

# 数据驱动的工业过程运行优化控制

李金娜<sup>1,2†</sup>, 高溪泽<sup>1</sup>, 柴天佑<sup>2</sup>, 范家璐<sup>2</sup>

(1. 沈阳化工大学 信息工程学院, 辽宁 沈阳 110142;

2. 东北大学 流程工业综合自动化国家重点实验室, 辽宁 沈阳 110819)

**摘要:** 现代工业过程机理复杂使得很难对生产过程以及运行指标与被控变量之间关系精确建模. 如何基于工业运行过程数据信息, 不依赖模型参数给出设定值设计方案, 优化运行指标是一挑战性难题. 本文针对在稳态附近可以线性化的一类工业过程, 考虑运行控制环和底层控制环不同时间尺度, 提出一种基于Q-学习方法的次优设定值学习算法. 此算法完全利用数据, 学习得到次优设定值, 实现运行指标以次优的方式跟踪理想值. 浮选过程仿真结果表明本文所提方法的有效性.

**关键词:** 运行优化控制; 设定值; 近似动态规划; Q-学习

**中图分类号:** TP273      **文献标识码:** A

## Data-driven operational optimization control of industrial processes

LI Jin-na<sup>1,2†</sup>, GAO Xi-ze<sup>1</sup>, CHAI Tian-you<sup>2</sup>, FAN Jia-lu<sup>2</sup>

(1. College of Information Engineering, Shenyang University of Chemical Technology, Shenyang Liaoning 110142, China;

2. State Key Lab of Synthetical Automation for Process Industries, Northeastern University, Shenyang Liaoning 110819, China)

**Abstract:** It is difficult to accurately model productive processes and describe relationship between operational indices and controlled variables for modern industrial processes. How to design the setpoints by using only data generated by operational processes for optimizing operational indices, without requiring the knowledge of model parameters of operational processes, poses a challenge on operational optimization control. This paper focuses on a class of industrial processes that can be linearized near the steady states and take different time scales adopted in the operational control loop and process control loop into account. In this context, a Q-learning based suboptimal setpoint learning algorithm is proposed to learn suboptimal setpoints by utilizing only data, such that the operational indices can track the desired values in a suboptimal manner. A simulation experiment in flotation process is implemented to show the effectiveness of the proposed method.

**Key words:** operational optimization control; setpoints; approximate dynamical programming; Q-learning

## 1 引言(Introduction)

在工业过程控制中, 设计的控制器不仅要镇定被控过程, 使被控变量很好地跟踪参考输入, 还要优化运行指标, 即表征产品在加工中的质量指标、效率指标、经济收益指标、能耗与物耗等相关的指标<sup>[1-3]</sup>, 这就需要研究运行优化控制. 如果设定值 (setpoints) 选取不适当, 即使被控变量跟踪设定值, 也不可能优化运行指标. 因此, 设定值设计是实现运行优化的一个关键问题<sup>[1-3]</sup>. 本文研究如何设计设定值, 迫使被控变量跟踪设定值, 保证运行指标以次优的方式达到理

想值或者被控制在指定的范围内, 即研究运行优化控制<sup>[1]</sup>.

目前, 国内外已经有众多学者面向工业过程开展了运行优化控制研究. 在20世纪50年代末期, 已经开始使用计算机为过程单元在线计算经济最优运行点. 同一时期, 美国一些大型公司(如: Texaco and Union Carbide)的计算机控制系统就已经能够实现系统的实时控制和优化<sup>[1]</sup>. 目前, 基于模型的运行优化控制已取得了大量研究成果. 文献[4-5]提出的单层动态实时优化 (dynamic real time optimization,

收稿日期: 2016-06-28; 录用日期: 2016-12-30.

†通信作者. E-mail: lijinna\_721@126.com, lijinna@syuct.edu.cn; Tel: +86 24-89387338.

本文责任编辑: 赵冬斌.

国家自然科学基金项目(61673280, 61104093, 61525302, 61333012, 61304028, 61590922, 61503257), 流程工业综合自动化国家重点实验室开放课题(PAL-N201603), 辽宁省高等学校杰出青年学者成长计划(LJQ2015088), 辽宁省自然科学基金项目(2015020164, 2014020138)资助.

Supported by National Natural Science Foundation of China (61673280, 61104093, 61525302, 61333012, 61304028, 61590922, 61503257), Open Project of State Key Laboratory of Synthetical Automation for Process Industries (PAL-N201603), Program of Distinguished Scholars Growing of Liaoning Provincial Universities (LJQ2015088) and Natural Science Foundation of Liaoning Province (2015020164, 2014020138).

DRTO)方法, 是通过设计被控过程的最优控制器, 来实现运行指标的最优化. 不过, 运行过程与被控过程时间尺度没有分离使得求解运行优化问题变得很困难.

与单层架构相比较, 双层层级架构运行优化策略更受欢迎. 基本思想为: 上层决策设定值, 优化运行指标; 底层为被控过程设计控制策略, 并保证被控变量跟踪设定值. 文献[6-7]采用实时优化(real time optimization, RTO)方法, 基于底层稳态的过程模型求解上层的优化问题, 为底层计算设定值, 底层基于模型预测控制(model predictive control, MPC)控制被控变量到设定值. 在文献[8-9]中, 上层分别采用 RTO与MPC集成技术和RTO与输出反馈集成技术补偿最优设定值, 底层控制环采用比例积分(proportional integral, PI)控制器, 使控制输出跟踪设计值. 正如文献[1,3]指出, RTO使用时, 要求仅当底层达到稳定状态时才能执行上层的经济优化策略, 这样导致上层优化决策等待时间较长, 且由于运行条件动态变化及一些具有不确定性的干扰因素, 在上层的等待时间中, 设定值可能发生改变, 此时底层仍然跟踪原先的设定值, 导致抗干扰性差.

为克服基于双层架构的RTO运行优化控制存在的问题, 文献[10]和文献[11]提出基于双层架构的DRTO方法, 将DRTO与MPC两种技术集成整合. 文献[12]中, 将经济指标动态优化问题与MPC集成, 在上层设计经济MPC最优化的时变运行轨迹, 底层采用传统的MPC方法跟踪设定值. 文献[13]集成RTO, DRTO和非线性的MPC(nonlinear model predictive control, NMPC)三种技术, 针对污水处理工厂运行优化问题, 给出最优运行控制策略. 但是, 目前现有的基于模型的RTO, DRTO和MPC方法往往需要参数校正, 且在被控动态过程与运行指标动态模型不精确的情况下, 运行指标优化的结果可能不能达到理想状态.

现有的基于模型的运行优化控制方法在实际的复杂工业过程运行优化控制应用中仍存在很大的局限性, 如: 要求能够精确建模被控过程和运行指标生成轨迹、涉及参数校正等, 导致计算量过大、设定值选择不优、运行指标不能保证等问题. 实际工业过程, 如浮选过程、磨矿过程等原料成份与性质反复波动、设备状况以及其他生产环境变化的影响其数学模型很难被精确构建<sup>[14]</sup>. 工业过程产生和存储的大量生产、设备和过程数据包含系统大量信息, 部分学者已经开始尝试基于数据驱动的运行优化控制研究<sup>[1-2, 14-16]</sup>. 文献[1-2, 14, 16]为复杂工业过程提出数据驱动的混杂智能运行优化控制方法, 基于案例推理(case-based reasoning, CBR)算法和PI反馈控制, 执行前馈补偿和反馈补偿, 优化运行指标. 但是, 文献[1-2, 14, 16]中没有给出设定值校正算法的收敛性证明. 文献[15]以

最小化运行指标与理想运行指标偏差为目标, 尝试采用神经网络估计最优设定值. 注意到, 文献[15]没有考虑多时间尺度问题.

数据驱动的工业过程运行优化控制研究还处于初级阶段. 如图1所示. 本文采用双层层级架构, 融合动态规划(dynamic programming, DP), Q-学习(Q-learning)方法, 采用近似动态规划(approximate dynamic programming, ADP)方法, 为工业过程提出一种基于Q-学习方法的数据驱动次优设定值设计方案. 底层采用PI控制器, 迫使控制输出跟踪上层设计的设定值, 实现运行指标以次优的方式跟踪理想值.

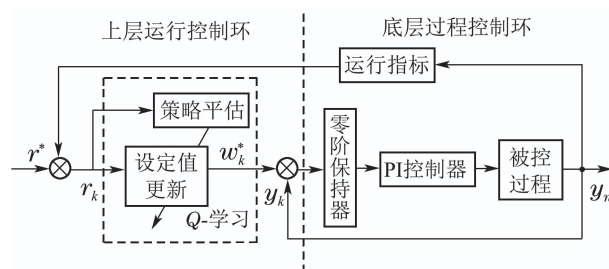


图1 工业过程运行优化控制

Fig. 1 Operational optimization control for industrial processes

本文贡献如下: 1) 不同于基于模型的运行优化控制方法, 本文提出一种Q-学习算法, 不依赖于被控过程模型参数和运行指标动态参数, 完全基于数据学习次优设定值, 避免模型参数不精确, 甚至未知给运行优化控制问题研究带来的消极影响; 2) 注意到文献[1-2, 14, 16]采用设定值修正策略, 目的是补偿不恰当的初始设定值. 不同于文献[1-2, 14, 16], 本文采用Q-学习方法, 直接给出次优设定值学习方案. 到目前为止, 采用本文方法研究数据驱动的工业过程运行优化控制问题还没有相关报道.

## 2 问题描述(Problem descriptions)

本部分首先建模底层控制环, 然后刻画运行指标生成轨迹, 最后阐述运行优化控制问题.

### 2.1 底层过程控制环建模 (Modeling the lower-layer process control loop)

实际工业过程往往为非线性动态模型, 然而它们通常在运行点附近稳态运行, 能在稳态附近线性化.

考虑如下被控过程模型:

$$\begin{cases} x_{n+1} = Ax_n + Bu_n, \\ y_n = Cx_n, \end{cases} \quad (1)$$

其中:  $x_n = x(nT_0) \in \mathbb{R}^{n_x}$ ,  $u_n = u(nT_0) \in \mathbb{R}^{n_u}$  和  $y_n = y(nT_0) \in \mathbb{R}^{n_y}$  分别为被控对象状态、控制输入和控制输出.  $A$ ,  $B$ 和 $C$ 为适维矩阵.  $T_0$ 为底层过程控制环采样周期,  $n(n = 1, 2, \dots)$ 为正整数. 如图2所示, 采用如下PI控制器:

$$u_n = K_p e_n + K_i E_n, \quad (2)$$

其中:  $K_p$ 为PI控制器的比例系数;  $K_i$ 为PI控制器的积分系数;  $e_n = w_n - y_n$ ,  $e_n$ 表示设定值与控制输出的误差;  $E_n = \sum_{i=0}^{n-1} e_i$ . 设计的PI控制器需要保证底层过程控制环的控制输出稳态跟踪设定值.

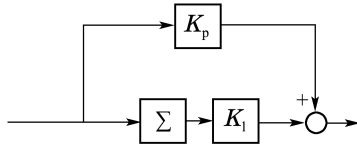


图2 PI控制器

Fig. 2 PI controller

引入增广向量  $\eta_n = (x_n^T \ E_n^T)^T$ , 可得底层控制环闭环控制系统:

$$\begin{cases} \eta_{n+1} = A_h \eta_n + B_h w_n, \\ y_n = C_h \eta_n, \end{cases} \quad (3)$$

其中:  $A_h = \bar{A} + \bar{B} \bar{K} \bar{C}$ ,  $C_h = [C \ 0]$ ,

$$\bar{A} = \begin{bmatrix} A & 0 \\ -C & I \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} B \\ 0 \end{bmatrix},$$

$$\bar{C} = \begin{bmatrix} -C & 0 \\ 0 & I \end{bmatrix}, \quad B_h = \begin{bmatrix} B K_p \\ I \end{bmatrix}, \quad \bar{K} = [K_p \ K_i].$$

## 2.2 运行指标生成轨迹(Generation trajectory of operational index)

工业过程的运行指标通常与底层控制环的控制输出有关<sup>[8-9,17]</sup>. 图1中运行指标为根据底层控制输出计算的经济收益指标, 其与底层控制输出的关系模型为

$$r_k = M y_k, \quad (4)$$

其中:  $r_k = r(kT) \in \mathbb{R}^{n_r}$  为运行指标,  $T$  为运行指标更新周期,  $k(k=1, 2, \dots)$  为正整数. 矩阵  $M$  为适维矩阵. 实际工业过程, 上层运行控制与底层过程控制环往往采用不同的时间尺度. 假设上层运行指标  $r_k$  的计算周期  $T$  为底层过程控制环采样周期  $T_0$  的  $N_0$  倍, 即  $T = N_0 T_0$ . 图1中设置零阶保持器(zero order holder, ZOH)的目的是为了使经过零阶保持器前后的信号频率相同. 由于底层采样速度快, 而上层输出的设定值信号为慢信号, 所以有如下关系成立:

$$w_k = w(kT) = w(N_0 k T_0) =$$

$$w((N_0 k + 1)T_0) = \dots =$$

$$w((N_0 k + N_0 - 1)T_0). \quad (5)$$

## 2.3 运行优化控制问题阐述(Operational optimization control problem formulation)

由式(3)(5), 有

$$\eta((k+1)T) =$$

$$\eta(N_0(k+1)T_0) = \eta((N_0 k + N_0)T_0) =$$

$$A_h \eta((N_0 k + N_0 - 1)T_0) + B_h w(kT) =$$

$$A_h^2 \eta((N_0 k + N_0 - 2)T_0) + A_h B_h w(kT) +$$

$$B_h w(kT) = \dots =$$

$$A_h^{N_0} \eta(kT) + \sum_{i=0}^{N_0-1} A_h^i B_h w(kT), \quad (6)$$

改写式(6)为

$$\eta_{k+1} = A_0 \eta_k + B_0 w_k, \quad (7)$$

其中:  $A_0 = A_h^{N_0}$ ,  $B_0 = \sum_{i=0}^{N_0-1} A_h^i B_h$ . 由式(3)-(4), 有

$$r_k = C_0 \eta_k, \quad (8)$$

其中  $C_0 = M C_h$ . 为实现运行指标跟踪理想运行指标  $r^*$ , 定义如下性能指标:

$$\frac{1}{2} \sum_{k=0}^{\infty} \gamma^k ((r_k - r^*)^T Q (r_k - r^*) + w_k^T R w_k), \quad (9)$$

其中:  $\gamma$  ( $0 < \gamma < 1$ ) 为折扣因子, 矩阵  $Q$  和  $R$  分别为半正定矩阵和正定矩阵. 由式(7)-(9), 得到如下线性二次跟踪(linear quadratic tracking, LQT)问题:

$$\begin{cases} \min_{w_k} \frac{1}{2} \sum_{k=0}^{\infty} \gamma^k ((r_k - r^*)^T Q (r_k - r^*) + w_k^T R w_k), \\ \text{s.t. (7) - (8)}. \end{cases} \quad (10)$$

**注1** 注意到, LQT问题(10)与传统的LQT问题不同, 其约束条件中增广系统的状态  $\eta_k$ ,  $k$  趋于无穷大时不趋于零. 设定值  $w_k$  往往依赖增广系统状态  $\eta_k$  和理想运行指标  $r^*$ , 运行指标通常表征产品质量、生产效率、能耗与物耗等<sup>[1-3]</sup>, 所以通常情况下  $r^* \neq 0$ . 那么如果  $\gamma=1$  时, 性能指标(9)可能无界, 因此取  $0 < \gamma < 1$ .

## 3 次优设定值设计(Suboptimal setpoint design)

本文最终目标是设计设定值, 迫使底层控制输出跟踪设定值, 实现运行指标以次优的方式跟踪理想值. 因此本文需要求解优化问题(10). 本部分首先将求解LQT问题转化为求解具有折扣因子的线性二次调节(linear quadratic regulator, LQR)问题, 然后引入贝尔曼方程及哈密尔顿函数, 基于动态规划方法和增强学习方法给出次优设定值设计方案.

### 3.1 LQR问题阐述(LQR problem formulation)

**引理1** 令  $X_k = [\eta_k^T \ r_k^*]^T$ , LQT问题(10)等价于如下具有折扣因子的LQR问题:

$$\begin{cases} \min_{w_k} \frac{1}{2} \sum_{k=0}^{\infty} \gamma^k (X_k^T A_J X_k + w_k^T R w_k), \\ \text{s.t. } X_{k+1} = \tilde{A}_0 X_k + \tilde{B}_0 w_k, \end{cases} \quad (11)$$

其中:  $A_J = \tilde{C}^T Q \tilde{C}$ ,  $\tilde{C} = [C_0 \ -I]$ ,

$$\tilde{A}_0 = \begin{bmatrix} A_0 & 0 \\ 0 & I \end{bmatrix}, \quad \tilde{B}_0 = \begin{bmatrix} B_0 \\ 0 \end{bmatrix}.$$

证 由式(8), 并且根据增广向量  $X_k$  的定义, 改写式(10)为

$$\frac{1}{2} \sum_{k=0}^{\infty} \gamma^k (X_k^T A_J X_k + w_k^T R w_k). \quad (12)$$

由于本文考虑理想运行指标为常数, 所以  $r_{k+1}^* = r_k^*$ . 那么, 参考式(7), 有

$$X_{k+1} = \tilde{A}_0 X_k + \tilde{B}_0 w_k. \quad (13)$$

由式(12)–(13), 得到式(11). 证毕.

根据式(11)中的性能指标, 定义值函数为

$$V(X_k, w_k) = \frac{1}{2} \sum_{k=0}^{\infty} \gamma^k (X_k^T A_J X_k + w_k^T R w_k). \quad (14)$$

**引理 2** 针对具有折扣因子的 LQT 问题(10), 设计设定值

$$w_k = -K X_k, \quad (15)$$

其中:  $-K X_k = K_\eta \eta_k + K_r r^*$ ,  $K_\eta$  和  $K_r$  为设定值反馈控制增益, 则值函数(14)可以表示为

$$V(\eta_k, r_k^*) = V(X_k) = \frac{1}{2} X_k^T P X_k, \quad (16)$$

其中  $P$  为正定矩阵.

证 证明方法类似文献[18], 略.

### 3.2 求解 LQR 问题(Solving LQR problem)

为求解 LQR 问题(11), 需做如下假设:

**假设 1**  $(\tilde{A}_0, \sqrt{A_J})$  可观测.

由式(14), 得到如下贝尔曼(Bellman)方程:

$$V(X_k) = \frac{1}{2} (X_k^T A_J X_k + w_k^T R w_k) + \gamma V(X_{k+1}, w_{k+1}). \quad (17)$$

将式(16)代入式(17), 有

$$X_k^T P X_k = X_k^T A_J X_k + w_k^T R w_k + \gamma X_{k+1}^T P X_{k+1}, \quad (18)$$

得到如下哈密尔顿函数:

$$H(X_k) = X_k^T A_J X_k + w_k^T R w_k + \gamma X_{k+1}^T P X_{k+1} - X_k^T P X_k. \quad (19)$$

**定理 1** 对于 LQR 问题(11), 最优设定值为

$$w_k^* = -K^* X_k, \quad (20)$$

其中  $K^* = (R + \gamma \tilde{B}_0^T P \tilde{B}_0)^{-1} \gamma \tilde{B}_0^T P \tilde{A}_0$ . 并且矩阵  $P$  满足如下代数黎卡提方程(algebraic riccati equation, ARE):

$$A_J - P + \gamma \tilde{A}_0^T P \tilde{A}_0 - \gamma^2 \tilde{A}_0^T P \tilde{B}_0 \cdot (R + \gamma \tilde{B}_0^T P \tilde{B}_0)^{-1} \tilde{B}_0^T P \tilde{A}_0 = 0. \quad (21)$$

证 基于最优性的必要条件<sup>[18–19]</sup>, 有

$$\begin{aligned} \frac{\partial H(X_k, w_k)}{\partial w_k} &= \\ 2R w_k + \gamma \left( \frac{\partial X_{k+1}}{\partial w_k} \right)^T \frac{\partial V(X_{k+1})}{\partial X_{k+1}} &= \\ 2R w_k + 2\gamma \tilde{B}_0^T P X_{k+1} &= 0. \end{aligned} \quad (22)$$

将式(13)代入上式, 得到式(20). 将式(13)(20)代入贝尔曼方程(18)中, 得到式(21). 证毕.

**注 2** 考虑 LQT 问题(10), 定义  $\bar{e}_{r,k} = \gamma^{\frac{k}{2}} e_{r,k}$ , 其中  $e_{r,k} = C_0 \eta_k - r_k^*$  为跟踪误差. 文献[18]已经证明 ARE(21) 得到的设定值保证  $\lim_{k \rightarrow \infty} \bar{e}_{r,k} = 0$ , 并且最小化 LQT 问题(10) 中性能指标.

**注 3** 参考文献[18], 可以得到  $V(\bar{X}_{k+1}) - V(\bar{X}_k) = \frac{1}{2} \bar{X}_k^T (-A_J - K^T R K) \bar{X}_k < 0$ . 由  $A_J = \tilde{C}^T Q \tilde{C}$  可知, 半正定矩阵  $Q$  的取值越大, 值函数变化越大,  $\bar{e}_{r,k}$  趋于零的速度越快. 因此, 可以取比较小的  $\gamma$  和适当大的  $Q$  使  $e_{r,k}$  在  $\gamma^k$  变得很小前尽可能小.

为求解 ARE (21) 中的正定矩阵  $P$ , 可以设计如下算法, 以便获得 LQR 问题(11) 的解.

**算法 1** 基于模型的策略迭代算法.

1) 策略评估(计算矩阵  $P^{j+1}$ ,  $j = 1, 2 \dots$ ).

$$P^{j+1} = A_J + (K^j)^T R K^j + \gamma (\tilde{A}_0 - \tilde{B}_0 K^j)^T \cdot P^{j+1} (\tilde{A}_0 - \tilde{B}_0 K^j). \quad (23)$$

2) 策略改进.

$$K^{j+1} = \gamma (R + \gamma \tilde{B}_0^T P^{j+1} \tilde{B}_0)^{-1} \cdot \tilde{B}_0^T P^{j+1} \tilde{A}_0. \quad (24)$$

**注 4** 算法 1 为策略迭代算法, 要求初始值  $K^0$  能镇定系统(13)<sup>[18–19]</sup>. 通过迭代 ARE(23), 计算正定矩阵  $P^{j+1}$ , 改进设定值  $w_k^{j+1} = -K^{j+1} X_k$ , 直到收敛, 即  $\lim_{j \rightarrow \infty} w_k^{j+1} = w_k^* = -K^* X_k$ . 对于模型参数未知的增广系统(13), 假设系统有形式模型  $\tilde{A}_0 = A_N + \Delta A$ ,  $\tilde{B}_0 = B_N + \Delta B$ , 其中  $\Delta A$  和  $\Delta B$  未知. 在这种情况下, 可以采用鲁棒控制等方法, 获得可镇定控制器, 即得到初始值  $K^0$ <sup>[20]</sup>.

**注 5** 应该指出的是, 用线性模型刻画工业过程是真实过程的次优表达形式, 那么算法 1 学习得到的是次优设定值, 是凸优化问题的全局解. 非线性模型的最优运行控制是一个非凸优化问题, 求解困难. 线性化处理方式对于工业运行优化控制是一种简单实用、易用实现的方法<sup>[21–23]</sup>.

**注 6** 算法 1 在学习次优设定值时要求底层控制环动态和上层运行控制环模型参数完全已知. 然而, 实际的工业运行过程, 由于过程动态复杂, 运行过程模型参数往往未知, 因此下文给出基于 Q-学习的数据驱动最优设定值学习算法.

#### 4 基于Q-学习的次优设定值估计 (Suboptimal setpoint approximation based on Q-learning)

本部分采用Q-学习方法,完全利用数据,给出不依赖模型参数的次优设定值学习算法。

##### 4.1 Q-函数(Q-function)

根据贝尔曼方程(17),定义Q-函数为

$$Q(X_k, w_k) = \frac{1}{2} X_k^T A_J X_k + \frac{1}{2} w_k^T R w_k + \frac{1}{2} \gamma X_{k+1}^T P X_{k+1}. \quad (25)$$

由增广系统(13),改写式(25)为

$$Q(X_k, w_k) = \frac{1}{2} \begin{bmatrix} X_k \\ w_k \end{bmatrix}^T \begin{bmatrix} A_J + \gamma \tilde{A}_0^T P \tilde{A}_0 & \gamma \tilde{A}_0^T P \tilde{B}_0 \\ \gamma (\tilde{A}_0^T P \tilde{B}_0)^T & R + \gamma \tilde{B}_0^T P \tilde{B}_0 \end{bmatrix} \begin{bmatrix} X_k \\ w_k \end{bmatrix}, \quad (26)$$

因此,定义

$$Q(X_k, w_k) = \frac{1}{2} \begin{bmatrix} X_k \\ w_k \end{bmatrix}^T H \begin{bmatrix} X_k \\ w_k \end{bmatrix} = \frac{1}{2} \begin{bmatrix} X_k \\ w_k \end{bmatrix}^T \begin{bmatrix} H_{X_k X_k} & H_{X_k w_k} \\ (H_{X_k w_k})^T & H_{w_k w_k} \end{bmatrix} \begin{bmatrix} X_k \\ w_k \end{bmatrix}, \quad (27)$$

其中:

$$\begin{cases} H_{X_k X_k} = A_J + \gamma \tilde{A}_0^T P \tilde{A}_0, \\ H_{X_k w_k} = \gamma (\tilde{A}_0^T P \tilde{B}_0)^T, \\ H_{w_k w_k} = R + \gamma \tilde{B}_0^T P \tilde{B}_0. \end{cases} \quad (28)$$

在式(27)中应用  $\frac{\partial Q(X_k, w_k)}{\partial w_k} = 0$ , 有

$$w_k^* = -(H_{w_k w_k})^{-1} H_{w_k X_k} X_k. \quad (29)$$

对应式(28),式(20)成立,即

$$w_k^* = -(R + \gamma \tilde{B}_0^T P \tilde{B}_0)^{-1} \gamma \tilde{A}_0^T P \tilde{B}_0 X_k. \quad (30)$$

##### 4.2 基于Q-学习方法求解LQT问题 (Solving LQT problem based on Q-learning method)

为实现本文研究目标:不依赖被控过程模型参数和运行指标生成轨迹模型参数估计次优设定值,根据定义的Q-函数(25),给出不依赖增广系统(13)模型参数的估计次优设定值的Q-学习算法。Q-函数(25)满足

$$Q(X_k, w_k) = \frac{1}{2} (X_k^T A_J X_k + w_k^T R w_k) + \gamma Q(X_{k+1}, w_{k+1}). \quad (31)$$

定义  $Z_k = [X_k^T \ w_k^T]^T$ , 考虑到  $Q(X_k, w_k) = \frac{1}{2} Z_k^T$

$H Z_k$ , 改写基于Q-函数的贝尔曼方程(31)为

$$Z_k^T H Z_k = X_k^T A_J X_k + w_k^T R w_k + \gamma Z_{k+1}^T H Z_{k+1}. \quad (32)$$

为计算Q-函数中矩阵H,以便根据式(29)计算最优设定值  $w_k^*$ , 采用策略迭代方法,给出算法2。

##### 算法2 Q-学习算法。

1) 策略评估。

$$Z_k^T H^{j+1} Z_k = X_k^T A_J X_k + (w_k^j)^T R w_k^j + \gamma Z_{k+1}^T H^{j+1} Z_{k+1}. \quad (33)$$

2) 策略改进。

$$w_k^{j+1} = -((H_{w_k w_k})^{j+1})^{-1} (H_{w_k X_k})^{j+1} X_k. \quad (34)$$

**注7** 算法2中步骤1评估策略  $w_k^j$ , 步骤2更新设定值。对比算法1,算法2不依赖工业过程和运行指标模型的参数信息,完全利用数据,以迭代的方式计算矩阵  $H^{j+1}$ , 更新设定值  $w_k^{j+1}$ 。更新后的设定值输入到底层过程控制环,在底层PI控制器作用下,被控过程的输出跟踪更新后的设定值,产生新的工业运行过程数据。利用新产生的数据,执行步骤1,步骤2,更新设定值。按照此种方法循环直到  $w_k^{j+1}$  收敛。

**注8** 以在线的方式执行算法2中步骤1,沿着增广系统(13)的动态轨迹收集数据  $X_k, Z_k, Z_{k+1}$ , 采用递推最小二乘法 (recursive least squares, RLS) 或者批最小二乘法 (batch least squares, BLS) 估计矩阵  $H^{j+1}$  的值<sup>[18,24]</sup>。矩阵H是  $(n_x + n_r + 2n_y) \times (n_x + n_r + 2n_y)$  维对称矩阵,具有  $(n_x + n_r + 2n_y) \times (n_x + n_r + 2n_y + 1)/2$  个独立元。因此在执行RLS或者BLS方法前至少需要收集  $(n_x + n_r + 2n_y) \times (n_x + n_r + 2n_y + 1)/2$  个数据。并且,为保证式(33)中  $H^{j+1}$  能被准确估计,要求数据充分,执行策略迭代时需要设定值输入持续激励。持续激励往往采用正弦信号<sup>[25]</sup>。

**注9** 如果式(33)中  $H^{j+1}$  能被准确估计,在初始值  $w_k^0$  保证运行指标稳定的前提下,基于策略迭代方法的Q-学习算法2能保证步骤2中的设定值  $w_k^{j+1}$  最终收敛于次优设定值<sup>[18]</sup>, 即  $\lim_{j \rightarrow \infty} w_k^{j+1} = w_k^*$ 。

**注10** 注意到,文献[1-2,14,16]为工业过程提出数据驱动的混杂智能运行优化控制方法,将案例推理算法和PI控制结合,执行前馈补偿和反馈补偿,调整设定值。但是,设定值修正的增量需要根据操作者经验确定。不同于文献[1-2,14,16],本文基于Q-学习方法,给出次优设定值学习算法,避免人工操作存在的偏差或错误。与文献[15]比较,文献[15]采用神经网络估计最优设定值,要求已知性能指标的最优值,而实际工业运行过程需要优化的性能指标最优值往往未知。并且,文献[15]没有考虑多时间尺度问题。

**注11** 底层过程控制环PI控制器可以采用极点配置

方法, Z-N 参数整定等方法设计. 或者实际应用中 PI 控制器参数可以由操作者经验获得. 因此, 本文不研究底层过程控制环 PI 控制器设计, 重点研究上层运行控制环次优设定值估计问题.

实际工业过程控制输出往往被约束在一定的范围内, 即  $y_{\min} \leq y_n \leq y_{\max}$ , 为满足这一约束, 用如下规则修正算法 2 中学习得到的设定值.  $\hat{w}_k^{j+1}$  表示通过如下规则得到的次优设定值.

**规则 1** 如果  $y_{\min} \leq w_k^{j+1} \leq y_{\max}$ , 那么  $\hat{w}_k^{j+1} = w_k^{j+1}$ ;

**规则 2** 如果  $w_k^{j+1} < y_{\min}$  或者  $w_k^{j+1} > y_{\max}$ , 那么  $\hat{w}_k^{j+1} = w_{k-1}^{j+1}$ .

**定理 2** 存在常数  $\lambda > 0$ , 由上述规则得到的设定值  $\hat{w}_k^{j+1}$  保证如下式子成立:

$$\|\hat{w}_k^{j+1} - w_k^*\| \leq \lambda. \quad (35)$$

证 由规则 1 和规则 2, 有

$$\hat{w}_k^{j+1} = w_k^{j+1} + \alpha(w_{k-1}^{j+1} - w_k^{j+1}), \quad (36)$$

其中  $\alpha = 0$  或  $1$ . 那么有

$$\begin{aligned} \|\hat{w}_k^{j+1} - w_k^*\| &\leq \\ \|\hat{w}_k^{j+1} - w_k^*\| + \|w_{k-1}^{j+1} - w_k^{j+1}\|. \end{aligned} \quad (37)$$

由注 8 可知, 对任意  $\varepsilon > 0$ , 有

$$\|w_k^{j+1} - w_k^*\| \leq \varepsilon. \quad (38)$$

鉴于策略迭代方法在初始值  $w_k^0$  保证运行指标稳定的前提下, 每次迭代  $j$  中  $w_k^{j+1}$  能镇定上层运行过程. 因此, 存在常数  $\sigma > 0$ , 有

$$\|w_{k-1}^{j+1} - w_k^{j+1}\| \leq \sigma. \quad (39)$$

取  $\lambda = \varepsilon + \sigma$ , 由式(37)–(39), 式(35)成立. 证毕.

### 5 实验仿真(Experiment simulations)

以浮选工业过程为对象, 验证本文方法的有效性. 对于单个槽体, 浮选过程的数学模型为<sup>[8-9, 17, 21]</sup>

$$\begin{cases} \dot{x} = \begin{bmatrix} -17.9624 & 0 & 65.6 & 0 \\ 0 & -0.1024 & 0 & 316 \\ 17.9 & 0 & -65.9474 & 0 \\ 0 & 0.04 & 0 & -316.247 \end{bmatrix} x + \begin{bmatrix} 0.3746 & 0.1549 \\ 18.3790 & 3.0484 \\ -3.96 & 0 \\ -0.0903 & 0 \end{bmatrix} u, \\ y = \begin{bmatrix} 0 & 0 & 0.0832 & -3.6501 \\ 0.0203 & -0.0004 & 0 & 0 \end{bmatrix} x. \end{cases} \quad (44)$$

槽体中存在大量泡沫, 那么在平衡点处精矿和尾矿流量之和与给矿流量不等. 对于底层闭环系统(3), 给定初始状态  $\eta_0 = [16.7 \ 1124 \ 4.56 \ 0.2 \ 0.1$

$$\begin{cases} \frac{dM_p^i}{dt} = -(k_p^i + \frac{q_T}{(1-\varepsilon_g)Ah_p})M_p^i + k_e^i M_e^i + q_a X_a^i, \\ \frac{dM_e^i}{dt} = -(k_e^i + \frac{q_c}{(1-\varepsilon_g)A(H-h_p)})M_e^i + k_p^i M_p^i, \end{cases} \quad (40)$$

其中: 精矿品位为

$$L_{cg} = \frac{M_e^1 g_{cp}^1 + M_e^2 g_{cp}^2}{M_e^1 + M_e^2} L_{cu}, \quad (41)$$

尾矿品位为

$$L_{tg} = \frac{M_p^1 g_{cp}^1 + M_p^2 g_{cp}^2}{M_p^1 + M_p^2} L_{cu}, \quad (42)$$

$M_e^i$  为泡沫质量(t),  $M_p^i$  为泥浆质量(t),  $q_a$  为给矿矿浆流量( $m^3 \cdot \min^{-1}$ ),  $h_p$  为矿浆液位高度(m).  $i = 1, 2$  分别表示矿物种类 1 (主要为黄铜矿) 和矿物种类 2 (主要为脉石). 槽子尺寸为  $53.2 \times 3.2 \text{ m}^3$ , 黄铜矿浆中黄铜品位为  $g_{cp}^1 = 0.417$ , 脉石矿浆中黄铜品位为  $g_{cp}^2 = 0.0034$ , 加料品位(黄铜矿浆和脉石矿浆中黄铜品位)为  $g_a = 0.0234$ , 黄铜矿中黄铜品位为  $l_{cu} = 0.421$ , 其他参数物理意义及取值见表 1<sup>[23]</sup>.

运行指标  $r_k$  表示黄铜矿浮选生产过程单位时间目标经济收益. 过程控制输入为泥浆高度  $h_p$  与给矿流量  $q_a$ , 即  $u = (h_p, q_a)$ ; 输出为精矿品位  $L_{cg}$  与尾矿品位  $L_{tg}$ , 即  $y = (L_{cg}, L_{tg})$ . 假设  $0 < L_{cg} < 1$ ,  $0 < L_{tg} < 1$ .  $w_k = (L_{cg}, L_{tg})$  的物理意义为黄铜矿浮选产品精尾矿品位设定值. 状态变量为泥浆质量与泡沫质量, 即  $x = (M_p^i, M_e^i)$ .

在表 1 所示的运行条件下, 黄铜矿浮选过程的一个平衡点为

$$\begin{cases} [h_p \ q_a] = [2.8 \ 17], \\ [M_p^1 \ M_p^2 \ M_e^1 \ M_e^2] = [16.8 \ 824.266 \ 4.56 \ 0.104]. \end{cases} \quad (43)$$

在该平衡点对式(40)–(42)线性化得到如下系统:

$-0.2]^T$ , 底层控制环采样周期  $T_0 = 6 \text{ s}$ . 取上层运行控制环采样周期  $T = 30 \text{ min}$ , 参考铜矿和尾矿的市场价格, 取  $M = [100 \ -10]$ , 理想经济收益  $r^* = 26$ .

$$\text{取 } Q = 100, \gamma = 0.6, R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

初始控制输入为

$$K^0 = \begin{bmatrix} 0.0001 & -0.0000 & -0.0006 \\ 0.0001 & -0.0000 & -0.0005 \\ 0.0239 & 0.0088 & -0.0095 & 0 \\ 0.0205 & 0.0073 & -0.0079 & 0 \end{bmatrix}.$$

执行算法2, 图3给出Q-函数矩阵H和设定值增益K的收敛结果, 得到的次优设定值增益为

$$K^* = \begin{bmatrix} 0.0003 & -0.0000 & -0.0003 \\ 0.0003 & -0.0000 & -0.0003 \\ 0.0211 & 0.0008 & 0.0001 & -0.0149 \\ 0.0176 & -0.0005 & 0.0014 & -0.0068 \end{bmatrix}.$$

表1 参数物理意义及取值

Table 1 Parameters and values

参数	物理意义	取值
$k_p^i$	浮选速率	$k_p^1 = 17.9 \text{ min}^{-1}$ $k_p^2 = 0.04 \text{ min}^{-1}$
$k_e^i$	排水速率	$k_e^1 = 65.6 \text{ min}^{-1}$ $k_e^2 = 316 \text{ min}^{-1}$
$q_T$	尾流	$9.3 \text{ m}^3/\text{min}$
$\varepsilon_g$	停滞常数	0
$A$	截面积	$53.2 \text{ m}^2$
$X_a^i$	矿物种类浓度	$X_a^1 = 0.1549 \text{ t/m}^3$ $X_a^2 = \frac{g_{cp}^1 - g_a}{g_a - g_{cp}^2} X_a^1 = 3.0484 \text{ t/m}^3$
$q_c$	精矿矿浆流量	$7.392 \text{ m}^3/\text{min}$
$H$	总高度	$3.2 \text{ m}$

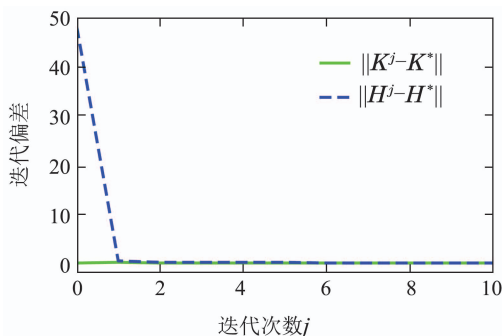


图3 Q-学习结果

Fig. 3 The results of Q-learning

由于用线性系统刻画非线性浮选过程存在偏差,

并且浮选过程原材料成分存在波动, 因此在底层控制环加入有界随机扰动, 图4和图5分别给出采用本文方法和文献[9]方法精矿品位和尾矿品位跟踪次优设定值曲线.

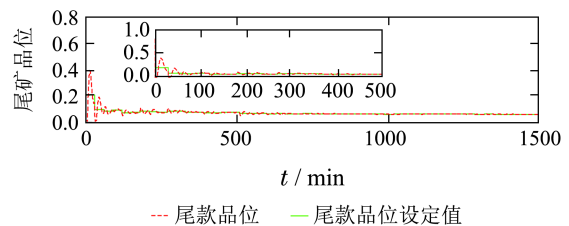
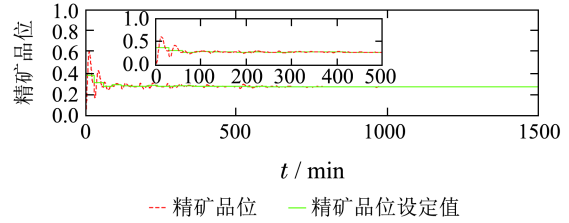


图4 矿物品位跟踪结果

Fig. 4 The tracking results of ore grades

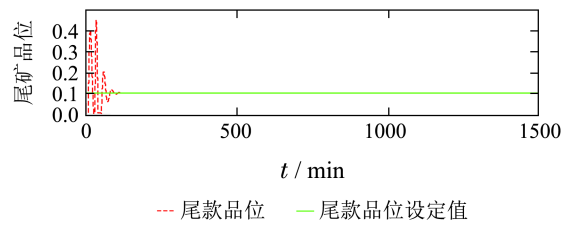
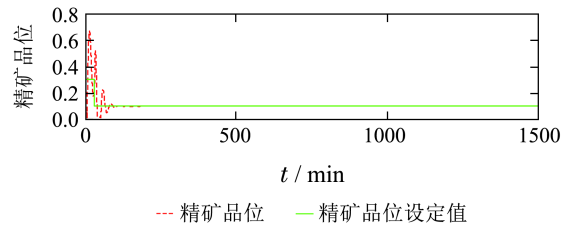


图5 矿物品位跟踪结果

Fig. 5 The tracking results of ore grades

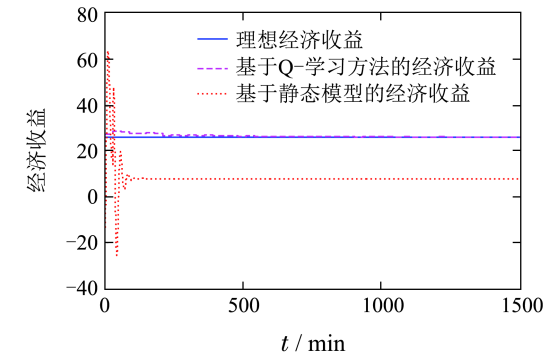


图6 运行指标跟踪结果

Fig. 6 The tracking results of operational index

图6给出本文方法和文献[9]方法下经济收益跟踪理想值的曲线. 仿真表明文献[9]采用基于RTO的补偿方法不能实时地给出适应当前工况或存

在外界有界干扰的最佳设定值. 本文采用数据驱动 Q-学习方法, 对于动态工况和外界干扰具有一定的鲁棒性.

考虑浮选过程具有两个槽体, 两个槽体的数学模型形如式 (40)–(42). 式 (43) 为其中一个槽体的平衡点, 另一槽体的平衡点为

$$\begin{aligned} [h_p \quad q_a] &= [2.8 \quad 14.5], \\ [M_p^1 \quad M_p^2 \quad M_e^1 \quad M_e^2] &= [7.5 \quad 1318 \quad 2.04 \quad 0.2]. \end{aligned}$$

设两个槽体的泡沫质量和泥浆质量为整个浮选过程的状态, 两个槽体的泥浆高度和给矿流量为整个浮选过程的控制输入, 控制输出为两个槽体的精矿品位和尾矿品位. 类似单槽体的浮选过程, 执行算法 2, 学习得到次优设定值. 取  $M = [100 \quad 0 \quad 100 \quad -10]$ , 即经济收益与两个槽体的精矿品位和后一个槽体的尾矿品位相关. 假设理想经济收益  $r^* = 30$ , 图 7 和图 8 分别给出槽体的精矿品位、尾矿品位和经济收益跟踪理想值的曲线.

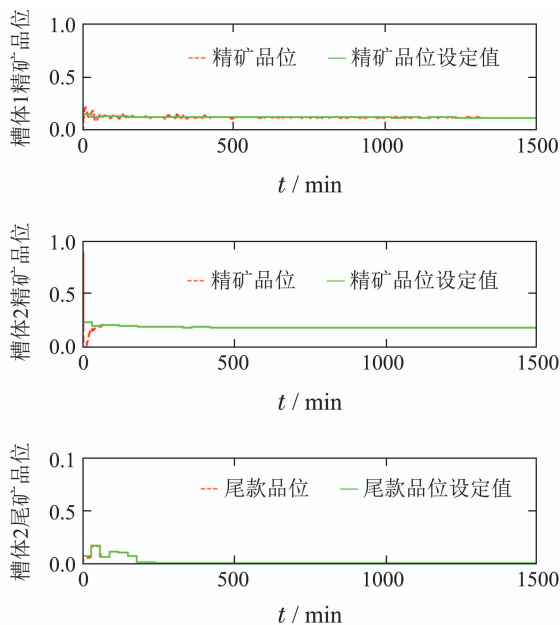


图 7 精矿品位、尾矿品位和经济收益跟踪结果

Fig. 7 The tracking results of concentrate grade, tail grade and economic benefit

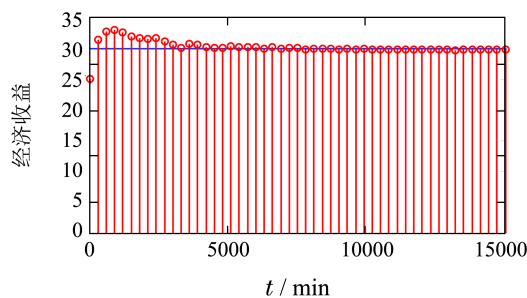


图 8 运行指标跟踪结果

Fig. 8 The tracking results of operational index

## 6 结语(Conclusions)

本文针对在运行点附近稳态运行, 能在稳态附近线性化的工业过程和一类经济收益指标, 考虑运行指标与控制变量之间关系以及被控过程很难被精确建模、甚至模型参数未知的问题, 提出了一种不依赖模型参数的 Q-学习算法, 估计次优设定值. 首先阐述运行优化控制 LQT 问题; 为实现运行指标以次优的方式跟踪理想值, 融合动态规划方法和 Q-学习方法, 给出基于 Q-函数的贝尔曼方程采用策略迭代, 提出 Q-学习算法, 估计次优设定值. 最后浮选过程的仿真实验和理论分析表明了本文方法的有效性.

## 参考文献(References):

- [1] CHAI T Y, QIN S J, WAND H. Optimal operational control for complex industrial processes [J]. *Annual Reviews in Control*, 2014, 38(1): 81 – 92.
- [2] DING J L, CHEN Q, CHAI T Y, et al. Data mining based feedback regulation in operation of hematite ore mineral processing plant [C] // *American Control Conference*. St. Louis: IEEE, 2009: 907 – 912.
- [3] CHAI T Y, DING J L, WU F. Hybrid intelligent control for optimal operation of shaft furnace roasting process [J]. *Control Engineering Practice*, 2011, 19(3): 264 – 275.
- [4] ENGELL S. Feedback control for optimal process operation [J]. *Journal of Process Control*, 2007, 17(3): 203 – 219.
- [5] ADETOLA V, DEHAAN D, GUAY M. Adaptive model predictive control for constrained nonlinear systems [J]. *Systems & Control Letters*, 2009, 58(5): 320 – 326.
- [6] ODLOAK D. Robust integration of RTO and MPC [J]. *Computer Aided Chemical Engineering*, 2009, 27(Part A): 119 – 126.
- [7] QIN S Z, BADGWELL T A. An overview of nonlinear model predictive control applications [M] // *Nonlinear Model Predictive Control*. Birkhäuser Basel: Springer, 2000: 369 – 392.
- [8] CHAI T Y, ZHAO L, QIU J, et al. Integrated network-based model predictive control for setpoints compensation in industrial processes [J]. *IEEE Transactions on Industrial Informatics*, 2013, 9(1): 417 – 426.
- [9] LIU F, GAO H, QIU J, et al. Networked multirate output feedback control for setpoints compensation and its application to rougher flotation process [J]. *IEEE Transactions on Industrial Electronics*, 2014, 61(1): 460 – 468.
- [10] DÜNNEBIER G, HESSEM D V, KADAM J V, et al. Optimization and control of polymerization processes [J]. *Chemical Engineering and Technology*, 2005, 28(5): 575 – 580.
- [11] TOSUKHOWONG T, LEE J H. Real-time economic optimization for an integrated plant via a dynamic optimization scheme [C] // *American Control Conference*. Boston, Massachusetts: IEEE, 2004: 233 – 238.
- [12] ELLIS M, CHRISTOFIDES P D. Integrating dynamic economic optimization and model predictive control for optimal operation of nonlinear process systems [J]. *Control Engineering Practice*, 2014, 22(22): 242 – 251.
- [13] VEGA P, REVOLLAR S, FRANCISCO M, et al. Integration of set point optimization techniques into nonlinear MPC for improving the operation of WWTPs [J]. *Computers and Chemical Engineering*, 2014, 68(8): 78 – 95.



- [14] ZHOU Ping, CHAI Tianyou. Intelligent operational feedback control for typical hematite grinding processes [J]. *Control Theory & Applications*, 2014, 31(10): 1352 – 1359.  
(周平, 柴天佑. 典型赤铁矿磨矿过程智能运行反馈控制 [J]. 控制理论与应用, 2014, 31(10): 1352 – 1359.)
- [15] DAI W, CHAI T Y, YANG S X. Data-driven optimization control for safety operation of hematite grinding process [J]. *IEEE Transactions on Industrial Electronics*, 2015, 62(5): 2930 – 2941.
- [16] WU Z, WU Y, CHAI T Y, et al. Data-driven abnormal condition identification and self-healing control system for fused magnesium furnace [J]. *IEEE Transactions on Industrial Electronics*, 2014, 62(3): 1703 – 1715.
- [17] FAN Jialu, ZHANG Yewei, CHAI Tianyou. Optimal operational feedback control for a class of industrial processes [J]. *Acta Automatica Sinica*, 2015, 41(10): 1754 – 1761.  
(范家璐, 张也维, 柴天佑. 一类工业过程运行反馈优化控制方法 [J]. 自动化学报, 2015, 41(10): 1754 – 1761.)
- [18] KIUMARSI-KHOMARTASH B, LEWIS F L, KARIMPOUR A, et al. Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics [J]. *Automatica*, 2014, 50(4): 1167 – 1175.
- [19] LEWIS F L, VRABIE D, VAMVOUDAKIS K G. Reinforcement learning and feedback control using natural decision methods to design optimal adaptive controller [J]. *IEEE Systems Magazine*, 2012, 32(6), 76 – 105.
- [20] MODARES H, LEWIS F L. Linear quadratic tracking control of partially unknown continuous-time systems using reinforcement learning [J]. *IEEE Transactions on Automatic Control*, 2014, 59(11): 3051 – 3056.
- [21] HARRIS C C. Multiphase models of flotation machine behavior [J]. *International Journal of Mineral Processing*, 1978, 5(2): 107 – 129.
- [22] MALDONADO M, SBARBARO D, LIZAMA E. Optimal control of a rougher flotation process based on dynamic programming [J]. *Minerals Engineering*, 2007, 20(3): 221 – 232.
- [23] 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.
- [24] HAYKIN S S, WIDROW B. *Least-mean Square Adaptive Filters* [M]. New Jersey: A John Wiley & Sons, 2003.
- [25] AL-TAMIMI A, LEWIS F L, ABU-KHALAF M. Model-free Q-learning designs for linear discrete-time zero-sum games with application to  $H_\infty$  control [J]. *Automatica*, 2007, 43(3): 473 – 481.

### 作者简介:

**李金娜** (1977–), 女, 副教授, 硕士生导师, 从事运行优化控制、

网络控制和多智能体协同控制等研究, E-mail: lijinna@syuct.edu.cn;

**高溪泽** (1994–), 女, 硕士研究生, 从事运行优化控制等研究, E-

mail: gaoxize666@163.com;

**柴天佑** (1947–), 男, 教授, 博士生导师, 中国工程院院士, IEEE

Fellow, IFAC Fellow, 从事流程工业一体化运行优化控制、自适应控制、多变量智能解耦控制等, E-mail: tychai@mail.neu.edu.cn;

**范家璐** (1985–), 女, 副教授, 硕士生导师, 从事工业过程运行控制、

工业无线传感器网络与移动社会网络等研究, E-mail: fanjialu@gmail.com.