

基于多智能体强化学习的焦炉集气管 压力多级协调控制

秦 斌^{1,2}, 吴 敏¹, 王 欣^{1,2}, 阳春华¹

(1. 中南大学信息科学与工程学院, 湖南长沙 410083; 2. 湖南工业大学电气工程学院, 湖南株洲 412008)

摘 要: 针对焦炉集气管压力这类多变量强扰动非线性耦合系统, 提出了一种基于 Multi-Agent system (MAS) 的焦炉集气管压力智能多级协调控制系统方案. 采用基于 Agent 单元系统梯级协调体系和基于任务分解的实时 Agent 的组织与演化机制, 通过 Agent 模态变迁进行模式切换, 以适应快速突变环境. 在控制 Agent 中采用 Actor-critic 强化学习方法, 运用 TS 回归模糊神经网络实现行动和评判模块, 使用分布式学习算法对多个 Agent 协调优化. 工程应用表明, 提出的控制策略有效地解决了高压氨水大干扰对集气管压力的冲击控制问题.

关键词: 焦炉集气管; 梯级协调; 多智能体; 强化学习; TS 回归模糊神经网络

中图分类号: TP273, TP18 **文献标识码:** A **文章编号:** 0372-2112 (2006) 10-1847-05

Multi-Level Coordination Control Based on Multi-Agent Reinforcement Learning for the Pressure of Gas Collectors of Coke Ovens

QIN Bin^{1,2}, WU Min¹, WANG Xin^{1,2}, YANG Chun-hua¹

(1. School of Information Science & Engineering, Central South University, Changsha, Hunan 410083, China;

2. School of Electrical Engineering of Hunan University of Technology, Zhuzhou, Hunan 412008, China)

Abstract: For the multi-variable nonlinear coupled system with strong disturbance such as the gas pressure of collectors of coke ovens, this paper proposes an intelligent multi-level coordinated control strategy based on multi-agent system. It adopts the multi-level coordination architecture with agent agency and the organization and evolution mechanism based on task decomposition. The system can be switched to different modes using the state change of agents in order to operate in rapidly time-varying environments. The reinforcement learning method is used in Agent learning, the TS type recurrent fuzzy neural network (TSRFNN) is employed to realize the actor-critic elements. The agents in system are optimized coordinately by using the distributed learning algorithm. The real-world application shows that the proposed control strategy has successfully solved the process coordination control problem of the gas pressure of collectors of coke ovens with the strong disturbance produced by high pressure ammonia.

Key words: gas collectors of coke ovens; cascade coordination; multi-Agent system; reinforcement learning; TS recurrent fuzzy neural network

1 引言

焦炉集气管压力控制是焦炉控制的关键参数之一. 稳定焦炉集气管压力, 对于改善环境、提高煤气回收量和质量、提高焦炉辅助产品的产量和质量、延长焦炉寿命, 具有重要的意义. 焦炉集气管被控对象是一个多扰动、变化大、耦合严重的分布式非线性系统, 经过多年的研究和探索, 基于专家规则和前馈解耦的综合控制算法取得较好的控制效果^[1]. 随着环境保护意识增强和法规建立, 在装煤期间, 采用高压氨水产生吸力以防止煤气外冒, 由于设备密封不严等原因, 往往对集气管压力产生高达几百帕的巨大冲击, 传统的自适应解耦难以适

应对对象参数快速变化情况, 而基于专家规则和前馈解耦控制方案存在规则获取困难, 难以适应时变和不确定性对象等缺点^[2], 因此需要寻找新的解决方案.

分布式人工智能研究的兴起与发展, 特别是 MAS 的发展, 为解决复杂工业过程智能控制提供了新的思路. 本文介绍一个基于 MAS 的焦炉集气管压力智能控制系统, 系统采用分层结构, 根据任务的分解, 组织成多 Agent 组系统, 形成梯级协调体系. 在控制 Agent 中采用强化学习结构, 在 actor-critic 离散框架的基础上, 引入回归模糊神经网络进行函数逼近. 运用定量计算和定性分析相结合的混合决策方法, 进行 Agent 模式切换, 有效地解决了集气管压力这类复杂对象的过程控

收稿日期: 2005-12-20; 修回日期: 2006-05-16

基金项目: 国家杰出青年科学基金 (No. 60425310); 湖南教育厅资助项目 (No. 04C718); 中国包装总公司重点科研项目 (No. 2005-83)

制问题.

2 工艺简介及控制要求

焦炉集气管系统的结构如图 1 所示. 焦炉煤气从各炭化室通过上升管被循环氨气冷却到 80~90, 然后进入集气管, 通过各自的蝶阀进入初冷器前主管道. 焦炉煤气从焦炉到初冷器分为两个独立系统, 即 1 号和 2 号焦炉为一个系统, 3 号和 4 号焦炉为一个系统. 通过初冷器被冷却到 35~40, 通过横管蝶阀进入鼓风机然后由鼓风机被送往下道工序.

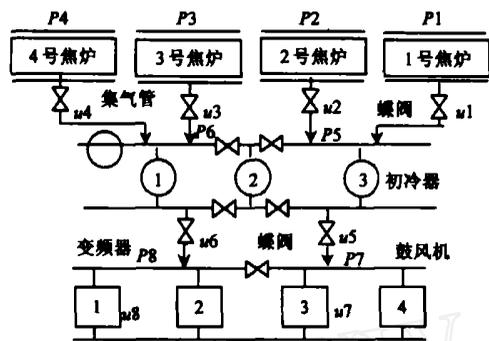


图 1 焦炉集气管工艺结构

焦炉集气管系统的主要控制要求将集气管压力稳定在 80~120 Pa 的范围内, 初冷器和鼓风机前压力满足安全运行要求. 影响集气管压力的因素很多, 如焦炉之间及与鼓风机相互间严重的耦合关系, 焦炉煤气生成量和用户使用量大范围变化与不平衡, 装煤推焦及高压氨水扰动等, 特别是在装煤期间, 高压氨水往往对集气管压力产生高达几百帕的巨大冲击, 而装煤完成后关闭高压氨水又会产生负压. 在高压氨水冲击下, 常规方法无法将集气管压力维持在工艺要求的范围之内. 因此如何协调各个控制部分使集气管压力解耦并维持稳定成为控制的关键. 控制手段主要有每座焦炉上集气管的蝶阀 ($u_1 \sim u_4$)、横管上的蝶阀 ($u_5 \sim u_6$) 和鼓风机变频调速 ($u_7 \sim u_8$), 其中鼓风机运行 2 台, 备用 2 台.

3 多级协调控制结构

根据 MAS 系统常用的任务分解思想, 首先将集气管压力控制问题分解成为三个部分: 焦炉集气管智能控制单元, 初冷器后横管智能控制单元和鼓风机智能控制单元, 如图 2 所示. 每座焦炉集气管本地控制 Agent 组负责两座焦炉集气管压力本地控制, 通过控制集气管收集管上的蝶阀保持集气管压力的稳定, 通过相互协作解决它们之间的耦合问题. 由横管初冷器压力控制 Agent 保证压力在安全范围内, 并根据集气管压力情况, 控制初冷器前压力处于最优状态, 使前一级焦炉集气管本地控制工作在有效范围内. 鼓风机变频调速控制 Agent 组采用相同的控制原理, 其主要目标是保证鼓风机的安全运行, 并与横管初冷器控制 Agent 组一道协调控制, 在更大范围内稳定工况, 焦炉集气管本地控制、横管蝶阀和鼓风机变频调速形成梯级协调体系, 调节范围逐步加大, 由后级保证前级的工作在有效范围内.



图 2 焦炉集气管压力、初冷器前压力和鼓风机前压力三级协调体系

4 多级协调控制算法

4.1 基于 Actor-critic 的 Agent 强化学习

通常, Agent 把一种状态表示映射为一个适当的动作或一组动作的概率分布, 这种映射称之为策略. 虽然函数逼近 Q 学习算法在强化学习中得到广泛的应用, 但是不能保证收敛, 此外当行动空间很大甚至无穷时, 需要解决每一步估计 Q 值的非线性规划问题, 这限制了该算法在实时控制中的应用.

自适应启发评判 (AHC) [3] 提供了一种试图同时找到最优动作和期望值的方式. 典型的自适应启发评判由两个基本模块组成: 评判模块 (ACE) 和动作模块 (ASE), 如图 3 所示.

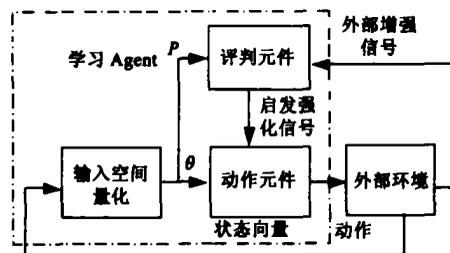


图 3 自适应启发评判结构

由状态向量和外界或环境提供的外部强化信号作为输入, 对于连续空间由输入空间量化成确定状态空间, 评判生成价值或评价的估计, 将状态映射为期望的价值. 评判模块采用 TD 方法进行更新, 从而产生自适应评判. 动作模块产生一个可能合理的动作, 将状态映射为期望行动, 实际行动输出按期望行动的 Boltzman 概率分布进行搜索. 在内部强化信号的指导下, 动作模块试图学习最优控制或决策能力.

在本项目中, 我们采用 Konda 和 Tsitsiklis 提出的 actor-critic 离散框架, 已经证明在满足一定的条件下行动元件参数收敛到局部优化值 [4, 5]. AC 算法具有两种不同的形式, 相对应于 TD() 评判和 TD(1) 评判, 在两种形式中, 评判是实际 Q 函数对行动特征空间投影的线性参数化逼近结构:

$$Q_p(s_{t+1}, a) = \sum_{j=1}^m p_j^t \frac{\partial}{\partial j} \ln(s, a) \quad (1)$$

其中 $p = (p^1, \dots, p^m)$ 为评判参数向量, $a = (a^1, \dots, a^m)$ 为行动参数向量, (s, a) 为在 (s, a) (参数策略和状态 s 下采取行动 a 的概率. 在这里评判元件结构完全取决于行动元件的结构, 因此只需要确定一个独立的函数逼近结构. 参数向量 p 按下式

进行更新:

$$p_{t+1} = p_t + \alpha z_t (r_t + \max_a Q(s_{t+1}, a, s_t) - Q(s_t, a, s_t)) \quad (2)$$

其中 α 为评判学习率, z_t 为特征向量, r_t 为瞬时奖赏, β 为折扣率, 这里取 $\beta = 0.9$. TD(1) 评判按下式更新 z_t :

$$z_{t+1} = \begin{cases} z_t + \nabla \ln \mu_{\beta}(s_{t+1}, a_{t+1}) & s_t = s_0 \text{ or } a_t = a_0 \\ \nabla \ln \mu_{\beta}(s_{t+1}, a_{t+1}) & \text{otherwise} \end{cases} \quad (3)$$

式中 ∇ 为梯度运算符, TD(1) 按下式更新 z_t :

$$z_{t+1} = z_t + \nabla \ln \mu_{\beta}(s_{t+1}, a_{t+1}) \quad (4)$$

行动元件根据下式更新参数:

$$p_{t+1} = p_t + \alpha Q_{\beta}^r(s_{t+1}, a_{t+1}) \nabla \ln \mu_{\beta}(a_{t+1} | s_{t+1}) \quad (5)$$

针对 MAS 控制系统在线自组织的特点, 本文采用 TS 回归模糊推理系统来构造行动元件(控制器算法)和评判元件(估计行动 Q 值), 行动元件和评判元件采用相同的推理结构. 采用 TS 回归模糊推理系统的最大特点是不需要系统的动态阶次信息, 在学习中系统能够在线自组织, 不断优化结构和参数, 学习后的规则易于理解和判定, 并且操作经验和专家知识能嵌入到模糊系统的初始规则库中, 作为初始模型进行学习提炼, 从而加快学习速度. 所采用的 TS 回归模糊推理模型可用如下模糊规则描述^[6]: (为简单起见, 采用 2 输入 1 输出系统)

Rule1 IF $x_1(t)$ is A_{11} and $x_2(t)$ is A_{12} and $h_1(t)$ is G_{11} and $h_2(t)$ is G_{12}
THEN $y(t+1)$ is $a_{01} + a_{11}x_1(t) + a_{21}x_2(t) + a_{31}h_1(t)$ and $h_1(t+1)$ is w_{11} and $h_2(t+1)$ is w_{21}

该模型可用图 4 所示的 6 层 TS 回归模糊神经网络来表示, 除 5 层外, 层 1-6 与一般的模糊神经网络相同, 第 1 层为输入节点, 第 2 层为模糊量化层, 采用两种隶属函数, 对于外部变量 x 采用高斯隶属函数, 对于内部反馈变量 h 采用全局 sigmoid 函数, 第 3 层为规则层, 该层输出由模糊 AND 操作确定, 在本系统中产生的输出为每个规则的激励强度

$$O_i^{(3)} = \prod_{j=1}^{n+1} O_j^{(2)} = b_i \exp \left\{ - \sum_{j=1}^{n+1} \frac{(x_j - c_{ij})^2}{2 \sigma_{ij}^2} \right\} \quad (6)$$

其中:

$$b_i = \frac{1}{1 + \exp(-O_i^{(5)})}$$

c_{ij} 和 σ_{ij} 分别为第 j 个外部输入第 i 项高斯隶属函数的中心和宽度.

第 4 层输出为输入的线性组合, 实现模糊推理规则的结论部分:

$$O_i^{(4)} = \sum_{j=0}^{n+1} a_{ij} u_j^{(4)} = \bar{a}_i = a_{0i} + \sum_{j=1}^n a_{ij} x_j + a_{i, n+1} h_i \quad (7)$$

第 5 层为前后关系结点, 该层相当于规则层推理输出的反模糊化, 连接权重代表内部规则的单值输出部分, 延迟值反馈到第 1 层.

$$O_i^{(5)} = h_i = \sum_{j=1}^r O_i^{(3)} w_{ij} \quad (8)$$

每一个规则都有一个内部变量, 用来确定过去历史对当前规则的影响程度.

第 6 层根据规则层的激励程度输出对第 4 层的输出进行加权运算, 最后得到系统输出.

$$y = O_i^{(6)} = \frac{\sum_{j=1}^r O_i^{(3)} O_i^{(4)}}{O_i^{(3)}} \quad (9)$$

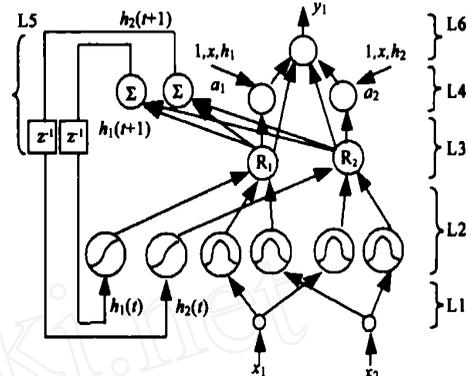


图 4 TS 类型回归模糊神经网络 (TSRFNN)

4.2 基于强化学习多级协调

多 Agent 强化学习问题可用随机竞赛框架来描述, 用数组 $(n, S, A_1, \dots, A_n, \dots, r_1, \dots, r_n)$ 表示, 其中 S 为环境状态集, A_i 为 Agent i 能得到的行动集, A 为联合行动空间 $A = A_1 \times \dots \times A_n$. $r_i: S \times A \rightarrow R$ 为 Agent i 的回报函数, $T: S \times A \rightarrow PD(S)$ 为状态传递函数, 将当前状态和 Agent 的联合行动映射为在状态空间概率分布集.

每个 Agent 的目标是找到一个策略 $\pi_i: S \rightarrow PD(A_i)$, 将联合状态映射为单独行动空间的概率分布, 来最大化该 Agent 在某一时间段获得的平均或折扣奖赏. 非零和对策中, 多个 Agent 可以在协作中获得比独立行动所获得的最大赢得更多的利益, 并且 Nash 平衡点在每个 Agent 都不能得到更优的行动策略时达到. Hu 和 Wellman^[7] 提出了 Nash-Q 算法, 并证明 Nash-Q 学习算法收敛到平衡点的 Q 值, 即对策值. 在本研究中, 由于可以通过通信获得各个 Agent 观察到的状态, 我们采用直接协调的方法, 将联合状态通过回归模糊网络直接映射到单个 Agent 行动空间.

4.2.1 单个 Agent 的全局奖赏

取期望解耦输出(解耦参考模型输出)与学习后的对象模型输出(根据需要也可能是实际输出)之差的平方和作为系统的性能指标函数, 表达式为:

$$e_i = \frac{1}{2} \sum_{k=1}^m (y_{d_i}^k(t) - \hat{y}_i^k(t))^2 \quad (10)$$

$i = 1, 2, 3, 4$ 分别对应两座焦炉集气管压力、初冷器前压力和鼓风机前压力控制性能指标, m 为控制周期内的采样数. 不考虑团队行为时 Agent 获得局部奖赏取性能指标的倒数 $\beta_i = 1/e_i$, 采取行动后的一步时间 Agent j 获取的奖赏取为团体和个体奖赏的凸结合:

$$r_i = \beta_i + (1 - \beta_i) \frac{1}{K-1} \left(\sum_{k=1}^K \beta_k \right) \quad (11)$$

其中 $K=4$ 为参与联合行动 Agent 的个数, i 为平衡参数, $i=1$ 对应于完全自私行为, $i=0$ 对应于完全团队意志行为. 根据控制要求取 $i_1 = i_2 = 0.9, i_3 = 0.3, i_4 = 0.2$.

4.2.2 回归模糊网络结构确定

回归模糊网络的输入为所有 Agent 观察到的状态, 在这里为各级压力误差值, 采用高斯隶属函数, 由于采用图 4 结构的回归网络因而不需要压力高阶变化率输入. 结论部分输出有两个节点, 分别是期望行动输出和 Q 值估计输出, 相应的可调参数集为 \bar{a} 和 P, 为减少计算量, 在本项目中采用 Actor-critic 算法调整输出部分参数. 采用聚类方法离线或在线确定模型结构, 结构学习可分为外部规则和内部规则学习两部分, 外部规则学习采用进化聚类算法 (ECA), 其主要思想是动态估计数据集的聚类数, 确定输入数据空间的当前聚类中心. 一般地规则层的每个规则都具有其相应的内部变量, 但如果规则具有相同的结论部分 (如连接到相同输出节点) 则相同的内部变量赋给这些规则, 从而有效减少参数数量. 初始权重采用 $[-1, 1]$ 区间的随机数, 通过经验数据监督学习可得到初始规则, 避免陷入局部最小和加快收敛速度.

根据回归模糊模型, 我们可得到 actor 元件的行动输出:

$$a(s) = f(s) = \frac{\sum_{i=1}^M \bar{a}_i b_i \exp\left[-\sum_{j=1}^n \frac{(s_j - c_{ij})^2}{2\sigma_{ij}^2}\right]}{\sum_{i=1}^M b_i \exp\left[-\sum_{j=1}^n \frac{(s_j - c_{ij})^2}{2\sigma_{ij}^2}\right]} \quad (12)$$

actor-critic 算法需要确定行动在整个行动空间的概率分布, 在本算法中实际行动 a 为以 $a(s)$ 为均值、 σ 为方差的高斯概率分布搜索, 因此在状态 s 采取行动 a 的概率分布为:

$$p(s, a) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\bar{a}_i - a(s))^2}{2\sigma^2}\right] \quad (13)$$

结构确定后将可调参数构成参数集 \bar{a} 和 P, 按式 (2) 和 (5) 进行更新. 如果需要调整前提部分参数可用基于 TD 误差的实时回归网络算法进行调整.

$$r_{t+1} = r_t + \max_a Q(s_{t+1}, a, i_t) - Q(s_t, a, i_t) \quad (14)$$

$$E_{t+1} = \frac{1}{2} \frac{r_{t+1}}{i_{t+1}} \quad (15)$$

如对于高斯隶属函数中心参数 c_{pq} 有^[6]:

$$c_{pq, t+1} = c_{pq, t} - \frac{\partial E_{t+1}}{\partial c_{pq, t}} \quad (16)$$

其中:

$$\frac{\partial E_{t+1}}{\partial c_{ji, t}} = \frac{1}{i_{t+1}} \left[\frac{O_p^{(4)} - a}{r} * \frac{\partial O_p^{(3)}}{\partial c_{pq}} + \sum_{i=1}^r O_i^{(3)} a_{in+1} \frac{\partial O^{(5)}}{\partial c_{pq}} \sum_{i=1}^r O_i^{(3)} \right]$$

$$\frac{\partial O_p^{(3)}}{\partial c_{pq}} = \frac{\partial O_p^{(5)}}{\partial c_{pq}} \cdot F^p + O_p^{(5)} \cdot F^p \cdot 2 \frac{x_{p, t} - c_{pq}}{2\sigma_{pq}^2}$$

$$\frac{\partial O_p^{(5)}}{\partial c_{pq}} = O_p^{(5)} (1 - O_p^{(5)})$$

$$\left\{ \sum_{i=1}^r W_{pi} \left[\frac{\partial O_{p, t-1}^{(5)}}{\partial c_{pq}} F_{i-1}^l + F_{i-1}^r \cdot O_{q, t-1}^{(5)} \cdot 2 \frac{x_{p, t-1} - c_{pq}}{2\sigma_{pq}^2} \right] \right\}$$

$$F^i = \sum_{k=1}^n O_k^2 = \exp\left[-\sum_{k=1}^n \frac{(x_j - c_{ij})^2}{2\sigma_{ij}^2}\right] \quad [0, 1]$$

为加快收敛速度采用 Delta-Bar-Delta 规则确定学习率^[8].

在实际运行中, 首先根据专家和熟练操作工人的经验学习构成初始规则, 将多个 Agent 观察的状态组成联合状态, 按行动网络产生行动, 并在模型上执行行动, 如评估风险较低, 则向实际对象输出行动. 计算各 Agent 获得的局部奖赏和全局奖赏, 更新参数向量 P 和 \bar{a} . 如向实际对象输出行动, 获得的实际经验可用来更新对象模型.

4.3 控制 Agent 的模式切换

结合一定的先验知识, 我们采用定量计算和定性分析相结合的混合决策方案, 设多个控制 Agent 对被控对象的辨识模型为 M_j , 相应的控制算法为 C_j , 模型辨识误差为 e_j , 选择指标函数为^[9]:

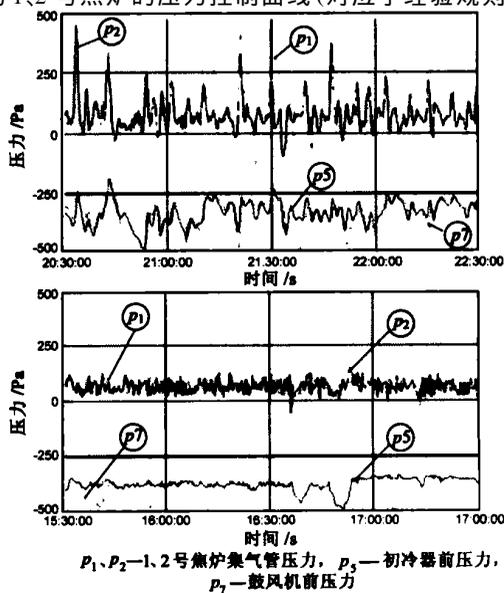
$$J_i(t) = e_j^2(t) + \int_0^t e^{-\alpha(t-\tau)} e_j^2(\tau) d\tau, \quad 0 \leq t \leq T, \quad \alpha > 0 \quad (17)$$

式中 α 和被选择作为即时和长期精度测量的一个理想结合, 遗忘因子 α 决定变迁环境下指标的内存和保证 $J_i(t)$ 有界. 定量决策主要是根据指标函数, 判断当前的激活 Agent 是否需要变迁和转换到哪一个 Agent 所决定的模态上去, 选择指标函数最小的 Agent 作为切换对象.

5 系统实现

根据上述方案实现的焦炉集气管压力集散控制系统, 分别由 1 号炉 SIEMENS S7400 PLC 系统, 3、4 号炉 AB PLC 系统, 鼓风机和冷却回收系统 SIEMENS S7300 PLC 及智能变送器和智能执行器构成. 各 PLC 控制器分别实现本地 Agent 模糊协调解耦控制, 考虑到运算量较大, 对象模型和各 Agent 参数的协调学习由监控站学习 Agent 完成. 应用 JADE 平台开发上层 Agent^[2], 采用 OPC (OLE for Process Control) 接口来实现控制 Agent 与上层 Agent 间的信息交互.

现场运行的历史趋势曲线如图 5 所示. 图中上方为学习前的 1、2 号焦炉的压力控制曲线 (对应于经验规则专家控



P_1, P_2 —1、2号焦炉集气管压力, P_5 —初冷器前压力, P_7 —鼓风机前压力

图 5 实际运行曲线

制),下方为学习后压力控制曲线,由图 5 可知,学习后当出现高压氨水扰动时,通过多 Agent 的协调控制,基本能在 2 分钟内快速调节达到设定值附近,高压氨水扰动对焦炉集气管压力的影响较学习前明显削弱,具有良好的协调解耦性能,保证焦炉集气管压力稳定在工艺要求的范围内.3、4 号与 1、2 号可得到同样的结论.

6 结论

焦炉集气管压力是一个多变量耦合、非线性严重的复杂被控对象,特别是高压氨水的强冲击干扰,应用常规控制方法很难实现稳定的自动控制,本系统提出了基于 MAS 的多级协调控制系统,形成集气管、初冷器和鼓风机梯级协调体系,通过 Agent 的模式变迁实现运行模式的切换,以适应快速变化的环境,相互协调合作完成大冲击干扰下的稳定控制.在控制 Agent 中,以回归模糊网络为核心,实现对象模型估计、Actor-critic 强化学习功能,利用对象模型有利于加速强化学习过程,减少探索风险.实际运行结果表明系统设计是成功的.

参考文献:

- [1] Yang Cunhua, Wu Min, Shen Deyao, et al. Hybrid intelligent control of gas collectors of coke ovens[J]. Control Engineering Practice 2001, 9(7): 725 - 733.
- [2] 秦斌,吴敏,王欣,等.基于 MAS 的分布式智能控制系统原型系统开发与应用[J].小型微型计算机系统,2006,27(7):1405 - 1408.
Qin Bin, Wu min, Wang Xin, et al. Development and application of distributed integrated intelligent control system based on multi-Agent system[J]. Mini-Micro System, 2006, 27(7): 1405 - 1408. (in Chinese)
- [3] Barto A G, Sutton R S, Anderson C W. Neural-like adaptive elements that can solve difficult learning control problem[J]. IEEE Trans on Systems, Man, and Cybern, 1983. 13(5): 834 - 846.
- [4] Konda V R, Tsitsiklis J N. Actor-critic algorithms [A]. Advances in Neural Information Processing Systems [C]. Cambridge MA: MIT Press, 1999, 12(11): 1008 - 1014.
- [5] Berenji H R, Vengerov D. On convergence of fuzzy reinforcement learning[A]. In proceedings of the 10th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE '01) [C]. Melbourne Australia, 2001. 618 - 621.
- [6] Juang Chia Feng. A TSK-type recurrent fuzzy network for dynamic systems processing by neural network and genetic algorithms[J]. IEEE Trans on Fuzzy Systems, 2002, 10(2): 155 - 170.
- [7] Hu Jun ling, Welhaan M P. Multi-gent reinforcement learning: Theoretical framework and an algorithm[A]. proceedings of the 15th International Conference of Machme Learning [C]. Madison Wisconsin, 1998. 115 - 122.
- [8] Jacobs R A. Increased rates of convergence through learning rate adaptation[J]. Neural Networks, 1988, 1(4): 295 - 307.
- [9] Kumpati S, Balakrishnan N J. Adaptive control using multiple models[J]. IEEE Transactions on Automatic Control, 1997, 42(2): 171 - 187.

作者简介:



秦 斌 男,教授,1963 年 1 月出生于河北承德,1988 在中南大学获工学硕士,现为湖南工业大学电气学院测控技术研究所副所长,主要研究领域为复杂系统建模与优化控制、多智能体应用、电力系统优化.

E-mail: qinbin99@163.com



吴 敏 男,教授,1963 年出生于广东化州,1986 年毕业于中南工业大学自动化专业,获硕士学位,1999 年获日本东京大学博士学位,主要研究方向为鲁棒控制、非线性控制和过程控制.



王 欣 女,副教授,博士生,1971 年 8 月出生于湖南株洲,1999 年、2004 年分别在国防科技大学、中南林业大学获工学学士、工学硕士学位,现为中南大学信息科学与工程学院博士生,主要研究方向为复杂工业过程建模与智能控制、机器学习、优化调度.