MOb-GRU神经网络工业软测量建模方法与输出预测

王 珠,刘佳璇†

(中国石油大学(北京) 自动化系,北京 102249)

摘要:由于工业过程具有强非线性、动态特性与慢时变性,其完整性建模相对较难.针对工业过程的现有软测量 技术并未综合考虑过程的非线性和动态特性,本文提出了一种依赖模型阶次的GRU(MOb-GRU)神经网络软测量模 型,针对非线性动态过程进行全动态建模.首先,在MOb-GRU的结构选择上,本文根据所研究实际对象的动态特性 复杂程度确定网络的总模块数.另外,MOb-GRU能灵活设置反向更新的单元数,这种设置打破了传统GRU只能从 第1个模块开始输出的限制.其次,为使记忆网络以较快的速率收敛到最优,本文分别设计了基于自适应学习率和学 习率矩阵的网络训练算法.接着,仿真实验分别选取了典型的单变量与多变量非线性动态过程,并采用MOb-GRU 神经网络对其进行建模和预测.最后,仿真结果证实了MOb-GRU网络结构的合理性以及训练算法的高效性.

关键词: 软测量技术; MOb-GRU; 非线性动态; 自适应学习率; 神经网络

引用格式: 王珠, 刘佳璇. MOb-GRU神经网络工业软测量建模方法与输出预测. 控制理论与应用, 2022, 39(9): 1758 – 1768

DOI: 10.7641/CTA.2022.10533

MOb–GRU neural network for industrial soft sensor modeling method and output prediction

WANG Zhu, LIU Jia-xuan[†]

(Department of Automation, China University of Petroleum, Beijing 102249, China)

Abstract: Modeling the integrity of industrial process is a relatively difficult task due to its strong nonlinearity, dynamic characteristics and slow time variability. Though there exist some soft sensing technologies for industrial process, they fail to consider the nonlinear and dynamic characteristics comprehensively of the process. Therefore, this paper proposes a model order based gated recurrent unit (MOb–GRU) neural network soft sensor model for fully-dynamic modeling of nonlinear dynamic process. Specifically, firstly, in terms of the MOb–GRU structure selection, this paper determines the total module number of the network according to the complexity of dynamic characteristics of the actual object. Moreover, the MOb–GRU can flexibly set the number of units for reverse update, which breaks the limitation that the traditional GRU can only output from the first module. Secondly, in order to make the memory network converge to the optimal at a faster rate, this paper designs the network training algorithms based on the adaptive learning rate and the learning rate matrix, respectively. Then, the typical univariate and multivariable nonlinear dynamic processes are selected in the simulation experiment, and the MOb–GRU neural network is used to model and predict them. Finally, the rationality of MOb–GRU network architecture as well as the high efficiency of the training algorithms is demonstrated through the simulation results.

Key words: soft sensing technology; MOb-GRU; nonlinear dynamic; adaptive learning rate; neural networks

Citation: WANG Zhu, LIU Jiaxuan. MOb–GRU neural network for industrial soft sensor modeling method and output prediction. *Control Theory & Applications*, 2022, 39(9): 1758 – 1768

1 引言

随着自动化水平的提升、质量要求的完善、生产 规模的扩大,现代工业过程往往存在强非线性、动态 特性与慢时变等本质特点,众多参变量中普遍存在不 确定、多层次与强耦合关系,因此传统的机理模型难

以准确地描述实际工业过程.

软测量技术^[1]的产生与发展为解决上述问题提供 了一种有效方法,其核心思想是利用易于测量的过程 变量(辅助变量)建立可以表征过程变量和质量变量 (主导变量)之间关系的软测量模型.软测量模型为后

收稿日期: 2021-06-21; 录用日期: 2022-01-24.

[†]通信作者. E-mail: 2020211232@student.cup.edu.cn; Tel.: +86 15611003986.

本文责任编委: 阳春华.

国家自然科学基金项目(61703434),中国石油大学(北京)科研基金项目(2462020YXZZ023)资助.

Supported by the National Natural Science Foundation of China (61703434) and the Scientific Research Foundation of China University of Petroleum, Beijing (2462020YXZZ023).

续过程控制[2-3]、在线估计[4-6]以及故障诊断[7]等方 面产生了很大的影响,发挥了必要且重要的作用.早 期的软测量是基于机理分析的建模,需要对工业过程 内部机理有充分的了解.其中:微分方程与代数方程 能够用于表示工业过程动态机理^[8],卡尔曼滤波^[9-10] 常用于软测量中对过程参数进行建模.但由于实际工 业过程极为复杂,模型结构的形式难以确定,很难通 过机理建模描述过程规律、反映过程特性.基于数据 驱动[11]的软测量建模方法解决了上述问题.数据驱动 仅依靠现场采集的大量历史输入输出数据建立质量 变量与过程变量之间的数学关系,因此非常适合于复 杂工业过程的软测量建模.早期的数据驱动建模 方法包含主成分分析法(principal component analysis, PCA^[12])、偏最小二乘法(partial least squares, PLS^[13]) 等回归分析法与人工神经网络(artificial neural network, ANN)等机器学习模型^[14]. 由于实际工业过程 具有动态特性,但上述回归分析法与多数人工神经网 络模型仅能反映过程的非线性特性而缺少对动态特 性的体现,因此非线性动态建模成为软测量的主要研 究方向,不少研究针对时序数据具备的特性提出了非 线性动态软测量模型[15-17].

近年来,随着神经网络理论的不断发展和完善,神 经网络软测量模型主要包含普通神经网络[18]、径向 基神经网络(radial basis function network, RBF^[19])、 生成对抗神经网络(generative adversarial network, GAN^[20])等模型.但上述提出的模型都是静态软测量 模型,在工业过程中具有一定的局限性.动态软测量 模型的相继提出解决了静态模型在实际应用中估计 精度低、鲁棒性差等问题. 回声状态网络(echo state network, ESN^[21])、卷积神经网络(convolutional neural network, CNN^[22])与循环神经网络(recurrent neural network, RNN^[23-25])等常作为动态软测量模型应用到 实际的工业过程.其中:循环神经网络的发展为时间 序列的建模提供了优良选择.因此,对于工业过程的 非线性全动态建模问题,常使用工业时序数据作为循 环神经网络的输入或采用固定结构的记忆神经网络 对非线性动态过程进行有效模拟. 记忆神经网络分为 循环神经网络、长短期记忆神经网络(long-short term memory, LSTM^[26])和门控循环单元(gated recurrent unit, GRU^[27]). 其中: 循环神经网络是一种短记忆模 型,不适合处理过长的时间序列;LSTM的提出解决了 上述问题,在记忆方面得以较大提升,对任意长度的 时序数据均能进行很好的训练及预测,但该网络的结 构过于复杂,加重了计算负担.为解决上述问题,GRU 的提出得到了广泛的应用.Fu等[28]使用GRU神经网 络对交通流进行了预测,实验表明,GRU在交通流 量预测上的性能优于LSTM与自回归积分移动平均 (autoregressive integrated moving average, ARIMA)模

型. Pavithra等^[29]将门控循环单元应用于医学领域,基 于GRU在预测糖尿病疾病的发展上取得了良好的预 测效果. Siwagorn等^[30]采用GRU 预测飞机垂直速度 的下降幅度,使飞机能够有效着陆,提高了飞机的着 陆效率. 倪维成^[31]建立了一种基于GRU的航空发动 机剩余寿命预测模型,实验表明该模型在预测精度上 高于多数浅层机器学习方法和部分深度学习方法. 虽 然目前已有大量学者在不同领域验证了GRU神经网 络较其他预测模型在时序预测问题上的优越性,但基 于GRU对工业领域中非线性动态过程的预测研究却 屈指可数,并且已有研究没有对GRU神经网络反向更 新单元数与实际非线性动态过程阶次之间的关系进 行研究.基于以上分析,本文提出一种依赖模型阶次 的GRU(model order based-GRU, MOb-GRU)软测量 模型,基于该模型对工业领域中单变量与多变量非线 性动态过程进行全动态建模,本文用带有输出非线性 的非线性动态过程代替实际非线性动态过程产生过 程数据,进而进行分析与建模.

Lynn等人^[32]的研究表明, GRU结构的更新门和输 出激活函数是GRU网络最关键的组件,学习率是门控 循环单元最重要的超参数.因此,建立神经网络软测 量模型的关键是选择合适的学习率优化算法使网络 以较快速度达到收敛.目前已有一些关于深度学习模 型中学习率策略的研究. Ranjeeth等^[33]提出了具有最 优随机梯度下降(stochastic gradient descent, SGD)的 多层感知器机器学习模型, SGD的引入提高了感知器 的性能与数据分类准确度,但SGD中学习率是固定的, 收敛速度慢且容易陷入局部最优解. Ralf等^[34]设计 了一种应用于复杂深度强化学习(deep reinforcement learning, DRL)问题的循环学习率方法,该方法较固定 学习率方法能达到更好的结果,但循环学习率是在两 个有理边界值的范围内变化,而不是单调衰减的.为 解决上述问题,本文设计了一种简单而有效的阶跃衰 减类(step attenuation class. SAC)自适应学习率算法 与学习率矩阵算法,两种方法均保证了整个系统更快 地收敛和稳定,提高了预测的准确率.

2 针对非线性动态过程的MOb-GRU网络 结构与信息流向图

GRU由Cho等人提出,是LSTM的一种变体模型, 不仅能够解决RNN存在的梯度消失问题,还简化了 LSTM的网络结构、提高了收敛速度.目前最常用的 GRU模型主要由更新门和重置门构成,图1为其内部 结构示意图.

每个GRU单元能够根据当前时刻的输入 x_t 和上一时刻隐藏层输出的激活值 h_{t-1} ,计算得到当前时刻隐藏层输出值 h_t 和候选激活值 \tilde{h}_t .相比于RNN,GRU通过引入门控机制,能够同时对不同时刻、不同长度的

时序关系进行记忆和学习.相比于LSTM, GRU减少 了网络参数数量,加快了训练的收敛速度,具有简单 的单元结构和高效处理数据的能力.目前GRU已被广 泛应用于机器翻译和序列生成等众多领域.



Fig. 1 GRU internal structure diagram

本文所提出的MOb-GRU软测量模型适用的场合 需满足以下两个条件: 1) 由于工业过程的复杂性, 过 程内部机理不清楚, 无法运用机理建模对其进行精确 建模; 2) 过程数据在时间上连续, 满足一定的时序关 系. 文中分别采用 u_t 和 y_t 来表示过程变量和质量变量, 软测量模型基于数据驱动, 通过学习得到过程变量与 质量变量间的映射关系, 即 $f: y_t \rightarrow u_t$.

MOb-GRU神经网络与传统GRU相比,网络结构 复杂度与训练所需计算量均较小,原因在于:a)从结 构上看. MOb-GRU能够根据实际过程的大致阶次调 节反向更新单元数,其数量可少于网络中的总单元数, 与传统GRU从第1个模块开始输出相比,既保证了长 期和短期的记忆性,又在结构设置上变得更加灵活, 需要注意的是, MOb-GRU的反向更新单元数是指包 含最后一个模块开始反向向前传播的单元数, 与训练 算法中权重梯度在时间上的叠加数量相等,以此保证 了权重在更新过程中不随噪声发生显著波动; b) 从训 练算法上看,在用随时间反向传播(back propagation through time, BPTT) 算法进行训练时, MOb-GRU综 合梯度量的确定依赖于反向更新单元数,而GRU综合 梯度量的确定依赖于全部模块数.基于此, MOb-GRU 训练时历经时间确定梯度所需的循环数量较小,计算 量较低,减轻了模型的计算负荷.但MOb-GRU神经 网络与RNN相比,网络结构复杂度与训练所需计算量 又是偏大的,原因在于: a) 从结构上看, RNN中间层神 经元的状态是由上一层过程输入的状态与自身前一 时刻的状态决定的,意味着与GRU, MOb-GRU相比, RNN总模块数只有2个; b) 从训练算法上看, RNN综 合梯度量的确定只需计算当前和前一时刻的梯度量, 即反向传播单元数为1,因此训练所需的计算量相对

更低. 综上所述, RNN与GRU, MOb-GRU相比, 网络结构更简单, 训练时的计算时间复杂度更低. 3个模型的计算时间复杂度关系如表1所示.

基于对模型计算时间复杂度与训练性能的考虑. 本文采用MOb-GRU模型对非线性动态过程进行建 模,将过程的输入数据u_t和输出数据y_t作为神经网络 的学习数据,无需明确过程内部机理和参数变量.本 文设计的单变量MOb-GRU (SISO-MOb-GRU)信息 流向图和单变量GRU (SISO-GRU)信息流向图如图2 所示.其中: i为前向传播单元数; i为反向更新单元数; 以SISO-MOb-GRU信息流向图为例,每个单元下方 直连的变量为MOb-GRU单元的输入变量,上方直连 的变量为MOb-GRU单元的预测输出. 定义x(t) = $[u(t-1) y(t-1)]^{T}$ 为MOb-GRU单元的输入; $\chi(t)$ $= [x(t-i) \cdots x(t-i) \cdots x(t-1)], \hat{\boldsymbol{v}}(t) = [\hat{y}(t)]$ -i+1) ··· $\hat{y}(t)$]分别为网络的输入和输出.为体 现过程的动态特性,本文将采用递归的方式给网络的 输入信号赋值,使MOb-GRU模型呈现出一种动态效 果. SISO-GRU信息流向图中的变量同理.

表13个模型的计算时间复杂度

 Table 1 Computational time complexity of three models



高维、高阶多变量过程普遍存在于现代工业过程中,因此基于MOb-GRU对多变量过程进行动态建模具有重要的实际意义.由于本研究限于理论分析层面,在单变量过程建模的基础上仅通过拓展维度便能实现对多变量过程的模拟及预测.图3为本文设计的多变量MOb-GRU (MIMO-MOb-GRU)信息流向图.其中: s和v表示多变量非线性动态过程输入与输出变量

的维度.每个MOb-GRU单元的输入变量引出的小圆 圈数量代表网络的输入维度.对于多变量过程,每增 加一个输入或输出变量,便会多一个黄色圆圈与网络 的输入变量相连,以表示网络输入维度的拓展;输出 维度的拓展同理.多变量GRU (MIMO-GRU)信息流 向图与MIMO-MOb-GRU信息流向图的区别和单变 量一样,在此不呈现具体的MIMO-GRU信息流向图.



图 3 MIMO-MOb-GRU信息流向图 Fig. 3 MIMO-MOb-GRU information flow diagram

3 方法

3.1 算法流程与训练算法

基于MOb-GRU神经网络对非线性动态过程进行 建模与预测的整体算法流程如下:

步骤1 输入输出数据的归一化处理.本研究将 根据过程变量与质量变量的量程范围进行归一化和 反归一化处理.

$$\bar{u}_t = \frac{u_t - u_{t,\min}}{u_{t,\max} - u_{t,\min}} \cdot (1 - 0) + 0, \qquad (1)$$

$$\bar{y}_t = \frac{y_t - y_{t,\min}}{y_{t,\max} - y_{t,\min}} \cdot (1 - 0) + 0, \qquad (2)$$

其中: $u_{t,\max}$ 和 $u_{t,\min}$ 为输入变量量程范围内的最大值与最小值; $y_{t,\max}$ 和 $y_{t,\min}$ 为输出变量量程范围内的最大值与最小值.

步骤2 初始化网络结构选择.

步骤 3 训练网络.本研究将采用BPTT对MOb-GRU网络进行训练.

步骤4 步长及步长矩阵的选取.本文采用SAC 自适应学习率算法与学习率矩阵算法选取合适的步长与步长矩阵对网络参数进行更新,以确定最优的网

络结构.

步骤 5 预测输出.

BPTT算法是记忆神经网络训练时进行权重更新的一种基于时间的反向传播算法,其本质为梯度下降法,因此求各参数的梯度成了该算法的关键.首先定义*t*时刻的损失函数为

$$E_t = \frac{1}{2}(y_e - y_t)^2,$$
 (3)

其中: y_e 表示t时刻的实际输出; y_t 表示t时刻的预测输出. BPTT训练算法具体见文献[35].

3.2 学习率优化算法

学习率对神经网络的学习有很大的影响. 学习率 过高, 易使网络参数优化过度, 导致训练变得发散; 学 习率过小, 虽然网络训练更加可靠, 但所需时间过长. 因此选择合适的学习率优化算法显得尤为重要.

3.2.1 阶跃衰减类SAC自适应学习率算法

对于不同波动程度的非线性动态过程会对应不同的最优学习率(optimal learning rate, Olr),以保证网络的预测精度和收敛速度.因此本文设计了1种SAC自适应学习率算法.

引入平均绝对百分比误差(mean absolute percentage error, MAPE)作为评价指标,用于判断模型训练 效果的好坏,即

MAPE =
$$\frac{1}{L} \sum_{c=1}^{L} |\frac{y_c - \hat{y}_c}{y_c}|,$$
 (4)

其中: L为训练数据的长度; y_c 为c时刻的真实输出; \hat{y}_c 为c时刻的估计输出.

将训练数据按时间连续划分为n个长度为l的阶段,将预选学习率中的固定学习率从大到小依次分配 给[2,n-1]的每个阶段.第1阶段采用初始学习率作 为网络的Olr进行训练;第2阶段到最后一个阶段之间 的每一阶段,首先确保网络在该固定学习率下收敛, 接着取后0.25*l*的数据通过递推式(5)计算MAPE数值, 并将该值作为本阶段的相对误差标准值,即

$$MAPE_{q} = \sum_{c=0.75*l+1}^{l} b^{l-c} d_{l-0.75*l} \cdot |\frac{y_{c} - \hat{y}_{c}}{y_{c}}| = (1 - d_{l-0.75*l})MAPE_{q-1} + d_{l-0.75*l} \cdot |\frac{y_{c} - \hat{y}_{c}}{y_{c}}|,$$
(5)

其中: $d_{l-0.75*l} = (1-b)/(1-b^{l-0.75*l})$, b表示遗忘因 子; $q \in [2, n-1]$, 表示某一阶段. 最后记录每一阶段 最后一个时刻的MAPE和学习率.

在判断阶段,将当前阶段的MAPE_q与最优MAPE 进行比较,若MAPE_q < MAPE,则更新最优MAPE的 值为MAPE_q,同时更新最优学习率Olr为Olr_q;反之, 评价指标保持不变作为后面阶段的相对误差标准值 进行判断,Olr也保持不变.Olr随着最优MAPE的变化 不断调整,到最后一个阶段采用最优学习率进行最优 神经网络的确定.

通过不断调整最优MAPE确定最优学习率的取值, 具体实现流程如下:

步骤1 判断当前时刻c处于哪个阶段.

步骤 2 如果q = 1,则设置初始学习率为最优 学习率,用于神经网络的稳定和收敛.如果q = 2,则 在该阶段的最后通过递推公式计算初始MAPE作为相 对误差标准值.

步骤 3 如果2 < q < n, 通过式(5)计算各阶段的 MAPE并将其与相对误差标准值进行对比, 按照上述 判断方法更新最优MAPE以及最优学习率Olr.

步骤 4 如果*q* = *n*,使用最优学习率训练神经 网络并进行最优网络结构的确定.

利用SAC自适应学习率算法确定网络的最优学习率,相应的MOb-GRU反向更新算法形式为

$$\boldsymbol{W}_{(\cdot)}(kT) = \boldsymbol{W}_{(\cdot)}((k-1)T) - \\ \text{Olr} * \sum_{\epsilon=0}^{i-1} \frac{\partial E(kT)}{\partial \boldsymbol{W}_{(\cdot)}((k-\epsilon)T)}, \quad (6)$$

其中: $W_{(.)}$ 代表MOb-GRU模型中需要学习的权重参数 W_{o} , W_{rh} , W_{rx} , $W_{\bar{h}h}$, $W_{\bar{h}x}$, W_{zh} , W_{zx} , 后面该变量含义与此保持一致; k为采样时刻; T为采样周期, 设置T = 1 min; i为反向更新单元数. 为方便起见, 后续形式上将采样周期T省略, 即用k表示kT.

注 1 本文根据数据集的长度将其划分为n段,预选学 习率由(n-2)个从小到大取值的固定学习率组成 $[\alpha_1, \alpha_2, \cdots, \alpha_{n-2}], \alpha_1 < \alpha_2 < \cdots < \alpha_{n-2}.$

注 2 为保证网络训练时不会发散且使网络较快地达 到收敛,本文将初始学习率设置为预选学习率数组中的中间 数值.

3.2.2 学习率矩阵算法

Hessian矩阵^[36–37]常用于优化问题,是用一个标量 对一个向量的二阶导数组成的方阵.对于实际问题 Hessian矩阵可能会很难计算,因此通常采用近似二阶 Hessian矩阵法代替计算,如 BFGS 算法^[38]、DFP 算 法^[39]与Levenberg–Marquardt (LM)算法^[40].本研究将 采用类LM算法作为学习率矩阵算法,利用近似二阶 Hessian矩阵的逆的方法代替步长,对权重进行更新. 该算法结合了梯度下降法和高斯–牛顿法的优点,使 网络能较快且稳定地找到参数的最优值.

利用学习率矩阵算法对MOb-GRU网络参数进行 反向更新时,算法形式如下:

$$\boldsymbol{W}_{(\cdot)}(k) = \boldsymbol{W}_{(\cdot)}(k-1) - [\mathrm{d}\boldsymbol{W}_{(\cdot)}^{\mathrm{T}}(k)\mathrm{d}\boldsymbol{W}_{(\cdot)}(k) + \mu\boldsymbol{I}]^{-1}\mathrm{d}\boldsymbol{W}_{(\cdot)}^{\mathrm{T}}(k),$$
(7)

其中: d $W_{(.)}(k) = \sum_{\epsilon=0}^{i-1} \frac{\partial E(k)}{\partial W_{(.)}(k-\epsilon)}$. **I**为与各权重 同维的单位矩阵; $W_{(.)}$ 代表模型中需要学习的权重参数; μ 为阻尼系数, 本文设置 $\mu = 0.5$; *i*为反向更新单 元数; *k*为采样时刻; 本文设置采样周期T = 1 min.

注3 两种学习率优化算法的区别在于:如果训练网 络时已经能够确定基础步长的范围,则采用SAC自适应学习 率算法;当无法确定基础步长范围时,采用学习率矩阵算法. SAC自适应学习率算法能够提高网络的整体运算效率,而学 习率矩阵算法只需提前选择合适的阻尼项,便能对权重进行 较好的调整,是能够保证神经网络达到收敛的一种较为稳妥 的训练方法.

4 仿真实验

对于真实的工业过程,可以通过一个较快的采样 频率采集过程数据,再进行神经网络建模.而本文给 出带有输出非线性的仿真系统,是为了代替实际工业 过程而产生过程数据,再根据所产生的数据对仿真系 统进行建模.本节基于MOb-GRU软测量模型分别对 单变量与多变量非线性动态过程的仿真系统进行模 拟,旨在验证本文提出的MOb-GRU模型的有效性. 在单变量非线性动态过程的建模中,讨论了关键网络 结构参数以及系统波动程度对MOb-GRU训练性能的 影响.另外,仿真时将MOb-GRU与RNN,GRU两个基 线模型进行了对比,同时将SAC自适应学习率算法、 学习率矩阵算法分别与固定学习率算法进行了比较.

4.1 单变量非线性动态过程的建模与预测

该实验中,单变量非线性动态过程的仿真系统可 表示为

$$\begin{cases} x(k) = -\sum_{c_1=1}^{6} \kappa_{c_1} x(k-c_1) + \\ \sum_{c_2=1}^{6} \tau_{c_2} u(k-c_2) + w(k), \\ y(k) = \begin{cases} 52 - \sqrt{52 - x(k)}, \ 0 \le x(k) \le 52, \\ x(k), \qquad 52 < x(k) < 68, \\ 68 + \sqrt{x(k) - 68}, \ x(k) \ge 68, \end{cases} \end{cases}$$

$$(8)$$

注 4 u(·)代表非线性动态过程的输入信号,作为软测量中的辅助变量,取为多正弦信号

$$u(k) = 80 + 15\sin(0.2k) + 2\cos(0.5k),$$
(9)

其中: k为采样时刻; 取采样周期T = 1 min; 本小节使用的数据集是通过式(8)–(9)仿真生成的, 共81000个输入输出样本数据, 以3:1:1的比例将数据集划分为训练集、验证集和测试集. $x(\cdot)$ 表示中间状态变量; $y(\cdot)$ 表示输出变量, 作为软测量中的主导变量; $w(\cdot)$ 表示过程噪声, 是一种分布服从均值为0、方差为 ϑ 的高斯随机噪声, 即

$$w(\cdot) \sim N(0, \vartheta), \tag{10}$$

注5 本小节动态线性环节由下列η阶后向差分方程 描述

$$x(k) = -\sum_{c_1=1}^{\eta} \kappa_{c_1} x(k-c_1) + \sum_{c_2=1}^{\zeta} \tau_{c_2} u(k-c_2), \quad (11)$$

其中: 参数 ζ 和 η 取决于实际系统动态阶次的选择. 定义 $\gamma = [\kappa_1 \cdots \kappa_\eta \tau_1 \cdots \tau_\zeta]^T$ 表示线性环节的参数向量, 随机参数 κ_{c_1}, τ_{c_2} 的分布均服从正态分布, 即

$$\gamma = \begin{bmatrix} \kappa_1 & \cdots & \kappa_\eta & \tau_1 & \cdots & \tau_\zeta \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} N(\kappa_1, \lambda_1) \\ \vdots \\ N(\kappa_\eta, \lambda_\eta) \\ N(\tau_1, \lambda_{\eta+1}) \\ \vdots \\ N(\tau_\zeta, \lambda_{\eta+\zeta}) \end{bmatrix}, \quad (12)$$

其中λ_(·)表示参数的波动方差,后续实验中λ用于表示系统内部波动程度.λ越大,表示系统内部波动程度越大.

本小节静态非线性环节的具体表示如下:

$$y(k) = \begin{cases} 52 - \sqrt{52 - x(k)}, & 0 \le x(k) \le 52, \\ x(k), & 52 < x(k) < 68, \\ 68 + \sqrt{x(k) - 68}, & x(k) \ge 68. \end{cases}$$
(13)

MOb-GRU神经网络模型的超参数包括前向传播 单元数、反向更新单元数以及预选学习率间隔.首先 基于实验确定模型的关键结构参数.参数初始化设置 如表2所示.

对于单变量非线性动态过程的仿真系统,考虑动 态较为丰富且复杂的情况,将输入输出阶次ζ,η均设 置为6; 预选学习率间隔lrg暂时设置为0.08, 即预选 学习率为[0.08, 0.16, 0.24, 0.32, 0.40, 0.48, 0.56, 0.64]; 式(5)中遗忘因子d设置为0.95;系统内部波动程度 λ 暂时设置为0.0015²;外部波动程度∂暂时设置为4.0². 为确定MOb-GRU的前向传播单元数i与反向更新单 元数i,在验证集上进行对比实验,将MAPE作为评价 指标,能够表征预测值与真实值之间偏差的实际水平, MAPE越小,模型性能越好.由于网络在训练过程中 具有一定的随机性,每次训练得到的结果均有所不同, 为保证实验结果的可靠性,对每组参数实验均进行50 次重复,并将结果取平均,如表3所示.其中:下标a表 示采用SAC自适应学习率算法的网络训练结果;下 标m表示采用学习率矩阵算法的网络训练结果;无下 标表示采用固定学习率算法的网络训练结果. 后面以 表格形式呈现的实验结果同理.

表 2 参数初始化设置

Table 2 Parameter initialization settings

ζ	η	lrg	d	λ	θ
6	6	0.08	0.95	0.0015^{2}	4.0^{2}

表 3 不同模型结构参数下的MAPE

 Table 3 MAPE under different model structure parameters

前向单元数j	反向单元数i	MAPEa	$MAPE_{\mathrm{m}}$	MAPE
17	6	0.0625	0.0661	0.0694
18	6	0.0618	0.0658	0.0687
19	6	0.0581	0.0648	0.0658
20	6	0.0594	0.0665	0.0685
21	6	0.0602	0.0678	0.0696
19	4	0.0600	0.0677	0.0692
19	5	0.0584	0.0653	0.0669
19	7	0.0602	0.0659	0.0661
19	8	0.0640	0.0677	0.0688

从表3可见,在3种学习率优化算法下,前向传播单 元数为19、反向更新单元数为6时,MAPE的值最小. 当前向传播单元数小于19时,性能指标随前向传播单 元数的增加而减小;当前向传播单元数大于19时,性 能指标随前向传播单元数的增加而增加.这是由于当 前向传播单元数增加到一定数目时,整个模型的参数 爆炸增长,模型复杂度变大的同时预测精度降低.当 反向更新单元数小于6时,性能指标随反向更新单元 数的增大而减小;当反向更新单元数大于6时,性能指 标随反向更新单元数的增大而增加,验证了当反向更 新单元数接近模型阶次时,网络具有更好的性能.因此,将MOb-GRU模型的前向传播单元数设置为19, 反向更新单元数设置为6进行后续实验.

进而基于实验确定合适的预选学习率间隔.在保证网络跟踪精度的基础上,lrg从0.01到0.3的范围内选取.在验证集上对每个lrg均进行50次实验,实验结果取平均,依旧采用MAPE作为评价指标.

如图4所示,对于[0.01,0.3]的预选学习率间隔而 言,MAPE的值集中在0.0532~0.0712之间.实验表明, 预选学习率间隔过大或过小都会使MOb-GRU模型的 训练性能变得相对较差,当lrg = 0.19时,网络训练性 能达到最优,此时MAPE = 0.0532.因此本文将预选 学习率间隔选为0.19,进行后续研究.





为研究系统波动对MOb-GRU网络性能的影响,本文选择MAPE和均方根误差(root mean square error, RMSE)作为评价指标,RMSE由式(14)计算得到.为了对比模型的预测效果,选择RNN与GRU模型作为对

比基线模型.为了验证本文所设计的两种学习率优化 算法的有效性,将其分别与固定学习率算法进行比较

RMSE =
$$\sqrt{\frac{1}{L} \sum_{c=1}^{L} (\hat{y}_c - y_c)^2},$$
 (14)

其中: *L*表示数据集长度; \hat{y}_c 为网络预测输出; y_c 为系 统真实输出. RMSE越小, 模型训练性能越好. 本文将 系统波动分为系统内部波动与系统外部波动进行研 究. 基于以上实验, 设置MOb-GRU结构参数j = 19, i= 6; GRU结构参数j = 19, i = 18; 根据第2节的理论 分析, 设置RNN的结构参数j = 2, i = 1.

a) 系统内部波动对预测模型预测效果的影响.

系统内部波动是指动态线性环节参数向量的波动 程度,每个参数可能具有不同的波动方差(如式(12)). 设置式(8)中 $\kappa_1 = 0.12, \kappa_2 = 0.03, \kappa_3 = 0.05, \kappa_4 = 0.02, \kappa_5 = 0.01, \kappa_6 = 0.01, \tau_1 = 0.61, \tau_2 = 0.21, \tau_3 = 0.06, \tau_4 = 0.02, \tau_5 = 0.01, \tau_6 = 0.02以及系统外部$ $波动方差<math>\vartheta = 4.0^2$.为研究模型预测效果与系统内部 参数波动程度(用 λ 表示)之间的关系,假设所有参数 的波动方差均相同,考虑如下几种情况:

- 情况 1 $\lambda_1 = \lambda_2 = \cdots = \lambda_{12} = \lambda = 0.001^2.$
- 情况 2 $\lambda_1 = \lambda_2 = \cdots = \lambda_{12} = \lambda = 0.005^2.$
- 情况 3 $\lambda_1 = \lambda_2 = \cdots = \lambda_{12} = \lambda = 0.01^2.$

对应不同的参数波动方差,实验分别进行50次重 复并将结果取平均.在验证集上,运用MOb-GRU与 两种基线模型分别对3种情况的非线性动态过程进行 预测,表4给出不同模型在不同学习率优化算法下的 预测结果.由表4可知,随着内部参数波动程度的增加, 各模型在3种学习率优化算法下的预测精度均有所降 低.原因是随着内部参数波动程度的增加,系统的随 机性增强,在固定的网络参数设置下,易导致网络模 型的预测精度降低,对实际系统的跟踪性能变差.

	表 4 不同系统内部波动程度下的MAPE	
Table 4	MAPE under different degrees of internal system fluctua	ation

	$\lambda = 0.001^2$			$\lambda = 0.005^2$			$\lambda = 0.01^2$		
	MAPEa	$\operatorname{MAPE}_{\mathrm{m}}$	MAPE	MAPEa	$\operatorname{MAPE}_{\mathrm{m}}$	MAPE	MAPEa	$\operatorname{MAPE}_{\mathrm{m}}$	MAPE
RNN	0.0605	0.0612	0.0624	0.0828	0.0925	0.0965	0.1088	0.1133	0.1281
GRU	0.0545	0.0553	0.0576	0.0793	0.0828	0.0846	0.1052	0.1089	0.1100
MOb-GRU	0.0523	0.0538	0.0547	0.0759	0.0802	0.0827	0.0988	0.1059	0.1074

b) 系统外部波动对预测模型预测效果的影响.

系统外部波动是指高斯随机噪声的波动程度,即 过程噪声的波动方差.设置内部波动程度λ = 0.001², 考虑噪声波动程度∂分别为4.0²,6.0²,8.0²3种情况, 在验证集上针对每种情况均进行50次独立重复实验 并将结果取平均,结果如表5所示.由表5可以看出,在 3种算法下,随着噪声波动程度的增加,各模型的预测 精度均发生了不同程度的降低.原因是噪声的存在会 导致系统的随机变化程度增加,且噪声波动程度越大, 系统变化频率越大,导致在相同的结构参数下预测曲 线难以跟上实际系统的变化,预测精度下降.

基于上述分析, 在测试集上, 运用MOb-GRU模型

与两种基线模型分别对参数λ = 0.001², θ = 4.0² 的 非线性动态过程进行预测,表6给出3种模型在不同算 法下的预测结果,采用RMSE评价指标评价模型的预 测精度.图5为测试集上3种预测模型采用SAC自适应 学习率优化算法对单变量非线性动态过程进行模拟 的预测曲线与真实曲线对比,取测试集中100个数据 进行预测.由图5可知,当合理设置模型参数后,MOb-GRU的预测曲线更接近实际曲线,能更好地跟踪真实 曲线的变化.从表6可知,在3种学习率算法下,相比于 基线模型RNN与GRU,本文提出的模型具有更高的预 测精度.这是因为RNN模型只有短记忆性,与长短期 记忆模型MOb-GRU相比,无法捕捉长距离依赖关系,预测效果较差.本文将传统GRU模型的反向更新单元数设置得较多,使得整个模型的记忆范围变得较大,但与MOb-GRU相比,GRU计算量变大的同时预测效果反而变得不好.由此说明,反向更新单元数不是越多越好,当反向更新单元数接近实际过程的动态阶次时,预测效果更好.另外,在3种预测模型中,SAC自适应学习率算法与学习率矩阵算法下的预测结果均优于采用固定学习率算法所预测的结果,说明了本文设计的SAC自适应学习率算法与学习率矩阵算法的合理性.

表 5 不同系统外部波动程度下的MAPE Table 5 MAPE under different degrees of external system fluctuation

	$\vartheta = 4.0^2$		$\vartheta = 6.0^2$			$\vartheta = 8.0^2$			
	MAPEa	$\operatorname{MAPE}_{\mathrm{m}}$	MAPE	MAPEa	$\operatorname{MAPE}_{\mathrm{m}}$	MAPE	MAPEa	$\operatorname{MAPE}_{\mathrm{m}}$	MAPE
RNN	0.0605	0.0612	0.0624	0.0898	0.0922	0.0972	0.1165	0.1273	0.1552
GRU	0.0545	0.0553	0.0576	0.0808	0.0875	0.0932	0.1112	0.1142	0.1182
MOb-GRU	0.0523	0.0538	0.0547	0.0743	0.0819	0.0904	0.1043	0.1077	0.1120

表 6 $\lambda = 0.001^2$, $\vartheta = 4.0^2$ 情况下3种预测模型的RMSE

Table 6 RMSE of three prediction models in case of $\lambda = 0.001^2$, $\vartheta = 4.0^2$

	, , , , ,		
模型对比	$\mathbf{RMSE}_{\mathrm{a}}$	RMSE_{m}	RMSE
RNN	0.3017	0.3120	0.3249
GRU	0.2648	0.2965	0.3082
MOb-GRU	0.2445	0.2616	0.2854



测曲线与真实曲线



4.2 多变量非线性动态过程的建模与预测

在单变量非线性动态过程的研究基础上,将其拓

展为多变量非线性动态过程进行研究.本文考虑三输 入单输出过程,该过程的仿真系统可表示为

$$\begin{cases} \varepsilon_g \frac{\mathrm{d}}{\mathrm{d}t} x_g(t) + x_g(t) = \beta_g u_g(t), \ g = 1, 2, 3, \\ y(t) = \sum_{g=1}^3 f_g(x_g(t)) + w(t), \\ \sum_{g=1}^3 f_g(x_g(t)) = \\ x_1(t) + 1.2x_1^2(t) - 0.3x_1^3(t) + 17.76x_2(t) - \\ 103.79x_2^2(t) + 229.62x_2^3(t) + 1.5x_3(t) - 0.3x_3^3(t), \end{cases}$$
(15)

其中:本小节用时间域的单变量高阶微分方程描述线 性环节各通道的输入输出关系,系数 $\varepsilon_1 = 1.5, \varepsilon_2 = 2, \varepsilon_3 = 1, \beta_1 = 0.5, \beta_2 = 0.3, \beta_3 = 0.7; 非线性环节$ 用多项式形式表示; 多变量非线性动态过程的阶次设置为1;本小节所用数据集通过式(15)(17)–(19)仿真生成,共生成81000个输入输出数据样本点,数据集划分 $同单变量非线性动态过程; <math>u_g(\cdot)$ 代表非线性动态过程 的输入信号,作为软测量中的辅助变量,对于每个输 入信号,均取为多正弦信号; $x_g(\cdot)$ 表示系统的中间状 态变量; $y(\cdot)$ 表示系统的输出变量,作为软测量中的主 导变量;输入输出变量的采样周期T = 1 min; $w(\cdot)$ 表 示分布服从均值为0、方差为 ϑ 的高斯随机噪声,即

$$w(\cdot) \sim N(0, \vartheta).$$
 (16)

考虑1个混频输入多变量非线性动态过程,即输入 信号中含有低、中、高3种频率的正弦信号

$$u_1(t) = 80 + 15\sin(0.02t) + 2\cos(0.2t) +$$

$$0.8\sin(8t),$$
(17)
$$u_2(t) = 80 + 15\sin(0.05t) + 2\cos(0.5t) + 0.8\sin(5t),$$
(18)

$$u_3(t) = 80 + 15\sin(0.08t) + 2\cos(0.8t) + 0.8\sin(2t).$$
(19)

预选学习率是以0.19为基础增量而构成的一个数 组,设置噪声波动程度 ϑ = 4.0²进行后续问题的研究. 为了合理使用提出的MOb-GRU软测量模型,首先需 要对MOb-GRU的结构参数进行确定,选择MAPE作 为评价指标,进行网络训练性能的判断.与单变量非 线性动态过程确定网络结构参数的实验一样,在验证 集上进行对比实验,每组实验均进行50次并将结果取 平均,经过9组对比实验后,最终确定MOb-GRU结构 参数j = 40,i = 2; GRU结构参数j = 40,i = 39; 根 据第2节理论分析,设置RNN结构参数j = 2,i = 1. 在测试集上,运用MOb-GRU模型与两种基线模型分 别对混频输入多变量非线性动态过程进行预测,结果 如表7所示.

在实验中,用到的计算机CPU主频为1.80 GHz,仿 真软件为MATLAB R2020a.表8显示了3种预测模型 分别采用SAC自适应学习率算法对混频输入多变量 非线性动态过程进行训练的时间需求.由表7--8可知, 相比于RNN模型, MOb-GRU模型提高了网络的训练 性能与预测精度;相比于GRU模型, MOb-GRU模型 在提高预测精度的同时还一定程度上减轻了网络的 计算负荷.图6给出测试集上3种预测模型在SAC自适 应学习率优化算法下的预测曲线与真实曲线,取测试 集的200个数据进行预测.从图6可以看出,在存在高 斯噪声的情况下,基于MOb-GRU软测量模型得到的 估计值能更好地拟合实际值的变化趋势,具有更高的 建模精度,说明MOb-GRU模型能够充分捕获数据中 隐藏的信息,从而使预测运算能够达到更好的效果.

表 7 3种预测模型的RMSE Table 7 RMSE of three prediction models

模型对比	RMSE _a	$\mathrm{RMSE}_{\mathrm{m}}$	RMSE
RNN	0.3302	0.3437	0.3659
GRU	0.3166	0.3291	0.3593
MOb-GRU	0.3085	0.3155	0.3445

表 8 3种模型模拟所需的时间 Table 8 Time required for three models' simulation

	RNN	GRU	MOb-GRU
模拟所需时间	10.0313 s	110.3750 s	57.1563 s

对比表6和表7可以得到图7所示结果,图7展示了 用SAC自适应学习率算法进行训练时,MOb-GRU模

型相较于传统GRU与RNN模型的预测结果对比. 由图 7可看出,当实际过程动态阶次较高时,从预测效果上 更能体现MOb-GRU长短期记忆模型较GRU长短期 记忆模型与RNN短记忆模型的优越性;但当实际过程 动态阶次较低时, MOb-GRU模型与GRU, RNN的预 测精度相差不多,而基于RNN预测所需的时间较少. 综上所述,可以得到:1)实际过程动态阶次高→系统 动态特性丰富→包含前面时刻的u, y 多→所需存储 空间较大→适合选择MOb-GRU模型→模型具有长 短记忆性且能充分体现当前时刻与前面时刻丰富的 非线性动态关系→性能优于长短期记忆的传统GRU 模型与短记忆的RNN模型; 2) 实际过程动态阶次低 →系统动态特性贫乏→包含前面时刻的u,y较少→ 所需存储内存较小→系统记忆性较短→适合使用短 记忆的RNN进行处理.因此比较MOb-GRU与GRU, RNN时, 需选取动态阶次高的非线性动态过程才更能 体现MOb-GRU长短期记忆网络的优越性.





Fig. 6 Comparison between predicted curve and real curve of the actual process under SAC adaptive learning rate algorithm





5 结论

本文提出了一种依赖模型阶次的工业软测量网络 模型——MOb-GRU,介绍了MOb-GRU神经网络的训 练原理、算法及流程,设计了两种学习率优化方法—— 阶跃衰减类(SAC)自适应学习率算法与学习率矩阵算 法. 实验表明, 记忆神经网络总的模块数与描述实际 过程动态特性丰富程度的能力相关, 每一个模块内部 状态变量的维度与表示非线性的能力相关, 设置好合 适的参数后, 神经网络能够包含复杂且充分的非线性 动态特性, 以上两部分共同完成了记忆神经网络对非 线性动态过程的充分性建模. 在输出预测任务中MOb-GRU的预测精度高于RNN与GRU模型, 采用SAC自 适应学习率算法和学习率矩阵算法的网络训练结果 均优于采用固定学习率算法的训练结果, 但MOb-GRU的泛化能力和适用范围有待进一步确定. 未来作 者将收集实际现场数据, 利用自适应学习率或学习率 矩阵算法对MOb-GRU软测量模型进行高效训练, 并 给出合适的输出预测以构造更多有效的虚拟样本.

参考文献:

- HU Changsong, JIA Yangang, XIA Bokai. Current situation and development of soft sensor technology. *Petroleum Tubular Goods & Instruments*, 2003, 17(3): 4 6.
 (胡长松, 贾延刚, 夏伯楷. 软测量技术现状与发展. 石油管材与仪器, 2003, 17(3): 4 6.)
- [2] SHI X, XIONG W. Approximate linear dependence criteria with active learning for smart soft sensor design. *Chemometrics and Intelligent Laboratory Systems*, 2018, 180: 88 – 95.
- [3] GE Z. Active probabilistic sample selection for intelligent soft sensing of industrial processes. *Chemometrics and Intelligent Laboratory Systems*, 2016, 151: 181 – 189.
- [4] ZHANG X, ZHU Q, JIANG Z Y, et al. A novel ensemble model using PLSR integrated with multiple activation functions based ELM: Applications to soft sensor development. *Chemometrics and Intelligent Laboratory Systems*, 2018, 183: 147 – 157.
- [5] DUCHANOY C A, MORENO-ARMENDARIZ M A, URBINA L, et al. A novel recurrent neural network soft sensor via a differential evolution training algorithm for the tire contact patch. *Neurocomputing*, 2017, 235: 71 – 82.
- [6] CAO Pengfei, LUO Xionglin. Wiener structure based modeling and identifying of soft sensor systems. *Acta Automatica Sinica*, 2014, 40(10): 2179 – 2192.
 (曹鹏飞, 罗雄麟. 基于Wiener结构的软测量模型及辨识算法. 自动 化学报, 2014, 40(10): 2179 – 2192.)
- [7] YIN L, WANG H, FAN W. Active learning based support vector data description method for robust novelty detection. *Knowledge-based Systems*, 2018, 153: 40 – 52.
- [8] XIAO H, BAI B, LI X, et al. Interval multiple-output soft sensors development with capacity control for wastewater treatment applications: a comparative study. *Chemometrics & Intelligent Laboratory Systems*, 2019, 184: 82 – 93.
- [9] WANG Z, LUO X. Modeling study of nonlinear dynamic soft sensors and robust parameter identification using swarm intelligent optimization CS-NLJ. *Journal of Process Control*, 2017, 58: 33 – 45.
- [10] GAO Yue, GAO Zhenhai, LI Xiangyu. Soft measurement method for vehicle yaw rate based on adaptive Kalman filter. *Journal of Jiangsu University (Natural Science Edition)*, 2005, 26(1): 24 27.
 (高越,高振海,李向瑜. 基于自适应Kalman滤波的汽车横摆角速度 软测量算法. 江苏大学学报(自然科学版), 2005, 26(1): 24 27.)
- [11] ZHANG Lei. Research and application of data-driven soft sensor method for complex industrial processes. Changsha: Hunan University, 2019.
 (张雷. 基于数据驱动的复杂工业过程软测量方法研究与应用. 长沙:

(张雷.基于数据驱动的复杂工业过程软测量方法研究与应用.长沙: 湖南大学,2019.)

- [12] BAKSHI B R. Multi-scale PCA with application to multi-variate statistical process monitoring. *AIChE Journal*, 1998, 44(7): 1596 – 1610.
- [13] SOONS Z, STREEFLAND M, STRATEN G V, et al. Assessment of nearinfrared and software sensor for biomass monitoring and control. *Chemometrics and Intelligent Laboratory Systems*, 2008, 94(2): 166 – 174.
- [14] XU Xueliang. The development and present situation of artificial neural network. *Microelectronics*, 2017, 47(2): 239 242.
 (徐学良.人工神经网络的发展及现状. 微电子学, 2017, 47(2): 239 242.)
- [15] KANEKO H, FUNATSU K. Maintenance-free soft sensor models with time difference of process variables. *Chemometrics and Intelligent Laboratory Systems*, 2011, 107(2): 312 – 317.
- [16] LIN B, RECKE B, SCHMIDT T M, et al. Data-driven softsensor design with multiple-rate sampled data: acomparative study. *Industrial* & *Engineering Chemistry Research*, 2009, 48(11): 5379 – 5387.
- [17] OHL J, AYOUBI M, KURTH M. Identification of a high efficiency boiler based on neural networks with locally distributed dynamics. *Proceedings of the 1996 IEEE International Symposium on Intelligent Control.* Dearborn, MI, USA: IEEE, 1996, 9: 426 – 431.
- [18] ZHANG M, LIU X. A soft sensor based on adaptive fuzzy neural network and support vector regression for industrial melt index prediction. *Chemometrics and Intelligent Laboratory Systems*, 2013, 126: 83 – 90.
- [19] YU Jing, JIN Xiuzhang. PFC-PID main steam temperature cascade predictive control based on RBF neural network. *Journal of North China Electric Power University (Natural Science Edition)*, 2020, 47(6): 91 98.
 (千静, 金秀章. 基于RBF神经网络的PFC-PID主汽温串级预测控制. 华北电力大学学报(自然科学版), 2020, 47(6): 91 98.)
- [20] KONG Rui, CAI Jiachun, HUANG Gang. Defense to adversarial attack with generative adversarial network. *Acta Automatica Sinica*, 2020, 41: 1 17.
 (孔锐, 蔡佳纯, 黄钢. 基于生成对抗网络的对抗攻击防御模型. 自动 化学报, 2020, 41: 1 17.)
- [21] ZHAO Zhibiao, LIU Bin. Adaptive identification model for export clinker temperature in cement grate cooler. *Control Theory & Applications*, 2019, 36(4): 651 658.
 (赵志彪, 刘彬. 水泥篦冷机出口熟料温度自适应辨识模型. 控制理 论与应用, 2019, 36(4): 651 658.)
- [22] WANG J, ZHANG J, WANG Y. Nonlinear identification of one-stage spur gearbox based on pseudo-linear neural network. *Neurocomputing*, 2018, 308: 75 – 86.
- [23] KUMAR R, SRIVASTAVA S, GUPTA J R P, et al. Diagonal recurrent neural network based identification of nonlinear dynamical systems with Lyapunov stability based adaptive learning rates. *Neurocomputing*, 2018, 287: 102 – 117.
- [24] CHEN C, ZHUO R, REN J. Gated recurrent neural network with sentimental relations for sentiment classification. *Information Sciences*, 2019, 502: 268 – 278.
- [25] GRACHEV A M, LGNATOV D I, SAVCHENKO A V. Compression of recurrent neural networks for efficient language modeling. *Applied Soft Computing*, 2019, 79: 354 – 362.
- [26] MUZAFFAR S, AFSHARI A. Short-term load forecasts using LSTM networks. *Energy Procedia*, 2019, 158: 2922 – 2927.
- [27] ZHANG D, KABUKA M R. Combining weather condition data to predict traffic flow: A GRU-based deep learning approach. *IET Intelligent Transport Systems*, 2018, 12(7): 578 – 585.
- [28] FU R, ZHANG Z, LI L. Using LSTM and GRU neural network methods for traffic flow prediction. *The 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. Wuhan, China: IEEE, 2016, 11: 324 – 328.

- [29] PAVITHRA M, SARULADHA K, SATHYABAMA K. GRU based deep learning model for prognosis prediction of disease progression. *The 3rd International Conference on Computing Methodologies and Communication (ICCMC)*. Erode, India: IEEE, 2019, 3: 840 – 844.
- [30] PAVITPOK S, PHASUKKIT P, PRADABPET C. Vertical speed prediction for the efficient landing of aircraft using GRU. *The 5th International STEM Education Conference (iSTEM-Ed)*. Hua Hin, Thailand: IEEE, 2020, 11: 47 – 50.
- [31] NI Weicheng. Residual life prediction method based on GRU. Information Technology and Informatization, 2021, 3: 30 33.
 (倪维成. 基于门控循环单元的剩余寿命预测方法. 信息技术与信息 化, 2021, 3: 30 – 33.)
- [32] LYNN H M, PAN S B, KIM P. A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks. *IEEE Access*, 2019, 7: 145395 – 145405.
- [33] RANJEETH S, KANDIMALLA V A K. Predicting diabetes using outlier detection and multi-layer perceptron with optimal stochastic gradient descent. *IEEE India Council International Subsections Conference (INDISCON)*. Visakhapatnam, India: IEEE, 2020, 10: 51 – 56.
- [34] GULDE R, TUSCHER M, CSISZAR A, et al. Deep reinforcement learning using cyclical learning rates. *The 3rd International Conference on Artificial Intelligence for Industries (AI41)*. Irvine, CA, USA: IEEE, 2020, 9: 32 – 35.
- [35] CHEN Cong, HOU Lei, LI Lele, et al. Prediction of aircraft fuel flow based on recurrent neural network. *Science Technology & Engineering*, 2021, 21(27): 11663 11673.
 (陈聪, 候磊, 李乐乐,等. 基于GRU改进RNN神经网络的飞机燃油流量预测. 科学技术与工程, 2021, 21(27): 11663 11673.)

- [36] BAO Xi, LI Feng. Light stripe center extraction method based on Hessian matrix combined with gradient variance. *Computer & Digital Engineering*, 2020, 48(8): 2018 – 2023.
 (鲍茜,李锋. Hessian矩阵结合梯度方差的光条纹中心提取方法. 计 算机与数字工程, 2020, 48(8): 2018 – 2023.)
- [37] DENZEL A, KASTNER J. Hessian matrix update scheme for transition state search based on gaussian process regression. *Journal of Chemical Theory and Computation*, 2020, 16(8): 5083 – 5089.
- [38] GENG Hongmei. A summary of BFGS algorithm. Popular Science & Technology, 2011, 11: 3 - 4.
 (耿红梅. BFGS算法综述. 大众科技, 2011, 11: 3 - 4.)
- [39] NHAT-DUC H, NGUYEN Q L, TRAN V D. Automatic recognition of asphalt pavement cracks using metaheuristic optimized edge detection algorithms and convolution neural network. *Automation in Con*struction, 2018, 94: 203 – 213.
- [40] WANG Bingxuan. Research on LM optimization algorithm and neural network predictive control in nonlinear system. Taiyuan: Taiyuan University of Technology, 2016.
 (王炳萱. LM优化算法和神经网络预测控制在非线性系统中的研究.太原:太原理工大学, 2016.)

作者简介:

王 珠 博士, 副教授, 硕士生导师, 目前研究方向为系统辨识与 自适应控制, E-mail: jz21561@163.com;

刘佳璇 硕士研究生,目前研究方向为系统辨识与自适应控制,

E-mail: 2020211232@student.cup.edu.cn.