

基于改进TD3算法的青霉素发酵过程控制方法

王晓君¹, 徐涛¹, 赵春丽², 彭亮亮², 杜越¹

(1. 河北科技大学信息科学与工程学院, 河北石家庄 050000;

2. 山西国药集团威奇达药业有限公司, 山西大同 037000)

摘要: 为提高复杂场景下的青霉素产量, 提出一种基于改进TD3算法的发酵控制模型 Res-OD-TD3, 以实现对青霉素发酵过程的优化控制。首先, 将残差网络融合到 Actor 和 Critic 网络中, 用于提升其收敛性; 其次, 使用 OU 噪声代替传统高斯噪声, 使动作值能够在连续控制任务中以更平滑的方式进行探索, 减少探索过程中的剧烈波动; 最后, 通过动态调整的 Huber 损失函数对原有损失函数进行改进, 在训练的不同阶段提供更合适的损失度量, 使算法在处理异常值时更为鲁棒。仿真实验结果表明, 与传统的 TD3 算法和 DDPG 算法相比, Res-OD-TD3 算法对于温度精度和稳定性的控制效果更好, 溶解氧浓度分别提升了 4.1%、5.5%, 产量分别提升了 12.8%、27.8%, 证实了其在青霉素发酵优化控制中的显著优势。

关键词: 深度强化学习; 控制策略; TD3 算法; 残差网络; 青霉素发酵

DOI: 10.11907/rj.dk.241805

中图分类号: TP273

文献标识码: A

文章编号: 1672-7800(2025)010-0104-07

扫描二维码阅读全文:



Penicillin Fermentation Process Control Based on Improved TD3 Algorithm

WANG Xiaojun¹, XU Tao¹, ZHAO Chunli², PENG Liangliang², DU Yue¹

(1. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050000, China;

2. Shanxi National Pharmaceutical Group Weiqida Pharmaceutical Co., Ltd., Datong 037000, China)

Abstract: To improve the production of penicillin in complex scenarios, a fermentation control model Res-OD-TD3 based on an improved TD3 algorithm is proposed to achieve optimized control of the penicillin fermentation process. Firstly, the residual network is integrated into the Actor and Critic networks to enhance the convergence of the network; Secondly, using OU noise instead of traditional Gaussian noise allows for smoother exploration of action values in continuous control tasks, reducing drastic fluctuations during exploration; Finally, by dynamically adjusting the Huber loss function, the original loss function is improved to provide more suitable loss metrics at different stages of training, making the algorithm more robust in handling outliers. The simulation experiment results show that compared with traditional TD3 algorithm and DDPG algorithm, Res-OD-TD3 algorithm has better control effect on temperature accuracy and stability, with dissolved oxygen concentration increased by 4.1% and 5.5% respectively, and yield increased by 12.8% and 27.8% respectively, confirming its significant advantages in optimizing penicillin fermentation control.

Key Words: deep reinforcement learning; control strategy; TD3 algorithm; residual network; penicillin fermentation

0 引言

青霉素是一种重要的抗生素, 其发酵过程涉及多种难以预测的因素^[1]。如何设计青霉素发酵过程的控制策略, 实现生产工艺的优化, 是一项非常关键的任务。青霉素发

酵过程具有显著的不确定性和非线性^[2]。在该过程中, 微生物的生长、代谢以及产物形成均受到多种因素的影响, 如温度、pH 值、溶解氧浓度等^[3]。为优化微生物发酵环境、确保发酵过程的有效性, 需要对各种参数进行实时监测与调整。其中, 温度和溶解氧浓度对青霉菌的生长和代谢至关重要。温度过高或过低都会影响菌体活性, 进而影

收稿日期: 2024-09-17

作者简介: 王晓君(1973-), 男, 博士, 河北科技大学信息科学与工程学院教授、硕士生导师, 研究方向为导航基带信号处理技术、数字信号处理算法及工程实现、嵌入式开发与移植; 徐涛(1998-), 男, 河北科技大学信息科学与工程学院硕士研究生, 研究方向为强化学习及其在青霉素发酵过程控制中的应用。本文通讯作者: 徐涛。

响青霉素的生成速率^[4]。溶解氧浓度则直接决定青霉菌的有氧呼吸效率, 进而影响青霉素产量。因此, 精准控制温度和溶解氧浓度能够最大限度地优化发酵环境, 提升青霉素产量。

发酵工艺具有强非线性、时滞性(如 pH 变化滞后于补料操作)和参数时变性(如菌种代谢随批次衰减), 传统 PID 控制或模型预测控制策略难以应对。作为机器学习的新兴分支, 强化学习通过“智能体—环境交互”框架, 无需精确数学模型即可实现动态优化^[5]。将强化学习算法与青霉素发酵工艺过程相结合, 对发酵过程的相关参数进行控制, 可优化发酵流程, 提高生产效率和质量。

1 相关工作

为对非线性、滞后性和强耦合的工业过程进行稳定精确的控制, 邓绍斌等^[6]提出一种基于局部策略交互探索的深度确定性策略梯度算法(Local Policy Interaction Exploration-based Deep Deterministic Policy Gradient, LPIE-DDPG)。该算法采用深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)作为控制策略, 并引入原控制器的控制策略作为局部策略进行搜索, 通过交互探索机制进行学习。

实验结果表明, 该算法将青霉素的发酵产量提高了 3.8%。Yifru 等^[7]提出一种基于拉格朗日变体的控制方法, 通过将约束策略优化与贝叶斯优化相结合, 实现了对预设参数的高效寻优以及工业过程的稳定控制。Du 等^[8]系统探讨了结合数学建模与深度强化学习算法优化发酵过程的方法。该研究融合机制建模与数据驱动建模, 重点阐述了混合建模在发酵工艺优化与放大过程中的应用, 展示了其在提升工艺控制效率和决策智能化程度方面的重要价值。

孪生延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient, TD3)算法及其改进算法, 以及其他一些强化学习算法也被用于发酵等工业控制领域^[9]。例如, Shi 等^[10]将 TD3 算法与 PID 控制相结合, 实现了对发酵过程中参数的动态调整, 但其使用的传统均方误差(Mean Squared Error, MSE)损失函数在处理异常数据点时表现不佳, 可能导致梯度爆炸问题, 进而影响算法的稳定性。Zhong 等^[11]提出一种基于双经验回放池 TD3 算法的 PID 参数优化方法。该方法利用深度强化学习方法自主优化 PID 控制器的参数, 并结合双经验回放池机制有效提升了算法前期的训练效率。Chai 等^[12]使用深度强化学习算法分批优化发酵过程, 重点优化底物的供给效率。但其探索策略通常依赖于高斯噪声, 这可能导致动作值的剧烈波动。周娴玮等^[13]提出一种带 Q 网络过滤的两阶段 TD3 深度强化学习方法, 通过预训练和 Q 网络过滤有效缓解了遗忘现象, 但同样存在探索策略波动较

大的问题。

为解决 MSE 损失可能导致梯度爆炸、传统 TD3 算法探索策略动作波动较大等问题, 本文提出一种结合残差 OU 噪声与动态 Huber 损失的改进 TD3 算法(TD3 Algorithm with Residual OU Noise and Dynamic Huber Loss, Res-OD-TD3)用于青霉素发酵过程的优化控制。首先, 引入残差网络结构, 用于缓解梯度消失和爆炸问题, 加速网络收敛^[14]; 其次, 引入 OU 噪声和动态调整的 Huber 损失函数, 使生成的噪声序列更加平滑, 对不同范围内损失的处理更加灵活, 从而大大提升算法的鲁棒性和稳定性^[15-16]。通过精准控制温度和溶解氧浓度, Res-OD-TD3 算法可有效提升青霉素产量。

2 基于改进 TD3 算法的控制器设计

2.1 状态、动作、奖励函数定义

2.1.1 状态空间

在发酵过程中, 温度(T)、溶解氧浓度(DO₂)、底物浓度(S)、发酵液中的青霉素浓度(P)共同组成状态空间, 每回合状态空间中的初始状态由仿真模型随机产生, 然后运行控制算法, 智能体与当前环境交互获得下一组状态和奖励。

2.1.2 动作空间

动作空间由冷水水量(F_c)、热水水量(F_h)、通气速率(F_g)、罐压(Pressure)、补糖量(F_s)构成。

2.1.3 奖励函数

奖励函数决定着策略网络选取策略的优劣, 对于发酵过程有重要的指导与反馈作用^[17]。青霉素发酵分为启动、生长和产物积累、衰退 3 个阶段。本文模拟控制青霉素发酵生长最旺盛和产物积累阶段的温度和溶解氧浓度。在该阶段, 应当保证温度稳定维持在 298 K 左右, 因此以 298 K 为基准来计算当前温度与其差值的绝对值, 并按照该绝对值的不同范围分别给予对应的奖励, 以此来鼓励温度回归至 298 K 的标准。氧气浓度高于 20%(即 0.2)时才能促进青霉素最终产量提高, 因此当溶解氧浓度小于 20% 时, 直接给予负奖励, 来引导浓度朝着 20% 以上发展; 当浓度大于 20% 时, 氧气浓度越高, 越分段给予更高的正向奖励, 引导氧气浓度适当上升。

经过多次实验验证, 最终设计的奖励函数分为温度奖励和氧气浓度奖励两部分, 采用当前温度与 298 K 的差值 E1 来设计温度的奖励函数 R1, 采用当前溶解氧浓度减去 0.2 的差值 E2 来设计溶解氧浓度的奖励函数 R2。根据二者的权重将奖励函数设置为 R1+3R2, 可以在温度稳定在目标区域时着重提升溶解氧浓度。通过精细化的分段奖励机制, 同时调整多个目标, 可有效引导模型朝着最优方向进化。温度、溶解氧的奖励函数分别表示为:

$$R_1 = \begin{cases} 100 - |E_1|, & |E_1| \leq 0.01 \\ 80 - |E_1|, & 0.01 < |E_1| \leq 0.05 \\ 60 - |E_1|, & 0.05 < |E_1| \leq 0.1 \\ |E_1|, & 0.1 < |E_1| \leq 0.5 \\ -10 + |E_1|, & 0.5 < |E_1| \leq 1 \\ -20 + |E_1|, & 1 < |E_1| \end{cases} \quad (1)$$

$$R_2 = \begin{cases} 100 + E_2, & 0.1 \leq E_2 \\ 80 + E_2, & 0.05 \leq E_2 < 0.1 \\ 50 + E_2, & 0.01 \leq E_2 < 0.05 \\ 30 + E_2, & 0.001 \leq E_2 < 0.01 \\ 10 + E_2, & 0 \leq E_2 < 0.001 \\ -10 + E_2, & E_2 < 0 \end{cases} \quad (2)$$

2.2 TD3算法原理

TD3算法是一种增强版的DDPG算法,通过引入3个关键改进来提升算法性能:①引入双重Q网络。具体来说,TD3算法维护两个独立的Q网络(Q_1 和 Q_2),并在更新策略时使用较小的Q值,以减小Q值过高估计的风

险;②引入目标策略平滑技术^[18]。在计算目标Q值时,TD3算法会在目标动作上加入一个服从均匀分布的高斯随机噪声,然后将其裁剪到一个合理的范围内,从而使目标Q值函数更平滑和稳健,避免过度乐观的估计;③引入策略网络与价值网络的延迟更新策略^[19]。具体而言,在Q网络更新后,每隔固定步数才会进行策略网络的更新,这种延迟更新机制有助于降低策略更新频率,从而提高算法稳定性。综合以上改进,TD3算法在处理连续动作空间的强化学习任务时能有效减少误差累积,提高策略学习的稳定性和效率。

2.3 改进TD3算法框架

原始TD3算法在深层网络中存在梯度消失或爆炸,高斯噪声可能导致动作剧烈波动以及固定损失函数在处理误差方面缺乏灵活性等问题。为解决这些问题,本文提出改进TD3算法,通过引入残差网络、OU噪声以及动态调整的Huber损失函数来提高模型的训练稳定性、动作连续性和处理异常数据点的鲁棒性。改进TD3算法框架如图1所示。

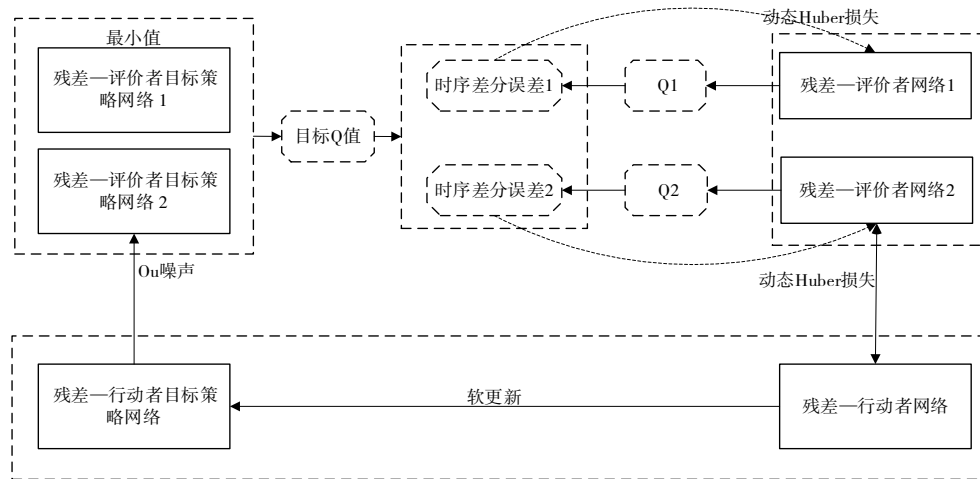


Fig. 1 Framework of improved TD3 algorithm

图1 改进TD3算法框架

2.3.1 残差网络

在TD3算法的Actor网络和Critic网络中引入残差块,以增强网络的表达能力。残差块采用两个相同输入和输出的全连接层,每个全连接层后均进行层归一化处理,并应用ELU激活函数。残差块结构如图2所示。

Actor网络输入4维状态,经过第一个全连接层映射到256维特征空间,使用ReLU激活后进入残差块结构,再经

过第二个全连接层将输出动作映射到5维动作空间,限制到0~1之间输出,方便后续使用。Critic网络将4维状态向量与5维动作向量拼接,通过第一个全连接层映射到256维特征空间,激活函数为ReLU。经残差块结构,再通过第二个全连接层映射到64维特征空间,最后经输出层将64维特征映射为一个标量Q值。Actor网络和Critic网络的结构分别如图3和图4所示。

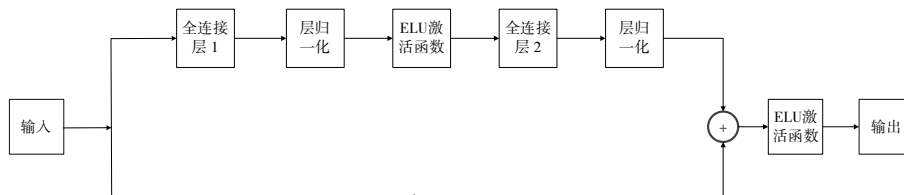


Fig. 2 Residual block structure

图2 残差块结构

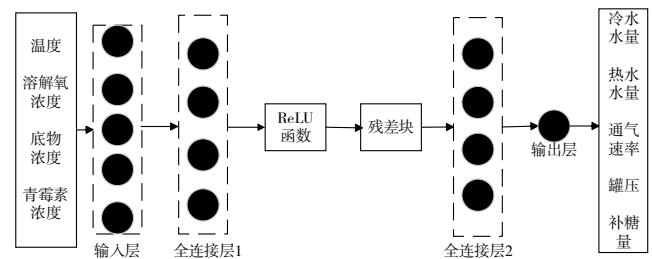


Fig. 3 Actor network structure
图 3 Actor 网络结构

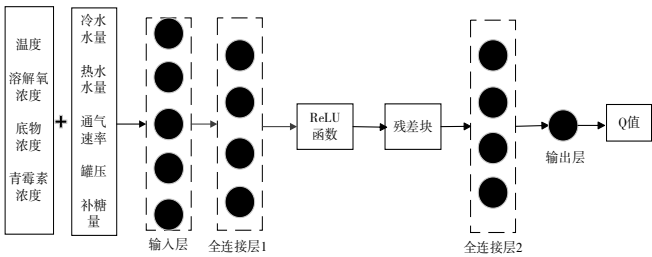


Fig. 4 Critic network structure
图 4 Critic 网络结构

通过引入恒等映射(即将输入向量直接添加到经过处理后的输出上,使得网络可以直接学习输入与输出之间的残差,而不是学习完整的映射),残差块允许梯度直接通过短连接反向传播,从而缓解梯度消失和爆炸问题,使得更深层次的网络能够被有效训练;而层归一化则进一步稳定了训练过程。将残差块改进的算法与原始算法分别运行 200 回合,对相关指标进行比较,结果见表 1。可以看出,改进算法平均每回合运行时间略有缩短,温度稳定时所需运行回合数和产量稳定时所需运行回合数相较原始算法明显减少,说明算法收敛性能得到了显著提升。

Table 1 Comparison of indicators before and after algorithm improvement
表 1 算法改进前后指标比较

指标	平均每回合运行	温度稳定时所	产量稳定时所
	时间/s	需回合数	需回合数
残差块改进算法	13.55	145	155
原始算法	13.62	165	170

2.3.2 OU 噪声

在原始算法中,噪声是通过在目标动作上添加独立分布的高斯随机噪声来实现的。这种方法的噪声独立于时间序列,缺乏时间相关性,有时会导致动作序列的不连续性。因此,本文采用 OU 噪声,其具有时间相关性和自回归特性,可有效改善模型的收敛性和稳定性。OU 噪声的生成公式为:

$$dx_t = \theta(\mu - x_t) + \sigma N_t$$
 (3)

式中: x_t 为当前噪声状态; θ 用于控制噪声回归到均值的速度; μ 为噪声均值; σ 为噪声的标准差; N_t 为正态分布噪声。

OU 噪声具有时间相关性,生成的噪声序列较为平滑,有助于在连续动作空间的任務中保持动作的平滑性和自

然性。

2.3.3 动态 Huber 损失函数

在强化学习中,损失函数的选择与设计对于算法性能至关重要。原始 TD3 算法采用传统的损失函数 MSE,其在处理噪声和异常值时性能有待提升。平均绝对误差(Mean Absolute Error, MAE)虽然对异常值更加鲁棒,但在误差较小时梯度不够平滑^[20]。因此,本文提出一种动态调整的 Huber 损失函数,其兼具 MSE 和 MAE 的优点,可增强算法在不同误差范围内的鲁棒性和稳定性。

Huber 损失函数通过参数 δ 控制行为。当误差较小时,Huber 损失函数表现为 MSE;当误差较大时,则表现为 MAE。为使损失函数在训练过程中更具适应性,本文引入动态调整机制,根据预测值与真实值的绝对误差均值大小来动态调整 δ 的值,然后根据 δ 值和误差值来计算最终的网路损失值。 δ 参数的调整公式为:

$$\delta = \begin{cases} 0.1, & \mu \leq 0.1 \\ 0.5, & 0.1 < \mu \leq 0.5 \\ 0.8, & 0.5 < \mu \end{cases}$$
 (4)

式中: μ 表示预测值与真实值绝对误差的均值。

计算出 δ 值后,最终网路损失值可通过以下公式计算:

$$loss = 0.6 * min^2 + \delta * diff$$
 (5)

式中: min 为预测值和真实值的绝对误差与 δ 二者之间的最小值; $diff$ 为预测值和真实值的绝对误差减去 min 的差值; $loss$ 为最终 Critic 网络的损失值。

动态调整 δ 参数使得 Huber 损失函数能够在不同训练阶段灵活调整其行为,兼顾对小误差的敏感性和对大误差的鲁棒性。这种自适应机制不仅可以提高模型的收敛性,而且能增强模型的泛化能力和稳定性,在实际应用中具有显著优势。

对噪声与损失函数共同改进前后的算法进行比较实验,得到第 160 回合的温度比较结果如图 5 所示。为方 便展示,图片中每隔 10 个时间步输出一次 X 值和 Y 值。可以看出,在该回合的 40 个时间步,改进算法的温度已经趋于稳定,而原始 TD3 算法的温度仍处于局部不稳定的跳跃状态,说明改进算法能更快达到理想的温度状态,控制稳定性更高。

3 仿真实验

3.1 仿真环境

本文采用的发酵仿真环境为 PenSimPy。其是一个专门用于模拟工业规模青霉素发酵过程的 Python 库,广泛应用于工业发酵过程的研究。该平台通过整合物质平衡、能量平衡和青霉素发酵动力学方程,不仅能够模拟温度、溶解氧浓度等关键参数的动态变化,而且可以评估不同控制策略对于生产过程的影响,为用户提供一个精确的发酵过

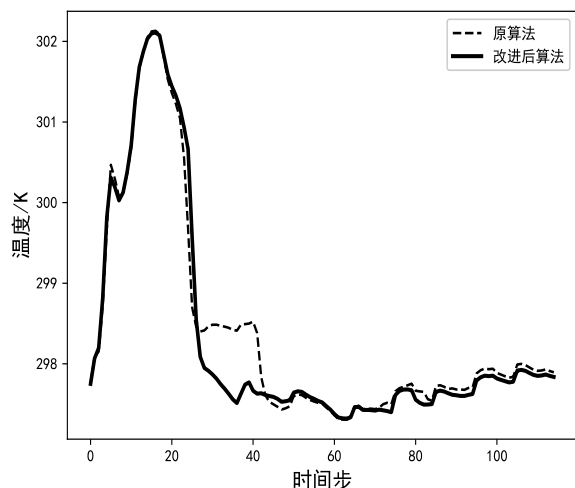


Fig. 5 Comparison of temperature before and after improvement

图5 改进前后温度比较

程建模与仿真环境。PenSimPy 支持基于 PID 控制的顺序批量控制策略,其开放接口亦允许与强化学习算法进行无缝结合,进一步提升系统的自适应性和控制精度,最终实现青霉素产量的提升与生产成本的降低。PenSimPy 的灵活性使其能适应不同的控制需求,成为研究工业发酵过程的重要工具,极大推动了基于数据驱动方法的发酵工艺改进与创新。

3.2 实验流程

实验由 PenSimPy 环境随机初始化出各状态,然后运行本文所提算法进行控制,每回合具体结束时的时间步由环境自身进行判断。为方便图表展示,在每个回合内均为每隔 10 个时间步输出一次 X 值和 Y 值。运行 200 回合后,当温度稳定在 298 K、溶解氧浓度稳定在 20% 以上且有所提升、产量得到提升且大致稳定时,比较改进前后算法的运行结果。经过多次实验发现,当回合数小于 40 或每回合时间步为 10 的整数倍时,设置动作标志为“0”,并为动作空间赋值为全零数组;其他情况设置动作标志为“1”,并由 Actor 网络产生动作空间值时效果最好。随后执行动作获取批次数据和终止标志等,并计算温度与溶解氧浓度的误差奖励。若动作标志为“0”,则采用批次数据中对应的各动作值代替全零数组,再进行存储经验和学习的过程。在初始化阶段或定期时间步使用全零的标准输入,模拟一种特定的行为状态,可以使初始化不受智能体控制策略的影响。在保存经验时使用批次中的历史数据,能确保智能体在存储状态转移时准确记录各种信息,更好地进行学习。以上方式可以提高智能体的学习效果和策略的泛化能力。实验流程如图 6 所示。相关参数设置见表 2。这些参数值基于强化学习算法中的经验、多次实验调优以及相关文献中的标准作法设置,可以很好地适应发酵过程,维持系统的稳定性。

3.3 实验结果与分析

若要同时实现温度稳定、溶解氧浓度达标、回合累计

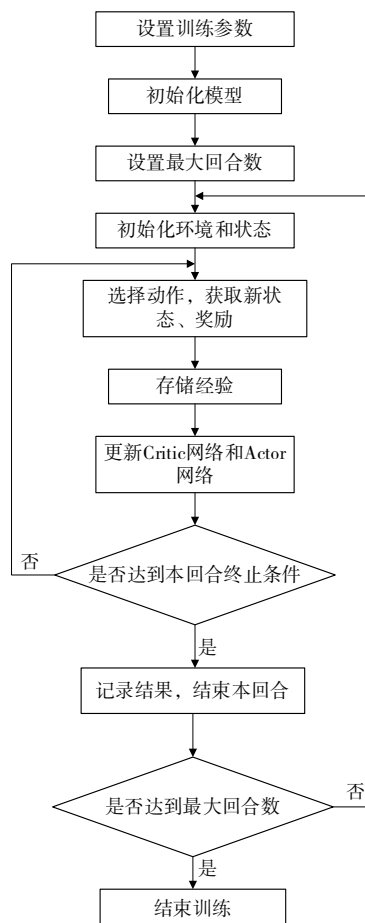


Fig. 6 Experiment flow

图6 实验流程

Table 2 Parameters setting

表2 参数设置

参数	数值
折扣因子	0.99
Actor 学习率	1e-5
Critic 学习率	1e-3
奖励衰减系数	0.95
软更新系数	0.005
策略网更新频率	2

总产量达到峰值 3 个目标,DDPG 算法需要约 180 个回合,TD3 算法需要约 170 个回合,Res-OD-TD3 算法需要约 130 个回合。3 种算法的温度、溶解氧浓度和青霉素产量比较分别见图 7、图 8、图 9。由图 7 可以看出,在初始阶段,Res-OD-TD3 算法的温度上升较为平稳;在后期,Res-OD-TD3 算法的平稳性优势表现得最为显著,曲线更加平滑且温度维持在 298K 左右,控制精度更高。

由图 8 可知,在 400 个时间步后,Res-OD-TD3 算法明显提升了溶解氧浓度,相较 DDPG、TD3 算法分别最高可提升 5.5%、4.1% 左右。此外,在溶解氧浓度提升过程中,Res-OD-TD3 算法的线条波动性明显低于其他两种算法,说明其控制稳定性更好。

由图 9 可以看出,在 200 个时间步后,Res-OD-TD3 算法的青霉素产量提升程度明显优于其他两种算法。具体

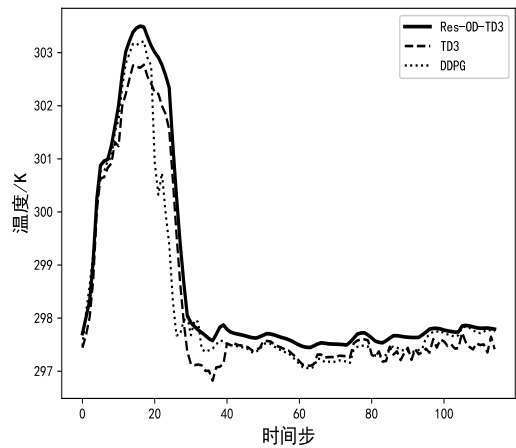


Fig. 7 Comparison of temperature indicators among three algorithms
图 7 3 种算法温度指标比较

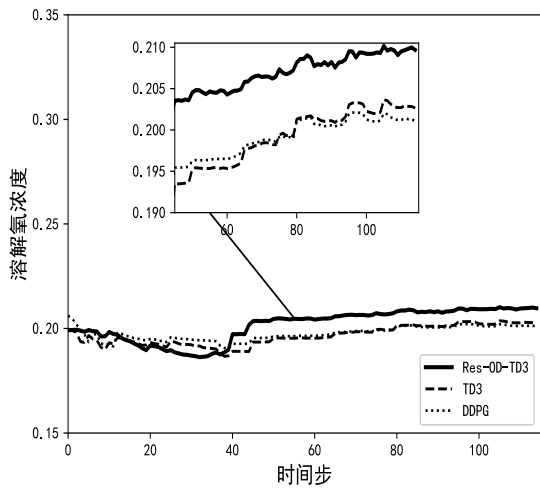


Fig. 8 Comparison of dissolved oxygen concentration among three algorithms
图 8 3 种算法溶解氧浓度比较

来说, Res-OD-TD3 算法的青霉素产量可达到 2 535 kg, 而 TD3 算法的产量为 2 247 kg, DDPG 算法的产量为 1 983 kg, 分别提升了 12.8% 和 27.8%。

将目前强化学习领域的一些主流算法, 如柔性动作—评价算法(Soft Actor-Critic, SAC)^[21]、TD3 算法、近端策略优化算法(Proximal Policy Optimization, PPO)^[22]、DDPG 算

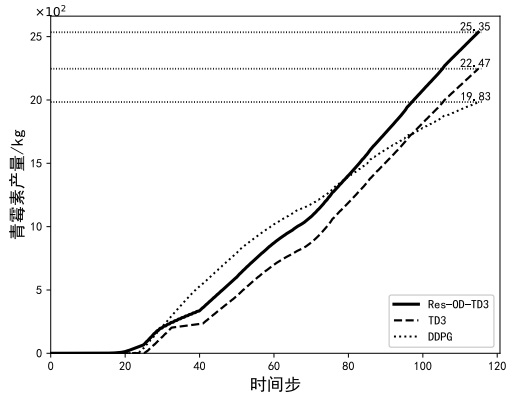


Fig. 9 Comparison of penicillin production using three algorithms
图 9 3 种算法青霉素产量比较

法、双重深度 Q 网络算法(Double Deep Q-Network, DDQN)^[23]等, 与 Res-OD-TD3 算法进行比较, 结果如表 3 所示。可以看出, Res-OD-TD3 算法在收敛速度上表现最为优秀。在各算法达到收敛时, Res-OD-TD3 算法的回合累计奖励最高, 表明其能够作出最优决策, 使训练效率得到大幅提升。因此, Res-OD-TD3 算法在最终收敛时的青霉素产量也是所有算法中最高的。

Table 3 Comparison between mainstream reinforcement learning algorithms and Res-OD-TD3 algorithm

表 3 主流强化学习算法与 Res-OD-TD3 算法的比较			
算法	收敛所需 回合数	回合累计奖 励/ $\times 10^3$	青霉素产量/kg
Res-OD-TD3	130	8.87	2 535
SAC	165	8.23	2 266
TD3	170	8.19	2 247
PPO	175	8.13	2 025
DDPG	180	7.32	1 983
DDQN	200	6.57	1 870

4 结语

针对青霉素发酵这一复杂的生物化学反应, 本文提出 Res-OD-TD3 算法对其过程进行控制与优化, 以提高青霉素产量。该算法在 Actor 和 Critic 网络中均引入残差块结构, 用于加速算法收敛; 同时通过 OU 噪声改善动作序列的连续性, 并提出一种动态调整的 Huber 损失函数来提高算法的鲁棒性和稳定性。强化学习算法的引入相当于设计了一个智能控制器, 可以动态调节温度和溶解氧浓度, 使青霉素产量获得显著提升。

然而, 本研究在发酵环境多样化测试方面仍显不足, 实验场景的复杂度与实际工业环境存在差距。此外, 模型超参数选择也有待优化, 尤其是在不同规模的发酵系统中可以通过动态调整学习率、探索率和模型复杂度等来平衡鲁棒性与效率。后续将在更大规模的工业发酵设备上验证, 同时结合神经架构搜索技术为发酵过程提供更强的灵活性和鲁棒性, 从而提升算法的泛化能力和工业实用性。

参考文献:

[1] ZHANG B, PAN F C, ZHOU X F, et al. An industrial process control method based on reinforcement learning of second-order value gradient model[J]. Computer Application Research, 2024, 41(8): 2434–2440.
张博, 潘福成, 周晓锋, 等. 基于二阶价值梯度模型强化学习的工业过程控制方法[J]. 计算机应用研究, 2024, 41(8): 2434–2440.

[2] WANG T B. Analysis of penicillin fermentation process[J]. World Digest of Latest Medical Information, 2018, 18(20): 103, 106.
王铁斌. 浅析青霉素发酵工艺[J]. 世界最新医学信息文摘, 2018, 18(20): 103, 106.

[3] GAO X J, LI X F, QI Y S. Information entropy FCM multi-stage LSTM-AE intermittent process fault monitoring based on diffusion distance[J]. Journal of Chemical Engineering in Higher Education, 2023, 37(1):

- 120-130.
- 高学金,李学凤,齐咏生. 基于扩散距离的信息熵FCM多阶段LSTM-AE间歇过程故障监测[J]. 高校化学工程学报, 2023, 37(1): 120-130.
- [4] LI B Q. Talk about the production process of penicillin[J]. Heilongjiang Science and Technology Information, 2015(5): 97.
- 李宝全. 谈青霉素的生产工艺过程[J]. 黑龙江科技信息, 2015(5): 97.
- [5] GAO Y Z, NIE Y M. A research review of deep reinforcement learning methods for multiple intelligences based on value function decomposition[J]. Computer Science, 2024, 51(S1): 34-42.
- 高玉钊,聂一鸣. 基于值函数分解的多智能体深度强化学习方法研究综述[J]. 计算机科学, 2024, 51(S1): 34-42.
- [6] DENG S B, ZHU J, ZHOU X F, et al. An industrial process control method based on local policy interaction exploration with deep deterministic policy gradient[J]. Computer Applications, 2022, 42(5): 1642-1648.
- 邓绍斌,朱军,周晓锋,等. 基于局部策略交互探索的深度确定性策略梯度的工业过程控制方法[J]. 计算机应用, 2022, 42(5): 1642-1648.
- [7] YIFRU L, BAHARI A. Concurrent learning of policy and unknown safety constraints in reinforcement learning [DB/OL]. <https://arxiv.org/abs/2402.15893>.
- [8] DU Y H, WANG M Y, YANG L H, et al. Optimization and scale-up of fermentation processes driven by models [J]. Bioengineering, 2022, 9(9): 473.
- [9] FUJIMOTO S, HOOFF H, MEGER D. Addressing function approximation error in actor-critic methods [C]//Stockholm: International Conference on Machine Learning, 2018.
- [10] SHI Q, LAM H K, XUAN C, et al. Adaptive neuro-fuzzy PID controller based on twin delayed deep deterministic policy gradient algorithm[J]. Neurocomputing, 2020, 402: 183-194.
- [11] ZHONG H, WANG Z. TD3 algorithm of dynamic classification replay buffer based PID parameter optimization [J]. International Journal of Control, Automation and Systems, 2024, 22(10): 3068-3082.
- [12] CHAI W Y, TAN M K, TEO K T K, et al. Optimization of fed-batch-Baker's yeast fermentation using deep reinforcement learning[J]. Process Integration and Optimization for Sustainability, 2024, 8(2): 395-411.
- [13] ZHOU X W, BAO M H, YE X, et al. A two-stage TD3 deep reinforcement learning approach with Q-network filtering[J]. Computer Technology and Development, 2023, 33(10): 101-108.
- 周娴玮,包明豪,叶鑫,等. 带Q网络过滤的两阶段TD3深度强化学习方法[J]. 计算机技术与发展, 2023, 33(10): 101-108.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [15] WANG T, ZHANG W H, PU Y F. Fractional-order improved OU noise for reinforcement learning inertial environments [J]. Journal of Sichuan University (Natural Science Edition), 2023, 60(2): 63-69.
- 王涛,张卫华,蒲亦非. 适用于强化学习惯性环境的分数阶改进OU噪声[J]. 四川大学学报(自然科学版), 2023, 60(2): 63-69.
- [16] YANG L, JIANG S, WEI X, et al. EEG feature learning model based on intrinsic time-scale decomposition and adaptive Huber loss [J]. Chinese Quarterly Journal of Mathematics, 2022, 37(3): 281.
- [17] ZHANG Y H, GUAN X. Optimization study of neural network fuzzy testing technique based on reinforcement learning algorithm [J]. Computerized Measurement and Control, 2024, 32(3): 131-137.
- 张宇豪,关昕. 基于强化学习算法的神经网络模糊测试技术优化研究[J]. 计算机测量与控制, 2024, 32(3): 131-137.
- [18] LI Y, WANG W G, ZHANG Y, et al. Offloading decision of in-vehicle edge computing tasks based on improved TD3 algorithm [J]. Electronic Measurement Technology, 2024, 47(6): 64-70.
- 李亚,王卫岗,张原,等. 基于改进型TD3算法的车载边缘计算任务卸载决策[J]. 电子测量技术, 2024, 47(6): 64-70.
- [19] HAO Y Z, WANG Z L. TD3 algorithm for cooperative multi-intelligence under security constraints [J]. Computer Application Research, 2023, 40(6): 1692-1696, 1701.
- 郝禹哲,王振雷. 安全约束下合作型多智能体TD3算法[J]. 计算机应用研究, 2023, 40(6): 1692-1696, 1701.
- [20] LIU Q, SUN Z L, JIANG R X. Research on magnetic field inversion and magnetic field evaluation algorithm based on minimum mean square error [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2023, 51(8): 142-148.
- 刘琪,孙兆龙,姜润翔. 基于最小均方误差的磁场反演及磁场评估算法研究[J]. 华中科技大学学报(自然科学版), 2023, 51(8): 142-148.
- [21] LIU Y, WANG C, ZHAO C, et al. A soft actor-critic deep reinforcement-learning-based robot navigation method using LiDAR [J]. Remote Sensing, 2024, 16(12): 2072.
- [22] ZHANG J, ZHANG Z, HAN S, et al. Proximal policy optimization via enhanced exploration efficiency [J]. Information Sciences, 2022, 609: 750-765.
- [23] ZHANG S, ZHANG L, XU F, et al. Dynamic deployment method based on double deep Q-network in UAV-assisted MEC systems [J]. Journal of Cloud Computing, 2023, 12(1): 130.

(责任编辑:尹晨茹)