

基于混合模型的强化学习在浮选过程 优化控制中的应用

贾润达¹, 张东豪¹, 郑君¹, 李康²

(1. 东北大学 信息科学与工程学院, 辽宁 沈阳 110819;

2. 矿冶科技集团有限公司 矿冶过程自动控制技术国家(北京市)重点实验室, 北京 100160)

摘要: 传统的优化控制方法很难在浮选过程状态发生变化时准确、快速做出决策, 导致精矿品位和尾矿品位大幅度波动、出现产品质量不稳定。此外, 浮选过程难以对精矿品位进行在线检测, 导致其实用性下降。针对上述问题采用混合模型对浮选过程建模, 并基于示例的安全增强值评估(safety augmented value estimation from demonstrations, SAVED)的强化学习算法, 控制浮选溢出气泡的尺寸分布, 从而间接实现对精矿品位和尾矿品位的控制。通过仿真实验验证了所提算法的有效性。与人工经验和数据驱动模型相比, 基于混合模型的SAVED算法在保证安全约束的条件下能够实现更好的控制效果。

关键词: 浮选过程; 强化学习; 混合模型; 安全约束; 优化控制

中图分类号: TP 273 文献标志码: A 文章编号: 1005-3026(2024)10-1386-08

Application of Reinforcement Learning Based on Hybrid Model in Optimal Control of Flotation Process

JIA Run-da¹, ZHANG Dong-hao¹, ZHENG Jun¹, LI Kang²

(1. School of Information Science & Engineering, Northeastern University, Shenyang 110819, China; 2. National (Beijing) Key Laboratory of Mining and Metallurgical Process Automatic Control Technology, Mining and Metallurgical Technology Group Co., Ltd., Beijing 100160, China. Corresponding author: JIA Run-da, E-mail: jiarunda@ise.neu.edu.cn)

Abstract: Traditional optimization control methods are difficult to make accurate and rapid decisions when the state of the flotation process changes, resulting in significant fluctuations in the concentrate grade and tailings grade, and unstable product quality. In addition, the flotation process is difficult to detect the concentrate grade online, leading to a decrease in its practicality. In response to the above problems, a hybrid model is used to model the flotation process and a reinforcement learning algorithm based on safety augmented value estimation from demonstrations (SAVED) is used to control the size distribution of flotation overflow bubbles to indirectly control the concentrate grade and tailings grade. The effectiveness of the proposed algorithm is verified through simulation experiments. Compared with artificial experience and data-driven models, SAVED based on hybrid models is used to model the flotation process and control the size distribution of flotation overflow bubbles. The algorithms can achieve better control effects while ensuring safety constraints.

Key words: flotation process; reinforcement learning; hybrid model; safety constraints; optimal control

我国的矿产资源种类多、储量丰富, 但普遍具有品位低、杂质含量高的特点, 所以不得不面临从贫杂矿石中回收矿产资源的问题^[1]。选矿生

产过程是目前富集低品位有色矿产资源的典型工业生产过程。浮选是选矿过程中的重要工艺, 其根据矿物颗粒表面物理化学性质的不同, 从矿

石中分离有用矿物.随着我国对资源需求不断增加,生产规模不断扩大,对浮选的性能指标有新的要求^[2-3].现有的浮选过程中,精矿和尾矿品位很难在线检测,所以传统的优化控制方法不能形成闭环控制,导致其实用性低.

强化学习(reinforcement learning, RL)作为一种机器学习算法,是以目标为导向的学习工具,主要解决控制领域中的最优控制问题,包括最优调节、最优跟踪以及最优协同等.通过设定奖励函数,RL算法可以自动学习满足该目标的控制策略,有效解决了浮选过程控制问题,可替代传统控制方法.

Rojas等^[4]针对浮选过程提出了一种基于模型预测控制(model predictive control, MPC)的方法.李金娜等^[5]针对模型未知时具有双时间尺度浮选过程的经济指标控制问题,提出了Q学习(Q-learning)的控制方法. Jiang等^[6]针对复杂浮选过程控制回路输出之间复杂的动态特性提出了一种基于RL的数据驱动浮选过程.然而上述算法为无模型算法,其学习过程需要大量的交互数据,数据效率非常低.此外,基于模型强化学习算法对浮选过程控制往往由于模型的预测准确性较差,导致其控制性能较差且难以满足浮选过程约束.

针对上述浮选过程问题,本文使用混合模型对浮选过程建模进而提高模型的预测准确性.采用SAVED算法来实现浮选过程优化控制的同时使其满足概率约束.此外,本文通过控制浮选溢出气泡的尺寸分布^[7]来间接实现对精矿和尾矿品位的控制.

1 浮选过程概述

浮选是利用矿物本身的亲水性或化学反应将有用的矿物从脉石中分离的物理过程,该过程所用设备包括浮选槽、充气机等.浮选过程主要在浮选槽中进行,浮选槽中按照功能分为充气搅拌区、选矿区和泡沫区.图1为浮选槽内部结构示意图.

在充气搅拌区,压缩空气被充气机通入并且在浮选槽搅拌电机的作用下破碎形成大小不一的气泡.在选矿区,由下而上的气泡在经过此区域时,由于给药剂已经加入了浮选的捕收剂,捕收剂对矿物有选择性,能够增加矿物的疏水性,有时也会加入抑制剂来抑制无用矿物与气泡相

结合,从而使矿物能够被通入的气泡所携带,进入到泡沫区.在泡沫区使用刮板将泡沫刮取,从而实现有用矿物和脉石的分离.

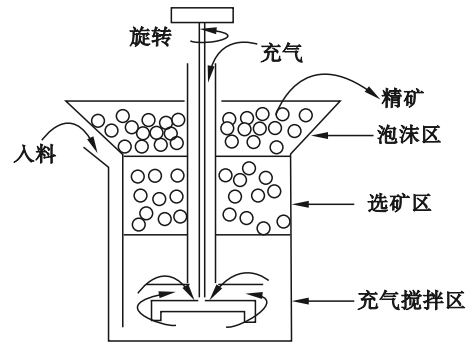


图1 浮选槽内部结构示意图

Fig. 1 Internal structure diagram of flotation cell

浮选过程工艺复杂,各个环节之间有很强的耦合性,影响浮选工艺的因素有:①不可调节因素如原矿的矿物组成和含量,矿石的氧化和泥化程度以及生产用水质量等;②可调节因素如充气量、磨矿细度、矿浆质量浓度、浮选时间、药剂制度、浮选回路结构和浮选设备类型等.各种可调节因素最佳条件的选择通常是经过对原矿性质的研究以及实验室和半工业性实验来确定.

在浮选过程中,气泡是最主要的载体,矿物颗粒附着于气泡表面,从而实现富集,浮选中气泡的尺寸以及上升频率会直接影响浮选的速率以及矿物整体的回收率;溢出气泡的尺寸分布以及气泡回收率能够反映浮选的精矿品位、尾矿品位等主要性能指标,所以气泡的相关变量是浮选过程的重要参数^[7].本文将气泡相关变量加入浮选机理模型的构建中,从而使机理模型的准确性得到提升.

2 混合模型的建立

在基于模型的强化学习算法中,建立准确的模型拟合真实浮选过程是优化矿物回收率、提高精矿品位的基础^[8-9].本文提出混合模型的建模方法来提高模型的准确性.具体如下:①浮选过程机理模型,通过机理模型来描述浮选过程的基本物理原理,为模型整体精度提供可靠依据;②残差补偿模型,考虑到机理模型与实际浮选过程中的误差以及各种不明确影响因素,本文采用概率集成模型(probabilistic ensemble models, PEM)作为数据残差模型来补偿因不确定性而导致的误差.图2为本文提出的浮选混合模型结构图.

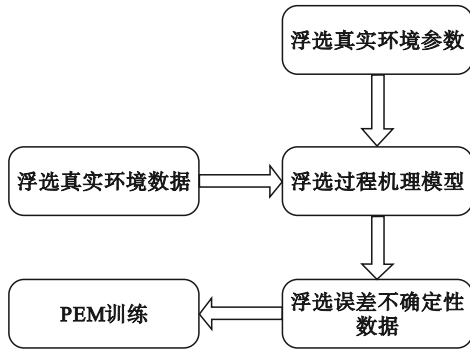


图 2 浮选混合模型结构图
Fig. 2 Hybrid model structure diagram of flotation

2.1 浮选过程机理模型

在浮选过程中大多数控制模型只关注矿浆相,而不包括泡沫相,但是泡沫稳定性会显著影响浮选的整体性能.本文采用了 Quintanilla 等^[10-11]提出的动态浮选机理模型,该模型包括了与泡沫稳定性相关的重要变量,如:爆裂率(泡沫上浮过程中破碎泡沫与总的泡沫的数量比)、空气回收率(未破碎泡沫与总的泡沫的数量比)以及计算泡沫回收和夹带的简化方程.这些方程能够用于估计矿物品位和回收率.此外,模型中还包括了矿浆与泡沫的界面,这样可更准确地预测相关的浮选变量.模型有 2 个变量可作为决策变量,分别是空气流量和尾矿流量.模型的整体物料转移微分方程如下:

$$\frac{dM}{dt} = m_f - m_t - m_E - m_T \quad (1)$$

式中: M 为浮选槽中初始矿石质量; m_f 为单位时间内进料的固体质量; m_t 为单位时间内尾矿流出的固体质量; m_E 为单位时间内精矿中夹带的固体质量; m_T 为单位时间内精矿中的真实浮选质量.

图 3 为该模型的物料转移示意图.

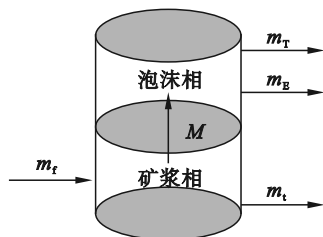


图 3 物料转移示意图
Fig. 3 Schematic diagram of the material transfer

2.1.1 矿浆层

单位时间内进料的固体质量为

$$m_f = Q_f C_f \quad (2)$$

式中: Q_f 为进料流量; C_f 为进料矿物质量浓度.

单位时间内尾矿流出的固体质量为

$$m_t = Q_t C_t \quad (3)$$

式中: Q_t 为尾矿流量; C_t 为尾矿中的固体质量浓度.

$$C_t = \frac{M}{V_p} \quad (4)$$

其中, $V_p = h_0 A_c + V_g$, $V_g = \sum_{k=1}^K V_{g,k}$, $V_{g,k} = \frac{\varepsilon_{0,k}}{1 - \varepsilon_{0,k}} h_0 A_c$.

式中: V_p 为浮选槽中的矿浆体积; h_0 为矿浆高度; A_c 为浮选槽的横截面积;气体体积 V_g 可表示为单个气泡尺寸 k 的体积之和; $V_{g,k}$ 为单个气泡尺寸等级的体积, $\varepsilon_{0,k}$ 是每个气泡尺寸等级 k 的气体滞留量; K 为总的气泡等级数量.

总气体滞留量 $\varepsilon_{0,t}$ 和矿浆的总高度 h_p 分别为

$$\varepsilon_{0,t} = \sum_{k=1}^K \varepsilon_{0,k} \quad (5)$$

$$h_p = \frac{h_0}{1 - \varepsilon_{0,t}} \quad (6)$$

矿浆高度随时间变化的导数为

$$\frac{dh_p}{dt} = \frac{1}{A_c} (Q_f - Q_{p,o}) \left(\frac{1}{1 - \sum_{k=1}^K \varepsilon_{0,k}} \right) + \frac{h_0}{\left(1 - \sum_{k=1}^K \varepsilon_{0,k} \right)} \sum_{k=1}^K \frac{d\varepsilon_{0,k}}{dt} \quad (7)$$

式中, $Q_{p,o}$ 为浮选槽溢流的总和, $Q_{p,o} = Q_t + Q_c$, Q_c 为精矿流量.

矿浆中气体向上的速度为

$$v_{g,k} = \frac{g \rho_p (d_{b,p,k})^2}{18 \mu_p (1 - \varepsilon_{0,k})^{1.39}} \quad (8)$$

其中

$$\rho_p = \phi \rho_s + (1 - \phi) \rho_w,$$

$$\mu_p = \mu_w \exp\left(\frac{2.5\phi}{1 - 0.609\phi}\right),$$

$$\phi = \frac{\rho_w}{\rho_w - \rho_s + \frac{\rho_s}{w_s}}$$

式中: g 是重力加速度; $d_{b,p,k}$ 是气泡尺寸; ρ_p 为矿浆的密度; ρ_w 和 ρ_s 分别为水和固体矿料的密度; μ_p 为矿浆的黏度; μ_w 为水的黏度; ϕ 是矿浆中固体的体积分数; w_s 为矿浆中固体矿料的质量分数.

出浆的总气体速度 $v_{g,t}$ 是每个气泡尺寸的加权和.

$$v_{g,t} = \sum_{k=1}^K v_{g,k} \varepsilon_{0,k} \quad (9)$$

2.1.2 矿浆泡沫界面层

界面气体速度 $v_{g,v}$ 和界面处的气泡平均尺寸 $d_{b,i}$ 分别为

$$v_{g,v} = \frac{dh_p}{dt} + v_{g,t} \quad (10)$$

$$d_{b,i} = \frac{\sum_{k=1}^K v_{g,k} \varepsilon_{0,k}}{\sum_{k=1}^K \frac{v_{g,k} \varepsilon_{0,k}}{d_{b,p,k}}} \quad (11)$$

2.1.3 泡沫层

精矿中真实浮选质量为

$$m_T = V_c K R_f C_T \quad (12)$$

式中: V_c 为浮选槽体积; K 为浮选速率常数; R_f 为泡沫回收系数.

$$R_f = \begin{cases} \left(\frac{a_r(1-a_r)v_{g,v}}{v_s}\right)^{F/2} \left(\frac{d_{b,i}}{d_{b,o}}\right)^F, & a_r < 0.5; \\ \left(\frac{v_{g,v}}{v_s}\right)^{F/2} \left(\frac{d_{b,i}}{d_{b,o}}\right)^F, & a_r \geq 0.5. \end{cases} \quad (13)$$

其中

$$d_{b,o} = \left(nC \frac{h_T - h_p}{v_{g,v}} + d_{b,i}^n\right)^{1/n},$$

$$v_s = \frac{g(\rho_s - \rho_w)d_p^2(1-\phi)^{4.65}}{54\mu_p},$$

$$a_r = \frac{v_{g,v} - v_b}{v_{g,v}},$$

$$v_b = a + b \frac{Q_a}{A_c} + cj_g^2.$$

式中: F 为校正常数; $d_{b,o}$ 溢出气泡大小; C, n 为校正常数; h_T 为浮选槽高度; v_s 为颗粒沉降速度; d_p^2 为矿物粒度; a_r 为空气回收率; v_b 是泡沫顶部的破裂速率; a, b, c 为校正常数; Q_a 为通入的空气流量; j_g 为输入气体流速.

精矿中夹带浮选质量为

$$m_E = Q_c E C_i \quad (14)$$

其中

$$Q_c = \begin{cases} \frac{A_c v_{g,v} \lambda_o (1-a_r) a_r}{k_1}, & a_r < 0.5; \\ \frac{A_c v_{g,v}^2 \lambda_o}{4k_1}, & a_r \geq 0.5. \end{cases}$$

$$k_1 = \frac{\rho_p g}{3C_{PB} \mu_p}, \lambda_o = \frac{k_\lambda}{d_{b,o}^2},$$

$$E = \begin{cases} \exp\left(\frac{-v_s^{1.5}(h_T - h_p)}{D_a \sqrt{v_{g,v}(1-a_r)}}\right), & a_r < 0.5; \\ \exp\left(\frac{-2v_s^{1.5}(h_T - h_p)}{D_a \sqrt{v_{g,v}}}\right), & a_r \geq 0.5. \end{cases}$$

$$D_a = \frac{j_g^{1.5}}{\sqrt{k_1(\sqrt{3} - \frac{\pi}{2})Pe}}$$

式中: Pe 为佩克莱数; C_{PB} 为阻力系数常数; k_λ 为物理常数; E 为浮选矿物的矿物学夹带因子; D_a 为轴向渗透系数.

与传统的机理模型不同,上述机理模型不仅关注矿浆相,同时也结合了描述浮选泡沫动力学方程,使得模型能够更好体现浮选真实过程,更准确地预测相关浮选性能指标.

2.2 基于PEM的残差补偿模型

结合泡沫物理学的动态浮选机理模型虽然能够在操作条件变化时正确反映浮选趋势,但机理模型不能完全表征真实的浮选过程,所以需要补偿模型拟合浮选中的以下不确定性:①预测不确定性,源于浮选固有的随机性,例如观测噪声和过程噪声;②认知不确定性,对应于浮选动力学函数的主观不确定性,这是因为缺乏足够的数据来唯一准确地确定浮选过程.然而确定性的神经网络模型容易受到过拟合的影响,导致预测不准确,所以本文采用概率集成模型PEM^[12-13]作为补偿模型.

不确定性感知模型由多个概率神经网络组成.图4为PEM网络结构.

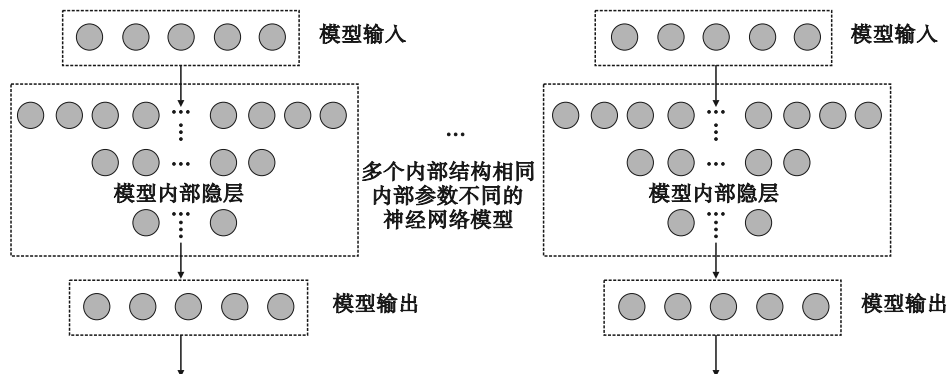


图4 PEM网络结构图

Fig. 4 PEM network structure diagram

2.2.1 激活函数的选择

传统的 sigmoid 激活函数,其饱和特性极易引起梯度丢失,进而造成训练失效. ReLu 激活函数在训练过程中不适合较大梯度输入,且存在神经元坏死现象,某些神经元可能永远不会被激活,进而导致相应参数一直得不到更新.所以本文模型采用了 Swish 激活函数 σ ,其计算公式如下:

$$\sigma(x) = \frac{x}{1 + e^{-x}}.$$

Swish 激活函数 $\sigma(x)$ 是平滑且非单调的函数,实验表明 Swish 激活函数在更深层次的模型上优化得比 sigmoid, ReLU 等激活函数更好,从而使模型具有更好的拟合性能.图 5 为 Swish 激活函数的示意图.

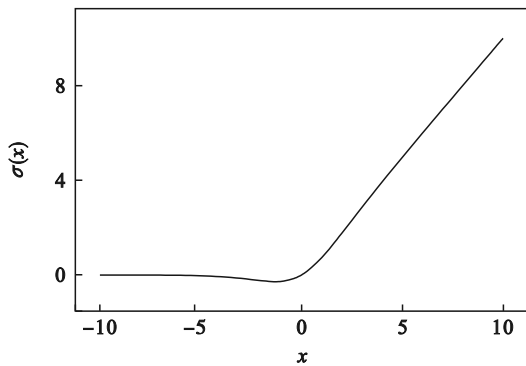


图 5 Swish 激活函数

Fig. 5 Swish activation function

2.2.2 损失函数的计算

整个 PEM 网络输入为 s_t (t 时刻浮选状态) 和 a_t (t 时刻浮选操作变量), 输出为 l_{t+1} ($t+1$ 时刻的模型预测误差) 的高斯分布, 而不仅是预测一个固定的误差值, 其可以更全面地表达模型的预测不确定性. 高斯分布的参数 (即均值和方差) 能够提供关于预测误差的有用信息, 其中均值代表了误差的期望大小. 而方差则反映了预测的不确定性或可靠性, 这种方法使得 PEM 网络不仅能够进行准确的预测, 还能够评估预测的可信度, 为决策提供更为丰富的信息.

$$\tilde{l}_{t+1}|s_t, a_t \sim G(\mu(s_t, a_t), \sigma(s_t, a_t)). \quad (15)$$

式中: $\tilde{l}_{t+1}|s_t, a_t$ 为单个模型的残差预测值; μ 为高斯分布均值; G 为高斯分布函数; σ 为高斯分布方差.

PEM 由多个不同模型组成集合体. 这种结构的设计允许它综合利用各个子模型的优势, 以达到更高的预测准确性和鲁棒性. 因此, 整体 PEM

的损失函数并不是简单地从单一模型中衍生出来的, 而是综合了集合中所有模型的损失函数. 这种方法确保整体模型可以在不同的数据集和预测任务上表现出最优性能. 整体 PEM 的损失函数为

$$l = -\frac{1}{B} \sum_{b=1}^B \tilde{f}_{b,\theta}(l_{t+1}|s_t, a_t). \quad (16)$$

式中: B 为模型的个数, l 为模型的最终残差预测值; l_{t+1} 为 $t+1$ 时刻模型预测值.

PEM 在不确定性的预测方面具有良好的性能, 在浮选模型精准度的提升中起着至关重要的作用.

3 基于 SAVED 算法的浮选过程优化控制

由于必须保证浮选过程的安全性, 所以在输入保持指定范围正确跟踪规定操作指标的同时, 必须保证其始终在安全范围之内. 但通过强化学习寻找最优策略时会使模型探索一些不安全的领域, 在将强化学习算法用于实际环境中时可能发生安全事故. 因此在探索过程中要限制不安全的动作, 本文采用基于混合模型的 SAVED 算法, 实现浮选过程的优化控制^[14-15], 同时使其满足概率约束. 图 6 为 SAVED 算法流程图.

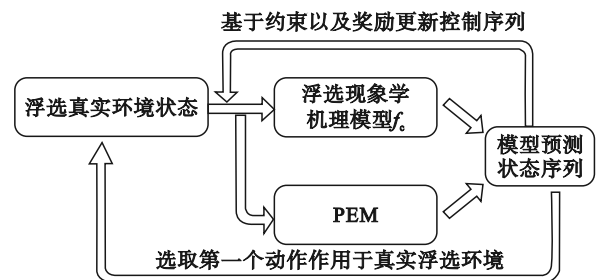


图 6 SAVED 算法流程图

Fig. 6 SAVED algorithm flowchart

该算法通过对任务完成的监督和一组适度的次优经验来约束探索, 并在处理复杂约束时高效地学习. 其根据当前状态 s_t (包括充气气泡尺寸等级、气体滞留量 ε , 精矿流量 Q_c , 矿浆高度 h_0) 在每个时间步数上解决 Thananjeyan 等^[16]提出的优化问题:

$$\left. \begin{aligned} a_{t+T-1}^* &= \arg \min \sum_{i=0}^{T-1} C(s_{t+i}, a_{t+i}), \\ s_{t+i+1} &\sim f(s_{t+i}, a_{t+i}) \quad \forall i \in \{0, \dots, T-1\}, \\ \rho(s_{t+T}) &> \delta, \\ P(s_{t+T} \in \chi_{T+1}) &\geq \beta. \end{aligned} \right\} \quad (17)$$

式中: $a_{t:t+T-1}^*$ 为 t 到 $t+T-1$ 时刻整条动作序列的评价值; s_{t+i+1} 为 $t+i+1$ 时刻的气泡尺寸; $f(s_{t+i}, a_{t+i})$ 为混合模型; T 为时刻; δ 和 β 为超参数, 人为设定.

式(17)中各个参数的构建方法如下: ①由于本文通过控制溢出气泡尺寸分布实现浮选精矿品位的控制, 故 $C(s_{t+i}, a_{t+i})$ 定义为 $t+i$ 时刻溢出气泡尺寸分布与设定均值的距离; ② $\rho(s_{t+T})$ 定义为 $t+H$ 时刻状态与经验状态的欧氏距离, 其中 H 为步长; ③ $P(s_{t:t+T} \in \chi_{T+1})$ 定义为浮选液面在合理的区间范围的概率. 表 1 为 SAVED 算法流程.

表 1 SAVED 算法
Table 1 Algorithm of SAVED

基于混合模型的 SAVED 算法	
1.	初始化: 基于浮选半实物仿真平台初始化数据 D ; 将半实物仿真平台模型参数输入浮选现象学机理模型 f_c ; 基于数据 D 中的状态以及动作输入机理模型生成误差数据 D_1 ;
2.	当 $k < K$, 循环执行;
3.	基于误差数据 D_1 训练 PEM;
4.	当 $t < P$ (P 为任务范围) 循环执行;
5.	当采样动作 a_t 小于超参数人为设定值 N 时, 循环执行;
6.	在动作范围内随机生成序列;
7.	基于回报与约束评价动作序列;
8.	选取动作序列更新交叉熵法分布;
9.	当 t 大于超参数人为设定值 z 执行;
10.	将动作加入到动作序列评价;
11.	执行 t 到 $t+T$ 之间的动作序列 $a_{t:t+T}^*$ 中的首个动作 a_t^* ;
12.	记录结果: $D_1 \leftarrow D_1 \cup \{s_t, a_t^*, s_{t+1}\}$.

SAVED 利用学习动力学中的不确定性对轨迹实施概率约束, 这使得 SAVED 能够处理复杂的状态空间约束. 在模型预测控制轨迹优化过程中, 控制序列从截断高斯分布中采样, 该截断高斯分布使用交叉熵方法 (cross entropy method, CEM) 迭代更新, 其更新过程如表 2 所示.

CEM 方法更新基于动力学模型对每个控制序列进行多次模拟, 并使用模拟的平均回报对序列进行评分.

如果超过 $100(1-\beta)\%$ 的模拟违反了约束, 通过丢弃控制序列来实现机会约束, 其中 β 是用户指定的容差. β 参数平衡了确保足够的探索来学习动力学和满足指定约束之间的关系. 同时它可以在存在动态不确定性、稀疏成本反馈和复杂约

束的情况下有效地实现浮选的控制任务, 方法是将其探索限制在有信心完成任务的区域. 从而使其选择的动作能够高概率满足浮选的约束控制. 图 7 为 SAVED 单步规划示意图.

表 2 CEM 迭代更新算法
Table 2 CEM iterative update algorithm

CEM 迭代更新算法	
1.	初始化参数: 根据浮选过程先验知识初始化参数集合;
2.	生成样本: 使用参数集合, 随机抽样生成样本;
3.	评估样本: 对于生成的样本, 使用式(17)的目标函数评价 x_n 到 x_{t+n} 时刻的样本性能;
4.	选择优秀样本: 根据评估的结果, 在当前样本中选择一定比例优秀样本;
5.	拟合新参数: 通过选出的优秀样本, 对参数集合进行更新;
6.	迭代优化: 重复上述步骤 2~5, 直到满足停止条件.

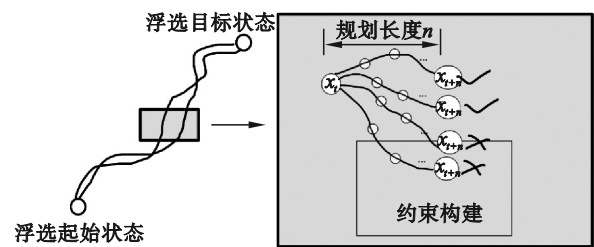


图 7 SAVED 单步规划示意图
Fig. 7 Schematic diagram of SAVED single step planning

4 仿真实验

4.1 浮选气泡状态采集及分析

基于浮选工业现场研究发现溢出气泡的尺寸分布与精矿品位在一定的范围之内存在负相关性. 目前难以对精矿品位在线检测, 所以本文采取控制更加容易测量的溢出气泡尺寸分布来实现精矿品位等浮选重要指标的间接控制, 从而使算法能够更好应用于真实浮选过程.

本文使用气泡观测器^[17]对浮选半实物仿真平台气泡尺寸分布进行测量. 其结构如图 8 所示. 该观测器能够在不干扰浮选过程的情况下, 实时捕捉浮选气泡尺寸分布. 基于气泡观测器采集的数据采用 Panjipour 等^[18]提出的方法实现对气泡尺寸分布的分析.

4.2 实验过程

4.2.1 PEM 参数及训练结果

PEM 中的网络模型的训练参数如表 3 所示.

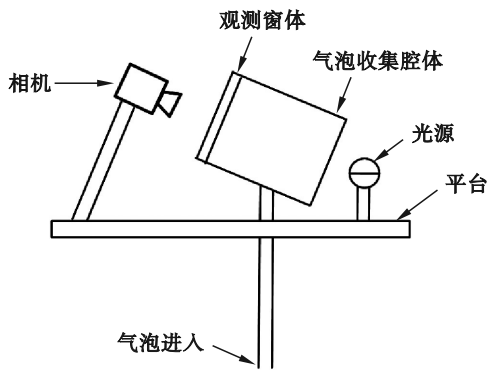


图 8 气泡观测器
Fig. 8 Bubble viewer

表 3 PEM 参数列表
Table 3 Parameter list of PEM

参数	参数值
PEM 模型数	5
模型层数	5
神经元数量	512
批次数量	64
迭代次数	200
β 参数	0.8

在优化器的选择方面,Adam 最优器具有实施简便、运算效率高等特点,所以本实验选取 Adam 优化算法. PEM 不确定性规划采用了 TS-1 轨迹采样方法(单位时间都从集合模型随机采样,用该模型进行预测). SAVED 算法中 PEM 训练结果如图 9 所示. 图 10 为真实值与模型预测值对比. 表 4 为预测精度对比.

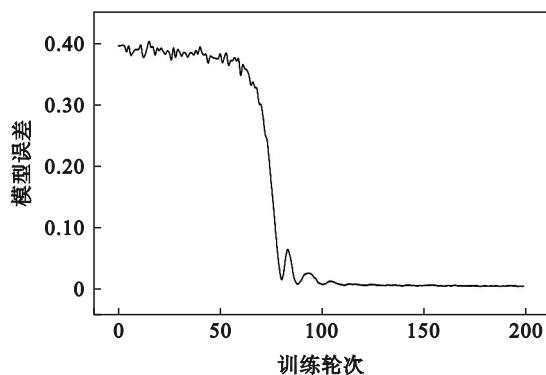


图 9 PEM 训练损失
Fig. 9 PEM training loss

由图 9 可知 PEM 在训练过程中的损失呈现出逐渐收敛的趋势.

由表 4 可知基于混合模型的预测状态 RMSE

(均方根误差)和 MAE(均方误差)要明显低于 PEM 模型和机理模型,表明混合模型的预测精度更高.

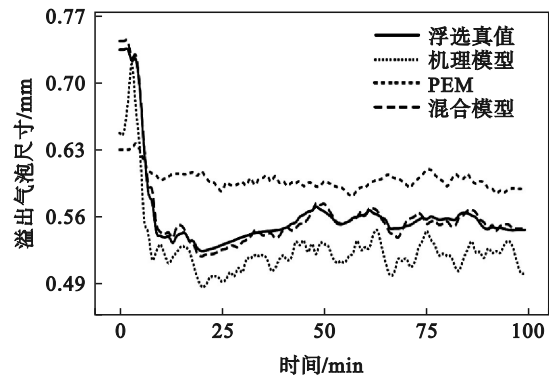


图 10 真实值与模型预测值对比

Fig. 10 Comparison of actual values and model prediction values

表 4 预测精度对比
Table 4 Comparison of prediction accuracy

误差类型	机理模型	PEM 模型	混合模型
RMSE	0.18	0.34	0.002 0
MAE	0.10	0.28	0.001 5

4. 2. 2 SAVED 算法参数及实验结果

在浮选的约束方面,本实验对浮选的液面高度进行约束,液面高度太高时矿浆直接从浮选槽中流出,无法进行浮选;液面高度太低时浮选的效率将严重降低,影响浮选质量.综合考虑将液面高度设定为(0.467~0.475). β 参数值设定为 0.8. 在算法进行状态规划时,一旦违反约束,算法直接将该控制序列丢弃.在气泡设定值方面,由于精矿品位与溢出气泡尺寸分布在一定的范围内存在负相关性,所以在本实验中选取溢出气泡尺寸分布相对较小的 0.09 cm 作为设定值.

将所提算法与以下算法对比:①人为经验选取动作:通过人工的经验来选取相应的动作;②来自数据 PEM:通过数据训练的模型直接用于预测下一时刻的状态,模型规划视界长度与 SAVED 规划视界长度相同.设定值完成情况与约束违反情况如图 11 和图 12 所示.图 11 为溢出气泡设定均值为 0.09 cm 时的完成情况.图 12 为设定溢出气泡尺寸均值为 0.09 cm 时液面高度控制图.

实验表明混合模型能够准确预测浮选的状态,同时基于混合模型的 SAVED 算法能够以较好的控制浮选达到设定值,控制效果稳定,且能够更好完成设定的约束条件,各方面都优于对比算法.

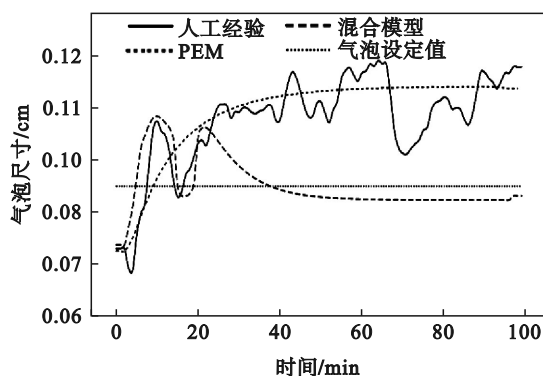


图 11 设定溢出气泡尺寸为 0.09 cm 的控制图

Fig. 11 Control diagram illustrating the overflow bubble size set at 0.09 cm

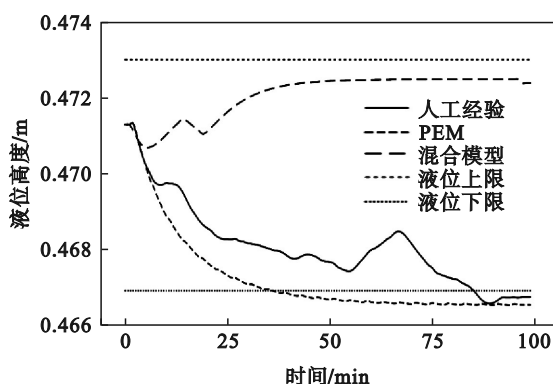


图 12 设定溢出气泡尺寸为 0.09 cm 的液面高度控制图

Fig. 12 Control diagram for liquid level height with an overflow bubble size set at 0.09 cm

5 结 语

本文提出了基于浮选混合模型的强化学习算法实现对溢出气泡尺寸分布的控制,从而间接实现对精矿品位和尾矿品位的优化控制.首先基于结合浮选现象学的机理模型实现对浮选过程趋势的大体描述;然后基于 PEM 实现浮选过程的误差以及各种不确定性的预测;最后基于训练好的混合模型 CEM 选取有利动作,实现对溢出气泡分布的优化控制.实验表明本文提出的基于混合模型的 SAVED 算法能够很好控制浮选达到设定值,且控制效果稳定不违反过程约束.未来的研究将致力于使用策略优化方法替代 CEM 方法,从而使算法能够满足更高实时性要求的环境.此外,基于机器视觉提取更多的泡沫信息与浮选优化控制相结合也是未来的研究重点.

参考文献:

[1] 范继涛,朱勃霖.矿产资源节约与综合利用先进技术推广的思考[J].中国矿业,2013,22(11):23-26.

(Fan Ji-tao, Zhu Bo-lin. Thinking on advanced technologies promotion of mineral resources saving and comprehensive utilization[J]. *China Mining Magazine*, 2013, 22(11): 23-26.)

[2] Xue W Q, Fan J L, Jiang Y. Flotation process with model free adaptive control [C]//IEEE International Conference on Information and Automation (ICIA). Macau: IEEE, 2017: 442-447.

[3] Osthuizen D J, le Roux J D, Craig I K. A dynamic flotation model to infer process characteristics from online measurements[J]. *Minerals Engineering*, 2021, 167: 106878.

[4] Rojas D, Cipriano A. Model based predictive control of a rougher flotation circuit considering grade estimation in intermediate cells[J]. *Dyna*, 2011, 78(166): 29-37.

[5] 李金娜,高溪泽,柴天佑,等.数据驱动的工业过程运行优化控制[J].控制理论与应用,2016,33(12):1584-1592. (Li Jin-na, Gao Xi-ze, Chai Tian-you, et al. Data-driven operational optimization control of industrial processes[J]. *Control Theory & Applications*, 2016, 33(12): 1584-1592.)

[6] Jiang Y, Fan J L, Chai T Y, et al. Data-driven flotation industrial process operational optimal control based on reinforcement learning [J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(5): 1974-1989.

[7] Nakhaei F, Irannajad M, Mohammadnejad S. A comprehensive review of froth surface monitoring as an aid for grade and recovery prediction of flotation process. Part B: texture and dynamic features[J]. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 2023, 45(3): 7812-7834.

[8] Quintanilla P, Neethling S J, Brito-Parada P R. Modelling for froth flotation control: a review [J]. *Minerals Engineering*, 2021, 162: 106718.

[9] Sun B, Yang W, He M F. An integrated multi-mode model of froth flotation cell based on fusion of flotation kinetics and froth image features [J]. *Minerals Engineering*, 2021, 172: 107169.

[10] Quintanilla P, Neethling S J, Navia D. A dynamic flotation model for predictive control incorporating froth physics. part I: model development [J]. *Minerals Engineering*, 2021, 173: 107192.

[11] Quintanilla P, Neethling S J, Mesa D, et al. A dynamic flotation model for predictive control incorporating froth physics. part II: model calibration and validation [J]. *Minerals Engineering*, 2021, 173: 107190.

[12] Abbeel P, Quigley M, Ng A Y. Using inaccurate models in reinforcement learning [C]//Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh: 2006: 1-8.

[13] Baranes A, Oudeyer P Y. Active learning of inverse models with intrinsically motivated goal exploration in robots [J]. *Robotics and Autonomous Systems*, 2013, 61(1): 49-73.

[14] Xu H, Wang K, Li X L. Multi-objective optimization control of flotation process based on policy iteration [C]//2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS). Xiangtan: IEEE, 2023: 417-422.

[15] Quintanilla P, Navia D, Neethling S J, et al. Economic model predictive control for a rougher froth flotation cell using physics-based models [J]. *Minerals Engineering*, 2023, 196: 108050.

[16] Thananjeyan B, Balakrishna A, Rosolia U, et al. Safety augmented value estimation from demonstrations (saved): safe deep model-based rl for sparse cost robotic tasks [J]. *IEEE Robotics and Automation Letters*, 2020, 5(2): 3612-3619.

[17] Bailey M, Gomez C O, Finch J A. A method of bubble diameter assignment [J]. *Minerals Engineering*, 2005, 18(1): 119-123.

[18] Panjipour R, Karamoozian M, Albijanic B. Bubble size distributions in gas-liquid-solid systems and their influence on flotation separation in a bubble column [J]. *Chemical Engineering Research and Design*, 2021, 167: 96-106.