基于强化学习的减少烘丝过程中烟丝 “干头” 量的方法

毕素环[1,2] 蒋一翔[3] 于树松[1] 丁香乾[1] 牟亮亮[1] 王 彬[4]

摘 要 针对烘丝开始阶段存在的烘丝温度超调、过干烟丝较多等问题, 提出一种基于强化学习 (Reinforcement learning, RL) 的减少烟丝“干头” 量的方法. 该方法利用生产实时数据作为输入特征向量感知烘丝生产过程的状态变化, 以烟丝含水率检测值为依据来评价、优化烘丝温度控制策略, 实现对烘丝机温度设定值的在线修正, 优化烘丝开始阶段的温度控制, 有效改善烟丝过干问题. 与烘丝机的自动控制模式和人工干预模式相比, 烟丝含水率的标准偏差比自动控制时降低了 44.7%, 比人工干预时降低了 14.3%. 实验结果表明烟丝含水率的稳定性有较大提高, 烟丝“干头” 量明显减少, 验证了所提方法的有效性和可行性.

关键词 烟丝含水率, 过干烟丝, 强化学习, 超调

引用格式 毕素环, 蒋一翔, 于树松, 丁香乾, 牟亮亮, 王彬. 基于强化学习的减少烘丝过程中烟丝 “干头” 量的方法. 自动化学报, 2023, **49**(8): 1679−1687

DOI 10.16383/j.aas.c190367

A Method for Reducing Over-dried Tobacco at Head Stage of Drying Process Based on Reinforcement Learning

BI Su-Huan[1,2] JIANG Yi-Xiang[3] YU Shu-Song[1] DING Xiang-Qian[1] MU Liang-Liang[1] WANG Bin[4]

Abstract To solve the problem of high overshoot of drying temperature and too much over-dried cut tobacco at head stage of drying process, a method for reducing over-dried tobacco based on reinforcement learning (RL) is proposed. The presented model detects dynamic performance of tobacco drying system relying on real-time production data, evaluates and optimizes the temperature control according to the amount of moisture content in tobacco, and performs real-time correction for the set value of dryer temperature. The control strategy optimizes the temperature control and effectively improves the over-dried problem. The proposed method is compared with the automatic control mode and manual intervention mode of dryer. The standard deviation of the moisture content in dried tobacco is reduced by 44.7% compared with automatic control, and decreased by 14.3% compared with manual intervention. The experimental results show that the stability of the moisture content level is improved, and the amount of over-dried tobacco is significantly reduced, which verify the effectiveness and feasibility of the proposed method.

Key words The amount of moisture content in tobacco, over-dried tobacco, reinforcement learning (RL), overshoot

Citation Bi Su-Huan, Jiang Yi-Xiang, Yu Shu-Song, Ding Xiang-Qian, Mu Liang-Liang, Wang Bin. A method for reducing over-dried tobacco at head stage of drying process based on reinforcement learning. *Acta Automatica Sinica*, 2023, **49**(8): 1679−1687

烟丝含水率是制丝生产中对烟丝质量评价的关键指标, 烘丝机通过控制温度对烘丝筒内的烟丝进行加热干燥, 使得烟丝含水率符合卷烟工艺的要求, 进而提高烟丝的填充能力和耐加工性, 改善烟丝质量[1]. 目前烘丝机的温度控制主要采用 PID 控制配合部分前馈控制, 在烘丝过程的开始阶段烘丝温度由于控制超调升高过快[2], 此时烘丝机内的烟丝量较少, 从而容易导致该阶段的烟丝水分散失过多使得烟丝过干, 即常见的“干头” 现象. 在实际生产中, 烘丝开始阶段人工对温度干预比较频繁, 中间阶段比较少, 尾部阶段一般采取关闭蒸汽阀门的方式减少烟丝水分散失, 因此开始阶段的烘丝温度控制是整个烘丝生产过程的关键环节. 由于人工操作经验的差异性和控制上的延时性, 使得对开始阶段烘丝温度的控制不稳定, 造成烟丝含水率批次间波动比较大. 以生产数据中随机选取的某个时间段内 223 个生产批次数据为例, 图 1 中实线是烟丝含水率标

收稿日期 2019-05-14 录用日期 2019-09-02
Manuscript received May 14, 2019; accepted September 2, 2019
国家重点研发计划 (2017YFA0700601) 资助
Supported by National Key Research and Development Program of China (2017YFA0700601)
本文责任编委 陈德旺
Recommended by Associate Editor CHEN De-Wang

1. 中国海洋大学信息科学与工程学院 青岛 266000 2. 青岛理工大学信息与控制工程学院 青岛 266520 3. 浙江中烟工业有限责任公司 杭州 310000 4. 中国海洋大学继续教育学院 青岛 266000
1. College of Information Science and Engineering, Ocean University of China, Qingdao 266000 2. School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520 3. China Tobacco Zhejiang Industrial CO., LTD., Hangzhou 310000 4. School of Continuing Education, Ocean University of China, Qingdao 266000

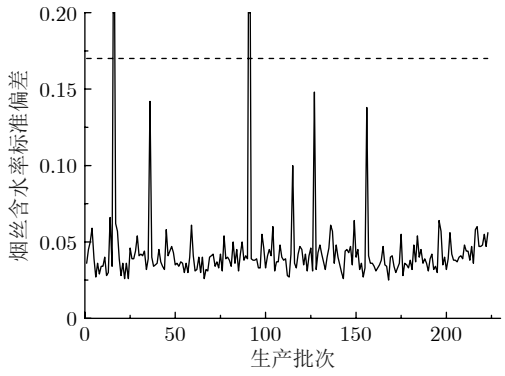


图 1 人工干预时烟丝含水率标准偏差

Fig. 1 Standard deviation of moisture content level in tobacco when in manual intervention mode

准偏差曲线, 虚线是《卷烟工艺规范》[3] 中对标准偏差规定的要求值 0.17, 从图中可看出, 虽然严重不合格的生产批次很少, 但是批次间波动比较大, 难以满足实际生产的精度需求.

烘丝过程控制主要是根据烟丝含水率检测值调整过程参数, 使得烘丝温度更好地匹配烟丝流量的变化, 使得烟丝含水率更接近目标值. 为提高控制精度、减少过干烟丝量, Pakowski 等[4] 提出了一种分布式参数模型描述烟丝温度、速度、含水率以及蒸汽温度等变量的关系, 通过迭代得到烟丝干燥过程的最佳工艺参数组合. Zhou 等[5] 基于多采样率的 RBF-ARX 模型对烘丝过程进行分阶段参数优化以减少烘丝过程中的过干烟丝量. 参数优化模型一旦建立, 最优参数设置就是固定值组合, 未能充分考虑生产系统的动态变化. 在炼焦生产过程中, 赖旭芝等[6] 为实现操作参数的动态调整, 提出了一种基于多目标遗传算法的优化控制策略, 求解多目标优化问题.

在研究过程优化控制策略中, 除采取参数优化设置外, 还有针对生产设备 PID 控制器的改善研究. 廖龙[7] 通过设计烘丝机模糊控制器来修正 PID 的三个参数实现温度控制的优化. 郑坤明等[8] 根据 Delta 机器人系统的结构特点, 采用模糊 PID 控制与传统 PID 控制器并联的形式对主动臂输入力矩进行控制, 提高控制的实时性和准确性. 王述彦等[9] 通过设计模糊 PID 控制器改善工程机械多功能试验台二次调节加载系统的动态响应, 优化控制效果. 文中所研究的烘丝机温度控制系统包括多个 PID 控制器, 然而同时优化多个 PID 控制器的参数实际操作难度比较大[10]. 在烘丝生产的开始阶段烟丝逐渐进入烘丝筒, 此时无法检测到烟丝实际含水率, 不能进行反馈控制, 烘丝机的温度控制主要依靠前馈控制和人工干预. 待烟丝流出烘丝筒才能进行含水率检测从而进行 PID 控制, 由于控制超调往往使得烘丝温度过高, 因此针对 PID 控制器参数调优的方法[11−12] 在烘丝生产最初阶段较难充分发挥作用.

随着工业生产中数据的大量积累, 机器学习算法逐渐应用于工业过程参数优化的预测控制[13−16]. Dai 等[17] 将支持向量机 (Support vector machine, SVM) 算法和遗传算法用于设计基于 PID 控制的谷物干燥控制器, 以改善干燥加工后的谷物质量. Li 等[18] 利用递归模糊神经网络进行复杂微波干燥过程中温度和水分的预测. 为预测烟叶烘烤过程中重要参数的变化, Wu 等[19] 采用了基于自适应模糊神经网络的方法. Balbay 等[20] 将极限学习机和人工神经网络用于黑孜然种子干燥过程中含水率的预测, 分析温度和干燥速率的关系.

研究者还尝试将强化学习算法用在工业控制中设计自适应 PID 控制器, 对 PID 控制器的参数进行在线调优[21−23]. 强化学习[24] 是机器学习的一个分支, 是从环境状态到动作映射的学习, 使得采取的动作从环境中获得的累积奖励最大. Günther 等[25] 利用深度神经网络和强化学习对激光焊接过程进行表征、预测, 对焊接功率进行实时控制, 提升激光焊接质量. Jiang 等[26] 研究了具有未知运行模型的浮选工业过程中双时间尺度运行最优控制问题, 提出了一种基于起重技术和强化学习 (Reinforcement learning, RL) 的双速率数据驱动算法, 使设备层的设定值保持在规定的范围内, 同时使精矿和尾矿的运行指标保持在目标范围内. Feng 等[27] 使用部分可观测马尔科夫决策过程的无模型逼近梯度来解决动力辅助轮椅的能量优化控制问题, 通过对具有 25 个控制参数的回报进行加权探索, 采用有限视界模糊 Q-迭代进行近最优性分析. Zhang 等[28] 针对部分未知模糊系统, 提出了一种基于模糊积分强化学习的跟踪控制算法, 将实际工作反馈控制策略的求解转化为虚拟最优控制问题, 克服了系统原始信息难以精确掌握的难题, 在机械系统中实现了目标跟踪、保证了系统的稳定性. 为解决烘丝开始阶段的烟丝“干头”问题, 本文结合实际生产需求, 利用强化学习策略对烘丝过程进行建模分析, 提出了一种基于数据驱动的烘丝温度优化控制方法.

本文主要贡献总结如下:

1) 提出了一种基于强化学习策略的减少烟丝“干头” 量的优化控制方法, 将生产状况的模糊经验判别转化为基于数据驱动的量化分析, 有效提升烟丝含水率的稳定性.

2) 提出了一种利用 Actor-Critic 算法优化烘丝温度控制的方法, 该算法通过 Actor 网络和 Critic 网络动态求解最优的烘丝机温度设定值.

3) 通过在实际生产线上进行与烘丝机自动控制和人工干预控制的对比实验, 验证了所提优化控

制方法的有效性. 实验结果表明, 本文所提方法在提高控制稳定性、降低烟丝含水率标准差方面均优于对比方法.

本文结构安排如下: 第 1 节描述烟丝含水率的控制流程; 第 2 节介绍优化控制模型的结构; 第 3 节详细描述烟丝含水率优化控制的建模过程, 并通过对比实验验证所提方法的有效性; 第 4 节对本文研究内容进行总结.

1　烟丝含水率控制

制叶丝工段是烟丝生产中关键的生产工序, 主要包括切丝、膨胀、干燥、风选等工序, 见图 2, 其中的 SIROX 是烟丝膨胀设备. 干烟丝的含水率是烟丝生产的关键指标, 受原料烟丝与加工过程参数的影响. 本文将切叶丝工序和贮柜缓存之间的加工过程作为烘丝生产系统进行温度优化控制的研究. 烘丝机出口的烟丝温度高、水分检测误差大, 经过风选后烟丝温度降低, 水分仪检测数据较准确, 因此烟丝含水率的检测点选在柔性风选之后, 即图 2 中的冷却水分检测点. 目前操作工对烘丝温度的干预是基于烘丝生产系统中的过程检测量、控制量、设备工作参数等生产状态数据以及个人经验而进行的.

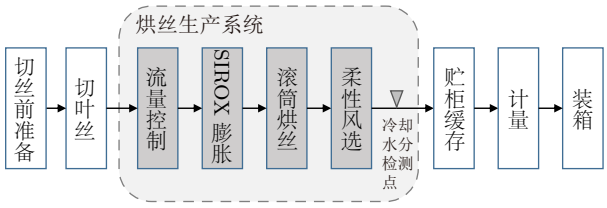


图 2　制叶丝工段

Fig. 2　The stage of cut tobacco processing

本文以德国 HAUNI 公司生产的 KLD 两段式滚筒烘丝机 (以下简称烘丝机) 为例进行研究, 该型号烘丝机包括两个相通的工作区, 目前一般采用相同的温度设置. 烘丝过程及温度控制如图 3 所示, 烟丝从入口主要经过 SIROX 增温增湿机和 KLD 烘丝机两个关键设备, 在烘丝机内进行干燥去湿. 烘丝机的控制系统主要包括前馈控制、烘丝温度控制、烟丝水分控制等部分. 烟丝含水率的控制最终是通过控制烘丝温度实现的, 烘丝温度 PID 控制器的输入量包括实际温度的偏差、前馈控制量和烟丝水分 PID 控制器的输出量, 烘丝温度的影响因素主要包括 KLD 烘丝流量、KLD 烘前水分、干燥因子等过程检测量和配方参数等, 在实际生产中人工干预修正 PID 温度控制器的设定值, 共同影响烘丝机的实际温度.

在目前人工干预模式下, 烘丝开始阶段的温度控制首先以前馈控制为主, 然后结合 PID 反馈控制

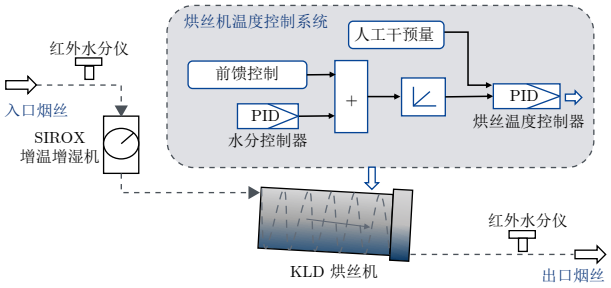


图 3　烘丝过程及温度控制

Fig. 3　Drying process and temperature control flow

共同进行烘丝温度的控制. 烘丝温度控制受多个状态参数的影响, 人工干预时主要依据烟丝冷却水分检测值以及对生产状态的预判进行烘丝温度的干预修正, 难以同时兼顾所有参数对烟丝含水率的影响. 烟丝一般在烘丝筒内滞留几分钟[29], 因此操作工对温度的干预控制具有延时性, 而该时间段内烘丝筒中的烟丝量较少以及由 PID 控制引起的温度超调, 两者叠加会导致烟丝水分散失过多, 使得达到稳定生产所需时间较长.

2　优化控制模型

2.1　优化控制策略

Actor-Critic 算法[30−31] 是强化学习中一种重要算法, 主要由策略网络 (Actor) 和评价网络 (Critic) 组成. 策略网络感知环境状态并选择控制策略输出动作, 评价网络根据来自环境状态的动作奖励和值函数计算时间差分误差, 用于优化 Actor 网络和 Critic 网络的网络参数. 基于 Actor-Critic 算法的烘丝温度优化控制策略见图 4, 主要包括烘丝机控制系统和优化控制模块. 其中, $f_1(x)\sim f_4(x)$ 是烘丝机进行温度控制时中间变量的计算函数, F_{tr}、F_{2w} 分别是计算得到的干烟丝流量和烟丝中的水分流量. 优化控制算法以原料参数、过程检测量、设备参数等作为状态输入, 以减小烟丝含水率检测值与目标值之间的差值为策略优化目标, 对烘丝温度设定值进行在线修正, 即使得式 (1) 中性能指标尽可能小:

$$J(t)=\sum_{i=1}^{N}|M_{\mathrm{act}_i}-M_{\mathrm{tar}}| \tag{1}$$

其中, M_{act_i} 和 M_{tar} 分别是烟丝含水率的实际值和目标值, 实际值是冷却水分检测值, 目标值是烟丝配方中的指标值, N 是一个生产批次中烟丝含水率采样次数.

2.2　优化控制流程

烘丝温度的优化控制流程如图 5 所示, 烘丝生

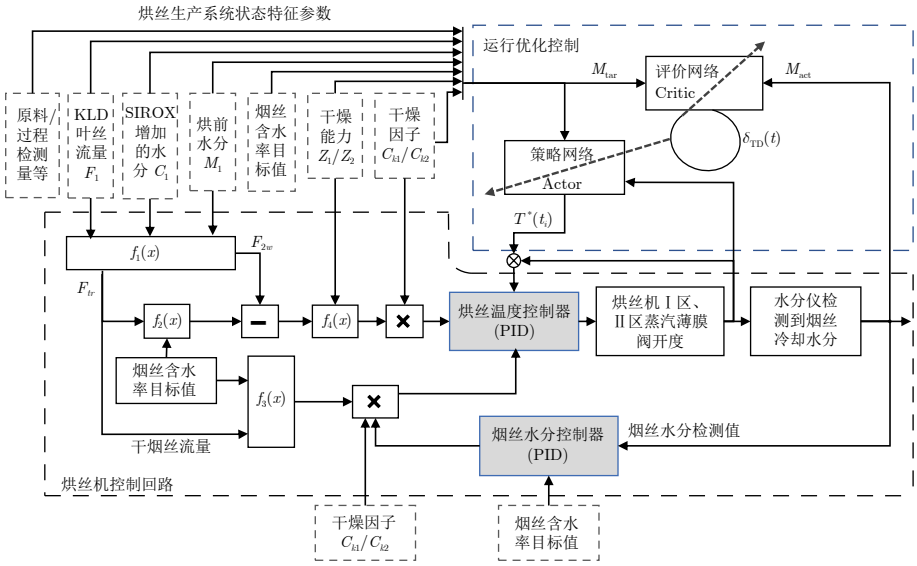


图 4 烘丝温度优化控制策略

Fig. 4 Optimal control strategy for drying temperature

产系统和优化控制模块之间通过 PLC (Programmable logic controller) 进行实时数据的同步, 优化控制模块主要包括数据采集、特征构建和 Actor-Critic 算法三部分. 实时生产数据通过 PLC 同步到优化控制模块, 数据采集部分根据算法需要尽可能多地获取表征生产状态变化的实时采样数据, 为算法感知生产环境提供数据支撑. 特征构建部分首先对采样数据进行预处理和特征参数筛选, 提取出算法可识别的状态特征和动作回报. Actor-Critic 算法感知生产状态的动态变化, 依据烟丝含水率检测值与目标值的差值对烘丝温度进行实时优化设置. 通过 PLC 将烘丝温度优化值同步到烘丝机控制系统, 进行烘丝温度设定值的在线修正, 实现烘丝温度的动态优化控制, 减少因温度超调导致的烟丝“干头”量.

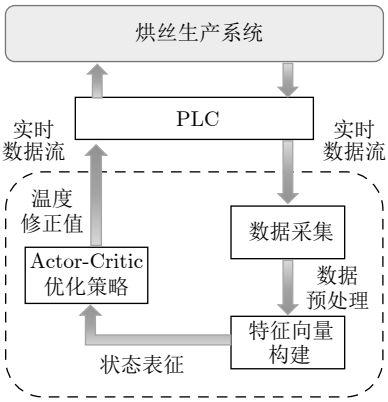


图 5 烘丝温度优化控制流程

Fig. 5 Optimization of temperature control flow

3 烟丝含水率的优化控制

3.1 生产数据采集

生产数据的采集往往是为了满足业务系统的需要而进行的, 而机器学习模型所需的重要数据采集不全面, 是人工智能技术在工业领域应用中的常见问题. 工业数据经过多年的积累往往体量很大, 但是机器学习算法需要的特征参数可能未包含在当前的数据采集系统中. 在研究之初发现对优化控制模型十分关键的烘丝机温度人工设定值并没有记录在当前的数据采集系统中, 因此增加了温度设定值的数据采集点, 并同时在三条制丝生产线上进行生产数据的采集.

烘丝生产是批次生产过程, 设定采样间隔是 10 s, 一个批次生产过程大约采样 750 次. 由于不同牌号烟丝的配方参数以及质量要求等不同, 选定产量较大的某一牌号的烟丝进行数据积累, 最终采用 2018 年 5 月 ~ 2018 年 12 月期间 800 个批次的约 60 万条生产数据进行研究.

3.2 状态特征构建

烘丝生产系统的状态感知是模型进行烘丝温度优化控制的前提[32−33], 特征向量能否尽可能全面地反映过程的状态变化直接影响控制的精度. 人工对温度干预控制时参考因素包括原料烟丝含水率、冷却水分检测值、含水率目标值等可检测量, 还包括对车间环境、生产线差异性的经验积累等潜在的影响因素. 优化模型要尽可能全面地将这些因素转化为模型可识别的特征参数 (见图 6), 与生产状态密

切相关的参数包括原料烟丝、配方参数、过程检测量以及设备参数等都应作为模型感知生产环境的输入量.

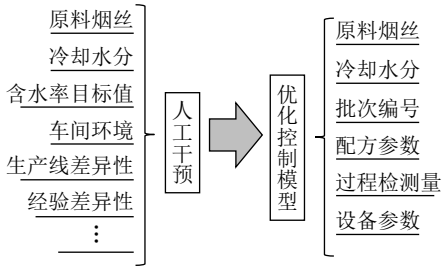


图 6　烘丝生产系统状态感知

Fig. 6　Tobacco drying system state perception

生产数据在输入模型前需要进行特征筛选和处理. 设备的频率、电流等相对固定的参数设置对模型优化作用不大, 不作为输入特征向量. 批次编号是多项信息的组合字符串, 每一位代表不同的生产信息, 例如编号 180823D205. 批次编号以字符串的形式作为输入参数, 其中隐含的信息不易被模型学习到, 因此将批次编号切分成年 (2018)、月 (08)、日 (23)、生产线编号 (D 线)、班组 (乙班)、生产序号 (05) 六项特征信息添加到输入向量中, 模型将能更好地学习生产相关信息. 人工干预时考虑的季节、生产线差异性、经验差异性等因素, 模型可以通过年份、月份、生产线编号等参数自动学习, 进而算法还能通过日期、班组、生产序号等学习到不同班组在不同生产时间段所采取的温度控制策略的差异性, 实现更精细的温度控制. 通过筛选和处理后的烘丝生产系统的状态特征见表 1, 包括原料烟丝、批次编号、配方参数、过程检测量和设备参数五大特征类别的 49 个生产状态特征.

烘丝生产数据是时间序列的数据, 采用强化学习算法进行优化控制需要保证当前的状态以及采取的动作能够决定下一时刻的状态, 因此在模型输入时需要对时间序列的数据进行预处理. 文中采用时间滑动窗口的方式进行数据截取, 当前 t 时刻的特征向量是 $t-n$ 到 t 时刻所有状态特征的串行组合, 如图 7 所示, 经过组合后形成具有 $49\times n$ 个元素的列向量作为模型的输入特征向量. 特征向量在输入模型前进行了 Min-Max 归一化处理, 将数据变为 (0, 1) 之间的小数, 即

$$x_{\text{norm}}=\frac{x-x_{\min}}{x_{\max}-x_{\min}} \tag{2}$$

其中, $x_{\min}$, $x_{\max}$ 分别为所采集生产状态数据的最小值与最大值.

3.3　Actor-Critic 优化控制设计

1) 系统状态和奖励函数

烘丝生产系统即 Actor-Critic 算法进行优化控制的环境, 上述构建的特征向量作为算法可识别的环境状态 $s(t)$. 烘丝生产系统的质量指标是干烟丝的含水率, 即冷却水分检测值. 奖励函数的设计是 Actor-Critic 优化控制的重要部分, 不同牌号的烟丝含水率要求不同, 在生产中最终目的是使烟丝含水率尽可能接近目标值, 且含水率的标准偏差越小越好. 在研究中参考 3 种常用激活函数分别设计关于烟丝含水率实际值和目标值的奖励函数, 3 种奖励函数计算式分别为

$$Sigmoid':\ r_1(t)=-\frac{1}{2}+\frac{1}{1+\mathrm{e}^{(|M_{\text{act}_i}-M_{\text{tar}}|)}} \tag{3}$$

$$tanh':\ r_2(t)=-\frac{\mathrm{e}^{(|M_{\text{act}_i}-M_{\text{tar}}|)}-\mathrm{e}^{(-|M_{\text{act}_i}-M_{\text{tar}}|)}}{\mathrm{e}^{(|M_{\text{act}_i}-M_{\text{tar}}|)}+\mathrm{e}^{(-|M_{\text{act}_i}-M_{\text{tar}}|)}} \tag{4}$$

$$ReLU':\ r_3(t)=-|M_{\text{act}_i}-M_{\text{tar}}| \tag{5}$$

其中, $tanh'$、$ReLU'$ 和 $Sigmoid'$ 分别表示参考双曲正切函数 tanh、ReLU 函数和 Sigmoid 设计的奖励函数.

3 种常用激活函数及所设计的奖励函数见图 8

表 1　烘丝生产系统状态特征

Table 1　The state features of tobacco drying system

特征类别	生产状态特征	特征数
原料烟丝	KLD 烘前水分、KLD 烘丝流量、叶丝累计量	3
批次编号	年、月、日、生产线编号、班组、生产序号	6
配方参数	KLD 除水量、含水率目标值、干燥能力、干燥因子	4
过程检测量	KLD 烘丝段蒸汽流量、SIROX 蒸汽流量、SIROX 烘丝分汽缸压力、SIROX 烘丝分汽缸温度、SIROX 排潮风机负压值、SIROX 后温度、SIROX 阀后蒸汽温度、SIROX 阀后蒸汽压力、SIROX 阀前蒸汽温度、SIROX 阀前蒸汽压力、KLD 一次减压后蒸汽压力、KLD 烘后水分、KLD 烘后温度、KLD 排潮温度、Ⅰ区工作蒸汽压力、Ⅱ区工作蒸汽压力、Ⅰ区回水温度、Ⅱ区回水温度、Ⅰ区筒壁温度、Ⅱ区筒壁温度、热风风速、热风温度、排潮负压、风选冷却排潮负压、冷却温度、冷却水分	26
设备参数	SIROX 蒸汽阀门开度、KLD 筒转速、KLDⅠ区蒸汽薄膜阀开度、KLDⅡ区蒸汽薄膜阀开度、Ⅰ区筒壁温度设定值、Ⅱ区筒壁温度设定值、热风蒸汽阀门开度、风门开度、排潮开度、风选冷却排潮开度	10

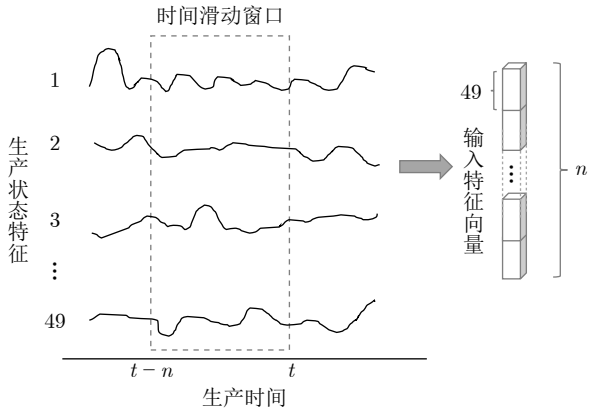


图 7 模型的输入特征向量

Fig. 7 Input feature vector of the proposed model

所示, 在绘制奖励函数曲线时以烟丝含水率目标值 12% 为例, 实际烟丝含水率在 12% 上下浮动, 因此重点绘制烟丝含水率在 12% 左右波动时的奖励函数取值. 由图看出奖励函数 $ReLU'$ 的曲线斜率不变, 当实际含水率偏离目标值或大或小时回报值的变化率是一样的, 而另外两个奖励函数具有不同的斜率, 在偏离目标值小时 $tanh'$ 的变化率比 $Sigmoid'$ 大. 根据操作经验, 越靠近烟丝含水率目标值, 温度设置幅度应该越小, 以免引起烟丝含水率波动太大, 结合实际稳定性需求本文采用 $Sigmoid'$ 作为研究的奖励函数 $r(t)$. 当实际值接近目标值时, 奖励函数接近于 0; 当实际值和目标值偏差很大时, 奖励函数取值为大于 $-1/2$ 的负值. 实际值越接近目标值, 回报值越高, $r(t)$ 是随烟丝含水率检测值变化的有界函数.

2) 烘丝温度优化控制

在实际生产中烘丝温度的正常范围是 129 ~ 134 ℃, 烘丝机可接收的最小调整幅度为 0.1 ℃, 优化控制模型实时输出烘丝温度的设定值, 通过 PLC 同步到烘丝机的 PID 控制器, 对烘丝温度设定值进行在线修正.

Actor 网络的输出动作是对烘丝温度的预测值 $a^*(t)$, 即烘丝温度控制的最优设定值. 其中策略函数为多层神经网络, 所采取动作限制在 129 ~ 134 ℃ 之间, 通过神经网络学习正态分布的均值和方差. 为探索人工干预之外的更多优化控制策略, 对神经网络输出的优化值叠加一个关于状态值函数的高斯随机噪声 η_k, 值函数小时噪声幅值增加, 值函数大时则抑制噪声, 对更大的动作空间进行探索. 烘丝温度优化值的修正计算式为

$$T(t) = a(t) + \eta_k(0,\, \sigma_{V(t)}) \tag{6}$$

高斯随机噪声的方差为

$$\sigma_{V(t)} = \frac{1}{1 + \mathrm{e}^{2V(t)}} \tag{7}$$

3) TD 误差和误差函数

基于 Actor-Critic 算法的神经网络更新采用 TD (Temporal difference) 误差算法, 计算式为

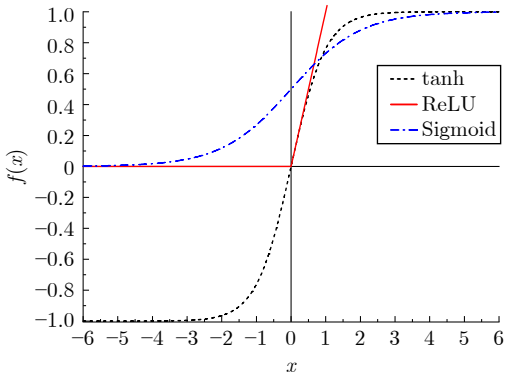
$$\delta_{\mathrm{TD}}(t) = r(t) + \gamma v_\pi(t+1) - v_\pi(t) \tag{8}$$

其中, $v_\pi(t)$ 和 $v_\pi(t+1)$ 分别为当前时刻和下一时刻的状态值函数; π 为采取的优化策略, 是对烘丝温度的修正. γ 为折扣因子, 表示未来的回报相对于当前回报的重要程度.

Actor 网络和 Critic 网络采用具有 3 层隐藏层的神经网络学习烘丝温度的优化值和状态值函数, 神经网络的输入为烘丝过程的系统状态特征 $s(t)$, 因此神经网络的输入神经元数目为时间滑动窗口内的特征采样数 m, 隐藏层神经元数目分别为 n_1, n_2, n_3, 输出分别是 $a(t)$ 和 $v_\pi(t)$.

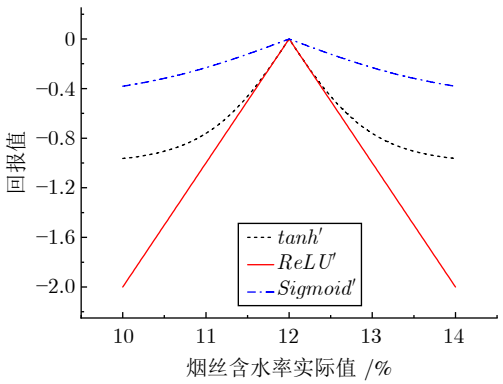
Actor 网络的迭代目标是优化网络参数输出回报值高的温度设定值, 误差函数 L_π 描述为

$$L_\pi = -\delta_{\mathrm{TD}} \ln \pi_\theta(s,\, a) \tag{9}$$



(a) 3 种常用激活函数

(a) Three commonly used activation functions



(b) 3 种本文设计的奖励函数

(b) The three designed reward functions

图 8 3 种常用激活函数及设计的奖励函数

Fig. 8 Three commonly used activation functions and the designed reward functions

其中, $\pi_\theta(s, a)$ 为网络采取的动作策略, θ 为神经网络的网络参数, 采用梯度下降法进行迭代更新, 更新式为

$$\theta^{(t)} = \theta^{(t-1)} + \alpha_A \nabla L_\pi = \theta^{(t-1)} + \alpha_A \nabla \ln \pi_\theta(s, a) \quad (10)$$

其中, α_A 为网络的学习速率.

Critic 网络基于状态值函数计算 TD 误差 δ_{TD}, Critic 网络的误差函数为

$$L_v = \frac{1}{2}\delta_{\mathrm{TD}}^2(t) \quad (11)$$

网络参数的更新式描述为

$$w^{(n)} = w^{(n-1)} + \alpha_C \nabla L_v \quad (12)$$

其中, w 为神经网络的网络参数, α_C 为网络的学习速率.

综上所述, 烘丝温度的优化控制过程具体如下:

步骤 1. 初始化模型的各个参数, 折扣因子 $\gamma = 0.9$, 学习速率 $\alpha_A = 0.02$, $\alpha_C = 0.1$, 时间滑动窗口 $n = 2$ min, 在 n 时间间隔内采样 12 次, 因此输入的特征数 $m = 588$, 隐藏层神经元数目设为 $n_1 = 256$, $n_2 = 128$, $n_3 = 64$.

步骤 2. 通过 PLC 获取生产数据并构建特征向量 $s(t)$, 筛选的特征参数 $x_i = \{x_{i1}, x_{i2}, \cdots, x_{i49}\}$, i 代表时刻 t_i, 模型的输入特征向量 $s(t_i) = \{x_{i-11}, x_{i-10}, \cdots, x_i\}$, 即具有 588 个元素的列向量. 因此输入模型的状态–烟丝含水率数据为 $\{(s(t_1), M_{\mathrm{act}_1}), (s(t_2), M_{\mathrm{act}_2}), \cdots, (s(t_i), M_{\mathrm{act}_i}), \cdots, (s(t_N), M_{\mathrm{act}_N})\}$, 根据式 (3) 计算回报值 $r(t)$. 如 t_i 时刻的实际烟丝含水率 M_{act} 为 11.9%, 目标值 M_{tar} 设为 12%, 则回报值 $r(t_i) = -1/2 + 1/(1 + \mathrm{e}^{|11.9-12|})$. Actor-Critic 优化模型根据输入特征向量 $s(t_i)$ 和回报值 $r(t_i)$ 输出烘丝温度的设定值 $a(t_i)$ 和状态值 $v_\pi(t_i)$. 为探索人工干预经验之外的温度设定值, $a(t_i)$ 叠加一个关于状态值函数的高斯随机噪声 η_k, 经过修正后的温度设定值为 $T(t_i) = a(t_i) + \eta_k(0, \sigma_{V(t_i)})$.

步骤 3. 温度优化值 $T(t_i)$ 通过 PLC 作用于烘丝机的 PID 温度控制器, 对实际烘丝温度进行在线优化控制, 进而影响烟丝含水率. 通过 PLC 获取新的生产状态数据来构建下一时刻的系统状态特征向量 $s(t_{i+1})$, 并根据实际烟丝含水率计算回报值 $r(t_{i+1})$. 将系统状态特征向量和回报值输入到网络后得到下一时刻的温度设定值 $a(t_{i+1})$ 和 $v_\pi(t_{i+1})$.

步骤 4. 由式 (8) 计算 TD 误差 $\delta_{\mathrm{TD}}(t)$, 通过损失函数 (9) 和 (11) 以及参数更新式 (10) 和 (12) 进行模型参数更新, 优化控制实验在烘丝开始阶段持续进行.

3.4　实验结果分析

为验证优化控制模型的有效性和稳定性, 在宁波卷烟厂进行了优化控制实验, 并与烘丝机自动控制和人工干预控制两种控制模式进行了对比. 实验时采用同一牌号的烟丝进行实验, 烟丝含水率的目标值是 12%, 烘丝机是 KLD2-2Z 薄板式滚筒烘丝机, 水分仪是 NDC 公司的 TM710 红外水分仪. 在制丝 D 线分别采用 3 种控制模式进行实验, 生产完成后对烘丝开始阶段 40 min 内的数据进行统计分析, 烘丝机温度和烟丝含水率曲线如图 9 所示.

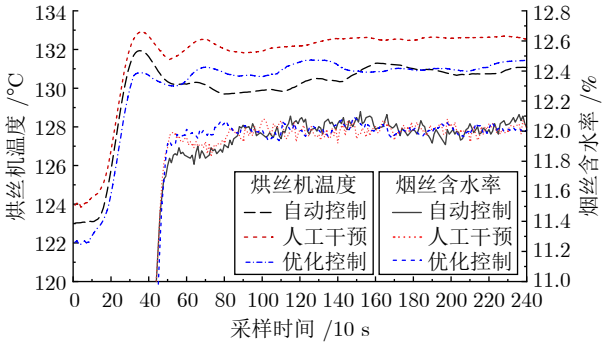


图 9　3 种模式下烘丝机温度和烟丝含水率曲线

Fig. 9　The dryer temperature and the moisture content level in tobacco when in three control modes

在烘丝机自动控制时, 烟丝开始进入烘丝机的阶段由 PID 控制引起的烘丝温度超调量较大, 此时烟丝较少、温度过高, 烟丝水分散失过多使得烟丝过干, 而且持续时间较长. 随着烟丝流量增大和 PID 控制器对温度的调整, 温度逐渐稳定, 烟丝含水率逐渐向目标值靠近. 人工干预控制时, 依据操作经验会提前对烘丝机温度进行预设置, 此时温度波动较小, 产生过干烟丝的持续时间缩短, 但由于实际生产中操作经验的差异性容易导致温度控制不稳定, 影响烟丝含水率的稳定性. 采用基于强化学习的优化控制方法时, 在烟丝进入烘丝机的阶段模型根据历史数据和实时生产数据对温度进行修正, 等检测到冷却水分时模型根据实时检测值对温度进行优化控制, 从图中可以看出烘丝机温度的波动较小, 烟丝含水率能较好地稳定在目标值附近.

实际生产中对烘丝机温度控制的目的是提高烟丝含水率的稳定性, 3 种控制模式下烘丝开始阶段的烟丝含水率标准偏差见表 2, 分别对比了在开机 20 min、30 min 和 40 min 的标准偏差. 烘丝机自动控制时, 开机 20 min 的标准偏差是 0.097, 40 min 时标准偏差为 0.076, 烟丝含水率在烘丝机的 PID 控制下逐渐稳定. 在人工干预和所提 Actor-Critic 优化控制时标准偏差明显降低, 人工干预和优化控制模型都是通过修正烘丝机的温度使得烟丝含水

表 2 烟丝含水率标准偏差
Table 2 Standard deviation of moisture content level in tobacco

控制模式	标准偏差		
	开机 20 min	开机 30 min	开机 40 min
自动控制	0.097	0.082	0.076
人工干预	0.056	0.053	0.049
Actor-Critic 优化控制	0.051	0.045	0.042

率接近目标值, 优化控制模型对烘丝机温度的调整更为精确, 烟丝含水率的标准偏差在开机 40 min 时达到 0.042, 比自动控制时降低了 44.7%, 比人工干预时降低了 14.3%. 综上所述, 基于 Actor-Critic 的烟丝含水率优化控制方法能够有效减小烘丝温度的超调量, 控制响应快, 烟丝含水率波动小, 烟丝"干头" 量大大减少.

4 结束语

在烘丝开始阶段由于烘丝温度的超调和人工干预的延时容易导致出现较多过干烟丝, 影响烘丝过程烟丝的整体质量. 本文提出了一种基于 Actor-Critic 算法的烘丝温度优化控制方法, 该方法利用连续采集的实时生产数据构建烘丝生产系统的状态特征向量, 优化控制模型通过特征向量感知实际生产环境, 依据烟丝含水率检测值进行烘丝温度的修正, 借助 PLC 对烘丝温度进行在线优化控制. 实验结果表明, 采用该优化方法烘丝温度几乎无超调, 控制响应快, 产生过干烟丝的持续时间短, 烟丝含水率的标准偏差比人工干预时降低了 14.3%, 烟丝"干头"量大大减少, 能够有效提升烘丝过程烟丝的整体质量.

下一步工作将深入研究提升烘丝全过程烟丝含水率稳定性的方法, 研究多种机器学习算法用于烘丝机温度的优化控制, 进行多参数控制和分布式、多步更新学习模式的探索. 对烘丝生产系统进行建模与仿真, 通过仿真分析与现场实验相结合验证所提算法, 提升实际烘丝生产中烟丝质量的优化空间.

References

1 Zhu W K, Wang Y, Chen L Y, Wang Z G, Li B, Wang B. Effect of two-stage dehydration on retention of characteristic flavor components of flue-cured tobacco in rotary dryer. *Drying Technology*, 2016, **34**(13): 1621−1629

2 Li Xiao-Li, Wang Kang, Yu Xiu-Ming, Su Wei. CPS-based multiple model adaptive control of GGBS production process. *Acta Automatica Sinica*, 2019, **45**(7): 1354−1365
(李晓理, 王康, 于秀明, 苏伟. 基于 CPS 框架的微粉生产过程多模型自适应控制. 自动化学报, 2019, **45**(7): 1354−1365)

3 State Tobacco Monopoly Administration. *Cigarette Making Process Specification*. Beijing: China Light Industry Press, 2016.
(国家烟草专卖局. 卷烟工艺规范. 北京: 中国轻工业出版社, 2016.)

4 Pakowski Z, Druzdzel A, Drwiega J. Validation of a model of an expanding superheated steam flash dryer for cut tobacco based on processing data. *Drying Technology*, 2004, **22**(1−2): 45−57

5 Zhou F, Peng H, Ruan W J, Wang D, Liu M Y, Gu Y F, et al. Cubic-RBF-ARX modeling and model-based optimal setting control in head and tail stages of cut tobacco drying process. *Neural Computing and Applications*, 2016, **30**: 1039−1053

6 Lai Xu-Zhi, Li Ai-Ping, Wu Min, Lei Qi. Optimization control based on the multi-objective genetic algorithm for coking plant production process. *Computer Integrated Manufacturing Systems*, 2009, **15**(5): 990−997
(赖旭芝, 李爱萍, 吴敏, 雷琪. 基于多目标遗传算法的炼焦生产过程优化控制. 计算机集成制造系统, 2009, **15**(5): 990−997)

7 Liao Long. Drying Machine Temperature Optimization Control Based on Fuzzy Control [Master thesis], Southwest University of Science and Technology, 2018.
(廖龙. 基于模糊控制的烘丝机温度优化控制 [硕士学位论文], 西南科技大学, 2018.)

8 Zheng Kun-Ming, Zhang Qiu-Ju. Fuzzy PID control of delta robot based on elastic dynamic model and genetic algorithm. *Computer Integrated Manufacturing Systems*, 2016, **22**(7): 1707−1716
(郑坤明, 张秋菊. 基于弹性动力学模型与遗传算法的 Delta 机器人模糊 PID 控制. 计算机集成制造系统, 2016, **22**(7): 1707−1716)

9 Wang Shu-Yan, Shi Yu, Feng Zhong-Xu. A method for controlling a loading system based on a fuzzy pid controller. *Mechanical Science and Technology for Aerospace Engineering*, 2011, **30**(1): 166−172
(王述彦, 师宇, 冯忠绪. 基于模糊 PID 控制器的控制方法研究. 机械科学与技术, 2011, **30**(1): 166−172)

10 Ang K H, Chong G, Li Y. PID control system analysis, design, and technology. *IEEE Transactions on Control Systems Technology*, 2005, **13**(4): 559−576

11 Skogestad S. Simple analytic rules for model reduction and PID controller tuning. *Journal of Process Control*, 2003, **13**(4): 291−309

12 Duma R, Trusca M, Dobra P. Tuning and implementation of PID controllers using rapid control prototyping. *Control Engineering and Applied Informatics*, 2011, **13**(4): 64−73

13 Murthy T P K, Manohar B. Microwave drying of mango ginger (Curcuma amada Roxb): Prediction of drying kinetics by mathematical modelling and artificial neural network. *International Journal of Food Science and Technology*, 2012, **47**(6): 1229−1236

14 Bravo S, Moreno A H. Prediction model based on neural networks for microwave drying process of amaranth seeds. In: Proceedings of the 3rd International Conference on Compute and Data Analysis. New York, USA: ACM, 2019. 88−93

15 Balbay A, Avci E, Sahin O, Coteli R. Modeling of drying process of bittim nuts (pistacia terebinthus) in a fixed bed dryer system by using extreme learning machine. *International Journal of Food Engineering*, 2012, **8**(4): 1−16

16 Chai Tian-You. Development directions of automation science and technology. *Acta Automatica Sinica*, 2018, **44**(11): 1923−1930
(柴天佑. 自动化科学与技术发展方向. 自动化学报, 2018, **44**(11): 1923−1930)

17 Dai A N, Zhou X G, Liu X D, Liu J Y, Zhang C. Intelligent control of a grain drying system using a GA-SVM-IMPC controller. *Drying Technology*, 2018, **36**(12): 1413−1435

18 Li J S, Xiong Q Y, Wang K, Shi X, Liang S. A recurrent self-evolving fuzzy neural network predictive control for microwave drying process. *Drying Technology*, 2016, **34**(12): 1434−1444

19 Wu J, Yang S X, Tian F C. An adaptive neuro-fuzzy approach to bulk tobacco flue-curing control process. *Drying Technology*, 2017, **35**(4): 465−477

20 Balbay A, Kaya Y, Sahin O. Drying of black cumin (Nigella sativa) in a microwave assisted drying system and modeling using extreme learning machine. *Energy*, 2012, **44**(1): 352−357

21 Sedighizadeh M, Rezazadeh A. Adaptive PID controller based on reinforcement learning for wind turbine control. In: Proceedings of World Academy of Science Engineering and Technology. Cairo, Egypt: 2008. 257−262

22 Chen Xue-Song, Yang Yi-Min. A novel adaptive PID controller

based on Actor-Critic learning. *Control and Decision*, 2011, **28**(8): 1187−1192
(陈学松, 杨宜民. 基于执行器−评价器学习的自适应 PID 控制. 控制理论与应用, 2011, **28**(8): 1187−1192)

23 Sun Jing-Gao, Yang Jia-Xiong, Wang Shuo, Xue Rui, Pan Hong-Guang. Design of closed-loop brain machine interface controller based on Actor-Critic and neural network. *Control and Decision*, 2018, **33**(11): 1967−1974
(孙京诰, 杨嘉雄, 王硕, 薛瑞, 潘红光. 基于 Actor-Critic 和神经网络的闭环脑机接口控制器设计. 控制与决策, 2018, **33**(11): 1967−1974)

24 Gao Yang, Chen Shi-Fu, Lu Xin. Research on reinforcement learning technology: A review. *Acta Automatica Sinica*, 2004, **30**(1): 86−100
(高阳, 陈世福, 陆鑫. 强化学习研究综述. 自动化学报, 2004, **30**(1): 86−100)

25 Günther J, Pilarski P M, Helfrich G, Hao S, Diepold K. Intelligent laser welding through representation, prediction, and control learning: An architecture with deep neural networks and reinforcement learning. *Mechatronics*, 2016, **34**: 1−11

26 Jiang Y, Fan J L, Chai T Y, Lewis F L. Dual-rate operational optimal control for flotation industrial process with unknown operational model. *IEEE Transactions on Industrial Electronics*, 2019, **66**(6): 4587−4599

27 Feng G X, Busoniu L, Guerra T M, Mohammad S. Data-efficient reinforcement learning for energy optimization of power-assisted wheelchairs. *IEEE Transactions on Industrial Electronics*, 2019, **66**(12): 9734−9744

28 Zhang K, Zhang H G, Mua Y F, Sun S X. Tracking control optimization scheme for a class of partially unknown fuzzy systems by using integral reinforcement learning architecture. *Applied Mathematics and Computation*, 2019, **359**: 344−356

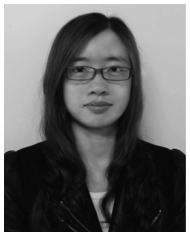
29 Zhou Kai-Min, He Jin, Sheng Ke, Yu Na, Zhu Sheng-Cai, Ji De-Xiang. Modeling and numerical simulation of residence time of cut tobacco in cylinder dryer. *Tobacco Science and Technology*, 2016, **49**(5): 94−99
(周凯敏, 何晋, 盛科, 余娜, 朱生才, 吉德祥. 滚筒烘丝机内烟丝滞留时间模型的建立及数值模拟. 烟草科技, 2016, **49**(5): 94−99)

30 Barto A G, Sutton R S, Anderson C W. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man and Cybernetics*, 1983, **13**(5): 834−846

31 Konda V R, Tsitsiklis J N. On Actor-Critic algorithms. *SIAM Journal on Control and Optimization*, 2003, **42**(4): 1143−1166

32 Liu Qiang, Zhuo Jie, Lang Zi-Qiang, Qin Si-Zhao. Perspectives on data-driven operation monitoring and self-optimization of industrial processes. *Acta Automatica Sinica*, 2018, **44**(11): 1944−1956
(刘强, 卓洁, 郎自强, 秦泗钊. 数据驱动的工业过程运行监控与自优化研究展望. 自动化学报, 2018, **44**(11): 1944−1956)

33 Chen Long, Liu Quan-Li, Wang Lin-Qing, Zhao Jun, Wang Wei. Data-driven prediction on performance indicators in process industry: A survey. *Acta Automatica Sinica*, 2017, **43**(6): 944−954
(陈龙, 刘全利, 王霖青, 赵珺, 王伟. 基于数据的流程工业生产过程指标预测方法综述. 自动化学报, 2017, **43**(6): 944−954)



毕素环 中国海洋大学信息科学与工程学院博士. 青岛理工大学信息与控制工程学院讲师. 主要研究方向为机器学习与智能控制.
E-mail: bisuhuan2016@163.com
(**BI Su-Huan** Ph.D. at the College of Information Science and Engineering, Ocean University of China. She is a lecturer at the School of Information and Control Engineering, Qingdao University of Technology. Her research interest covers machine learning and intelligent control.)

蒋一翔 浙江中烟工业有限责任公司工程师. 主要研究方向为信息系统应用, 信息安全管理.
E-mail: jiangyxlunwen@sina.com
(**JIANG Yi-Xiang** Engineer at the China Tobacco Zhejiang Industrial CO., LTD.. His research interest covers information system application and information security management.)

于树松 中国海洋大学信息科学与工程学院副教授. 主要研究方向为人工智能, 智能控制. 本文通信作者.
E-mail: yushusong@ouc.edu.cn
(**YU Shu-Song** Associate professor at the College of Information Science and Engineering, Ocean University of China. His research interest covers artificial intelligence and intelligent control. Corresponding author of this paper.)

丁香乾 中国海洋大学信息科学与工程学院教授. 主要研究方向为人工智能, 智能控制.
E-mail: dingxq1995@vip.sina.com
(**DING Xiang-Qian** Professor at the College of Information Science and Engineering, Ocean University of China. His research interest covers artificial intelligence and intelligent control.)

牟亮亮 中国海洋大学信息科学与工程学院博士研究生. 主要研究方向为深度学习与数据挖掘.
E-mail: merlin_mu@163.com
(**MU Liang-Liang** Ph.D. candidate at the College of Information Science and Engineering, Ocean University of China. His research interest covers deep learning and data mining.)

王 彬 中国海洋大学继续教育学院讲师. 主要研究方向为机器学习与数据挖掘.
E-mail: wangbin@ouc.edu.cn
(**WANG Bin** Lecturer at the School of Continuing Education, Ocean University of China. His research interest covers machine learning and data mining.)