doi:10.3969/j.issn.1672-1292.2023.01.010

煤-电双目标下基于有模型强化学习的回转窑工艺参数优化

张　翔[1],谢　天[1],曹　健[1],朱　毅[2]

(1.朗坤智慧科技股份有限公司,江苏 南京 210005)
(2.扬州大学信息工程学院,江苏 扬州 225000)

[摘要]　基于煤-电双目标下回转窑工艺参数优化问题,提出了有模型强化学习的解决方法. 首先,以固定时间间隔为单位对历史工艺参数和运行目标进行数据处理与聚合. 其次,搭建概率神经网络建立回转窑控制参数与影响参数、运行目标值的关系模型,该模型被用作为后期强化学习框架中的奖励模型. 然后,利用基于模型的离线策略优化的强化学习算法构建控制参数推荐智能体,同时优化回转窑生产过程的煤电消耗. 最后,给出一个案例证明所提方法对回转窑工艺参数优化的适应性、高效性.

[关键词]　回转窑,工艺参数优化,概率神经网络,基于模型的离线策略优化,煤-电双目标
[中图分类号]TP181　[文献标志码]A　[文章编号]1672-1292(2023)01-0075-09

Optimization of Process Parameters of Rotary Kiln Based on Model-Based Reinforcement Learning Under the Dual Objectives of Coal and Electricity

Zhang Xiang[1], Xie Tian[1], Cao Jian[1], Zhu Yi[2]

(1.Luculent Smart Technology Co., Ltd, Nanjing 210005, China)
(2.College of Information Engineering, Yangzhou University, Yangzhou 225000, China)

Abstract: Aiming at the optimization problem of rotary kiln process parameters under the dual objectives of coal and electricity, this paper proposes a model-based reinforcement learning solution. Firstly, data processing and aggregation were performed on historical process parameters and operating targets in units of fixed time intervals. Secondly, a probabilistic neural network is built to establish the relationship model between the control parameters of the rotary kiln, the influencing parameters, and the operating target value, which was used as the reward model in the later reinforcement learning framework. Then, a reinforcement learning algorithm based on model-based offline strategy optimization was used to construct a control parameter recommendation agent, and at the same time, the coal and electricity consumption of the rotary kiln production process was optimized. Finally, a case analysis was given to prove the adaptability and high efficiency of the proposed method for optimizing the process parameters of rotary kiln.

Key words: rotary kiln, process parameter optimization, probabilistic neural network, model-based offline strategy optimization, coal-electricity dual objective

水泥厂作为能源密集型行业之一,其核心生产装备回转窑消耗着大量的能源,能源的消耗量直接影响到成品的成本[1],能源的高效利用一直是水泥行业实现提高竞争力的优先事项. 其中,煤、电作为水泥厂的主要能源,占总生产总成本的 40%上[2],是节省能源的重点关注对象. 为了以更低的成本生产出更多的水泥,回转窑系统性能的优化具有很好的潜力[3]. 通过优化工艺参数降低系统煤、电能耗,是实现提高性能的有效方法. 该方法也具有需要资金更少、承担风险小、操作时间短的优势[4],深受广大水泥厂商的青睐.

为了研究工艺参数与对系统优化的影响关系,需要对运行过程进行建模,主要包括了构建机理模型和构建数据模型两类. 在构建机理模型研究回转窑工艺参数优化方面,一般从回转窑内发生的物理化学变化出发,基于传热机理、流体力学理论、物质守恒规律等建立数学模型分析生产过程,模型的可读性

收稿日期:2022-09-15.
通讯作者:朱毅,博士,研究方向:数据挖掘与知识图谱. E-mail:zhuyi@ yzu.edu.cn

强[5-6]. 张荣等[5]通过研究回转窑筒体辐射和对流机理,构建了筒体热损失的计算模型. 袁芷晨等[7]综述了回转窑运行机理建模过程,并列举了部分工艺参数对水泥回转窑生产过程的影响情况. 李庆峰[8]通过分析水泥回转窑系统的工艺和影响烧成带温度的主要工艺参数,建立了回转窑机理模型,进而辅助影响温度参数的控制. 上述研究主要是为回转窑某一部件或子系统建立数学机理模型,研究参数对运行过程中产生的现象,其研究价值多为生产过程输入优越的工艺参数提供决策支持,工艺参数的决策依然人为主导. 然而,受到水泥回转窑系统运行过程中参数的非线性、时变、行为不确定的特点[9],将导致构建回转窑运行过程的整体机理模型十分困难,甚至无法反应实际的运行状态,精确性不足.

在构架数据模型优化工艺参数优化的研究时,利用能反应工业系统运行状态的历史数据,不用考虑实际的运行机理[10-12],使得到的工艺参数更符合实际情况,也更加高效便捷. 郭飞等[13]基于历史工艺数据集构建了一个模糊规则网络模型,学习工艺参数优化规则,实现了塑料注射成型工艺参数优化. 李瑞[14]融合基于 k-means 聚类的多种群灰狼算法和极端学习机构建了水泥回转窑熟料游离钙含量预测模型,建立其与工艺参数之间的函数关系,在此关系的基础上,使用了粒子群优化算法,以水泥中游离钙含量为目标,优化了工艺参数. Hassan 等[15]研究了回转窑工艺参数与系统优化的关系,以降低电耗为目标,基于人工神经网络,分析工艺参数之间的关系后给出了一个回转窑运行过程窑速、风机转述和总炉篦流量的最优静态参数值. 上述研究虽然在优化工艺参数取得了长足的进步,但是在优化回转窑工艺参数时构建的数据模型多用于表达其系统的运行机理,在优化工艺参数时依然采用优化算法,模型输出的工艺参数的动态性和模型迁移能力差,且目标单一,没有考虑多目标联合的影响. 因此,利用构建数据模型的方法优化回转窑工艺参数的潜力有待进一步挖掘. 如何在优化多目标的前提下构建一个可实时推荐工艺参数的数据智能模型,是降低生产成本、提高水泥工厂竞争力的重中之重.

得益于大数据技术、人工智能算法的快速发展及在优化工业工艺参数中积累的丰富经验,使得构建数据模型在解决复杂系统工艺参数优化问题中具有显著的应用价值. 在回转窑工艺参数优化问题中,通过直接挖掘工艺参数和运行目标之间的关系进而对其模型化,然后以该模型为前提,学习参数推荐的智能代理,促进实现回转窑运行过程对工艺参数的自主决策.

本文针对回转窑工艺参数优化问题,在煤-电双目标下,提出了一种有模型强化学习优化回转窑工艺参数的方法. 在构建工艺参数与运行目标的关系模型时把参数的分布情况考虑在内,然后将该模型作为强化学习范式中的奖励模型,使用基于模型的离线策略优化强化学习算法学习工艺参数与系统优化的关系,最后构建一个自主推荐工艺参数的智能体.

1　工艺参数优化方法

鉴于水泥回转窑工艺流程复杂的运行机理不得而知,但经验样本数据可收集. 考虑到回转窑的工艺参数寻优需要根据设备的调节压力、不同工况条件、在熟料质量达标的情况实时寻找能使单位煤耗和电耗综合成本最低的最佳工艺参数组合,提出了一种基于有模型的离线强化学习方法[16]解决回转窑工艺参数的连续动态寻优问题.

优化流程如图 1 所示,针对回转窑能耗优化任务的整体建模可以分为离线训练和在线优化阶段,描述为:

(1)采集连续时间段内的水泥生产工艺流程数据,并对数据进行预处理.

(2)构建 Offline RL 需要学习的静态数据集形式,将采集的数据集进行划分,定义为元组表达(s_t,a_t,r_t,Episode).

(3)基于 BPNN 网络构建虚拟环境模型,学习真实水泥生产过程的状态变化规律.

(4)以静态数据集及(3)中训练的状态转移模型为基础,建立 RL 能耗优化模型. 智能体采用基于离线策略的 CQL 算法学习能够使系统能耗降低的最优策略,最大化获得奖励,最终训练后的 RL 模型具有良好的能耗优化能力,并输出智能控制策略.

(5)按照一定时间跨度采集在线的 s_t 数据.

(6)通过已训练的 RL 智能控制策略估计当前步 t 能效优化控制的推荐动作策略 a_t,并由现场生产环境的操作人员决定是否执行相应动作.

(7)根据是否停窑判断此次在线优化是否完成,若优化未完成则进行下一步的数据采集及优化. 另

外,考虑到水泥生产工艺流程过程所具有的时变时延性以及数据分布随时间有逐渐偏移的情况,每次的优化动作 a_t、状态 s_t 以及现场的煤耗电耗情况需要保存下来进行后续的 RL 模型更新,以适应实际情况下复杂的水泥生产过程.

其中,(1)~(4)为离线训练阶段,(5)~(7)为多步在线优化阶段.

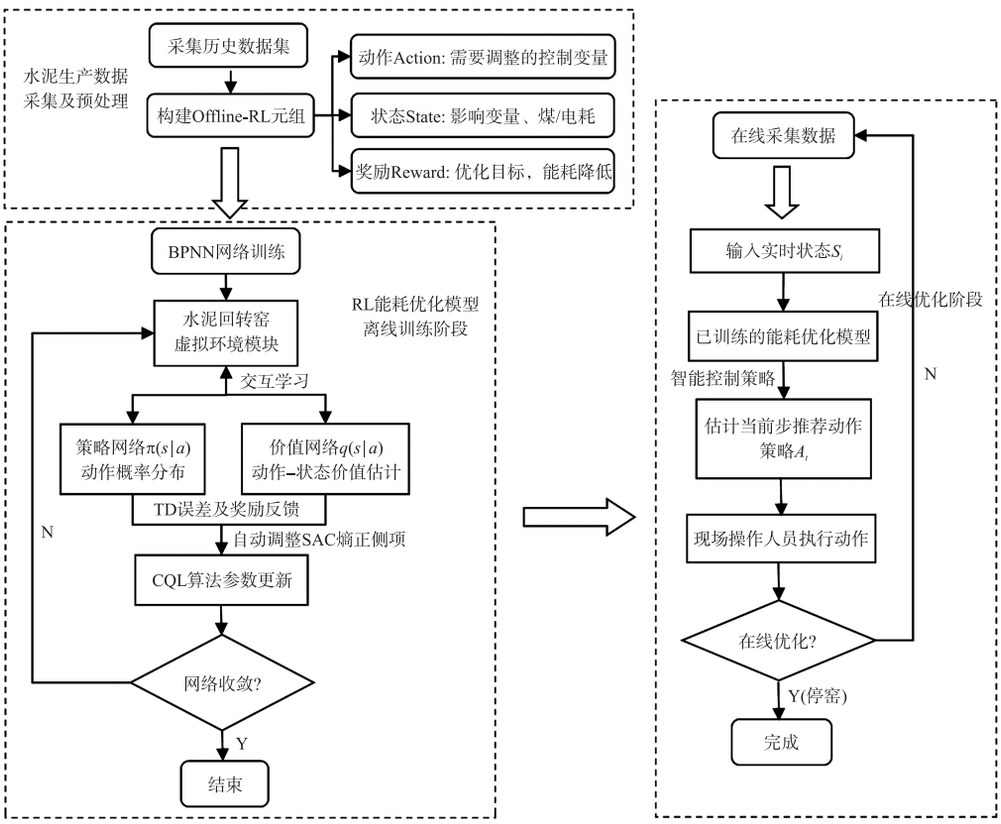

图 1　基于有模型强化学习的能耗优化流程

Fig. 1　Energy consumption optimization process based on model-based reinforcement learning

1.1　强化学习基本要素设计

将回转窑的能耗优化任务转化为一个强化学习问题,将整个水泥工艺生产流程作为系统环境. 为遵循强化学习所满足的马尔可夫决策过程,基于强化学习理论,需要对环境中智能体(Agent)相关的状态(State)、动作(Action)、奖励(Reward)进行详细定义和计算,具体如下:

(1)状态空间设计

状态空间代表了所感知到的水泥生产环境信息,以及因所选取的动作到来的变化的集合. 根据水泥生产工艺流程过程中采集的数据,环境可提供的信息包括系统内的各类影响变量 inf_t 和系统环境整体运行指标值煤耗 C_t 和电耗 E_t,则 t 时刻的综合输入状态为:

$$s_t=(inf_t,E_t,C_t). \tag{1}$$

(2)动作空间设计

选择需要人为调节的各类工艺控制变量作为动作空间中的动作向量,仅考虑连续的优化任务,即在连续域中进行优化而不是离散的,且约束各类动作的值范围为[-1,1],限制动作向量的变化在正常的范围内波动,避免非法动作. 则 t 时刻的动作空间为:

$$a_t=inf_t. \tag{2}$$

(3)奖励函数设计

由于优化目标为单位煤耗和电耗综合成本最低,即对系统性能指标来说期望的优化方向为煤耗及电耗量不断降低,则可以直接定义奖励函数为 t 时刻系统负的整体能耗值,即煤耗和电耗的加权和,并采用去均值和方差统一对奖励进行标准化:

$$r_t=-w_1C_t-w_2E_t. \tag{3}$$

式中,w_1 和 w_2 是 2 个能耗指标的权重值,需依据实际工业电价与煤价确定.

1.2　算法架构

本文采用的 model-based RL 方法借鉴了 Yu 等[17]所提出的 MOPO 算法思想. 首先,通过对采集的工艺数据建模,构造基于回转窑系统的虚拟环境模型,模拟水泥生产工艺过程的动态变化. 然后,借助强化学习方法,以系统能耗优化为目标,自动学习得到回转窑最优的工艺参数控制策略. 该方法主要分为 2 个模块,分别为仿真环境建模模块和系统能耗指标优化模块.

(1)仿真环境建模模块

针对真实水泥生产系统的虚拟环境学习,构建水泥生产过程的状态转移模型,挖掘生产过程相邻时间间隔的状态转变规律,拟合各类工艺参数、系统运行参数和能耗性能指标之间的潜在关系. 考虑到实际的水泥回转窑工艺[18]流程极为复杂,具有时变时延性,该系统过程中既涉及物理化学变化,又受人为操作、环境、工艺参数等多因素及变量的影响,为保证虚拟环境模型和真实环境的一致性,考虑采用集成建模方式的概率神经网络结构[19-20](ensemble-bootstraped probabilistic network,BPNN)来模拟水泥生产过程中所体现的随机不确定性和动力学规律.

BPNN 网络假设水泥生产系统的状态变化服从高斯分布,同时拟合多个 PNN 网络进行训练,每个 PNN 网络结构完全相同,由深层的全连接神经网络 FC 组成,采用 Swish 激活函数,基于 BPNN 网络的状态转移模型可表示为:

$$f(s,a)=\mathrm{BPNN}(\Delta s_{t,t+1},r_t\mid s_{t-L},\cdots,s_{t-1},s_t,a_t)=N(u_\theta(s_{t-L},\cdots,s_{t-1},s_t,a),\Sigma_\theta(s_{t-L},\cdots,s_{t-1},s_t,a)). \tag{4}$$

式中,网络输入当前时刻动作 a_t、状态 s_t 以及历史状态 $s_{t'}(t'\in[t-L,\cdots,t-1])$,$L$ 为历史时间步长度. 输出层为 two-head 结构,参数化估计相邻时刻状态差 $\Delta s_{t,t+1}$ 以及当前步奖励 r_t 的概率分布均值 μ_θ 和对角协方差 Σ_θ,θ 为 BPNN 网络参数,这意味着不同网络仅仅权重不同,从而通过 Boostrap 方式打乱数据顺序学习水泥生产数据中潜在的不同状态转移规律以及捕获真实水泥生产系统的任意不确定性. 因此,对于需要预测的下一状态 s_{t+1} 为通过重参数技巧从网络最终输出的状态差分布中采样得到的值与当前输入的当前步状态 s_t 之和,

$$s_{t+1}=\Delta s_{t,t+1}+s_t=\mu_\theta+\varepsilon*\Sigma_\theta+s_t,\varepsilon\sim N(0,1). \tag{5}$$

定义 BPNN 网络损失函数为负对数似然函数,网络训练时最小化该损失目标:

$$\underset{\theta}{\mathrm{argmin}}\mathrm{loss}=\sum_t^N[\mu_\theta-\Delta s_{t,t+1}]^{\mathrm{T}}\Sigma_\theta^{-1}[\mu_\theta-\Delta s_{t,t+1}]+\lg\det\Sigma_\theta. \tag{6}$$

同时,引入 L2 正则化损失缓解所训练的状态转移模型的过拟合行为. 此外,BPNN 网络训练后需要选取验证集 loss 较小的前 k 个 PNN 作为最优子模型库,并存储相应的网络参数,在预测时会随机从子模型库中选取一个子模型进行计算.

(2)系统能耗指标优化模块

由于水泥生产过程具有复杂的数据分布漂移情况,为使智能体在离线的水泥生产工艺数据集以及仿真的水泥生产环境模型中有效学习能够使系统能耗降低的解决策略,最大化累积奖励回报,采用基于最大熵强化学习 SAC[21-22]框架的保守 Q-Learning(CQL)算法[23]. 一方面,SAC 使用 2 个动作价值网络,Q 网络损失函数为:

$$L_Q=E_{(s_t,a_t,r_t,s_{t+1})\sim R}\left[\frac{1}{2}(Q(s_t,a_t)-(r_t+\gamma(\min Q(s_{t+1},a_{t+1})-\alpha\lg\pi(a_{t+1}\mid s_{t+1}))))^2\right]. \tag{7}$$

式中,R 是策略历史收集的数据,γ 为折算系数,α 为正则化系数,控制熵的重要程度. 鉴于需要控制的水泥生产工艺参数变量属于连续动作空间,对 SAC 算法的策略网络 π 输出动作概率分布的均值和标准差,并采用 tanh 激活函数对动作值进行映射. 策略网络的损失函数由 KL 散度得到,训练时需要最大化该函数:

$$L_\pi=E_{a_t\sim\pi}[Q(s_t,a_t)-\alpha\lg\pi(a_t\mid s_t)]. \tag{8}$$

另一方面,CQL 算法的引入有效解决了数据分布偏移问题,避免分布外动作出现 Q 值的过估计,减少 SAC 算法中因策略网络优化方向的错误改变而导致的外推误差变大,对能够准确学习水泥生产过程中的复杂和多模态数据分布非常有帮助. CQL 算法通过在原 TD 误差的基础上增加正则化项,学习真实 Q 网络的下界,从而不断更新 Q 值和改进策略:

$$\overline{Q}_{\mathrm{new}}\leftarrow\mathrm{argmin}_Q\max_\pi\beta\cdot E_{s\sim R}\left[\lg\sum_{a_t}\exp(Q(s_t,a_t))-E_{s\sim R,a\sim\overline{\pi}(a_t\mid s_t)}[Q(s_t,a_t)]\right]+L_{\mathrm{Bellman}}. \tag{9}$$

式中,$\bar{\pi}$ 为通过已有的真实工艺生产数据而近似得到的真实行为策略,β 是平衡因子,L_{Bellman} 为 TD 误差. 由于水泥工艺流程中所存在的大量不确定性,利用 Q 网络直接输出当前状态下所期望的确定动作值并不合理,在原 CQL 算法的基础上,引入隐式分位数网络 Q 函数 IQN[24] 估计动作值的分布.

此外,RL 模型训练采用引入带权重的优先级经验回放机制,指导抽取重要程度较高的经验样本学习,提高 RL 模型性能. 同时,考虑到智能体对虚拟的水泥生产环境模型的使用次数变多,仿真模型累积的复合误差会快速增加,且随着时间推移,水泥生产过程的数据分布发生较大变化,使得过早之前训练的环境模型得出的结果变得很不可靠,从而进一步影响智能体最优策略的学习. 针对当前有模型 RL 算法的局限性,基于 MBPO[16] 和 MOPO[17] 方法中的成功经验,本文同样设置虚拟环境模型可用较少的推演步数,并借助 Wasserstein distance 度量,对模型所预测得到的奖励进行惩罚,以此提高模型泛化能力.

2　案例研究

采集某水泥厂回转窑系统的 2022 年 1 月到 5 月的历史工艺参数数据,按 1 min 为数据间隔收集,涉及的工艺参数主要包括:人为控制变量有风机高压变频频率及电流、篦速、头排/尾排高压变频频率及电流,各类系统内部的影响变量为入窑二次风温度、三次风温度、窑尾烟室氧化氮含量、窑尾烟室温度、窑尾烟室压力、窑尾烟室氧含量、分解炉内/中部/出口温度、熟料温度、O_2 浓度、CO 浓度、窑头罩负压、窑尾负压、分解炉出口温度,需要优化的系统运行指标为单位煤耗和电耗.

数据按 7∶3 比例划分训练集与测试集,训练 BPNN 时设置 5 层全连接神经网络,隐含层神经元个数为 200,L2 权重衰减率为 0.000 25,采用 AdamW 优化器,学习率 lr 为 1e-3,epoch 为 1 000,设置初始集成 PNN 网络个数为 7 个,选取验证集上表现最好的 3 个 PNN 网络作为最优模型集.

对于 model-based CQL 算法:策略网络和 Q 网络学习率分别设置为 3e-4 和 3e-3,隐含层神经元个数为 128,贝尔曼折扣系数为 0.99,目标 Q 网络软更新系数 tau = 0.005,SAC 熵正则项学习率 alpha_lr 为 3e-4,CQL 损失修正系数 beta = 5,单步从数据池抽取数据大小 buffer_sample_size 为 256,epoch 为 2 000. 对虚拟环境模型的推演最大长度 rollout 限制为 5,模型不确定性基于 BPNN 网络训练得到的预测方差 L2 范数形式量化,预测奖励的惩罚系数固定为 0.8.

2.1　评价指标

采用平均绝对误差(MAE)、均方误差(MSE)、拟合优度 R^2 评估虚拟环境模型的预测效果,以反映与真实环境模型的差距:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i-y_i|, \tag{10}$$

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-y_i)^2. \tag{11}$$

n 是样本量,MAE 和 MSE 计算预测值 $\hat{y}_i$ 与实际值 y_i 间误差,值越小模型精度越高,

$$R^2=1-\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-y_i)^2\Big/\frac{1}{n}\sum_{i=1}^{n}(\bar{y}_i-y_i)^2. \tag{12}$$

$\bar{y}_i$ 是样本均值,R^2 能判断模型好坏,取值范围为($-\infty$,1],R^2 越大,模型效果越好.

2.2　BPNN 网络预测性能分析

准确的虚拟环境模型对后续 RL 能耗优化模型训练至关重要,包括真实水泥回转窑系统的状态变化机制是否被 BPNN 网络真正学习以及确定模型预测奖励的合理性. 图 2 显示了 BPNN 训练后在测试集上计算得到的部分工艺影响变量趋势.

从图 2 中可以看出,多数温度参数的拟合与实际值基本吻合,BPNN 网络有效学习到了此类影响变量的趋势变化规律,R^2 达到 0.9 以上,均方误差 MSE 和绝对值误差 MAE 较低. 氧含量、氧化氮以及压力参数的原始数据波动较大,BPNN 网络仍在一定程度上捕捉到了该类气体或压力参数与其他影响变量间的非线性关系和内部状态的转变机理. 对比各类影响变量的 95%置信区间范围,BPNN 网络表现了水泥生产过程中的不确定性,特别是在氧含量以及压力参数上,较大的方差表明该类参数在时间上数据分布更易发生漂移和数值会发生较大变化.

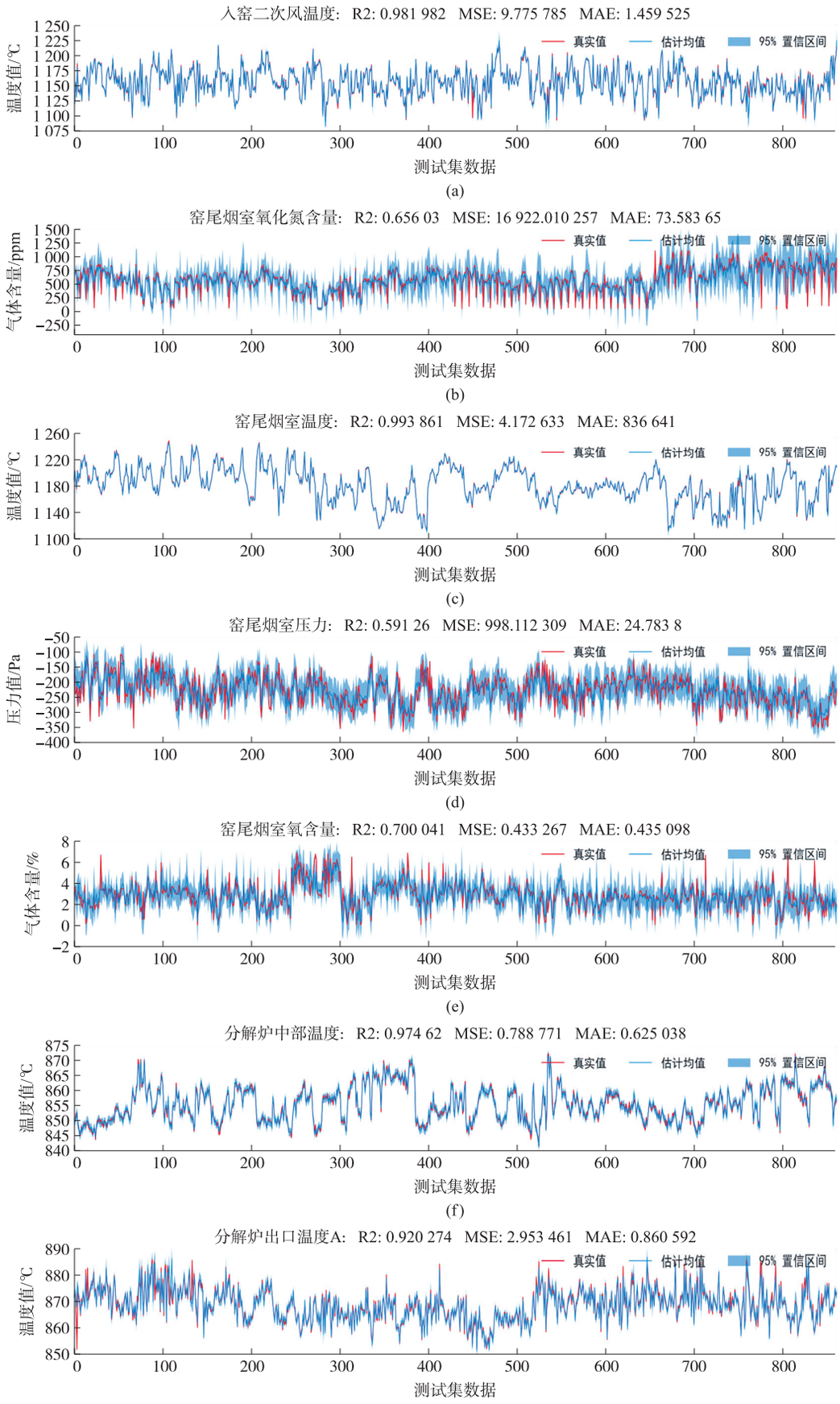

图 2　系统部分工艺影响变量预测曲线

Fig. 2　Prediction curve of some process influencing variables of the system

同样绘制了 BPNN 网络输出的奖励预测曲线. 尽管奖励的计算综合考虑了煤耗和电耗指标值,BPNN 网络学习并不困难,其对于奖励目标值与各类控制变量和影响变量间的潜在关系捕捉较为准确,从而保证了所训练的虚拟环境模型与实际的水泥回转窑系统环境模型具有一致性.

此外,为测试采用不同算法的虚拟环境模型效果,分别尝试了 LSTM[25]、自回归递归神经网络 DeepAR[26]、梯度提升树 XGBoost[27]、支持向量机 SVM[28]和高斯过程 GPR 模型[29]如表 1 所示. 由结果能够得到,传统的机器学习模型 SVM 和 GPR 模型较差. 虽然整体对影响变量的拟合 BPNN 网络不如 LSTM 和 DeepAR,但其训练速度快,仅依靠全连接层的网络结构更具有优势,且在奖励预测方面. 如图 3 所示,其 *MAE* 误差效果远低于其余基准模型,模型性能最佳.

表 1　模型效果对比

Table 1　Comparison of model's effect

虚拟环境模型	BPNN 网络	LSTM	DeepAR	XGBoost	支持向量机	高斯过程
影响变量 *MAE*	8.271	8.254	7.092	10.993	12.236	10.298
奖励 *MAE*	2.523	2.977	3.819	2.764	4.017	6.153

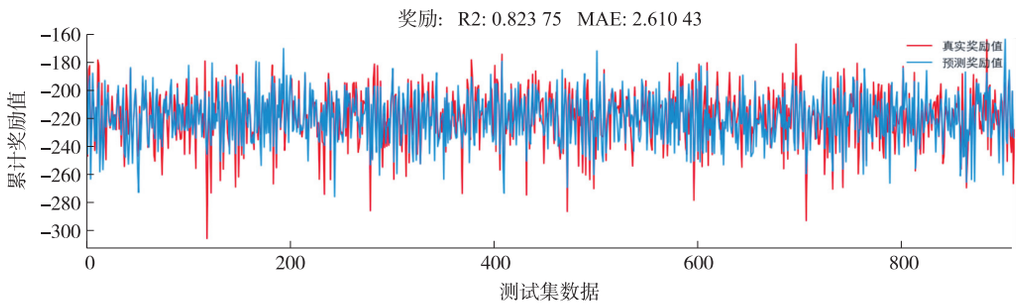

图 3　测试集奖励预测效果

Fig. 3　Test set reward prediction effect

2.3　能耗优化模型效果验证分析

RL 智能体需要通过与构建的虚拟水泥生产环境模型交互进行持续训练,由于强化学习本身存在不稳定性,采用多次仿真试验方式进行学习. 图 4 是 CQL 算法 1 000 次 Epochs 训练过程的各网络损失变化图,前 300 代策略网络损失呈现逐渐下降趋势,2 个价值网络的损失不断上升并趋于峰值,反映了智能体在虚拟环境中逐渐探索到了未知的状态空间,利用已获得的信息,记忆和理解水泥生产过程相关影响变量与控制变量间的潜在关系,并通过多步的价值迭代估计不断调整和寻找能使系统能耗降低的最佳优化方向. 最终通过训练迭代,各损失逐渐下降至收敛,智能体学习到了较优的工艺参数控制策略.

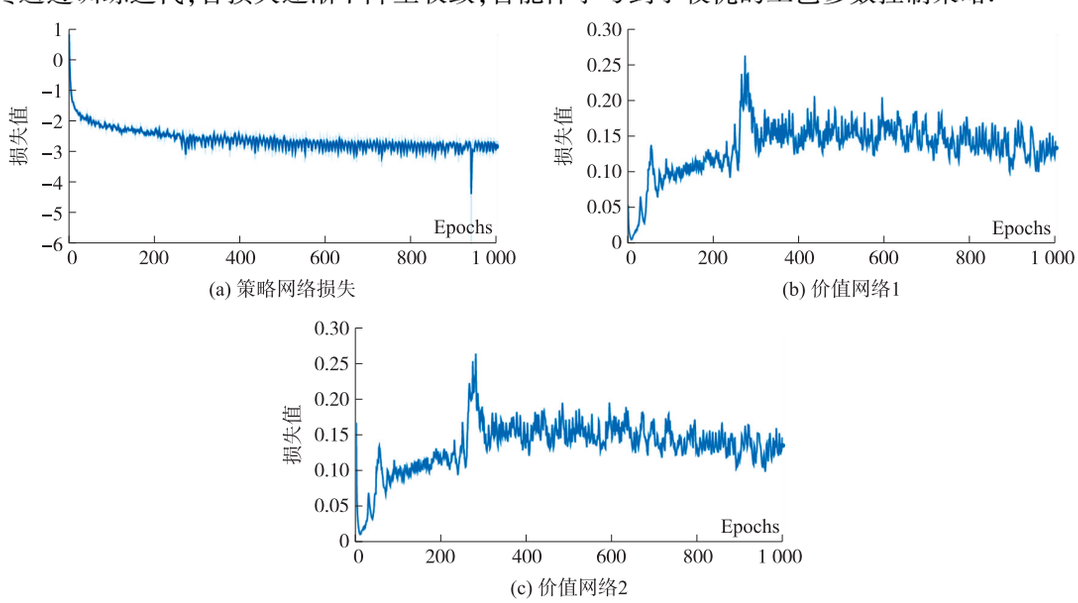

图 4　智能体训练损失曲线

Fig. 4　Agent training loss curve

为体现所用的 RL 算法能够针对真实的水泥回转窑系统有显著的能耗优化作用,最直接衡量策略学习好坏的性能指标是智能体在实际系统环境中的最大期望回报奖励. 基于水泥回转窑系统是持续优化的序列决策任务,选择以水泥生产的启停窑过程作为单个 RL 回合分别计算了每个训练 Epoch 下的总奖励值大小.

图 5 显示了智能体在训练过程中的累计奖励变化趋势. 针对离线 RL 环境,同样分析了有/无虚拟环境模型下 RL 算法从确定的数据集中学习到的策略行为和表现. 其中,CQL 模型为本文提出的方法,PPO 模型[30]算法框架类似于在线 RL 算法,采用替代的虚拟环境模型不断与智能体交互,并采用行为克隆 BC[31]方法作预训练. 在大多数 Epoch 内 SAC 和 CQL 算法的累积奖励都有较大的波动,其需要在虚拟的环境模型中不断试错并探索水泥生产环境内的潜在状态空间,而在线策略 PPO 算法受益于 BC 初始化以及每次迭代需要约束在信任域内进行策略更新的特性,训练相对稳定. 对比 4 类算法的收敛情况,有模型-SAC 算法收敛最慢,其平均累积奖励明显低于有模型 CQL 算法,表明 CQL 在 SAC 基础上通过进一步近似推断真实的行为策略下界和进行策略提升,切实学到了较优的策略. PPO 模型收敛速度最快,但在回转窑的优化场景下其仅采用当前迭代策略采样和学习的特点限制了其整体性能. 此外,由无模型的 CQL 算法结果验证了环境学习的必要性,虚拟环境模型可通过合理地生成更多数据从而辅助智能体学习到更优的策略,整体累积奖励相比无模型增益 8.27%.

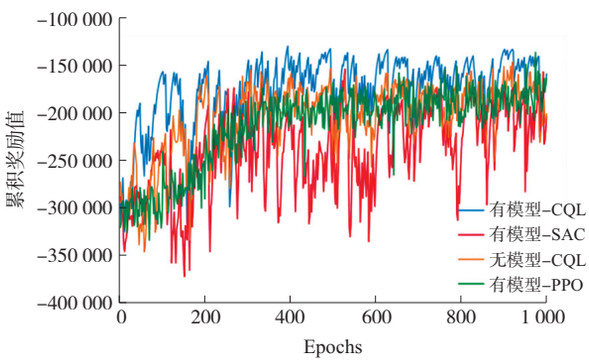

图 5　智能体训练过程累积奖励变化

Fig. 5　Cumulative reward changes during training

结果表明提出的有模型 CQL 算法达到了相对最优的效果,策略网络推荐的控制变量能够符合实际生产系统的进程,智能体已经学会了如何根据水泥回转窑系统内部不同的过程参数条件去控制相应的参数. 将利用构建的控制参数推荐模型在本文所研究案例的水泥制造商实现了应用,提取一个月的使用数据,煤耗与水泥产量比上个月降低了 2.3%,电耗与水泥产量比上个月降低的 3.4%,应用效果明显,起到了减少水泥生产成本、提高水泥制造商竞争力的作用.

3　结论

本文提出了一种面向煤-电双目标的工艺参数优化方法,该方法利用概率神经网络构建了回转窑仿真模型,挖掘回转窑各工艺参数和能耗的潜在非线性关系,预测性能优于 5 种基准模型. 同时基于有模型 RL 离线策略算法 MB-CQL,实现水泥回转窑生产工艺的能耗指标优化. 通过多次对比实验结果显示,MB-CQL 相比同类算法能达到更高的目标累积奖励值,证明了本文所提出方法的高效性,为水泥生产过程提供基础理论支撑.

[参考文献](References)

[1] RADWAN A M. Different possible ways for saving energy in the cement production[J]. Advances in Applied Science Research,2012,3(2):1162-1174.

[2] CHATTERJEE A,SUI T B. Alternative fuels— effects on clinker process and properties[J]. Cement and Concrete Research,2019,123:105777.

[3] ZHENG J Q,ZHAO L,DU W L. Hybrid model of a cement rotary kiln using an improved attention-based recurrent neural network[J]. ISA Transactions,2022,129:631-643.

[4] LV S Z,YU H L,WANG X H,et al. Multi-control strategy combinatorial control of burning temperature of cement rotary kiln[C]//2018 IEEE 4th Information Technology and Mechatronics Engineering Conference,2018:86-90.

[5] 张荣,刘小燕,武伟宁,等. 回转窑筒体热损失测量系统的研究[J]. 电子测量与仪器学报,2017,31(11):1843-1848.

[6] GENG F,LI Y M,WANG X Y,et al. Simulation of dynamic processes on flexible filamentous particles in the transverse

section of a rotary dryerand its comparison with ideo-imaging experiments[J]. Powder Technology,2011,207:175-182.
[7] 袁芷晨,杨永斌,李骞,等. 球团回转窑建模与仿真的研究进展[J]. 钢铁研究学报,2022,34(11):1187-1196.
[8] 李庆峰. 新型干法水泥回转窑烧成带温度建模与控制研究[D]. 合肥:合肥工业大学,2020.
[9] 殷润. 基于数据驱动的水泥生产能耗系统建模与优化[D]. 南京:南京邮电大学,2021.
[10] 林满山,梁欣. 回转窑煅烧配置参数的预测模型设计[J]. 科技创新与应用,2017,6(14):49-50.
[11] 张成华,雷玉成,刘伟. 应用遗传算法优化铝合金穿孔型等离子弧立焊工艺参数[J]. 扬州大学学报(自然科学版),2004(3):32-35.
[12] 曹丽茹,王晓强,王排岗,等. 基于 NSGAⅡ算法的超声滚挤压工艺参数优化[J]. 塑性工程学报,2022,29(7):19-25.
[13] 郭飞,汪汝健,张云,等. 塑料注射成型工艺参数优化的模糊规则网络模型[J]. 机械工程学报 2022,58(20):206-220.
[14] 李瑞. 多种群优化算法研究及在水泥回转窑中的应用[D]. 秦皇岛:燕山大学,2019.
[15] HASSAN A,SEYED S H,JAFAR H. Improvement of a cement rotary kiln performance using artificial neural network[J]. Journal of Ambient Intelligence and Humanized Computing,2021,12:7765-7776.
[16] JANNER M,FU J,ZHANG M,et al. When to trust your model:model-based policy optimiz-ation[C]//33rd Conference on Neural Information Processing Systems. Canada,Vancouver,2019.
[17] YU T H,THOMAS G,YU L,et al. MOPO:Model-based offline policy optimization[C]//34th Conference on Neural Information Processing Systems. Canada,Vancouver,2020.
[18] 周剑平. 水泥生产工艺[M]. 西安:西北大学出版社,2008.
[19] CHUA K,CALANDRA R,MCALLISTER R,et al. Deep reinforcement learning in a handful of trials using probabilistic dynamics models[C]//32nd Conference on Neural Information Processing Systems. Canada,Monteal,2018.
[20] LAKSHMINARAYANAN B,PRITZEL A,BLUNDELL C. Simple and scalable predictive uncertainty estimation using deep ensembles[C]//31st Conference on Neural Information Processing Systems. Long Beach,CA,USA,2017.
[21] HAARNOJA T,ZHOU A,Abbeel P,et al. Soft actor-critic:off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]//International Conference on Machine Learning. Sweden,Stockholm,2018:1861-1870.
[22] HAARNOJA T,ZHOU A,HARTIKAINEN R,et al. Soft actor-critic algorithms and applications[J]. arXiv Preprint arXiv:1812.05905,2019.
[23] KUMAR A,ZHOU A,TUCKER G,et al. Conservative Q-learning for offline reinforcement learning[C]//34th Conference on Neural Information Processing Systems. Canada,Vancouver,2020.
[24] DABNEY W,OSTRONSKI G,SILVER D,et al. Implicit quantile networks for distributional reinforcement learning[C]//International Conference on Machine Learning. Sweden,Stockholm,2018.
[25] YU Y,SI X,HU C,et al. A review of recurrent neural networks:LSTM cells and network architectures[J]. Neural Computation,2019,31(7):1235-1270.
[26] SALINAS D,FLUNKERT V,GASTHAUS J,et al. DeepAR:Probabilistic forecasting with autoregressive recurrentnetworks[J]. International Journal of Forecasting,2020,36(3):1181-1191.
[27] CHEN T,HE T,BENESTY M,et al. Xgboost:extreme gradient boosting[J]. R Package Version 0.4-2,2015,1(4):1-4.
[28] CHERKASSKY V,MA Y. Practical selection of SVM parameters and noise estimation for SVM regression[J]. Neural Networks,2004,17(1):113-126.
[29] CHALUPKA K,WILLIAMS C K I,MURRAY I. A framework for evaluating approximation methods for Gaussian process regression[J]. Journal of Machine Learning Research,2013,14:333-350.
[30] SCHULMAN J,WOLSKI F,DHARIWAL P,et al. Proximal policy optimization algorithms[J]. arXiv Preprint arXiv:1707.06347,2017.
[31] SYED U,BOWLING M,SCHAPIRE R E. Apprenticeship learning using linea programming[C]//Proceedings of the 25th International Conference on Machine Learning. Finland,Helsinki,2008.

[责任编辑:陈　庆]