

DOI: 10.13957/j.cnki.tcx.2023.04.019

引文格式:

程荣健, 方毅翔, 赵怡, 等. 基于强化学习的瓷砖产线智能生产调控算法研究[J]. 陶瓷学报, 2023, 44(4): 776–785.

CHENG Rongjian, FANG Yixiang, ZHAO Yi, et al. Intelligent control algorithm of ceramic tile production line based on reinforcement learning [J]. *Journal of Ceramics*, 2023, 44(4): 776–785.

基于强化学习的瓷砖产线智能生产调控算法研究

程荣健, 方毅翔, 赵怡, 张天助, 李俊, 王俊祥

(景德镇陶瓷大学 机械电子工程学院, 江西 景德镇 333403)

摘要: 现代瓷砖生产过程可被看作是一个复杂的系统, 其中一些调控行为会影响瓷砖最终的产品质量。传统瓷砖生产过程的工艺参数通常是通过多次实验确定的, 这些实验一般基于工程师的经验行为。然而, 经验行为往往难以准确确定工艺参数, 同时最优工艺参数会随着实际工况(例如外部环境)进行动态变化。在瓷砖产线中多变的工况影响下, 很难保证产品质量的稳定性。为了解决经验主导的人工调控问题, 并实现实际过程(工况)中的参数动态更新, 首次提出了一种基于深度强化学习(DRL)算法的瓷砖产线智能调控框架。所构建的框架包含环境(Environment)模块和智能体(Agent)模块。其中, 环境模块基于数据挖掘技术来模拟、更新瓷砖产线中的各种工况, 并通过随机森林(RF)预测模型及时预测相应的产品质量。智能体调控模块能够根据预测的产品质量快速自适应调整工艺参数, 使瓷砖产品达到预期产品质量。实验结果表明, 该方法构建的预测模型准确性相较于其他同类方法的产品质量预测模型性能更好, 平均提高率为 2%。同时, 经过多次迭代后, 瓷砖产线智能调控算法可以将生产工艺参数合格提高到 95%, 具有较好的调控效果。

关键词: 深度强化学习; 瓷砖生产; 工艺参数; 随机森林

中图分类号: TQ174.76

文献标志码: A

文章编号: 1000-2278(2023)04-0776-10

Intelligent Control Algorithm of Ceramic Tile Production Line Based on Reinforcement Learning

CHENG Rongjian, FANG Yixiang, ZHAO Yi, ZHANG Tianzhu, LI Jun, WANG Junxiang

(School of Mechanical and Electronic Engineering, Jingdezhen Ceramic University, Jingdezhen 333403, Jiangxi, China)

Abstract: The modern ceramic tile production process can be taken as a complex system, where various regulatory behaviors could affect the quality of the final ceramic tile products. The process parameters of traditional ceramic tile production are usually determined through multiple experiments, which are based on the experience and behavior of engineers. However, it is difficult to accurately determine process parameters based on empirical behavior. Also, optimal process parameters would dynamically change with actual operating conditions (such as external atmosphere). Under the influence of variable working conditions in ceramic tile production lines, it is difficult to ensure the stability of product quality. In order to solve the manual control problem dominated by experience and achieve dynamic parameter updates in actual processes (working conditions), an intelligent control framework is proposed for the first time for ceramic tile production lines, based on deep reinforcement learning (DRL) algorithm. The framework includes an Environment module and an Agent module. The Environment module simulates and updates various working conditions in the ceramic tile production line, based on data mining technology, where the corresponding product quality could be timely predicted through random forest (RF) prediction model. The intelligent

收稿日期: 2023-03-25。

修订日期: 2023-04-24。

基金项目: 国家自然科学基金(62062044, 62063010); 景德镇市级科技计划项目(2020ZDGG004)。

通信联系人: 方毅翔(1987-), 男, 硕士, 讲师。

Received date: 2023-03-25.

Revised date: 2023-04-24.

Correspondent author: FANG Yixiang (1987-), Male, Master, Lecturer.

E-mail: fangyixiang@jci.edu.cn

Agent control module can quickly adaptively adjust process parameters based on the predicted product quality, so that quality of the ceramic tile product meets the expectation. It is found that the accuracy of the prediction model constructed by using this method is higher than that of similar methods, with an average improvement rate of 2%. At the same time, after multiple iterations, the intelligent control algorithm for the ceramic tile production lines can increase the qualification of production process parameters to 95%, with desirable control effects.

Key words: deep reinforcement learning; ceramic tile production; production parameters; random forest

0 引言

中国作为世界上最大的瓷砖生产国、消费国和出口国,正以巨大的体量推动全球瓷砖生产的扩张。然而,瓷砖生产是一个相当复杂的过程,涉及许多设备参数和工艺参数。如图 1 所示,灰色方框表示生产线上的相关设备,白色方框表示设备输出的产品及产品质量评价指标。其中,泥浆质量评价指标包括容重、流速、含水率等;粉料质量评价指标包括含水率、粒径、容重等;生坯质量评价指标包括厚度、抗折强度等。其他如干坯、素坯、釉面砖、瓷砖半成品等中间产品同样也包含对应的质量评价指标。由于相关参数大多依赖工程师的经验并通过相关实验来确定,会导致很多不可控的产品消耗。因此,建立一个能更少依赖经验的瓷砖生产智能调控框架显得尤为重要。

瓷砖的生产过程涉及多个环节,此前的研究中大多是基于生产数据和人类经验知识,采用模糊系统^[1]和专家系统^[2]对生产过程工艺参数进行优化。当前有部分学者运用机器学习算法来解决相关问题。例如:Deng 等^[3]使用正交实验和反向传播人工神经网络(BP ANNs)方法来研究具有优异挤出和形状保持性能的氧化铝浆料;Ahmad 等^[4]采用随机森林算法根据新型氟氧玻璃的化学成分和离子半径预测其密度;Mu 等^[5]研究发现人工智能技术能够有效辅助鉴定中国古陶瓷。其他相关行业也有不少运用人工智能算法进行研究,瓷砖生产过程由于涉及环节多、不确定性较大。其中,工艺参数与产品性能之间的关系通常是复杂且模糊的,相关研究中使用人工智能算法较少。

在工业 4.0 时代,瓷砖制造过程涉及更多的工艺参数,而不同生产环节中不同工艺参数的多种组合形式可以构成巨大的搜索空间,实际生产中不同工艺参数对应不同的产品质量,因此在巨大的搜索空间中搜索最优的工艺参数从而达到期望的产品质量具有重要意义。在实现这一目标的传统方法一般通过简化某些不重要的细节,或加入专家经验和人工干预,导致无法在巨大的搜索

空间中灵活地进行参数调整以适应制造场景变化。强化学习^[6-7]由于可以在高维搜索空间中有效地学习马尔可夫过程的最优决策,已被广泛用于解决多个行业中的实际优化和决策问题。例如:Krashennnikova 等^[8]将保险行业的续期价格调整问题建模为基于马尔可夫决策过程(MDP)的顺序决策问题,并运用强化学习算法根据客户保留率优化收益;Han 等^[9]使用近端策略优化(PPO)算法构建公路养护计划的智能决策模型,可满足日益增长的公路养护领域的需求;Ren 等^[10]应用基于竞争构架 Q 网络(Dueling-DQN)的强化学习算法来优化调度家庭能源管理系统(HEMS);Guo 等^[11]采用强化学习框架和自预测人工神经网络模型来解决工艺窗口过窄的问题,并生产出超高精度的产品;He 等^[12]将纺织过程优化问题转化为随机博弈问题并使用深度 Q 网络(DQN)算法在多智能体系统中找到纺织臭氧氧化过程的最优解。越来越多的学者将强化学习用于解决相关行业决策问题。然而,目前还没有完整的研究来解决瓷砖制造行业中复杂的工艺参数调控问题。

通过以上研究的启发,同样可以将瓷砖生产中的工艺参数优化问题看作一个高度动态和复杂的调控过程。基于此,本文主要工作如下:

- (1) 提出了一种基于强化学习的瓷砖生产过程智能调控框架。
- (2) 构建产品质量预测模型。使用实际生产数据构建了一个 RF 快速预测模型,通过该模型可以映射出工艺参数和产品质量之间的复杂关系。然后将 RF 预测模型用于强化学习中环境模块的一部分。
- (3) 训练强化学习调控模型。依据强化学习算法建立并训练用于工艺参数调整策略的调控模型。

1 算法

1.1 随机森林算法

随机森林(RF)是由 bagging 集成学习算法演变而来的基于决策树的机器学习算法^[13],RF 包含分类和回归两种模型。其中,RF 分类模型利用

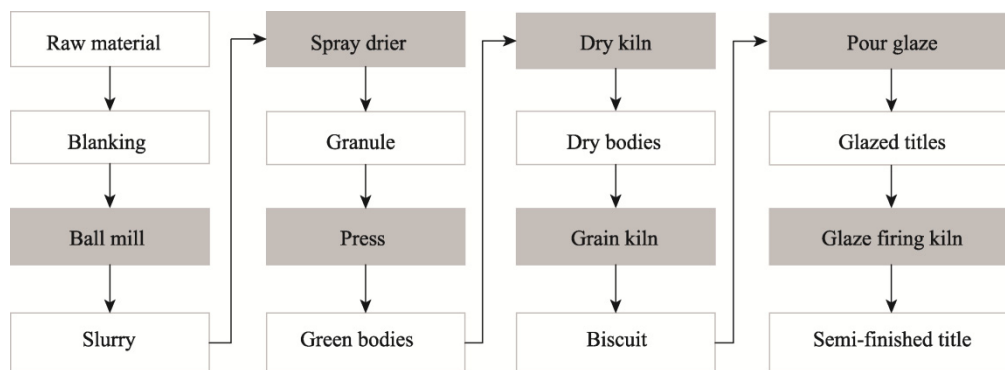


图1 瓷砖生产工艺流程图

Fig. 1 Flow chart of ceramic tile production

多个独立构建的决策树对样本进行训练与预测,运用投票机制进行决策得到样本的分类结果,最终的分类结果由决策树分类器输出结果中的票数决定。在 RF 中,后续生长的决策树不依赖于早期建立的树,它们通过从原始训练数据集中使用 bootstrap 方法(随机并且有放回抽样)独立构建决策树,因此每个构建的决策树对生长的树的分类结果都具有最终投票权。那些用于预测的生长的树被定义为 $\{h(X, \theta_k), k=1, 2 \dots K\}$ 。其中: K 表示随机森林中决策树的数量; X 表示输入向量; θ_k 是独立同分布的随机向量。分类模型可表示为:

$$H(x) = \max_Y \sum_{i=1}^k \mathbf{I}(h_i(x) = Y) \quad (1)$$

式中: $H(x)$ 表示分类结果; $h_i(x)$ 表示第 i 棵决策树的分类结果; x 是一组新的输入变量; Y 是目标类别; $\mathbf{I}(\bullet)$ 表示特征函数。

随机森林算法在一些行业中^[4, 12]被用于预测产品性能,并且发现 RF 模型在处理预测问题时可以实现较高的预测精度。因此,本研究使用 RF 构建预测模型。在 RF 模型中,以瓷砖生产过程的工艺参数为输入,以工艺参数相对应的产品质量参数为输出建立映射关系。

1.2 深度强化学习算法

强化学习作为一种重要的机器学习方法已被广泛应用于各个领域^[14-15],强化学习主要由能自主学习的智能体 (Agent) 模块和外部环境 (Environment) 模块组成。

本文使用深度强化学习中代表性算法,基于策略的 DQN^[16]来解决决策优化问题。DQN 算法与一些基础的强化学习算法不同的是其智能体模块,为了解决基础的强化学习算法如 Q-learning^[17-18]存在的维数灾难问题, DQN 算法在智能体模块使用了深度神经网络 (DNN)。其中,网络的权重值由 θ 表示, DNN 以连续的状态值 s_t 作为输入,输出为

有限个离散动作对应的有限个 Q 函数值,代替了 Q-learning 表格搜索的形式。智能体在学习最优策略时,输入状态 S_t 时智能体选择最优动作 A_t 是由策略函数 $p(S_t) = A_t$ 来决定的。智能体根据动作对应的 Q 值是否最大更改其选择动作的策略。当对环境执行动作 A_t 后环境模块能对基于当前动作的效果反馈一个奖励值 R_t ,并且使环境达到了一个新的状态 S_{t+1} ,然后智能体通过重复上述操作来学习最优决策策略。其中,环境的状态 S 是一个有限集合,由状态 $s \in S$ 表示,同理奖励 $r \in R$,动作 $a \in A$ 。

考虑到瓷砖生产过程中的动态优化过程是一个序列调控问题,可建模为 MDP,而 MDP 可以通过强化学习^[19]来求解。本文建模的 MDP 解决了瓷砖生产过程的工艺参数优化调控问题,以产品质量参数为目标,实现了工艺参数的智能调控。

2 基于深度强化学习算法的智能调控模型

强化学习框架的主体结构如图 2 所示,其主要由智能体模块和环境模块组成。其中,智能体充当调控者的角色学习最优策略,环境模块接收智能体给出的动作来调整生产过程工艺参数,通过预先训练的 RF 预测模型,并结合接收到的环境模块生产过程工艺参数,对产品质量进行预测,同时采用奖励函数计算奖励值并反馈给智能体。智能体通过不断地与环境交互训练在一定迭代步数后能够学习到调控工艺参数的最优策略。

2.1 问题定义

本小节定义了一些参数变量。将 $\{pv_1, pv_2 \dots pv_k\}$ 定义为瓷砖生产过程中的工艺参数,而工艺参数对应的多个产品质量参数由 $\{c_1, c_2 \dots c_m\}$ 来表示,这两组参数变量的集合可以表示为:

$$\begin{cases} PV = \{pv_1, pv_2 \cdots pv_k\} \\ C = \{c_1, c_2 \cdots c_m\} \end{cases} \quad (2)$$

调控模型首先要探索这些工艺参数变量是如何影响产品质量的，即某一组参数 PV 对相应产品质量 C 的影响，这一过程由预测模型来实现，可以表示为： $RF: PV \rightarrow C$ ，即

$$RF(pv_k) = C, pv_k \in PV \quad (3)$$

模型的目的是找到工艺参数变量设置的最优解，如式(4)所示：

$$\text{Argmax}_{pv_k \in PV} [RF(pv_k) = C] \quad (4)$$

在瓷砖生产中找到这一类最优解的传统方法主要依赖于经验和试错。下面将详细描述如何使用强化学习构建瓷砖生产智能调控模型。

2.2 预测模型

本文提出的框架中预测模型的构建分为两个步骤：

(1) 训练预测模型：在对实际生产数据进行预处理后，用 RF 构建并训练预测模型，训练好的预测模型可用于预测不同工艺参数条件下的产品质量特性。本文的 RF 模型由机器学习库 $Scikit-learn$ 来实现^[20]。

在将数据输入 RF 模型前先对其进行预处理，预处理后用 $Scikit-learn$ 的 `train_test_split` 函数将数据分类为训练集和测试集，其中训练集和测试集分别占 80% 和 20%。这一过程如图 3 所示。

通过使用瓷砖生产过程的部分环节的生

产数据来验证所提出的框架，当数据收集全面后可以考虑用于一整条生产线。依据实际情况预测模型分为两种模式。第一种模式是依据瓷砖生产线上的多个设备的工艺参数预测瓷砖的产品质量，包括喷雾塔、压机、素烧窑炉等。第二种模式是依据某个环节设备的工艺参数预测该环节输出产品的产品质量，以喷雾塔为例。

(2) 运用预测模型：使用训练好的随机森林模型来预测设备输出瓷砖产品的产品质量，具体是运用多元随机森林(MRF)算法^[21]来构建预测模型，在提出的框架中用该模型来模拟瓷砖生产过程。

2.3 基于深度强化学习 DQN 算法的瓷砖生产智能调控模型

基于 DQN 算法模型的各组成部分，具体框架如图 4 所示。 DQN 算法的智能体模块由框架中的当前 Q 网络和目标 Q 网络模块组成。这两个模块是由相同结构的深度神经网络设计的，环境模块则主要由 RF 预测模型构成。在训练 DQN 时，当前 Q 网络依据当前状态选择最优动作传递给环境模块，环境模块依据选择动作的效果更新状态并将奖励值反馈给智能体，通过这样不断地交互训练，让智能体学习喷雾干燥过程的最优控制策略。 DQN 为了降低样本之间的相关性，提升训练的稳定性，引入了经验回放机制和冻结参数机制，如图 4 所示，目标 Q 网络每隔 N 步更新参数。

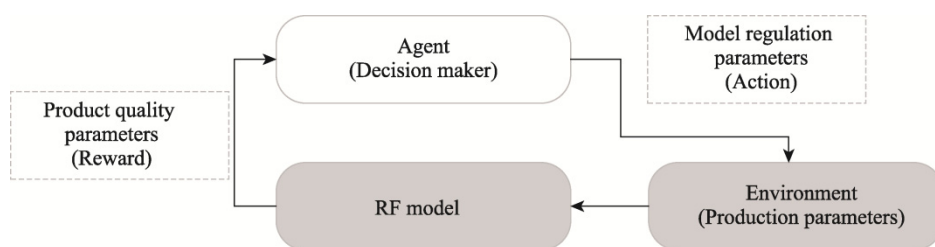


图 2 强化学习主体框架

Fig. 2 Main framework of reinforcement learning

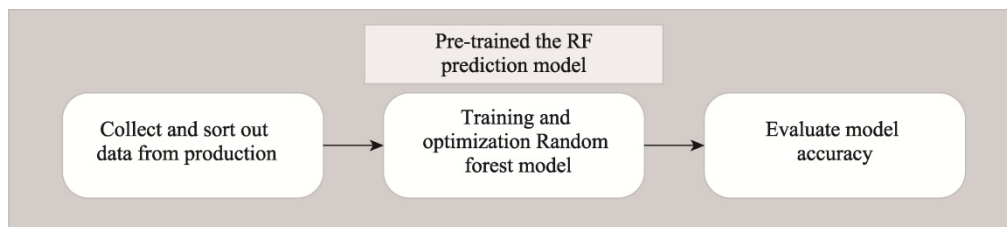


图 3 构建预测模型流程图

Fig. 3 Flow chart of building prediction model

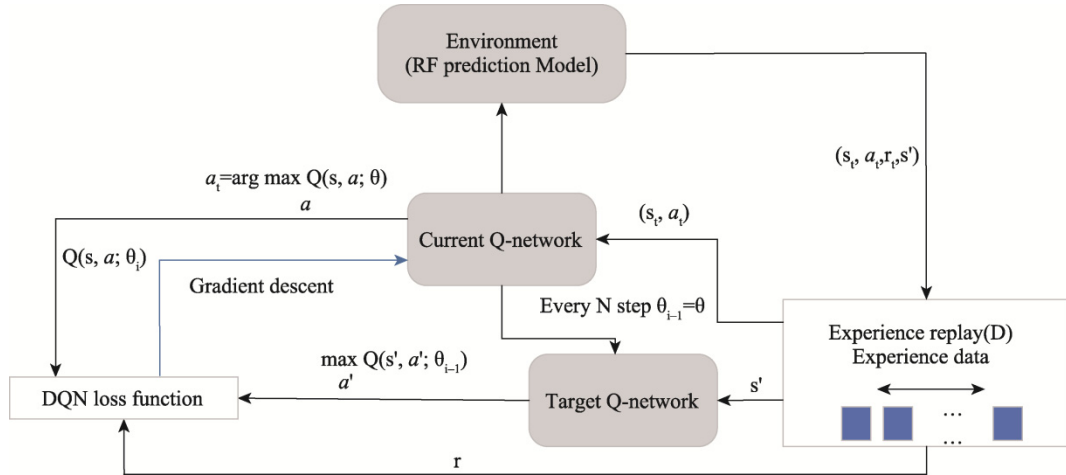


图 4 基于 DQN 算法的瓷砖生产工艺过程优化调控框架

Fig. 4 Optimization and control framework of ceramic tile production process based on DQN algorithm

将喷雾塔干燥过程建模为 MDP, 由一个五元组 (S, A, T, r, γ) 组成, 其中 T 是状态转移概率函数 $T(s_{t+1} | s_t, a_t)$, 对这些元素的详细介绍如下:

状态空间 S : 状态空间 $s_t \in S$ 由设备的四个工艺参数(燃烧温度、进风温度、排风温度、塔内温度)组成环境模块中 MRF 模型的输入参数 $\{pv_1, pv_2 \cdots pv_n\}$ 。状态空间表示为 $S_t = \{s_t^{pv_1}, s_t^{pv_2}, s_t^{pv_3}, s_t^{pv_4}\}$, 其中, $s_t^{pv_k}$ 代表第 k 个工艺参数值。

动作空间 A : 动作的选择是基于当前状态 s_t 的, 动作空间表示为 $A_t = \{a_t^{pv_1}, a_t^{pv_2}, a_t^{pv_3}, a_t^{pv_4}\}$ 。状态空间 S 中的元组由动作空间 A 来调控, 如 $a_t^{pv_k}$ 调控 $s_t^{pv_k}$ 。其中, 对于某一个变量 pv_k 在给定的范围内保持 (0) 或调整 (加减 u), 表示为 $a_t^{pv_k} \in \{-u_k, 0, +u_k\}$ 。

传递函数 P : 传递函数用于限制状态的转换, 在环境模块状态为 s_t 时收到动作 a_t , 当状态值在给定的范围之内时 P 为 1, s_t 能够实现转换到 s_{t+1} , 反之, P 为 0 时则不能实现转换。

奖励函数 R : $r_t(s_t, a_t)$ 表示智能体在状态 s_t 时执行 a_t 获得的即时奖励, 为了引导智能体学习工艺参数的调控策略实现目标优化, 本文设计的奖励函数如下:

$$r_t = \sum_{i=1}^k (f_i(s_{t+1}) - pc_i) - \sum_{i=1}^k (f_i(s_t) - pc_i) \quad (5)$$

式中: pc_i 表示喷雾干燥塔输出的期望产品质量参数; $f_i(s_t)$ 表示 MRF 模型预测输出的产品质量参数。

衰减率 γ : 衰减率 γ 是损失函数中的一个变量。当设置为 $\gamma=0$ 时, 智能体只考虑当前的即时奖励值; 当 $\gamma=1$ 时, 智能体会更多考虑未来的奖

励值。本文设置 γ 为 0.9。

Loss 函数: 当前 Q 网络在第 i 次迭代更新的 loss 函数为:

$$L_i(\theta_i) = E_{(s, a, r, s') \sim U(D)} [(r + \gamma \max_a Q(s', a'; \theta_{i-1}) - Q(s, a; \theta_i))^2] \quad (6)$$

式(6)中的相关参数如图 4 所示。由当前 Q 网络、目标 Q 网络 and 回放记忆单元在第 i 次迭代传递的值组成。目标 Q 网络中的神经网络每 N 步从当前 Q 网络复制更新神经网络参数值^[19]。

DQN 的相关参数经实验调整后设置如下, 迭代步数 T 设置为 5000 步, 经历回放记忆库大小 D 为 2000, 学习率为 0.01。特别地, 这里用于更新 DQN 的频率 N 表示在 100 步之后每 5 步更新一次 Q 网络参数。通过设定的 DQN 参数可以训练智能体学习工艺参数的调控策略指导实际生产过程。

3 实验和分析

3.1 瓷砖生产过程工艺参数定义

实验采用两组瓷砖生产过程的数据。其中, 实验一构建一个用于优化生产流程的多环节(喷雾塔、压机、素烧窑炉)工艺参数预测模型。实验二构建一个喷雾塔单环节预测模型, 用于揭示输入工艺参数和输出产品性能之间的关系。具体数据说明如下:

第一组数据为: 模式一的连续多环节工艺参数(喷雾塔、压机、素烧窑炉), 包括入窑前的几组参数作为输入变量(Input variable), 以及作为输出变量(Output variable)的瓷砖产品质量参数, 共组成 348 条数据样本作为数据集。其中, 输入变

量可分为“可控变量(Controllable variable)”和“状态变量(State variable)”。“可控变量”参数是指相关设备的可控运行参数。在实际生产中调节频率较高,“状态变量”指在生产中相对来说较稳定不会频繁进行调整的一类参数。参数的详细信息见表 1。

第二组数据为：模式二的单环节生产工艺参数，从喷雾塔环节收集相关设备参数和粉料检测参数。表 2 列出了收集的 203 组喷雾塔稳态工作的运行参数及对应检测的粉料产品质量参数的详细信息。

3.2 基于工艺参数的 MRF 预测模型构建

根据上述两组数据,分别训练两组预测模型,并对预测模型进行了验证。在实际生产过程中相关设备输出的产品一般具有多个评价指标。这些评价指标即输出变量有不同的合格范围,各指标在合格范围的区间内即认为是合格产品,调整输入变量的值并使输出变量在合格范围内称为工艺参数的调节。当实际生产过程中的工艺参数调整主要发生在输出变量处于合格范围的边界值时,

为了使预测模型更适合于实际应用,依据专家经验和实际生产状况,将输出变量分为两类,一类是即将超出合格范围的被分类为“0”,输出变量被分类为“0”则说明需要对输入变量进行调控,另一类是可维持操作条件的安全范围“1”,分类细节如图 5 所示。图中 $Q_{\min}$ 和 $Q_{\max}$ 分别表示合格范围的下边界和上边界。因为部分合格范围有上下边界,如产品合格范围在某最小值和最大值之间则依据公式(7)分类;部分具有下边界,如产品合格需大于某最小值则依据公式(8),因而有(a)、(b)两种分类方式。阈值 $\lambda(0.25)$ 和 $\delta(0.5)$ 是根据专家经验来设定的。

$$\begin{cases} 1, & Q_{\min} + \lambda(Q_{\max} - Q_{\min}) \leq Q \leq \\ & Q_{\max} - \lambda(Q_{\max} - Q_{\min}) \\ 0, & Q \leq Q_{\min} + \lambda(Q_{\max} - Q_{\min}) \text{ or } \\ & Q_{\max} - \lambda(Q_{\max} - Q_{\min}) \leq Q \end{cases} \quad (7)$$

$$\begin{cases} 1, & Q_{\max} - \delta(Q_{\max} - Q_{\min}) \leq Q \leq Q_{\max} \\ 0, & Q_{\min} \leq Q \leq Q_{\max} - \delta(Q_{\max} - Q_{\min}) \end{cases} \quad (8)$$

表 1 多环节工艺参数及范围
Tab. 1 Multi-link process parameters and scope

Parameter		Unit	Value range
Controllable variable	Granule moisture	%	6.0–6.7
	Granule unit weight	—	0.892–0.935
	Thickness of green bodies	mm	9.05–9.37
	Moisture of dried green	—	0.59–0.79
State variable	Temperature of kiln	°C	Several temperatures of firing curve
	Rupture modulus	—	17.65–24.43
Output variable	Water absorption	—	16.32–20.93
	Biscuit size	mm	607.53–609.14
	Biscuit thickness	mm	8.97–9.37

表 2 喷雾塔环节工艺参数及调整步长
Tab. 2 Process parameters and adjustment steps of spray towers

Parameter		Unit	Value range	Step
Controllable variable	Burning temperature	°C	1001–1044	2
	Inlet air temperature	°C	640–659	1
	Exhausted air temperature	°C	97–125	2
	Tower temperature	°C	428–460	2
State variable	Slip feeding pressure	MPa	30.0–33.2	—
	Slip feeding pressure	—	1.680–1.707	—
	Moisture content	%	5.2–6.4	—
Output variable	20 eyes	g	0.15–0.54	—
	40 eyes	g	48.20–56.37	—
	100 eyes	g	0.41–2.16	—

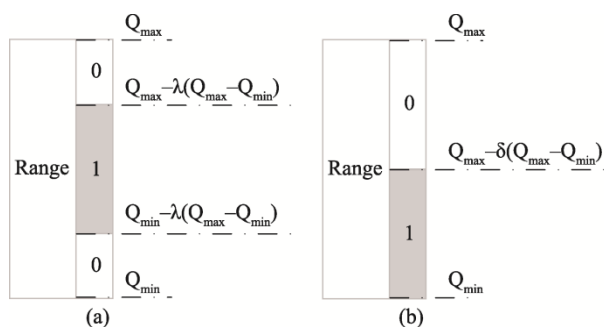


图 5 输出变量分类示意图

Fig. 5 Schematic diagram of output variable classification

3.2.1 多环节工艺参数预测模型

对于实验一,通过训练 MRF 预测模型来预测瓷砖的产品性能参数,模型评估指标采用 sklearn.metrics 中的评估方法 accuracy_score,其中 accuracy_score 是指所有分类正确的百分比,可由公式(9)来表示:

$$\text{Accuracy_score} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (9)$$

式中: TP 表示预测结果为真,实际结果是真; FP 表示预测结果为真,实际结果是假; TN 表示预测结果为假,实际结果是假; FN 表示预测结果为假,实际结果是真。

经过测试 MRF 预测模型的预测准确性如表 3 所示,可以看出实验结果符合预期要求,预测模型对实际应用具有指导意义。

表 3 多环节工艺参数模型预测准确率

Tab. 3 Prediction accuracy of the multi-link process parameter model

	Rupture modulus	Water absorption	Biscuit size	Biscuit thickness
Accuracy score	0.84	0.86	0.81	0.81

3.2.2 喷雾干燥过程建模与预测

对于实验二,训练 MRF 预测模型来预测喷雾干燥过程输出产品的四个目标颗粒性能指标。MRF 模型的预测性能如表 4 所示。表 4 列出了黄^[22]运用神经网络模型预测粉料工艺参数的结果,可以看出本文提出的 MRF 预测模型比神经网络模型的预测性能更优。训练的预测模型可用于构建强化学习的环境部分以训练本文提出的调控模型。

表 4 喷雾塔环节工艺参数模型预测准确率

Tab. 4 Prediction accuracy of the process parameter model of spray tower link

	Moisture content	20 eyes	40 eyes	100 eyes
Accuracy_score (proposed)	0.86	0.88	0.92	0.90
Accuracy_score (Huang ^[22])	0.89	—	0.86	0.88

3.3 DQN 优化控制实验分析

为了验证本文提出算法的有效性,将粉料工艺参数预测模型作为深度强化学习模型中的环境部分。由于喷雾干燥过程中的“可控变量”相比于其他参数调整的频率更高,粉料的产品性能参数也主要受“可控变量”的影响,所以将“可控变量”参数的值构成强化学习框架中的状态 $S_t = \{s_t^{pv_1}, s_t^{pv_2}, s_t^{pv_3}, s_t^{pv_4}\}$ 。其中, pv_k , $k \in \{1, 2, 3, 4\}$ 表示状态空间 S 的输入参数,在训练调控模型时可以由动作 $A_t = \{a_t^{pv_1}, a_t^{pv_2}, a_t^{pv_3}, a_t^{pv_4}\}$ 进行调整。由于单一动作向量表示为 $a_t^{pv_k} \in \{-u_k, 0, +u_k\}$, $k \in \{1, 2, 3, 4\}$,在 t 时刻,为了调整参数,对于动作空间 A ,元素 u_k 对应 $k \in \{1, 2, 3, 4\}$ 的大小分别设置为 2、1、2 和 2。显然,某一时刻 t 的动作向量可以是来自以下动作向量集的元素中的任意组合: $\{+2, 0, -2\}$ 、 $\{+1, 0, -1\}$ 、 $\{+2, 0, -2\}$ 和 $\{+2, 0, -2\}$ 。为了验证所提框架的适用性,接下来设计了一系列实验来证明提出的调控框架能够在喷雾干燥过程中通过优化工艺参数使粉料性能参数满足要求。从测试数据中挑选出 10 组样本数据如表 5 所示,可以发现输出的粉料参数的分类值有较多“0”,综合来看“1”所占比例仅为 37.5%,说明对应的输入设备工艺参数是需要调控的。实验目的是运用深度强化学习算法调整输入的设备运行参数值,使对应的工艺参数更多地被分类为“1”,训练强化学习模型实现调控过程。

本实验采用 TensorFlow^[23]来构建 Q 网络。网络由两层组成,分别具有 50 个和 3⁴ 个隐藏节点,其中,后一层对应于动作。实验过程中目标 Q 网络的损失函数的收敛图如图 6 所示,选择十个实验样本中的奇数位样本(即 1#、3#、5#、7#、9#)来展示训练过程中 loss 的收敛情况。

从图 6 可以看出,调控模型开始的时候不是很稳定,loss 值波动比较大,在大约 800 次迭代后开始收敛并逐渐稳定。这一时期的优化结果如表 6 所示。从表中可以发现,仍有一些优化结果未达到预期目标,而“1”所占比例为 77.5%。这表明需要更多的迭代步数来提高模型的稳定性。

实验的目的是运用构建的强化学习模型调控设备的工艺参数值,使输出的粉料性能参数更多地被分类为“1”,训练强化学习模型并实现优化调控过程。本实验的最终结果如表 7 所示。从表中可以看出,输出结果中“1”所占比例为 95%,表明模型优化的有效性。所构建的奖励函数可以有效地引导智能体寻找环境中的最优解,智能体能够通过与环境交互学习调整工艺参数。

表 5 待优化实验数据
Tab. 5 Experimental data to be optimized

Samples	Burning temperature	Inlet air temperature	Exhausted air temperature	Tower temperature	Slip feeding pressure	Slip specific gravity	Actual classification value of output			
1#	1030	655	105	436	30.3	1.700	0	1	0	0
2#	1030	655	104	442	30.5	1.697	1	0	0	1
3#	1031	650	100	437	30.9	1.690	0	1	0	0
4#	1032	655	101	449	31.1	1.680	0	0	0	1
5#	1037	655	102	451	31.1	1.680	1	0	0	1
6#	1022	655	104	438	31.4	1.680	1	1	0	0
7#	1035	645	108	460	32.5	1.696	0	0	1	0
8#	1035	650	110	457	31.9	1.696	1	0	1	0
9#	1036	650	111	450	31.5	1.680	0	1	0	0
10#	1024	640	111	438	30.8	1.700	0	0	1	1

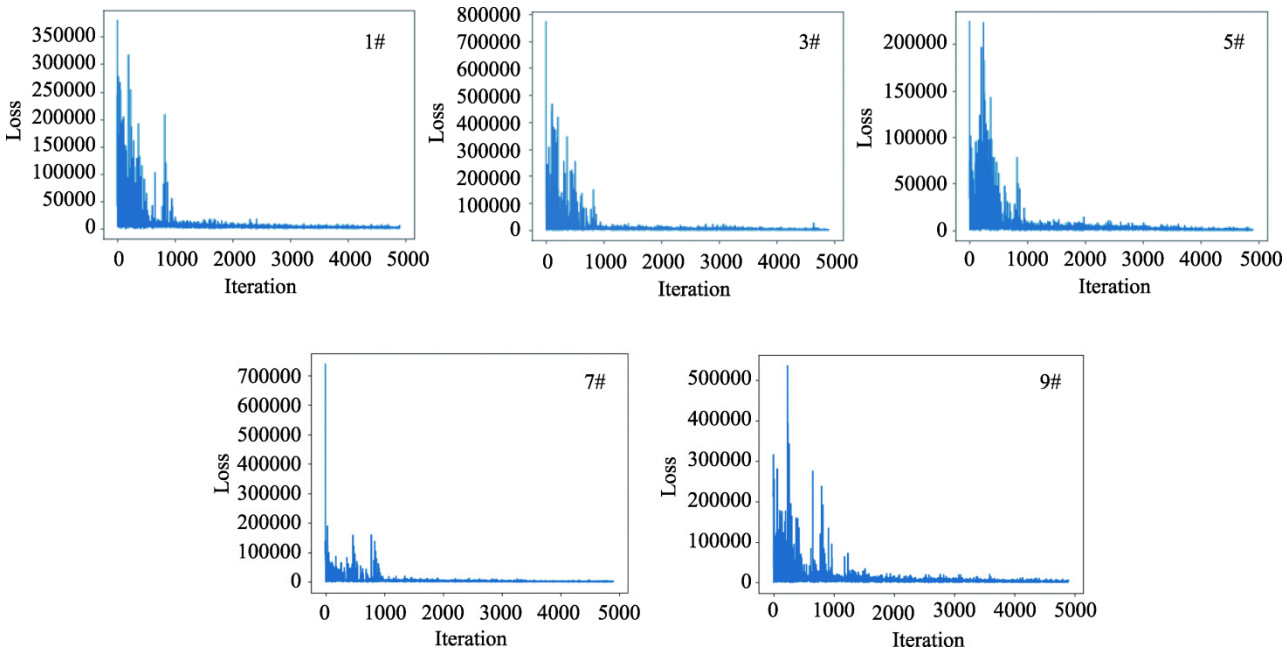


图 6 测试样本的 loss 收敛曲线图
Fig. 6 Loss convergence curve of the test samples

表 6 800 次迭代后模型优化结果
Tab. 6 Model optimization results after 800 iterations

Burning temperature	Inlet air temperature	Exhausted air temperature	Tower temperature	Slip feeding pressure	Slip specific gravity	Classification value after optimization			
1044	657	125	460	30.3	1.700	1	0	1	1
1004	659	120	428	30.5	1.697	1	0	1	1
1003	641	122	427	30.9	1.690	0	1	1	1
1030	642	97	427	31.1	1.680	1	1	1	1
1035	642	126	459	31.1	1.680	0	0	1	1
1018	647	118	430	31.4	1.680	1	1	1	1
1038	648	116	452	32.5	1.696	1	0	1	1
1003	650	115	443	31.9	1.696	1	0	1	1
1006	642	125	454	31.5	1.680	0	1	1	1
1036	646	100	454	30.8	1.700	1	0	1	1

表 7 5000 次迭代后模型优化结果
Tab. 7 Model optimization results after 5000 iterations

Burning temperature	Inlet air temperature	Exhausted air temperature	Tower temperature	Slip feeding pressure	Slip specific gravity	Classification value after optimization			
1018	655	99	430	30.3	1.700	1	1	1	1
1030	659	126	430	30.5	1.697	1	1	1	1
1045	645	110	461	30.9	1.690	1	0	1	1
1000	647	105	443	31.1	1.680	1	1	1	1
1039	658	110	461	31.1	1.680	1	0	1	1
1026	640	96	432	31.4	1.680	1	1	1	1
1023	640	122	428	32.5	1.696	1	1	1	1
1035	647	98	439	31.9	1.696	1	1	1	1
1008	642	107	428	31.5	1.680	1	1	1	1
1030	644	105	428	30.8	1.700	1	1	1	1

4 结 论

设计了一个基于强化学习算法的瓷砖生产过程智能调控框架。

(1) 构建的随机森林预测模型用于模拟瓷砖生产的各种周边环境,解决了瓷砖生产过程中的产品质量预测问题。训练的随机森林预测模型作为强化学习中环境模块的一部分,并基于深度强化学习算法快速自适应调整工艺参数,使瓷砖产品达到预期产品质量。

(2) 实验结果表明,构建的预测模型的平均准确率相比于其他同类方法的产品质量预测模型提高了 2%。同时,经过多次迭代后,瓷砖生产智能调控算法能够将生产工艺参数的合格率提高到 95%,具有较好的调控效果。

(3) 将来的工作应集中提升新预测模型下的工艺参数预测准确性,以改善本文算法在实际生产过程中的效果。通过大量数据训练模型以提升性能,为工程师调控瓷砖产线中的工艺参数提供一定的借鉴。

参考文献:

- [1] QIN Y, JIA L M. Fuzzy hybrid control and its applications in complex combustion processes [C]//Proceedings 2002 IEEE International Conference on Artificial Intelligence Systems (ICAIS 2002), Los Alamitos: IEEE Computer Society, 2002, 78–81.
- [2] ZHU Y H, ZHAO Y F. Hybrid intelligent control of ceramic shuttle kiln firing temperature [C]//International Conference on Applied Mathematics, Simulation and Modelling, Beijing, 2016: 247–251.
- [3] DENG L N, FENG B, ZHANG Y. An optimization method for multi-objective and multi-factor designing of a ceramic slurry: Combining orthogonal experimental design with artificial neural networks [J]. *Ceramics International*, 2018, 44(13): 15918–15923.
- [4] AHMMAD S K, JABEEN N, AHMED S T U, et al. Density of fluoride glasses through artificial intelligence techniques [J]. *Ceramics International*, 2021, 47(21): 30172–30177.
- [5] MU T H, WANG F, WANG X F, et al. Research on ancient ceramic identification by artificial intelligence [J]. *Ceramics International*, 2019, 45(14): 18140–18146.
- [6] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of Go without human knowledge [J]. *Nature*, 2017, 550(7676):354–359.
- [7] CHEN Y F, WANG Z, WANG Z J, et al. Automated design of neural network architectures with reinforcement learning for detection of global manipulations [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2020, 14(5): 997–1011.
- [8] KRASHENINNIKOVA E, GARCÍA J, MAESTRE R, et al. Reinforcement learning for pricing strategy optimization in the insurance industry [J]. *Engineering Applications of Artificial Intelligence*, 2019, 80: 8–19.
- [9] HAN C J, MA T, CHEN S Y. Asphalt pavement maintenance plans intelligent decision model based on reinforcement learning algorithm [J]. *Construction and Building Materials*, 2021, 299: 124278.
- [10] REN M F, LIU X F, YANG Z L, et al. A novel forecasting based scheduling method for household energy management system based on deep reinforcement learning [J]. *Sustainable Cities and Society*, 2022, 76: 103207.
- [11] GUO F, ZHOU X W, LIU J H, et al. A reinforcement learning decision model for online process parameters optimization from

- offline data in injection molding [J]. *Applied Soft Computing*, 2019, 85: 105828.
- [12] HE Z L, TRAN K P, THOMASSEY S, et al. Multi-objective optimization of the textile manufacturing process using deep-Q-network based multi-agent reinforcement learning [J]. *Journal of Manufacturing Systems*, 2022, 62: 939–949.
- [13] BREIMAN L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5–32.
- [14] QIN S J, CHENG L. A real-time tracking controller for piezoelectric actuators based on reinforcement learning and inverse compensation [J]. *Sustainable Cities and Society*, 2021, 69: 102822.
- [15] VINYALS O, BABUSCHKIN I, CZARNECKI W M, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning [J]. *Nature*, 2019, 575(7782):350–354.
- [16] MNH V, KAVUKCUOGLU K, SLIVER S, et al. Human-level control through deep reinforcement learning [J]. *Nature*, 2015, 518(7540): 529–533.
- [17] SUTTON R S, MCALLESTER D, SINGH S, et al. Policy gradient methods for reinforcement learning with function approximation [J]. *Advances in Neural Information Processing Systems*, 1999, 12: 1057–1063.
- [18] WATKINS C J C H, DAYAN P. Technical note: Q-learning [J]. *Machine Learning*, 1992, 8(3): 279–292.
- [19] SUTTON R S, BRATO A G. Reinforcement Learning: An Introduction [M]. Cambridge: The MIT Press, 2018.
- [20] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: Machine learning in python [J]. *Journal of Machine Learning Research*, 2011, 12: 2825–2830.
- [21] RAHMAN R, OTRIDGE J, PAL R. IntegratedMRF: Random forest-based framework for integrating prediction from different data types [J]. *Bioinformatics*, 2017, 33(9): 1407–1410.
- [22] 黄煌. 基于人工智能的陶瓷工业喷雾干燥模型研究[D]. 景德镇陶瓷大学, 2020.
- [23] ABADI M, AGARWAL A, BARHAM P, et al. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. ArXiv:1603.04467v2 [cs.DC] [Preprint]. Available from: <https://doi.org/10.48550/arXiv.1603.04467>.

(编辑 梁华银)