Advances in Energy and Power Engineering 电力与能源进展, 2026, 14(2), 103-113
Published Online April 2026 in Hans. https://www.hanspub.org/journal/aepe
https://doi.org/10.12677/aepe.2026.142012

Hans汉斯

# 基于强化学习的再热汽温系统优化控制

**郭子逸**

华北电力大学控制与计算机工程学院，北京

收稿日期：2026年2月9日；录用日期：2026年2月21日；发布日期：2026年4月28日

## 摘　要

**针对具有传统控制策略面对再热汽温这种大惯性、大迟延系统时控制性能不足的问题，提出了一种基于DQN前馈与PID复合控制的策略，以提高系统的控制精度。在控制过程中，根据系统状态选择最优动作对PID控制器输出进行补偿。并且针对强化学习面对迟延对象训练难以收敛的问题，设计多维奖励函数和迟延缓冲区。结果表明：所设计的控制方法在超调量、调节时间、鲁棒性方面均优于传统PID控制，为再热汽温的高效稳定控制提供了新方案。**

## 关键词

**再热蒸汽温度，强化学习，DQN算法，PID控制**

# Reheat Steam Temperature System Optimization Control Based on Reinforcement Learning

**Ziyi Guo**

School of Control and Computer Engineering, North China Electric Power University, Beijing

Received: February 9, 2026; accepted: February 21, 2026; published: April 28, 2026

## Abstract

**Aiming at the problem of insufficient control performance of traditional control strategies when dealing with the large inertia and large delay system of reheated steam temperature, a strategy based on DQN (Deep Q Network) feedforward and PID compound control is proposed to improve the control accuracy of the system. During the control process, the optimal action is selected according to the system state to compensate the output of the PID controller. Moreover, to address the issue of difficulty in convergence during the training of reinforcement learning for delay objects, a**

文章引用: 郭子逸. 基于强化学习的再热汽温系统优化控制[J]. 电力与能源进展, 2026, 14(2): 103-113.
DOI: 10.12677/aepe.2026.142012

**multi-dimensional reward function and delay buffer are designed. The results show that the designed control method outperforms traditional PID control in terms of overshoot, regulation time, and robustness, providing a new solution for the efficient and stable control of reheated steam temperature.**

## Keywords

**Reheated Steam Temperature, Reinforcement Learning, DQN Algorithm, PID Control**



## 1. 引言

在“双碳”目标的背景下，我国正在建立以新能源为主体的新型电力系统，风能、太阳能等时变特性强烈的可再生能源发电装机规模持续快速增长。燃煤发电作为我国电力供应安全的“压舱石”，逐渐从电力供应的主体向支撑性、调节性电源转变。提升机组运行灵活性已成为燃煤发电面临的最为迫切的技术需求[1] [2]。为了消纳可再生能源的发电量和用户侧的负荷变化，燃煤机组需要频繁参与一次调频和深度调峰，大大增加了锅炉主汽压力、蒸汽温度等运行参数的调控难度[3] [4]。

再热汽温作为锅炉机组的关键控制参数之一，主要采用烟气挡板调节、燃烧器摆角调节和喷水减温等调节方式。但是大量的减温喷水会降低机组的经济性，所以将喷水减温作为超温时的应急调整手段[5]。尽管烟气侧挡板和燃烧器摆角调节结构相对简单，但其固有的大滞后及非线性特征使得其现场自动投运率低，且调节不当易引起烟气挡板和喷水阀两个控制回路的失调震荡，并且频繁动作易损坏燃烧器摆角执行机构，严重影响再热汽温控制效果和设备寿命[6]。因此，研究先进的再热汽温优化控制方案，对火电机组安全经济运行具有重要意义。

针对这一问题，越来越多的科研人员进行了研究，王东风等人[7]使用燃烧器摆角与喷水减温作为调节手段，引入广义预测控制方法(GPC)作为控制策略，大幅提高了系统性能。赵东华等人[8]提出了一种基于综合加权多模型的改进预测函数控制算法，通过综合所有子控制器的综合加权系数得到最终输出，有效适应负荷变化。王焕敏等[9]和赵征等[10]通过强化学习实现脱硝系统优化控制中，显著提高了系统的调节性能和稳定性。Xie P 等人[11]通过贝叶斯神经网络监督器规划智能体的动作，克服了传统免模型强化学习算法无法利用历史数据的不足。设计了强化学习智能体脱硝控制器。减少还原剂的使用量，从而降低机组运行成本。

基于目前的研究现状，本文针对某 600 MW 超临界机组再热气温系统，利用海量实际运行的历史数据建立再热汽温模型，并设计了强化学习前馈控制器与 PID 控制器的复合控制策略，以建立的再热汽温模型为基础，对强化学习前馈控制器的训练，针对训练过程中存在的收敛困难和时滞敏感性问题，本文对 DQN 算法做出了以下针对性改进：首先设计了迟延缓冲区机制，有效抑制系统大迟延特性的影响。其次，构建多维状态观测空间并设计分段奖励函数，显著优化了策略的训练效率与稳定性。

## 2. 再热汽温系统特性分析与建模

### 2.1. 再热蒸汽温度控制方式

再热汽温受到锅炉烟气侧和蒸汽侧两方面的影响，喷水减温因其降低机组运行经济性的特点，通常

用于事故喷水以及负荷变化剧烈等危急情况。燃烧器通过改变燃烧器摆角来改变炉膛内火焰中心的位置，从而改变炉膛的出口烟温，这种控温方式想对烟气挡板来说，响应快，迟延小，能量损耗小，所以本文采用改变燃烧器摆角大小的方案控制机组再热汽温。

根据设定值与再热汽温之间的偏差，通过调节合理的 PID 控制器参数，通过调节燃烧器摆角来控制炉膛出口烟气温度，从而达到控制再热汽温的目的。

## 2.2. 基于粒子群算法的模型辨识

控制系统的设计是以被控对象的模型为基础设计的，只有准确描述被控对象的数学模型，才能设计出准确的控制器。本文利用粒子群算法辨识出精确的燃烧器摆角到再热汽温的传递函数模型，利用该模型完成控制器的设计并验证控制器性能，并且利用辨识出的负荷侧和燃料量侧的模型作为系统的扰动通道。

再热汽温系统是一个复杂控制系统，为了保证辨识模型的准确性，需要建立一个精确的数学模型，用于系统建模的数学模型很多，在电厂中常用传递函数模型，大多数热工对象均具有自平衡能力，一般为多阶惯性环节[12]，其传递函数如式(1)：

$$G(s)=\frac{K}{(T_1s+1)(T_2s+1)\cdots(T_ns+1)}e^{-\tau s} \tag{1}$$

式中：$S$——拉氏微分算子；$K$——放大系数；$T_n$ $(n=1,2,\ldots)$——时间常数；$\tau$——延迟时间。

该模型需辨识的参数数量较多，为了降低辨识的复杂度，可采用更简单的传递函数结构。建立其模型结构如式(2)：

$$G(s)=\frac{Ke^{-\tau s}}{(T_1S+1)^n} \tag{2}$$

式中：$S$——拉氏微分算子；$K$——放大系数；$T_1$——时间常数；$\tau$——延迟时间，$n$ 为系统阶数。

本文从电厂历史运行数据中选取所需数据进行分析处理，由于电厂热工过程的复杂性，实际数据中通常存在多种噪声干扰，为了防止噪声对模型辨识精度的干扰，需要对数据进行零初始化、滤波和粗大值剔除等处理，通过主成分分析法对影响再热汽温的因素进行筛选，确定建立再热汽温系统模型的主要输入量有 4 个，分别是负荷、燃料量、燃烧器摆角、减温水流量。对应的结构图如图 1 所示：

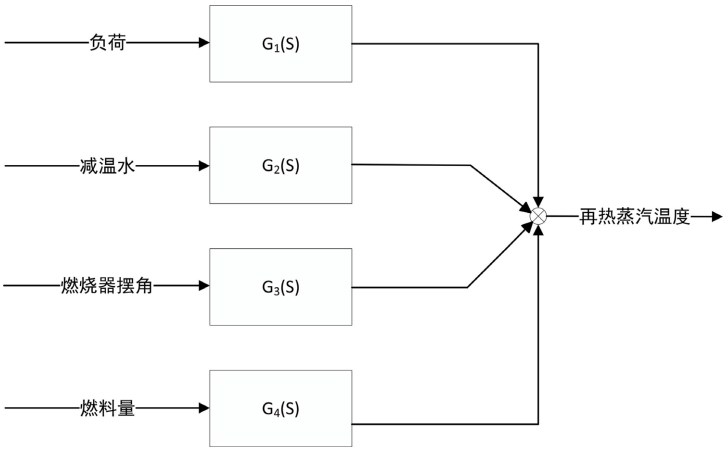


**Figure 1.** Model structure diagram of the reheated steam temperature system
**图 1.** 再热汽温系统的模型结构图

基于处理后的现场数据，使用粒子群算法辨识传递函数参数，实现再热汽温多变量系统建模，系统主要研究再热器出口温度，辨识出的控制通道传递函数如式(3)：

$$G(s)=\frac{0.01e^{-40s}}{(50S+1)^2} \tag{3}$$

辨识出的负荷侧传递函数模型如式(4)所示，燃料量侧的传递函数模型分别为式(5)所示。

$$G(s)=\frac{0.38e^{-80s}}{(10S+1)^2} \tag{4}$$

$$G(s)=\frac{0.15e^{-38s}}{(5S+1)^3} \tag{5}$$

## 3. 基于 DQN 前馈的控制系统设计

### 3.1. DQN_PID 复合控制系统设计

传统再热汽温由 PID 控制，控制算法示意图如图 2 所示。PID 算法是控制系统中最经典的算法，由比例、积分和微分三部分组成，以系统误差作为输入，三部分共同作用，使系统能够快速稳定地响应设定值。

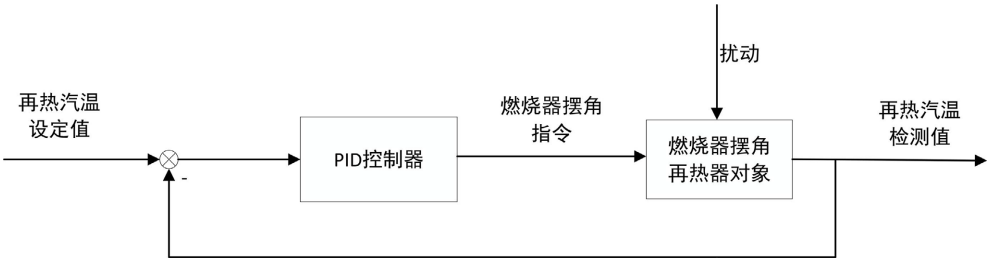


**Figure 2.** Reheat steam temperature control system
**图 2.** 再热汽温控制系统

然而燃煤机组再热汽温控制系统的大惯性大迟延的特点，传统的 PID 控制在面对机组大范围变负荷的情况，缺乏在线调整参数的能力，难以取得较好的控制效果。针对这个问题，本文引入了基于强化学习(DQN)的前馈控制器，设计了 DQN_PID 复合控制控制策略，控制流程图如图 3 所示：

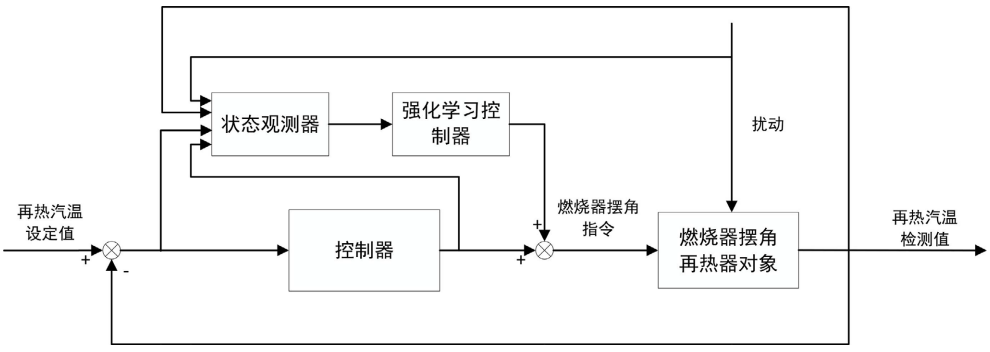


**Figure 3.** Composite control strategy
**图 3.** 复合控制控制策略

在控制过程中，PID 控制器与强化学习控制器一同工作，通过 PID 控制器实现系统的稳定，利用强化学习控制器优化系统的动态性能，提高系统的快速性，实现再热汽温的智能控制。

## 3.2. 强化学习控制器原理

由于整个强化学习是随着动作、回报和状态而更新并得到新的信息，因此可用马尔可夫决策过程(Markov Decision Process, MDP)对该过程进行表示。

### 3.2.1. 马尔可夫过程

马尔可夫性质是指下一个状态只取决于当前状态，不受过去的状态影响。一系列具有马尔科夫性质的随即状态组成的随机过程称为马尔可夫过程。通常由一个二元组表示$\langle S,P\rangle$，其中 $S$ 为有限状态集合，$P$ 为状态转移矩阵，定义了从当前状态到所有后继状态的转移概率，如式(6)所示：

$$P=\begin{bmatrix} P\left[S_1|S_1\right] & \cdots & P\left[S_n|S_1\right] \\ \vdots & \ddots & \vdots \\ P\left[S_1|S_n\right] & \cdots & P\left[S_n|S_n\right] \end{bmatrix} \tag{6}$$

矩阵 $P$ 中的元素如式(7)所示

$$P\left[s_i|s_j\right]=P\left[S_{t+1}=s_i|S_t=s_j\right] \tag{7}$$

### 3.2.2. 马尔可夫奖励过程

马尔可夫奖励过程是在马尔科夫链的基础上增加了奖励，加入奖励相当于给随机过程增加了导向，一个马尔可夫奖励过程可以用一个四元组表示马尔可夫奖励过程是一个元组$\langle S,P,R,\gamma\rangle$，其中：

$R$ 是一个奖励函数，$R_s=\mathbb{E}\left[R_{t+1}|S_t=s\right]$；

$\gamma$ 是一个折扣系数，$\gamma\in[0,1]$。

回报 $G_t$ 指的是从时间步 $t$ 开始累积的折扣奖励总和：

$$G_t=R_{t+1}+\gamma R_{t+2}+\gamma^2R_{t+3}+\cdots=\sum_{k=0}^{\infty}\gamma^kR_{t+k+1} \tag{8}$$

其中折扣系数 $\gamma$ 属于[0, 1]，当在第 $k+1$ 个时间步后获得奖励为 $\gamma^kR$，$\gamma$ 接近 0 会导致“近视”评估，$\gamma$ 接近 1 会导致“远视”评估。

有了回报的概念后，我们可以定义状态价值函数 $V(s)$，表示从状态 $s$ 出发，按照当前策略所能获得的期望回报：

$$\begin{aligned} v(s)&=\mathbb{E}\left[G_t|S_t=s\right] \\ &=\mathbb{E}\left[R_{t+1}+\gamma v\left(S_{t+1}\right)|S_t=s\right] \end{aligned} \tag{9}$$

### 3.2.3. 马尔可夫决策过程(MDP)

马尔科夫决策过程(MDP)引入了动作的概念，agent 通过做出动作来影响环境状态的转移。通常由一个五元组组成$\langle S,A,P,R,\gamma\rangle$。其中 $A$ 是有限动作的集合。MDP 的目标是找到一个最优策略 $\pi$，使得智能体在给定状态下选择最优的动作，从而最大化长期累积的折扣奖励。

策略 $\pi$ 是给定状态下动作的概率分布：

$$\pi(a|s)=\mathbb{P}\left[A_t=a|S_t=s\right] \tag{10}$$

策略完全定义了一个 agent 的行为，依照策略，我们可以得到动作价值函数，从状态 $s$ 开始依据策略采取动作的期望回报：

$$Q^{\pi}(s,a)=\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^{t}R(S_t,A_t)\middle|S_0=s,A_0=a\right] \tag{11}$$

用于指导智能体进行迭代训练，得到最优策略，原理如图 4 所示。

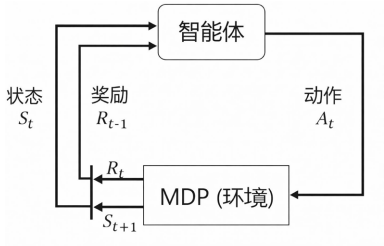


**Figure 4.** Reinforcement learning principle
**图 4.** 强化学习原理

### 3.3. DQN 算法

$Q$-learning 算法是强化学习中最为基础和重要的算法之一，其通过反映着状态和动作的价值关系构建 $Q$ 表格，通过查阅 $Q$ 表格获得在不同的状态下应该采取的动作，在智能体与环境的交互中更新 $Q$ 表格，$Q$ 值的更新方式如式(10)所示：

$$Q(s,a)\leftarrow Q(s,a)+\alpha\left[r+\gamma\max Q(s',a')-Q(s,a)\right] \tag{12}$$

式中：$Q(s,a)$是当前状态动作对$(s,a)$ $Q$ 值，$\alpha$ 是学习率，控制每次从新的状态信息中对 $Q$ 值更新的幅度，$r$ 表示在状态 $s$ 下执行动作 $a$ 后获得的即时奖励，$\gamma$ 是折扣因子，衡量未来奖励的重要性，max$Q(s',a')$表示在下一个状态 $s'$中所有可能动作中最大 $Q$ 值。

由于再热汽温的状态空间和燃烧器摆角的动作空间是连续的，如果使用 $Q$-Learning 算法会导致 $Q$ 表格出现维度爆炸的问题，导致算法收敛速度变慢或失效。DQN 算法将深度学习与 $Q$-Learning 算法结合，通过引入深度神经网络来近似 $q$ 值函数，从而处理了 $Q$-learning 算法面对高维状态空间时查找表存储和更新效率低下的问题[13]。

并且 DQN 算法通过引入经验回放(Experience Replay)和目标网络(Target Network)等技术，DQN 有效解决了深度学习训练过程中的数据相关性和不稳定性问题，使得智能体能够在动态环境中稳定地学习和优化策略。DQN 算法的训练框架图如图 5 所示。

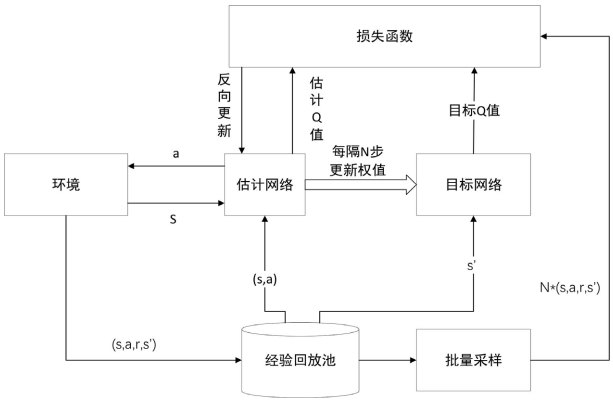


**Figure 5.** Schematic diagram of DQN framework
**图 5.** DQN 框架示意图

### 3.4. DQN_PID 复合控制策略

强化学习作为智能控制的重要方法之一，因其在智能控制方面表现出的优越性得到广泛应用。为实现火电机组的少人、无人值守运行提供借鉴。

#### 3.4.1. 状态观测空间

智能体通过与环境不断的交互获得最优控制策略，建立状态观测空间获取状态信息，当前状态信息是智能体对环境感知到的特征，通过感知到状态信息不断进行训练更新，更好地适应环境。从而不断地更新策略，实现最优控制。

本文设置的状态观测值有 7 个，分别为系统输出 y 与设定值之间的误差 error 的积分值，微分值、系统输出 y、PID 控制器的输出值，喷水扰动 D1，负荷扰动 D2。D1、D2 作为状态输入，在训练中对两个扰动采用归一化处理，采用不同扰动组合进行训练。

#### 3.4.2. 动作空间

DQN 算法一般是用来做离散空间的动作执行，从极限的理论看，增大离散动作空间的动作数量可以近似看作是连续的动作空间。其控制效果也趋近于连续动作区间的控制效果。为了使训练速度达到收敛，分析控制器输出，选择了合理的动作空间范围与维度。本文选取动作空间为[−10, 10]，步长为 0.1。

#### 3.4.3. 多维奖励函数

在强化学习中，奖励为智能体与环境交互的反馈，反应智能体在当前状态下的动作优劣，根据获得奖励来进行动作网络的更新，因此，奖励函数的设计是基于强化学习控制的重要部分，奖励函数设置的合理与否会直接影响控制算法的控制效果以及动作网络的能否收敛。

在再热汽温控制中，为实现再热汽温的平稳运行，设置误差奖励函数以及快速性奖励函数，在稳定的同时，追求一定系统响应的快速性。

1) 阶梯性误差奖励函数

$$R=\begin{cases}-5 & \text{error}<5\\ 1 & 1<\text{error}\le 5\\ 10 & \text{error}\le 1\end{cases} \tag{13}$$

2) 快速性奖励函数

在系统响应前 450 s 内设置快速性奖励

$$R=\begin{cases}-1 & \text{error}<20\\ 1 & 1<\text{error}\le 20\\ 10 & \text{error}\le 1\end{cases} \tag{14}$$

error 表示系统设定值与系统输出之间的偏差的绝对值，当误差大于 5%，误差过大，奖励为−5，当误差大于 1%且小于 5%时奖励为 1，当误差小于 1%获得奖励+10。针对大惯性大迟延系统设计快速性奖励，在系统输出前期误差小于 20 时获得奖励+1，误差小于 5 时获得奖励+5。并且设置提前终止当误差小于 0.1 并且连续保持时间大于等于 50s 时，完成控制目标，提前结束本回合训练，获得奖励+20。

3) 超调惩罚

$$R=\begin{cases}-1 & 10<\sigma\le 40, dt>0\\ -100 & \sigma>40, dt>0\end{cases} \tag{15}$$

$\sigma$ 为系统超调量，$dt$ 为输出曲线的斜率，当在系统上升阶段超调 $\sigma$ 小于 40%，奖励为−1，当超调 $\sigma$ 大于 40%时，奖励为−100，触发终止条件，终止训练。

#### 3.4.4. 建立迟延缓冲区

在面对带有迟延的被控对象时，强化学习控制器在训练时往往难以收敛，因为系统的迟延特性，动作执行后无法即时获得对应的奖惩反馈，导致动作与奖励之间的关联被割裂，造成了系统难以收敛，针对这个问题，本文设置了迟延缓冲区，首先通过模型辨识得到模型的迟延系数，设立合理的迟延补偿区间，将当前状态的动作与 $\tau$ 时刻后的奖励进行时序对齐，从而重建动作与反馈之间的因果关联，有效提升了控制器的训练效率和收敛稳定性。

#### 3.4.5. 动作网络设计

本文采用 MLP 网络结构，因 DQN 算法的特殊性，在训练中引入了经验回放机制，会打断数据间的关联性，若采用 RNN 结构的网络无法发挥其优势，所以本文采用 MLP 网络结构。本文设置了多维状态空间，其中系统输出 $y$ 与设定值之间的误差 error 总是在$[-2, 2]$之间，与 PID 控制器的输出以及误差的微分相差较大，为规避由此引发的训练震荡或收敛困难，本文在网络前端集成了标准化处理模块，消除量纲差异，确保各维度特征对损失函数的贡献均衡，从而优化控制器的训练效率。

## 4. 仿真结果

### 4.1. 实验参数设置

用辨识出的摆角值到再热气温的传递函数模型如上文式(3)所示，训练轮次设置为 40，每个训练轮次遍历不同扰动组合，若达到终止条件则提前结束训练，采用时间步长 $dt = 1$，折扣系数 $\gamma = 0.99$，优化器选择为 Adam，经验池大小为 10,000。

### 4.2. 实验仿真结果

为了验证本文所使用的复合控制策略的有效性，设计了两个实验，即引入强化学习前馈的控制策略与 PID 控制的控制策略的抗扰动实验和鲁棒性测试实验，与使用 PID 控制的再热汽温控制进行对比，根据两次实验结果中的性能指标来评价两种控制方案的优劣。

#### 4.2.1. 抗扰动实验

为了验证强化学习前馈的控制策略的抗扰动能力，将 D1 作为燃料量扰动，D2 作为负荷侧扰动，在初始阶段，将多种扰动组合作为扰动系统输入，扰动系统输出如图 6 所示：

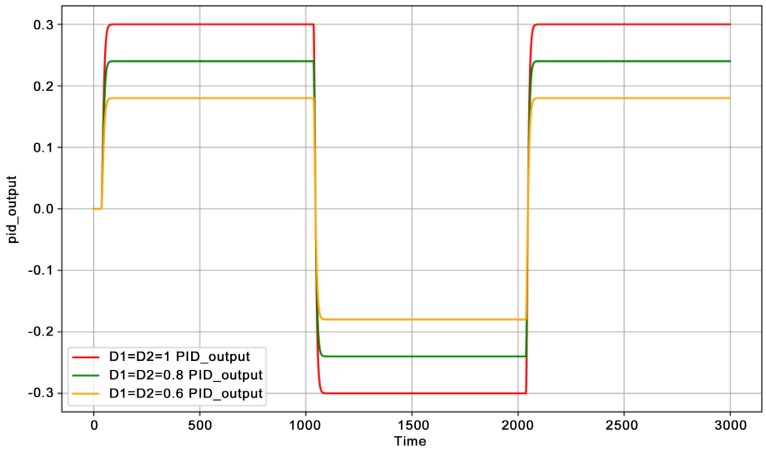


**Figure 6.** Output result of perturbed system
**图 6.** 扰动系统输出结果

将强化学习前馈的控制策略与 PID 控制策略进行对比，主系统输出结果如图 7 所示：

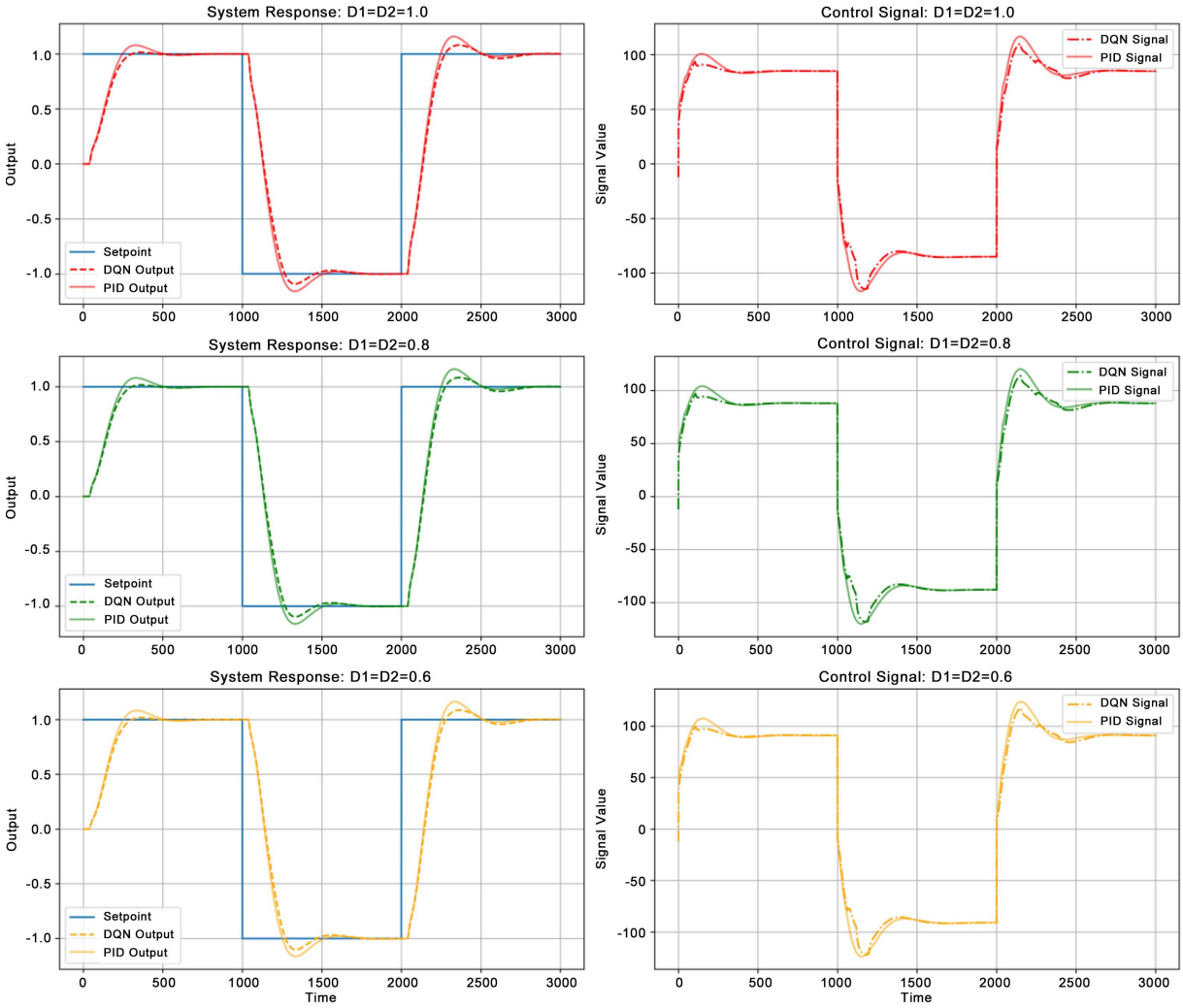


**Figure 7.** Anti-Interference test results
**图 7.** 抗干扰测试结果

引入不同的性能指标进行比较，包括超调量、调节时间、上升时间，ITAE，IAE，等多个控制性能指标。其中 IAE 主要反应系统的瞬态响应，ITAE 主要反映系统瞬态响应的振荡性能。相关性能指标从工况变化后的时间开始计算，计算结果为多个阶跃时间段内的平均值。结果如表 1 所示。

**Table 1.** Anti-Interference test performance index
**表 1.** 抗干扰测试性能指标

| 性能指标 | | 超调量/% | 调节时间/s | IAE | ITAE |
|---|---|---|---|---|---|
| D1 = 0.6<br>D2 = 0.6 | 传统 PID | 15.54 | 433 | 235 | 25,390 |
| | DQN-PID | 7.67 | 363 | 235 | 24,454 |
| D1 = 0.8<br>D2 = 0.8 | 传统 PID | 15.16 | 424 | 220 | 23,144 |
| | DQN-PID | 7.14 | 353 | 219 | 21,955 |
| D1 = 1.0<br>D2 = 1.0 | 传统 PID | 14.97 | 415 | 205 | 21,035 |
| | DQN-PID | 6.73 | 340 | 203 | 19,669 |

结果表明，传统的 PID 控制在稳定到目标值的过程中调节时间较长，超调量较大，相比较下，本文提出的复合控制策略在各方面指标均有优势，综合三种不同强度的扰动测试结果，基于强化学习前馈的复合控制策略表现出显著的抗干扰性能提升。超调量平均降低了 52.83%，调节时间平均降低了 16.98%。可以看出，基于强化学习前馈的复合控制策略抗干扰性优于传统控制策略。

#### 4.2.2. 鲁棒性仿真实验

为了验证本文控制方案的鲁棒性，在 PID 参数改变 10%的情况下，仍采用复合控制策略进行控制，加入多种扰动信号组合作为扰动系统输入作为扰动系统输入。

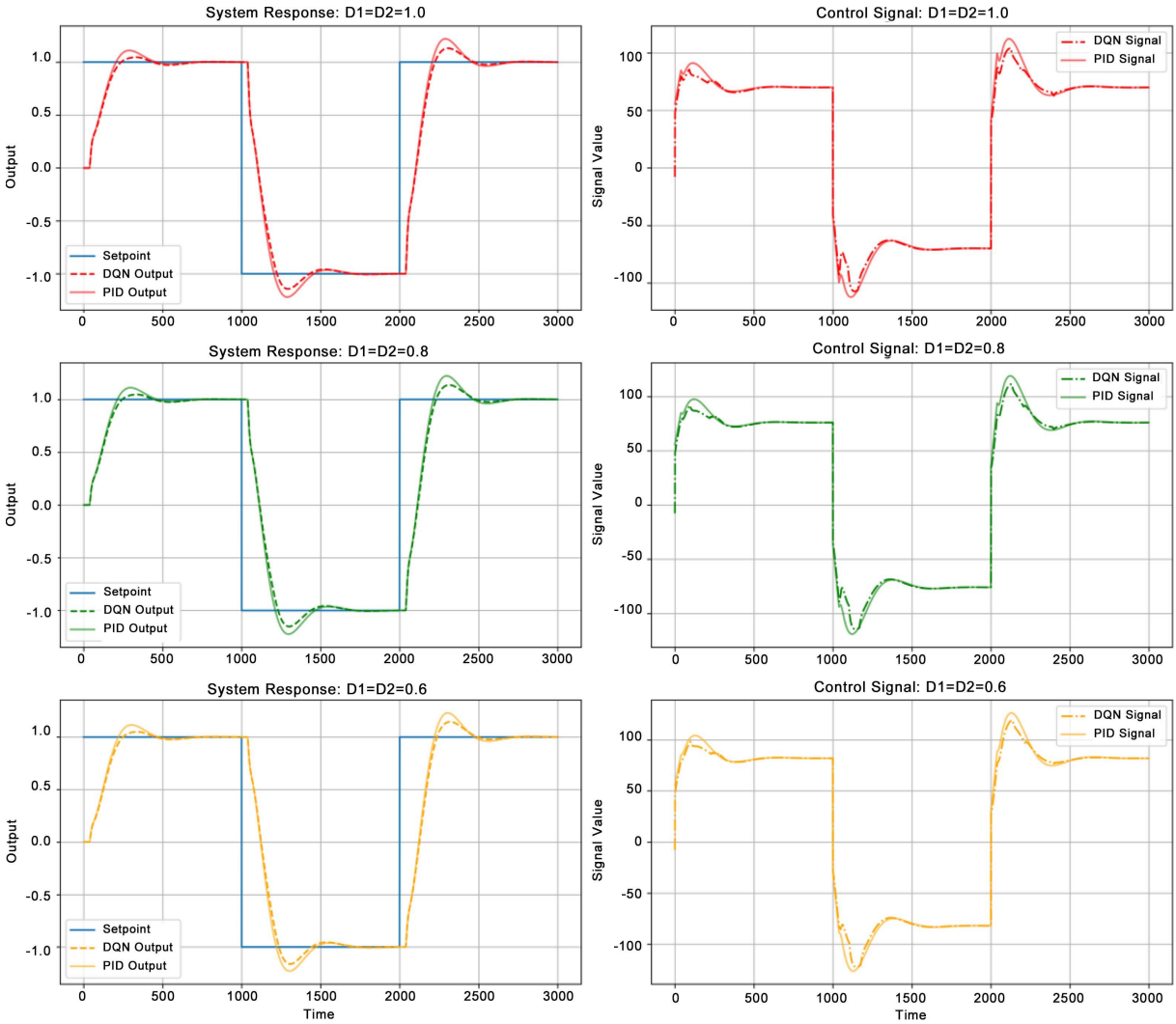


**Figure 8.** Robustness test results
**图 8.** 鲁棒性测试结果

对结果引入性能指标进行分析，结果如表 2 所示。

综合图 8 和表 2 的数据可以看出，系统在 PID 参数变化 10%的情况下采用复合控制策略优于传统控制策略，各项性能指标都优于传统控制策略，在扰动下超调更小，调节时间更短，充分说明本文所提出的控制方案鲁棒性更好。

**Table 2.** Robustness testing performance metrics
**表 2.** 鲁棒性测试性能指标

| 性能指标 | | 超调量/% | 调节时间/s | IAE | ITAE |
|---|---|---|---|---|---|
| D1 = 0.6<br>D2 = 0.6 | 传统 PID | 23.00 | 424 | 236 | 27,361 |
| | DQN-PID | 14.54 | 355 | 231 | 24,735 |
| D1 = 0.8<br>D2 = 0.8 | 传统 PID | 22.38 | 416 | 221 | 25,160 |
| | DQN-PID | 13.72 | 346 | 216 | 22,453 |
| D1 = 1.0<br>D2 = 1.0 | 传统 PID | 22.00 | 408 | 207 | 23,119 |
| | DQN-PID | 13.09 | 337 | 201 | 20,392 |

## 5. 结论

1) 针对强化学习算法面对迟延对象难以收敛问题，设置了多维奖励函数以及迟延缓冲区，有效提高了算法的训练效率。

2) 针对传统 PID 控制在大惯性、大迟延的再热汽温控制系统中难以控制的难点，提出了 RL_PID 复合控制策略，引入强化学习前馈控制器。结果表明此控制策略比传统 PID 控制策略速度更快，超调更小，具有更好的抗干扰能力和更强的鲁棒性。

## 参考文献


[1] 刘吉臻, 李云鹭, 宋子秋, 房方, 牛玉广, 曾德良. 灵活智能燃煤发电技术及评价体系[J]. 动力工程学报, 2022, 42(11): 993-1004, 1012.

[2] 周守为, 朱军龙, 李清平, 等. 科学稳妥实现“双碳”目标, 积极推进能源强国建设[J]. 天然气工业, 2022, 42(12): 1-11.

[3] 白玉峰, 孙伟鹏, 林楚伟, 等. 超超临界机组全负荷段再热汽温智能控制[J]. 电力与能源, 2019, 40(2): 254-257.

[4] 丁建良, 于国强, 罗建裕. 深度调峰下超超临界机组再热汽温控制优化[J]. 中国电力, 2020, 53(5): 143-149.

[5] 李旭. 再热汽温的动态特性与控制[J]. 动力工程, 2009, 29(2): 150-154.

[6] 吴吕斌, 罗自学. 汽温控制现状及其新方法应用研究[J]. 电站系统工程, 2009, 25(1): 5-7, 10.

[7] 王东风, 李玲, 王玉华. 电站锅炉再热蒸汽温度的燃烧器摆角和喷水减温协调预测控制[J]. 动力工程学报, 2018, 38(7): 558-563, 571.

[8] 赵东华, 潘维加, 刘攀, 等. 火电厂锅炉再热汽温优化控制仿真[J]. 计算机仿真, 2019, 36(9): 142-146.

[9] 王焕敏, 王沈振, 唐亮, 等. 基于强化学习的 SCR 脱硝系统优化控制[J]. 动力工程学报, 2024, 44(12): 1916-1922, 1934.

[10] 赵征, 刘子涵. 基于深度强化学习的 SCR 脱硝系统协同控制策略研究[J]. 动力工程学报, 2024, 44(5): 802-809.

[11] Xie, P., Zhang, G., Niu, Y. and Sun, T. (2021) Selective Catalytic Reduction System Ammonia Injection Control Based on Deep Deterministic Policy Reinforcement Learning. *Frontiers in Energy Research*, 9, Article 725353. https://doi.org/10.3389/fenrg.2021.725353

[12] 韩璞, 袁世通, 张金营. 超超临界锅炉主汽温控制系统的建模研究[J]. 计算机仿真, 2013, 30(12): 115-120.

[13] Mnih, V., Kavukcuoglu, K., Silver, D., *et al*. (2013) Playing Atari with Deep Reinforcement Learning. arXiv: 1312.5602.