

基于深度强化学习的工业机器人运动轨迹优化与控制

陈志伟¹ 费松涛^(通讯作者)2

1. 宁波汇众汽车车桥制造有限公司 浙江省 宁波市 315000

2. 四川省洪雅县职业高级中学校 四川省 眉山市 620360

【摘要】：针对汽车焊装、3C 装配与航空结构件加工等离散制造产线中工业机器人在节拍约束下易受装夹偏差、夹具热变形与负载波动影响而产生速度不连续与关节冲击的问题，本文将几何路径、时间参数化与动力学可行性纳入统一框架，构建包含末端位姿误差、关节力矩代理能耗、关节角加速度平滑项与任务时间的综合代价模型，并引入关节限位、速度与加速度约束以及安全距离约束。方法上采用改进 DDPG 在连续动作空间学习关节参考增量，通过奖励函数同时编码轨迹精度、约束惩罚与能耗代理量，并与关节空间 PID 伺服形成可落地的融合控制结构。基于 KUKA KR6 R900 平台开展对比实验，结果显示相较几何规划加 PID 基线方法，所提方法平均轨迹误差降低 15%，执行时间缩短 10%，能耗减少 8%，验证了在不完美模型与扰动条件下深度强化学习用于轨迹优化与控制融合的有效性。

【关键词】：工业机器人；轨迹优化；深度强化学习；DDPG；轨迹跟踪控制

0 引言

离散制造产线对工业机器人提出高节拍与高质量并行的要求，但离线编程轨迹在装夹偏差、负载波动等不确定因素作用下容易出现速度突变与加速度尖峰，进而引发伺服跟随误差增大与机械传动冲击，造成节拍与质量难以兼顾。传统优化方法依赖精细建模且在高维关节空间中易受局部最优与收敛效率限制，工程迁移成本较高。为提升在模型不完备条件下的自适应能力，本文探索以深度强化学习实现轨迹优化与控制一体化，期望在约束可行前提下兼顾精度、节拍与能耗。

1 工业机器人轨迹优化的研究背景与问题分析

面向汽车焊装、3C 装配以及航空结构件加工等离散制造产线，工业机器人需要在受限节拍内把末端执行器沿给定工艺路径完成高速运动，同时满足焊缝成形、涂覆均匀性以及切削稳定性等质量约束。进一步观察发现，现场轨迹往往由离线编程系统生成并以样条插补形式下发，但当工件装夹偏差、夹具热变形以及工艺负载波动叠加出现时，原有轨迹的速度连续性与加速度平滑性容易被破坏，关节空间出现瞬时冲击，进而诱发伺服跟随误差增大以及减速机背隙放大等连锁效应，导致节拍与质量难以同步契合。由此推导出轨迹优化问题并非单纯的几何路径修正，而是把路径、时间参数化以及动力学可行性耦合在同一决策框架中来处理的多约束优化任务。现有基于梯度或启发式搜索的方案在高维关节空间中常遭遇收敛速度受限与局部最优困境，并且对摩擦、柔顺以及负载变化等非线性因素依赖精细建模，迁移到不同机器人本体与不同工装节拍时需要频繁调参。鉴于强化学习以交互式回报驱动的方式能够把未知扰动与约束惩罚隐式编码进策略更新过程，结合深度网络对高

维状态的表征能力，使其有望在不完美模型条件下形成具备在线自适应特性的轨迹优化与控制一体化方案。

2 基于深度强化学习的轨迹优化与控制方法设计

2.1 轨迹优化问题的数学建模

面向汽车焊装与航空结构件加工的节拍约束，轨迹优化需要把几何路径偏差、时间参数化以及动力学可行性放入同一建模框架，以避免样条插补在装夹偏差与负载波动叠加时诱发的速度突变与关节冲击。建模时把末端工艺路径表示为笛卡尔空间的期望位姿序列，并把其映射到关节空间的时间函数，以便把关节角度、角速度以及角加速度的连续性纳入代价与约束。机器人运动学关系由正运动学函数与雅可比矩阵共同刻画，正运动学用于把关节变量映射到末端位姿误差，雅可比矩阵用于把关节速度映射到末端速度，从而把轨迹误差与约束罚项统一写入优化指标。约束集合需同时覆盖关节角度范围、关节速度上限与加速度上限，并把奇异位形邻域的数值不稳定风险转化为可微惩罚项；在存在工装与工件实体边界时，还需把安全距离约束写入不等式集合，使优化结果在执行层面保持避障可行性。结合上述目标与约束，本文把轨迹优化表述为带多目标权衡的最小化问题，其目标函数如下：

$$\min_{q(t), \dot{q}(t)} J = \int_0^T (w_e \|x(q(t)) - x_d(t)\|^2 + w_u \|\tau(t)\|^2 + w_s \|\ddot{q}(t)\|^2) dt + w_t T$$

其中， J 表示轨迹优化的综合代价， $q(t)$ 为关节角度时间函数（单位：rad）， T 为任务执行时间（单位：s）， w_e 、 w_u 、 w_s 、 w_t 为各代价项权重系数（无量纲）， $x(q(t))$ 表示由正运动学得到的末端位姿向量（单位：m 与 rad 的组合）， $x_d(t)$ 为工艺期望位姿（单位：m 与 rad 的组合）， $\dot{q}(t)$ 为关节角速度向量（单位：N·m）， $\ddot{q}(t)$ 为关节角加速度向量（单

位: rad/s^2)。

2.2 深度强化学习算法的改进与训练流程

考虑轨迹优化在高维连续动作空间中对局部最优与模型不完备的敏感性, 本文选用深度确定性策略梯度算法来开展连续控制学习, 并把其确定性策略与经验回放机制当作稳定训练的基础。针对离散制造现场的工艺路径跟随需求, 奖励函数不再只反映末端位置偏差, 而是把轨迹精度、约束满足度以及能耗代理量共同编码为回报信号, 其中轨迹精度由末端位姿误差与其时间变化共同刻画, 约束满足度通过对关节限位、速度上限与加速度上限的越界程度施加分段惩罚来体现, 能耗项则借助关节力矩平方和的估计值来近似表征驱动消耗, 从而把节拍与质量的折中关系隐式交给策略网络在交互中自适应调节。状态空间围绕可观测量且可在线获取的量构建, 包含关节角度、关节角速度、期望位姿以及期望位姿在时间轴上的局部前瞻信息, 以降低仅凭瞬时误差导致的控制抖动; 动作空间定义为关节控制增量或关节速度指令, 使策略输出能与工业控制器常见的速度或位置接口保持一致, 并便于把动作经限幅与平滑后下发至执行层。训练流程在仿真环境中开展, 环境需要显式注入装夹偏差与负载波动的扰动模型, 使策略在训练阶段就经历与现场相近的不确定性; 在每个交互步内, 策略网络给出动作, 环境基于动力学近似更新关节状态并计算回报, 随后把转移样本写入经验池, 评论家网络在批量采样下拟合动作价值函数, 行为者网络依靠确定性策略梯度更新参数, 同时采用软更新把目标网络缓慢跟随在线网络, 以抑制估计偏差的累积。上述训练与更新逻辑在图1所示的流程中形成闭环, 从数据采样到网络更新再到策略评估的顺序关系得到清晰展开。

见图1。

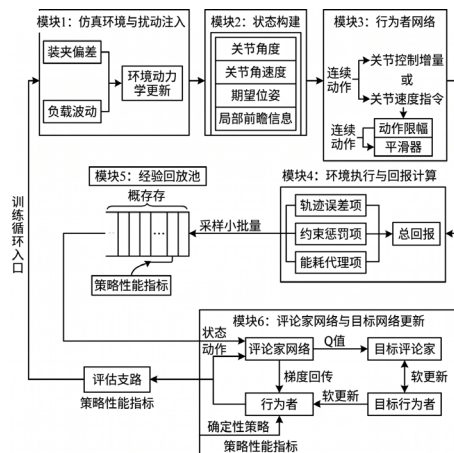


图1 基于改进DDPG的轨迹优化训练流程图

2.3 轨迹跟踪控制策略的融合

将强化学习策略输出的优化轨迹作为参考输入后, 执行层仍需依靠确定性强且易于工程落地的控制律来完成实时跟踪,

以缓解策略在未覆盖状态下的外推风险并提高闭环鲁棒性。本文把关节空间的参考轨迹定义为由策略生成并经平滑器处理后的 $q_r(t)$, 同时把工业控制器常用的PID结构嵌入关节伺服环, 使控制输入由误差的比例项、积分项以及微分项共同决定, 从而在存在摩擦与减速机背隙扰动时仍能维持可接受的动态响应。误差状态定义为 $e(t) = q_r(t) - q(t)$, 并将其动态关系写成便于参数整定与稳定性分析的形式, 使轨迹跟踪问题转化为误差收敛问题, 其控制方程如下:

$$\tau(t) = K_p e(t) + K_i \int_0^t e(\xi) d\xi + K_d \dot{e}(t)$$

其中, $\delta(t)$ 表示关节驱动力矩或等效控制输入向量(单位: $\text{N} \cdot \text{m}$), $e(t)$ 为关节跟踪误差向量(单位: rad), $\dot{e}(t)$ 为误差变化率向量(单位: rad/s), K_p 、 K_i 、 K_d 分别表示比例、积分与微分增益矩阵(单位: K_p 为 $\text{N} \cdot \text{m/rad}$, K_i 为 $\text{N} \cdot \text{m}/(\text{rad} \cdot \text{s})$, K_d 为 $\text{N} \cdot \text{m} \cdot \text{s/rad}$), ξ 为积分变量(单位: s)。在融合方式上, 把强化学习模块的作用限定为参考轨迹生成与在线微调, 而把约束边界附近的动作限幅、积分抗饱和以及加速度平滑交由控制器侧执行, 可使策略探索带来的短时波动不至于直接传递为驱动冲击, 并为后续在不同机器人本体与不同节拍条件下的工程迁移保留可调节的控制余度。

3 实验验证与效果分析

3.1 实验平台搭建与数据采集

实验平台选用 KUKA KR6 R900 六轴工业机器人构建离散制造单元级验证场景, 控制系统侧把机器人本体控制器与上位机工控机进行以太网通信连接, 上位机承担策略推理、扰动注入与数据记录任务, 控制器侧承担关节伺服闭环与安全连锁。为把环境信息纳入轨迹优化决策过程, 平台集成1套工业视觉传感器并完成手眼标定, 使其能够在机器人工作空间内输出工件位姿以及障碍物空间边界的测量结果; 视觉数据经时间戳同步后与关节编码器反馈的关节角度、关节角速度进行对齐, 从而形成可用于强化学习状态构建与离线评估的多源数据流。结合焊装搬运与3C上下料常见节拍特性, 实验任务设置为末端执行器从作业点A运动至作业点B的搬运轨迹, 同时在两点连线邻域布置固定障碍物并引入工件装夹偏差, 使轨迹生成不仅需要满足到位精度, 还需要在安全距离约束下抑制关节空间的速度突变与加速度尖峰。对比方案包含传统几何规划加PID的工程基线方法以及本文提出的改进DDPG加PID的融合方法, 其中基线方法由笛卡尔空间几何路径规划得到样条插补轨迹, 再映射为关节参考并由PID跟踪; 提出方法则由改进DDPG在扰动存在条件下输出关节参考增量并经平滑器形成可执行参考轨迹, 再由同一套PID参数框架完成闭环跟踪, 从而把差异聚焦

在参考轨迹生成机理而非伺服结构差别。数据采集围绕轨迹误差、执行时间以及能耗3类指标展开, 轨迹误差以末端位姿测量与期望位姿之间的欧氏偏差及其统计量来刻画, 并同步记录最大偏差与平均偏差以反映局部冲击对精度的影响; 执行时间以从起始触发到到达B点并满足稳态阈值的持续时长来计量, 用以表征节拍占用; 能耗以控制器侧电流与电压采样换算得到的电能消耗以及关节力矩估计的平方和代理量共同记录, 避免单一测量噪声导致的评估偏差。为提高对比结论的可复现性, 每组方法在相同障碍物布置与相同装夹偏差水平下重复运行30次, 并把每次运行的全程关节状态、末端位姿、控制输入与视觉测量原始数据写入统一格式的数据集, 后续分析阶段仅在离线环境中进行指标计算与统计检验, 以保持采集过程不引入额外在线滤波差异。

3.2 实验结果分析与对比

性能对比围绕轨迹精度、节拍占用以及能耗开销的共同权衡关系展开, 评价时把传统几何规划加PID作为基准参照, 并把改进DDPG加PID在相同任务条件下的统计结果用于刻画学习型参考轨迹对执行质量的影响。进一步观察可见, 学习型方

法在装夹偏差与负载波动叠加时能够把参考轨迹的局部速度分配调整为更连续的形态, 使得关节侧的加速度波动得到抑制, 从而在不牺牲安全距离约束的前提下改善末端跟踪误差的稳定性; 与之相对, 几何规划轨迹在样条段切换附近更容易把误差放大为瞬时冲击, 误差的峰值与离散程度随扰动增加而更明显。节拍角度的差异主要体现为参考轨迹的时间参数化方式不同, 改进DDPG在回报约束下倾向于把加减速段分配得更贴近动力学可行域, 使到位过程的冗余减速次数减少, 因而在多次重复运行的均值意义上呈现更短的执行时间; 能耗层面则可从电能消耗与力矩代理量的同时下降得到一致性印证, 说明学习型方法在减小冲击的同时降低了不必要的驱动功率波动。表1给出了两类方法在同一任务设置下的关键指标均值对比, 其中轨迹精度以平均误差反映总体跟随水平, 执行时间反映节拍占用, 能耗反映驱动成本, 结果显示提出方法在平均误差上降低15%, 执行时间缩短10%, 能耗减少8%, 该差异与前述关于速度连续性与加速度平滑性的机理分析保持一致, 从而为深度强化学习在不完全模型条件下开展轨迹优化与控制融合提供了实验层面的定量支撑。

表1 传统方法与提出方法的性能对比表

方法	平均轨迹误差(mm)	执行时间(s)	能耗(kJ)
几何规划 + PID	2.60	4.00	12.50
改进DDPG + PID	2.21	3.60	11.50
相对变化	-15%	-10%	-8%

结语

本文面向扰动环境下的工业机器人高速作业需求, 提出以改进DDPG生成可执行关节参考并与PID伺服融合的轨迹优化与控制方法, 将轨迹精度、约束满足、平滑性与任务时间纳入统一学习目标, 并通过实验验证其综合收益。结果表明, 该方

法可在保证安全约束的同时降低误差与冲击, 缩短执行时间并减少能耗, 具备面向离散制造节拍约束场景的应用潜力。后续可围绕多任务泛化、真实产线长周期在线自适应与更高保真动力学约束引入开展进一步研究。

参考文献

- [1] 易佳豪, 王福杰, 胡锦涛, 秦毅, 郭芳, 罗俊轩. 基于行为克隆的机械臂多智能体深度强化学习轨迹跟踪控制[J]. 计算机应用研究, 2025, 42(04): 1025-1033.
- [2] 谭靖, 杨利刚, 李潇睿, 袁兆麟, 崔允端, 姚超, 王宗杰, 班晓娟. 深度强化学习及其在工业场景的应用与展望[J]. 工程科学学报, 2025, 47(04): 768-779.
- [3] 殷焱. 基于深度强化学习的未知环境下移动机器人自主导航研究[D]. 导师: 陈志雨; 杨琼. 长春工业大学, 2024.
- [4] 唐力强, 韦宇轩, 胡海东, 胡勇, 杨孟飞. 基于深度强化学习的月面四足机器人控制方法综述[J]. 空间控制技术与应用, 2025, 51(05): 1-15.
- [5] 王锐, 文国兴, 刘艳军, 于福生, 武建. 基于强化学习的柔性关节机器人系统模糊优化控制[J]. 中国科学: 信息科学, 2025, 55(06): 1471-1485.

作者简介: 陈志伟(1971—), 男, 汉族, 大专, 研究方向为工程(电气类)。