机性. 为了使变异策略的选择独立于优化问题和初始化, 根据文献 [30], 本文采用自适应方法, 使变异策略在进化的过程中, 在指定的策略池中内自适应地选择, 策略池包括的策略: “DE/rand/1”、“DE/rand/2”、“DE/target-to-best/1”和 “DE/current-to-pbest/1”^[29], 如式 (4)、(5)、(8)、(9).

首先在 LP 代及 LP 代之前, 策略池中的每个策略被选择的概率是一样的, 假设策略池有 M 个策略, 那么初始时每个变异策略被选择的概率是 $1/M$. 每次迭代时, 将通过第 m 个策略生成进入下一代个体的数量记作 $n_{s,G}^m$, 将通过第 m 个策略生成的个体没有成功进入下一代的个体数量记作 $n_{f,G}^m$. 从 $LP + 1$ 代开始, 根据之前 $n_{s,G}^m$ 和 $n_{f,G}^m$ 的记忆, 来更新策略池中每个策略被选择的概率. 例如, 在 G 代时, 第 m 个策略被选择的概率为:

$$p_G^m = \frac{S_G^m}{\sum_{m=1}^M S_G^m} \quad (14)$$

$$S_G^m = \frac{\sum_{g=LP}^{G-1} n_{s,g}^m}{\sum_{g=LP}^{G-1} n_{s,g}^m + \sum_{g=LP}^{G-1} n_{f,g}^m} + \varepsilon \quad (15)$$

式中, $m = 1, 2, \dots, M$, $G > LP$, S_G^m 表征当前 G 代中第 m 个策略生成的试验矢量成功进入下一代的可能性大小. 为了使每个策略被选择的概率和为 1, 则将 S_G^m 除以各个策略的成功率的和. 为避免变异策略生成进入下一代个体的成功率为 0, 本文中 ε 的值设为 0.01.

3 改进的差分进化算法优化储备池参数

ESN 优点在于网络学习过程中仅需调节储备池到输出层之间的输出连接权值, 而其他连接权值一般都随机赋值后保持不变. 储备池参数设置直接影响 ESN 的预测性能, 人工调节参数既费时又不能选择出最适合的参数值, 所以本文提出基于改进的差分进化算法优化回声状态网络 (Improved differential evolution-echo state network, IDE-ESN) 的参数. 本文将实验数据分为训练集、测试集两部分, 目标函数为训练集误差最小. 具体阐述如算法 1 所示, 流程图如图 2 所示.

4 仿真实验

为验证本文所提模型的有效性, 本文选择 Lorenz 时间序列、大连月平均气温 - 降雨量数据集进行仿真实验. 同时, 在相同的数据集上, 与人工鱼群^[18] 优化 ESN (AF-ESN)、粒子群^[19] 优化 ESN (PSO-ESN)、教与学优化算法^[25] 优化 ESN (TLBO-ESN) 和极限学习机预测模型 (ELM)^[10] 的仿真结果进行比较. 模型预测性能好坏的评价指标为均方根误差 (Root mean square error, RMSE)、平均对称绝对误差 (Symmetric mean absolute percentage error, SMAPE) 和标准化均方根误差 (Normalized root mean squared error NRMSE). RMSE、SMAPE 和 NRMSE 的计算公式定义如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y'(t) - y(t))^2} \quad (16)$$

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y'(t) - y(t)|}{\frac{1}{2} |y'(t) + y(t)|} \quad (17)$$

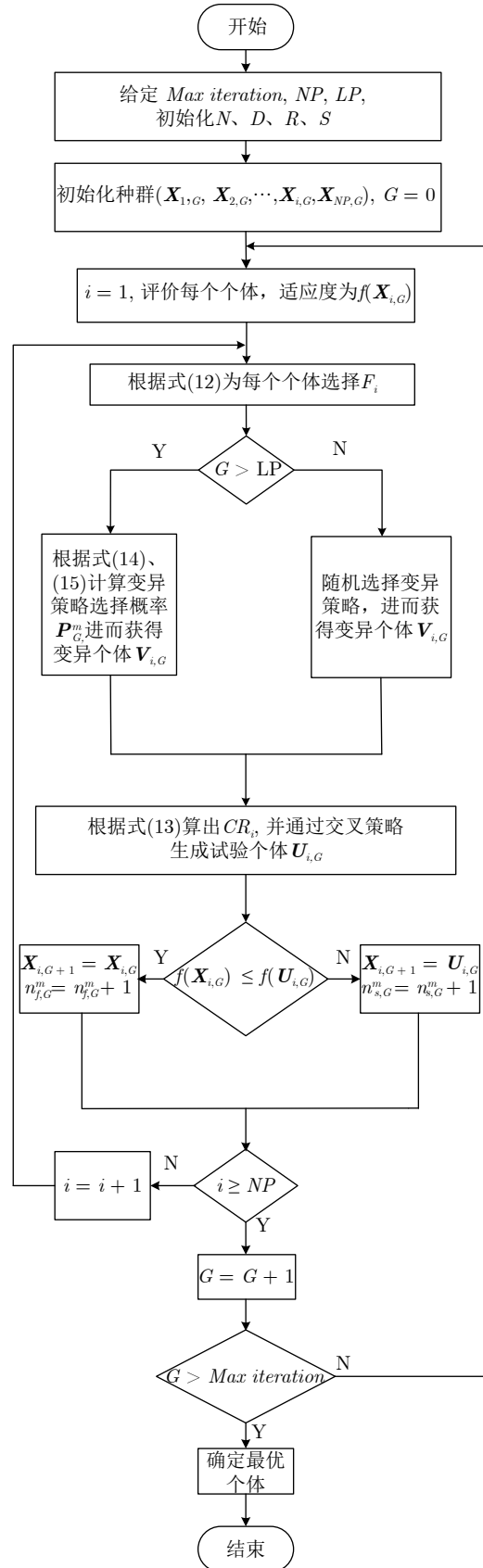


图 2 IDE-ESN 算法流程图
Fig.2 Flow chart for IDE-ESN

$$\text{NRMSE} = \sqrt{\frac{\sum_{t=1}^n [y'(t) - y(t)]^2}{\sum_{t=1}^n [\bar{y}(t) - y(t)]^2}} \quad (18)$$

式中, $y(t)$ 为测试数据的真实值, $y'(t)$ 为网络预测值, $\bar{y}(t)$ 为真实值的平均值, n 为测试集的大小.

4.1 Lorenz 时间序列

Lorenz 系统方程描述如下:

$$\begin{cases} \frac{dx}{dt} = a(-x + y) \\ \frac{dy}{dt} = bx - y - xz \\ \frac{dz}{dt} = xy - cz \end{cases} \quad (19)$$

当 $a = 10$, $b = 28$ 和 $c = 8/3$ 时, 系统是具有混沌特性. 采用龙格库塔方法产生 2 500 组离散时间序列.

算法 1. 改进差分进化回声状态网络

步骤 1. 随机初始化种群 $\mathbf{X} = (\mathbf{X}_{1,G}, \mathbf{X}_{2,G}, \dots, \mathbf{X}_{i,G}, \dots, \mathbf{X}_{NP,G})$, NP 为种群个体数, $G = 0$ 为初始代数. 种群个体的每一维代表储备池的一个参数, 每一个参数都有选择范围, 因此对个体的每一维进行约束.

步骤 2. 将个体的每一维分别赋值给储备池对应的参数: 储备池规模 N , 稀疏度 D , 谱半径 R 和输入变换因子 S , 进行适应度评价, 选出最优个体. 适应度评价为计算 ESN 模型在训练集上的均方根误差.

步骤 3. 根据式 (12) 来选择每个个体对应的缩放因子 F_i .

步骤 4. 当迭代次数 $G \leq LP$ 时, 在变异策略池中随机选择变异策略, 生成变异个体 $V_{i,G}$. 当迭代次数 $G > LP$ 时, 根据前 $G - LP$ 代得出的 $n_{s,G}^m$ 和 $n_{f,G}^m$, 依式 (14) 和 (15) 计算每个变异策略被选择的概率 p_G^m , 并选择概率最大的变异策略来产生变异个体 $V_{i,G}$.

步骤 5. 用式 (13) 为每个目标个体选择 CR_i , 通过交叉策略, 即式 (10) 来生成试验个体 $U_{i,G}$.

步骤 6. 通过对原个体与试验个体进行适应度评价比较, 选出适应度好的个体来进行种群更新. 即若 $f(\mathbf{X}_{i,G}) > f(\mathbf{U}_{i,G})$, 则 $\mathbf{X}_{i,G+1} = \mathbf{U}_{i,G}$, $n_{s,G}^m = n_{s,G}^m + 1$. 若 $f(\mathbf{X}_{i,G}) \leq f(\mathbf{U}_{i,G})$, 则 $\mathbf{X}_{i,G+1} = \mathbf{X}_{i,G}$, $n_{f,G}^m = n_{f,G}^m + 1$. 并更新最优个体.

步骤 7. 在迭代的过程中, 种群中的个体慢慢趋于相同, 为了增加种群的多样性, 避免陷入局部最

优, 在迭代次数大于 5 时, 随机选取 $NP/5$ 个体, 根据 Logistic 混沌映射产生与原个体差异较大的新的个体. 若迭代次数 G 大于最大迭代次数 *Max iteration*, 执行步骤 8; 反之执行步骤 3.

步骤 8. 输出最优个体, 在测试集数据上进行验证.

改进的差分进化算法优化 ESN 模型参数的范围设定: 储备池规模 N 范围设为 $[20, 100]$ 、稀疏度 D 范围设为 $[0.01, 0.5]$ 、谱半径 ρ 范围设为 $[0.1, 1]$ 及输入变换因子 S 范围设为 $[0.0001, 0.1]$; 种群大小 NP 设为 25, 最大迭代次数设为 30. 缩放因子 F 的范围设为 $[0.1, 0.9]$, 交叉概率 CR 的范围设为 $[0.1, 0.9]$. 前 70% 数据用于训练, 30% 数据用于测试, 其中训练样本中舍弃前 50 个样本产生的状态以消除初始暂态的影响. 人工鱼群优化 ESN 模型及粒子群优化 ESN 模型的参数设置同改进差分优化 ESN 模型. 表 1 给出了对于 Lorenz 序列 IDE-ESN 选出的最好的参数值.

表 1 Lorenz- $x(t)$ 序列: IDE-ESN 模型参数

Table 1 Lorenz- $x(t)$ series: parameters in IDE-ESN

储备池参数	取值
储备池规模	50
稀疏度	0.0210
谱半径	0.9589
输入变化因子	0.0600

表 2 给出了不同模型对 Lorenz- $x(t)$ 的单步预测结果, 可以看出本文所提模型在 RMSE、SMAPE 及 NRMSE 方面均好于其他对比模型, 更具优势. 其中, 人工鱼群优化 ESN 模型的预测精度不及粒子群优化 ESN 模型, 可以看出同时优化多个范围不同的参数时, 人工鱼群优化 ESN 模型存在局限性. 图 3 给出了 IDE-ESN 对 Lorenz- $x(t)$ 测试数据的预测曲线和误差曲线. 从图中可以看出, 预测误差很小, IDE-ESN 能够很好地拟合 Lorenz- $x(t)$ 序列. 得到各个算法误差指标如表 2 所示.

表 2 Lorenz- $x(t)$ 序列: 测试集仿真结果

Table 2 Lorenz- $x(t)$ time series: prediction results on the test dataset

模型	RMSE	SMAPE	NRMSE
AF-ESN	2.0850E-06	1.8571E-07	2.7992E-07
PSO-ESN	1.0139E-06	1.0211E-07	1.3613E-07
ELM	1.8422E-03	6.6638E-04	2.1061E-04
TLBO-ESN	7.7210E-07	1.6737E-07	1.0528E-07
IDE-ESN	3.2156E-07	9.8008E-08	4.3089E-08

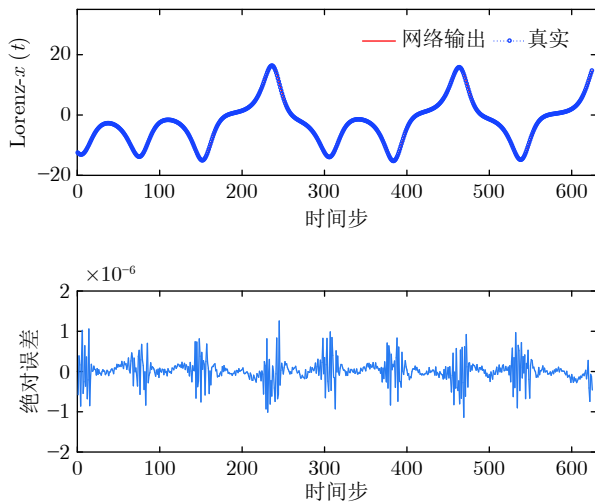


图3 Lorenz- $x(t)$ 序列: IDE-ESN 的预测曲线及误差曲线
Fig.3 Lorenz- $x(t)$ series: prediction and error curves obtained by IDE-ESN

表3给出对于Lorenz序列不同模型在30次迭代次数下运行时间,可以看出AF-ESN模型运行时间最长,该模型的时间复杂度最大. PSO-ESN的时间复杂度最小,本文所提模型IDE-ESN运行时间虽然比PSO-ESN长,但是预测精度高,综合来看,IDE-ESN模型更具优势.图4给出对于Lorenz时间序列,本文所提出的改进的差分进化算法优化ESN (IDE-ESN)、粒子群优化ESN (PSO-ESN)、人工鱼群优化ESN (AF-ESN)及教与学优化算法优化ESN (TLBO-ESN)在迭代过程中的适应度值(Fitness)的变化曲线图.为更能清楚地显示各个模型适应度曲线的差别,对适应度的值取以30为底的对数.从图中可以看出IDE-ESN算法在迭代的过程中,误差越来越小,收敛速度较快并最终稳定于较小的适应度值.

表3 Lorenz- $x(t)$ 序列: 不同模型的运行时间

模型	AF-ESN	PSO-ESN	TLBO-ESN	IDE-ESN
时间	1405.4289 s	47.6972 s	168.3124 s	102.8856 s

4.2 大连月平均气温 - 降雨量序列

大连市月平均气温和月降雨量数据集包括从1951年1月到2001年12月的数据记录值,采样间隔为月.大连月降雨量序列作为实验数据的第二维,来辅助大连月平均气温的预测.70%的数据用于训练,30%数据用于测试,其中训练样本中舍弃前50个样本产生的状态以消除初始暂态的影响.储备池规模 N 范围设为 $[20, 100]$ 、稀疏度 D 范围设为

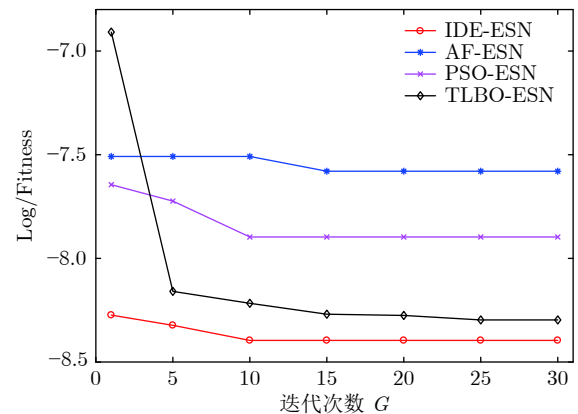


图4 Lorenz- $x(t)$ 序列: 不同模型的适应度曲线
Fig.4 Lorenz- $x(t)$ series: the curves of fitness for different models

$[0.01, 0.5]$ 、谱半径 R 范围设为 $[0.1, 1]$ 及输入变换因子 S 范围设为 $[0.0001, 0.1]$;种群大小 NP 设为25,最大迭代次数设为30.缩放因子 F 的范围设为 $[0.1, 0.9]$,交叉概率 CR 的范围设为 $[0.1, 0.9]$.为证明本文所提模型有效性,对于PSO-ESN、AF-ESN模型,设置同样的储备池参数范围、种群大小及迭代次数.表4给出本文所提出的模型选出的最适合大连月平均气温 - 降雨量序列的储备池参数.

表4 大连月平均气温: IDE-ESN 模型参数

储备池参数	取值
储备池规模	47
稀疏度	0.0206
谱半径	0.9802
输入变化因子	0.0459

图5给出了IDE-ESN模型对大连月平均气温 - 降雨量序列测试数据的预测曲线和误差曲线.从图中可以看出,本文模型能够很好地拟合大连月平均气温 - 降雨量序列曲线,绝对误差较小.表5给出了不同模型对大连月平均气温 - 降雨量序列的单步预测结果,可以看出本文所提模型在RMSE、NRMSE方面均好于其他对比模型.从表5可以看出虽然PSO-ESN模型的RMSE小于AF-ESN,但是SMAPE却很大,说明PSO-ESN模型不能很好地预测出大连月平均气温的序列趋势.而AF-ESN虽然能够较好地拟合出序列的变化趋势,但是在RMSE和NRMSE方面误差较大,说明在某些点上的预测上存在较大偏差.本文所提IDE-ESN模型无论在RMSE方面还是NRMSE方面都小于其他

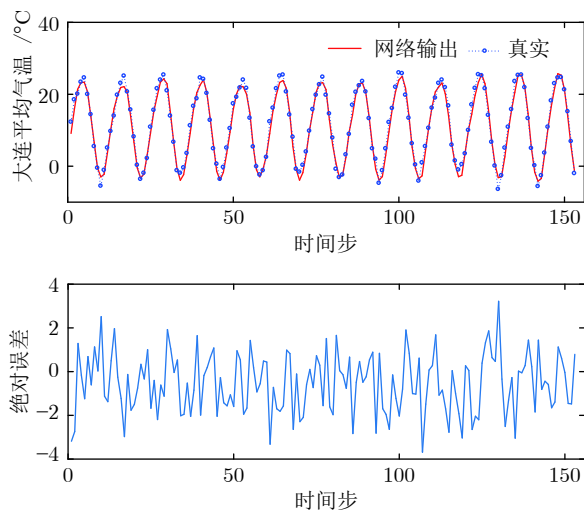


图 5 大连月平均气温: IDE-ESN 的预测曲线及误差曲线
Fig. 5 Dalian monthly average temperature series: prediction and error curves obtained by IDE-ESN

表 5 大连月平均气温: 测试集仿真结果
Table 5 Dalian monthly average temperature series: prediction results for the test dataset

模型	RMSE	SMAPE	NRMSE
AF-ESN	1.8042	0.2902	0.1820
PSO-ESN	1.6511	0.2956	0.1666
ELM	5.4235	0.6704	0.5520
TLBO-ESN	1.6726	0.2088	0.1708
IDE-ESN	1.4215	0.2741	0.1440

对比模型, 具有明显优势.

图 6 给出对于大连月平均气温数据集, IDE-ESN、PSO-ESN、AF-ESN 及 TLBO-ESN 在迭代过程中的适应度值的变化曲线图, 从图中可以看出 IDE-ESN 算法较快收敛, 且误差较小. 表 6 给出对于大连月平均气温数据集不同模型在 30 次迭代次数下的运行时间, 可以看出 AF-ESN 模型运行时间最长, 该模型的时间复杂度最大. IDE-ESN 模型运行时间略大于 PSO-ESN 模型, TLBO-ESN 模型运行时间是本文提出模型运行时间的两倍. PSO-ESN 模型虽然运行时间最短, 但预测精度较差.

5 结论

本文针对不同的时间序列, 利用改进的差分进化算法来动态选择回声状态网络的参数, 以适应于不同时间序列的动力学特性, 从而提高预测精度和泛化性能. 对于标准的差分进化主要有两方面的改进: 1) 对于算法的控制参数的自适应选择, 主要包括缩放因子 F 和交叉概率 CR ; 2) 变异策略的自适

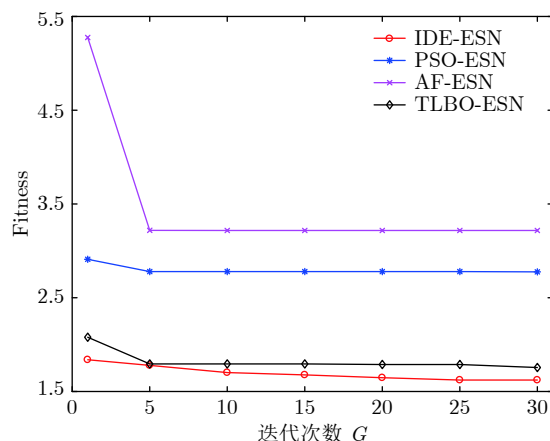


图 6 大连月平均气温: 不同模型的适应度曲线
Fig. 6 Dalian monthly average temperature series: the curves of Fitness for differential models

表 6 大连月平均气温: 不同模型的运行时间
Table 6 Dalian monthly average temperature series: run time of different models

模型	AF-ESN	PSO-ESN	TLBO-ESN	IDE-ESN
时间	347.1955 s	10.5115 s	31.1971 s	15.1921 s

应选择. 对两组时间序列进行预测, 仿真实验结果表明, 本文所提出的模型较其他预测模型有了很大的改善, 既具有较高的预测精度, 又具有较快的收敛速度和运行速度, 在时间序列预测分析中具有实用性和有效性、普遍性.

References

- Su C H, Cheng C H. A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock. *Neurocomputing*, 2016, **205**: 264-273
- Haidar A, Verma B. A novel approach for optimizing climate features and network parameters in rainfall forecasting. *Soft Computing*, 2018, **22**(24): 8119-8130
- Arthun M, Eldevik T, Viste E, Drange H, Furevik T, Johnson H L, et al. Skillful prediction of northern climate provided by the ocean. *Nature Communications*, 2017, **8**: 15875
- Wang X, Jiang R, Li L, Lin Y, Zheng X, Wang F. Capturing car-following behaviors by deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19**(3): 910-920
- Zhou Ping, Liu Ji-Ping. Data-driven multi-output ARMAX model for online estimation of central temperatures for cross temperature measuring in blast furnace ironmaking. *Acta Automatica Sinica*, 2018, **44**(3): 552-561
(周平, 刘记平. 基于数据驱动多输出 ARMAX 建模的高炉十字测温中心温度在线估计. *自动化学报*, 2018, **44**(3): 552-561)
- Zhang Yan-Yan, Tang Li-Xin. Improved data-driven subspace algorithm for energy prediction in iron and steel industry. *Control Theory and Application*, 2012, **29**(12): 1616-1622
(张颜颜, 唐立新. 改进的数据驱动子空间算法求解钢铁企业能源预测问题. *控制理论与应用*, 2012, **29**(12): 1616-1622)
- Zheng Nian-Zu, Ding Jin-Liang. Regression GAN based prediction for physical of total hydrogen in crude oil. *Acta Automat-*

- ica Sinica, 2018, **44**(5): 915–921
(郑念祖, 丁进良. 基于 Regression GAN 的原油总氢物性预测方法. 自动化学报, 2018, **44**(5): 915–921)
- 8 Hearst M A, Dumais S T, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intelligent Systems and Their Applications*, 1998, **13**(4): 18–28
 - 9 Trappey A J, Hsu F C, Trappey C V, Lin C I. Development of a patent document classification and search platform using a back-propagation network. *Expert Systems with Applications*, 2006, **31**(4): 755–765
 - 10 Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and application. *Neurocomputing*, 2006, **70**: 489–502
 - 11 Jaeger H, Haas H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 2004, **304**(5667): 78–80
 - 12 Han Min, Ren Wei-Jie, Xu Mei-Ling. An improved echo state network via L1-norm regularization. *Acta Automatica Sinica*, 2014, **40**(11): 2428–2435
(韩敏, 任伟杰, 许美玲. 一种基于 L1 范数正则化的回声状态网络. 自动化学报, 2014, **40**(11): 2428–2435)
 - 13 Lun Shu-Xian, Lin Jian, Yao Xian-Shuang. Time series prediction with an improved echo state network using small world network. *Acta Automatica Sinica*, 2015, **41**(9): 1669–1679
(伦淑娴, 林健, 姚显双. 基于小世界回声状态网的时间序列预测. 自动化学报, 2015, **41**(9): 1669–1679)
 - 14 Han M, Xu M. Laplacian echo state network for multivariate time series prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(1): 238–244
 - 15 Li De-Cai, Han Min. Chaotic time series prediction based on robust echo state network. *Acta Physica Sinica*, 2011, **60**(10): 108903
(李德才, 韩敏. 基于鲁棒回声状态网络的混沌时间序列预测研究. 物理学报, 2011, **60**(10): 108903)
 - 16 Han Min, Wang Ya-Nan. Multivariate time series online predictor with Kalman filter trained reservoir. *Acta Automatica Sinica*, 2010, **36**(1): 169–173
(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. 自动化学报, 2010, **36**(1): 169–173)
 - 17 Xu D, Lan J, Principe J C. Direct adaptive control: an echo state network and genetic algorithm approach[C]. In: Proceedings of the 2005 IEEE International Joint Conference on Neural Networks. Montreal, Canada: IEEE, 2005. 3: 1483–1486
 - 18 Wang J S, Han S, Guo Q P. Echo state networks based predictive model of vinyl chloride monomer convention velocity optimized by artificial fish swarm algorithm. *Soft Computing*, 2014, **18**(3): 457–468
 - 19 Chouikhi N, Ammar B, Rokhani N, Alimi A M. PSO-based analysis of echo state network parameters for time series forecasting. *Applied Soft Computing*, 2017, **55**: 211–225
 - 20 Sun Xiao-Yan, Chen Shan-Shan, Gong Dun-Wei, Zhang Yong. Weighted multi-output Gaussian process-based surrogate of interactive genetic algorithm with individual's interval fitness. *Acta Automatica Sinica*, 2014, **40**(2): 172–184
(孙晓燕, 陈姗姗, 巩敦卫, 张勇. 基于区间适应值交互式遗传算法的加权多输出高斯过程代理模型. 自动化学报, 2014, **40**(2): 172–184)
 - 21 Du W, Ying W, Yan G, Zhu Y, Cao X. Heterogeneous strategy particle swarm optimization. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2017, **64**(4): 467–471
 - 22 Mavrovouniotis M, Muller F M, Yang S. Ant colony optimization with local search for dynamic traveling salesman problems. *IEEE Transactions on Cybernetics*, 2017, **47**(7): 1743–1756
 - 23 Zhang Z, Wang K, Zhu L, Wang Y. A Pareto improved artificial fish swarm algorithm for solving a multi-objective fuzzy dis-assembly line balancing problem. *Expert Systems with Applications*, 2017, **86**: 165–176
 - 24 Hu Rong, Qian Bin. A hybrid differential evolution algorithm for stochastic flow shop scheduling with limited buffers. *Acta Automatica Sinica*, 2009, **35**(12): 1580–1586
(胡蓉, 钱斌. 一种求解随机有限缓冲区流水线调度的混合差分进化算法. 自动化学报, 2009, **35**(12): 1580–1586)
 - 25 Rao R V, Savsani V J, Vakharia D P. Teaching-learning-based optimization: A novel method for constrained mechanical design optimization problems. *Computer-Aided Design*, 2011, **43**: 303–315
 - 26 Xu M L, Han M, Lin H F. Wavelet-denoising multiple echo state networks for multivariate time series prediction. *Information Sciences*, 2018, **465**: 439–458
 - 27 Zhou Xiao-Gen, Zhang Gui-Jun, Hao Xiao-Hu. Differential evolution algorithm with local abstract convex region partition. *Acta Automatica Sinica*, 2015, **41**(7): 1315–1327
(周晓根, 张贵军, 郝小虎. 局部抽象凸区域剖分差分进化算法. 自动化学报, 2015, **41**(7): 1315–1327)
 - 28 Ding Jin-Liang, Yang Cui-E, Chen Li-Peng, Chai Tian-You. Dynamic multi-objective optimization algorithm based on reference point prediction. *Acta Automatica Sinica*, 2017, **43**(2): 313–320
(丁进良, 杨翠娥, 陈立鹏, 柴天佑. 基于参考点预测的动态多目标优化算法. 自动化学报, 2017, **43**(2): 313–320)
 - 29 Zhang J Q, Sanderson A C. JADE: Adaptive differential evolution with optional external archive. *IEEE Transactions on Evolutionary Computation*, 2009, **13**(5): 945–958
 - 30 Zhao S Z, Suganthan P N, Das S. Self-adaptive differential evolution with multi-trajectory search for large-scale optimization. *Soft Computing*, 2011, **15**: 2175–2185



许美玲 东北大学智能工业数据解析与优化教育部重点实验室讲师, 大连理工大学电子信息与电气工程学部博士后. 主要研究方向为神经网络和多元时间序列预测. 本文通信作者.

E-mail: xuml@dlut.edu.cn

(XU Mei-Ling Lecturer at the Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Ministry of Education. Postdoctor at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. Her research interest covers neural networks and multivariate time series prediction. Corresponding author of this paper.)



王依雯 大连理工大学电子信息与电气工程学部硕士研究生. 主要研究方向为时间序列预测.

E-mail: wangyiwen0802@163.com

(WANG Yi-Wen Master student at the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. Her main research interest is time series prediction.)