

控制与决策

Control and Decision

基于强化学习的遗传算法求解一种新的钻削路径优化问题

朱光宇, 张德颂

引用本文:

朱光宇,张德颂. 基于强化学习的遗传算法求解一种新的钻削路径优化问题[J]. *控制与决策*, 2024, 39(2): 697–704.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1112>

您可能感兴趣的其他文章

Articles you may be interested in

[基于机床超低待机状态的流水车间能耗调度](#)

Energy consumption scheduling in flow shop based on ultra-low idle state of numerical control machine tools

控制与决策. 2021, 36(1): 143–151 <https://doi.org/10.13195/j.kzyjc.2019.0433>

[基于平衡鲸鱼优化算法的无人车路径规划](#)

Path planning of unmanned ground vehicle based on balanced whale optimization algorithm

控制与决策. 2021, 36(11): 2647–2655 <https://doi.org/10.13195/j.kzyjc.2020.0416>

[超启发式交叉熵算法求解模糊分布式流水线绿色调度问题](#)

Hyper-heuristic cross-entropy algorithm for green distributed permutation flow-shop scheduling problem with fuzzy processing time

控制与决策. 2021, 36(6): 1387–1396 <https://doi.org/10.13195/j.kzyjc.2019.1681>

[基于MCPDDPG的智能车辆路径规划方法及应用](#)

The method and application of intelligent vehicle path planning based on MCPDDPG

控制与决策. 2021, 36(4): 835–846 <https://doi.org/10.13195/j.kzyjc.2019.0460>

[基于 \$\sigma\$ 正态概率区间分族遗传蚁群算法的机器人路径规划](#)

Path planning of mobile robot based on σ normal probability interval population division using genetic ant-colony algorithm

控制与决策. 2021, 36(12): 2861–2870 <https://doi.org/10.13195/j.kzyjc.2020.0745>

基于强化学习的遗传算法求解一种新的 钻削路径优化问题

朱光宇[†], 张德颂

(福州大学 机械工程及自动化学院, 福州 350108)

摘要: 孔加工是机械制造的基本工序之一. 针对数控机床的刀具路径优化问题, 提出一种新颖的孔加工刀具路径优化模型——带可决策孔的孔加工多刀具路径优化问题(MTdDPO). 在该模型中, 工件上的孔分为两类: 固定孔和可决策孔. MTdDPO 的目标是通过判断可决策孔的路径归属和路径内各孔的加工顺序来实现加工路径长度的最小化. 为实现 MTdDPO 的优化, 提出基于强化学习的分段遗传算法(RLSGA). 在 RLSGA 中, 种群被视为智能体, 智能体的状态是种群的多样性系数, 3 种不同的分段交叉算子是智能体的动作, 智能体的奖励与种群的适应度值和多样性系数的变化有关. 针对 MTdDPO, 新建 5 个基准测试问题, 并在测试问题上将 RLSGA 与其他 4 个算法进行对比. 结果表明, RLSGA 的表现明显优于其他算法, RLSGA 能够有效地解决 MTdDPO 问题.

关键词: 孔加工路径优化; 遗传算法; 强化学习; 组合优化

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1112

引用格式: 朱光宇, 张德颂. 基于强化学习的遗传算法求解一种新的钻削路径优化问题[J]. 控制与决策, 2024, 39(2): 697-704.

Genetic algorithm based on reinforcement learning for a novel drilling path optimization problem

ZHU Guang-yu[†], ZHANG De-song

(School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China)

Abstract: Hole-making is one of the basic processes of mechanical production. For the problem of toolpath optimization of computer numerical control (CNC) machine tools, a novel toolpath model for hole-making called as multi-tool drilling path optimization problems with decidable holes (MTdDPO) is proposed. In the MTdDPO, holes on workpieces are divided into two categories: fixed holes and decidable holes. The goal of the MTdDPO is to minimize the length of the machining path by judging the path ownership of decidable holes and the machining sequence of all holes in each path. To realize the optimization of the MTdDPO, a segmented genetic algorithm based on reinforcement learning (RLSGA) is proposed. The population of the RLSGA is regarded as the agent, the states of the agent are the intervals of the diversity coefficient of the population, three different segmental crossover operators are the actions of the agent, and the reward of the agent is related to the changes in fitness value and diversity coefficients of the population. Based on the MTdDPO, 5 benchmark test problems are designed, and the RLSGA is compared with other 4 algorithms on these test problems. Results show that the performance of the RLSGA is significantly better than other algorithms, which means the RLSGA can effectively solve the MTdDPO problems.

Keywords: drilling path optimization; genetic algorithm; reinforcement learning; combinatorial optimization

0 引言

孔加工是机械零件制造的基础工序之一. 为了实现自动化生产, 企业通常使用数控机床对工件进行

孔加工^[1]. 在数控机床上, 孔加工过程的时间成本主要包括两部分: 切削时间和非切削时间. 对于一个特定的孔, 不论其工艺规划如何, 它的第 1 步加工通常

收稿日期: 2022-06-25; 录用日期: 2022-11-14.

基金项目: 工信部智能制造综合标准化与新模式应用项目(工信部联装(2016)213号).

责任编辑: 刘士新.

[†]通讯作者. E-mail: zhugy@fzu.edu.cn.

*本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

是钻削. 在钻削阶段, 非切削时间平均占总加工时间的70%左右^[2], 因此它存在很大的优化空间. 数控机床非切削时间的优化可以归结为一个组合优化问题: 钻削路径优化问题 (drilling path optimization, DPO)^[3].

根据待加工孔的种类和数量, 工件可分为两类: 第1类工件上, 只有一种待加工孔, 即所有待加工孔的直径和工艺规划完全相同; 第2类工件上, 有多种待加工孔, 即待加工孔的直径或者工艺规划并不完全相同. 对于第1类工件, 其任意一个工艺要求的加工只使用一种刀具; 而对于第2类工件, 任意一个工艺要求的加工通常都需要使用多种刀具. 第1类工件的DPO被称为单刀具DPO, 学者们通常将其建模为旅行商问题进行求解^[4]. 本文作者^[3,5]也针对单刀具DPO问题开展了研究. 然而, 在现实的生产活动中, 大多数工件上都有多种待加工孔, 属于第2类工件. 针对第2类工件, 学者们基于实际情况提出了不同的数学模型予以解决, 如: 基本多刀具DPO^[6], 带优先约束的多刀具DPO^[7], 时间顺序相关的多刀具DPO^[8]等. 本文所提出的模型正是基于第2类工件.

在数控机床上, 孔的加工方法有钻、扩、铰和镗等, 分别对应不同工序. 对于不能通过一次钻削成型的孔, 为了保证加工精度, 减轻后续刀具负担, 需要在该孔的位置预先钻削一个底孔^[9]. 针对包含钻削底孔的工件, 为实现刀具钻削路径优化, 本文提出一种带可决策孔的多刀具钻削路径优化模型 (multi-tool DPO with decidable holes, MTdDPO). 在MTdDPO中, 可行解的编码方式复杂, 常规的启发式算法或路径规划算法应用于该模型的优化效果不佳. 为了优化MTdDPO, 本文设计出基于强化学习的分段遗传算法 (segmented genetic algorithm based on reinforcement learning, RLSGA). 在不同的多样性状态下, 种群通过尝试与积累选择最佳的交叉算子以获得刀具的最短路径.

遗传算法 (genetic algorithm, GA) 是解决组合优化问题的常用方法之一, GA被应用于各类DPO模型的求解^[2]. Zhu等^[10]使用多级方法对GA进行改进, 实现了单刀具DPO的优化; Khalkar等^[11]使用GA对带优先约束的多刀具DPO进行求解. 虽然GA在求解问题时收敛速度快, 但是随着迭代的进行, GA种群的多样性往往无法保持, 进化中后期种群搜索趋于缓慢, 容易陷入局部最优. 为了改善GA存在的缺陷, 增强算法的各项性能, 在组合优化领域, 基于强化学习理论的GA得到了广泛的应用^[12], 如: 李润佛等^[13]使用

强化学习自动调整GA的参数, 求解船舶调度问题.

本文首先介绍所提出的MTdDPO模型; 然后介绍所设计的RLSGA算法; 最后, 对比并分析RLSGA与其他4个算法在5个新建的测试问题上的表现, 实验结果表明了该方法的有效性.

1 带可决策孔的多刀具钻削路径优化模型

1.1 实际生产与数学模型

在汽车、航空航天和医疗等常出现的平板类、箱体类零件加工的生产中, 会遇到如下情况: 假设在某个工件上需要加工若干个孔, 需要加工的孔分为3种类型: 类型I (孔径3 mm), 数量若干, 仅需钻削工序加工; 类型II (孔径5 mm), 数量若干, 仅需钻削工序加工; 类型III (孔径12 mm), 数量若干, 需预钻削底孔、扩孔的多道工序加工. 对于I孔和II孔, 使用符合孔径的刀具钻削一次即可完成加工; 对于III孔, 预钻削底孔时, I孔和II孔所使用的刀具都可以作为III孔预钻削底孔时的刀具.

从上述实例可以引申出一个优化问题: 当III孔在预钻削时使用I孔或II孔中的哪一把刀具, 才能使得I孔和II孔的刀具移动路径之和最短? 换言之, 预钻削底孔时使用的刀具是可决策的, 因此, 将需要预钻削的孔定义为“可决策孔”. 根据这一定义, 本文建立带可决策孔的多刀具钻削路径优化模型 (MTdDPO), 其详细描述如下.

假设工件上, 需要加工若干个不同直径、精度或表面粗糙度的孔, 这些孔在MTdDPO中被分为固定孔、可决策孔两类: 固定孔仅需一次钻削成型, 单个工序即可制成; 可决策孔需要经过多道工序, 使用不同的刀具多次加工成型. 在MTdDPO中, 加工过程被分为粗加工和后续加工两个阶段. 粗加工阶段包括固定孔的钻削和可决策孔的底孔预钻削; 后续加工阶段包括可决策孔的其他工序 (如半精加工或精加工工序). 在粗加工阶段, 在对可决策孔的底孔进行预钻削时, 允许使用固定孔中的任意一直径刀具进行加工. 当某个工件上有可决策孔, 固定孔中直径类型数目大于等于2 (即此时至少有2把刀具), 且可决策孔中最小的孔径大于固定孔中最大的孔径时, 孔加工的钻削路径优化模型可视为MTdDPO. MTdDPO的目标是最小化粗加工阶段 (固定孔和可决策孔的底孔加工阶段) 中所有刀具路径的总长度, 故MTdDPO需要解决两方面问题以最小化加工的路径: 1) 粗加工阶段中的可决策孔在预钻削底孔时使用固定孔的哪把刀具, 即可决策孔钻削底孔时的路径归属问题; 2) 同一刀具的孔钻削顺序.

MTdDPO 只涉及粗加工阶段刀具路径的优化, 至于可决策孔的后续加工阶段, 可以建模为基本多刀具 DPO 或单刀具 DPO, 学者们已经开发了多种先进的算法来解决这类问题^[3-5], 因此本文不作讨论。

假设某工件上需要加工 N 个孔, 其中固定孔有 J 种直径类型, 即在可决策孔的底孔预钻削时有 J 把刀具可供选择. 若将粗加工阶段中每把刀具的路径标记为 path_j ($j = 1, 2, \dots, J$), 则 MTdDPO 的目标函数可表达为

$$\min f = \sum_{j=1}^J \left(\sum_{i=1}^{|N_j|} d(i, i+1) + d(|N_j|, 1) \right). \quad (1)$$

其中: $d(a, b)$ 表示孔 a 和孔 b 之间的欧氏距离; $N_j = m_j \cup n_j$ 是路径 j 上所有孔的集合, m_j 和 n_j 分别表示 path_j 上固定孔的集合和可决策孔的集合。

1.2 MTdDPO 的可行解

在 MTdDPO 中, 一个 J 条路径问题的可行解如下式所示:

$$X = \begin{cases} \text{path}_1 : [x_1^1, x_2^1, \dots, x_{|N_1|}^1], \\ \text{path}_2 : [x_1^2, x_2^2, \dots, x_{|N_2|}^2], \\ \vdots \\ \text{path}_J : [x_1^J, x_2^J, \dots, x_{|N_J|}^J]. \end{cases} \quad (2)$$

一个可行解分为 J 段, 每段分别表示在各个路径上的孔的加工顺序。

本文作者设计了一个 14 孔工件^[3,5], 被一些学者用作单刀具 DPO 的测试问题^[1,14]. 本文将此工件修改为带可决策孔的零件模型, 如图 1 所示. 每个孔的编号后面括号内为孔坐标, 其中可决策孔为孔 4、6、7、8、9、11、12. 固定孔的钻削需要两把不同直径的刀具, 即 $J = 2$. 图 1 所展示的可行解为

$$X = \langle \text{path}_1 : [2, 4, 6, 5, 11, 12, 13], \\ \text{path}_2 : [1, 9, 10, 14, 8, 7, 3] \rangle.$$

图 1 中, 可决策孔内圆圈表示在该孔位置预钻削的底孔。

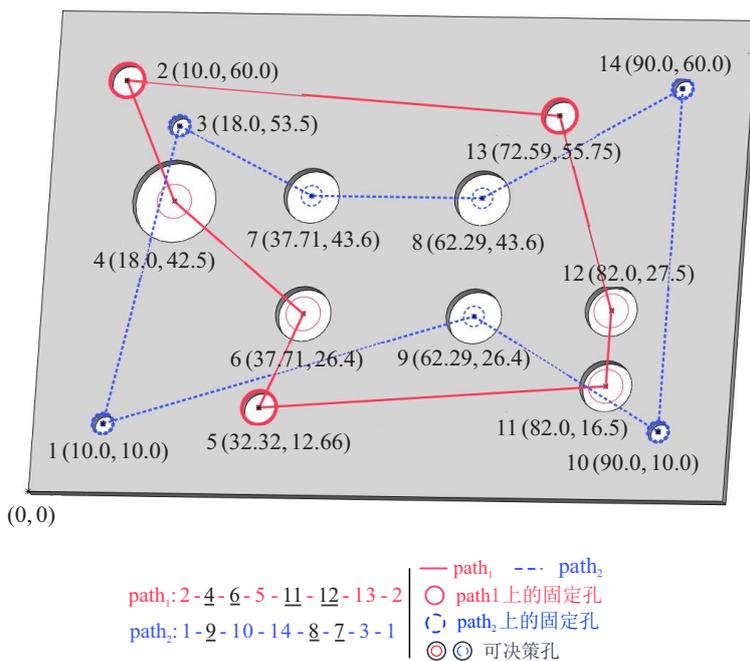


图 1 MTdDPO 的一个可行解

2 基于强化学习的分段遗传算法

GA 将问题的可行解视为染色体, 多个染色体构成种群. 种群在选择、交叉、变异等操作下不断迭代进化直至收敛。

强化学习理论的核心思想是“尝试”与“积累”, 智能体通过根据自身状态和积累的奖励在环境中执行动作, 获取即时奖励并更新积累奖励。

为求解 MTdDPO, 本文为 GA 设计 3 种分段交叉算子, 加入迁移操作, 并基于强化学习, 以种群为智能

体, 设计了状态 (state, s)、动作 (action, a)、即时奖励 (reward, R), 根据种群的状态选择不同的分段交叉算子, 在保证收敛性的同时, 增强种群的多样性。

在 GA 中, 染色体的优劣表达为适应度值 fit (fitness), 适应度值越高, 染色体越好. MTdDPO 的目的是最小化粗加工阶段刀具路径总长度 TL (所有路径长度之和). 对于一个含有 n 个染色体的种群, 设 L_j^i 为第 i ($i = 1, 2, \dots, n$) 个染色体第 j 把刀具的路径长度, 则第 i 个染色体的适应度值为

$$\text{fit}^i = \frac{1}{\text{TL}^i} = \frac{1}{\sum_{j=1}^J L_j^i}. \quad (3)$$

2.1 状态

为了增加GA种群的多样性,在 t 时刻,种群 pop_t 的状态由其多样性系数所决定.种群的多样性系数 div 由下式计算:

$$\text{div}(\text{pop}_t) = \left(1 - 0.5 \frac{t}{t_{\max}}\right) \cdot e^p. \quad (4)$$

其中: t_{\max} 是最大迭代数;系数 p 由下式计算:

$$p = -\frac{(\text{mean}[\text{fit}_t])^2 \cdot \sum_{i=1}^n (D[i] - 1)}{n \cdot (\text{best}[\text{fit}_t] - \text{mean}[\text{fit}_t])}, \quad (5)$$

$\text{mean}[\text{fit}_t]$ 和 $\text{best}[\text{fit}_t]$ 分别是 t 时刻种群适应度值的平均值和最优值, $D[i]$ 表示与第 i 个染色体适应度值相同的染色体的个数(包括自身).多样性系数 $\text{div} \in [0, 1]$, $\text{div} \rightarrow 1$ 时种群的多样性越好;反之,则越差.

定义: $s_t^1: \text{div} \in [0, 0.25)$ 、 $s_t^2: \text{div} \in [0.25, 0.75]$ 、 $s_t^3: \text{div} \in (0.75, 1]$ 是种群在 t 时刻的3个状态.

2.2 动作

可行解的编码方式决定染色体的交叉方式.如式(2)所示,在MTdDPO中,可行解采用分段向量形式进行编码,因此GA应用于MTdDPO的求解时,本文设计了分段交叉算子(segmented crossover, SX),包括3种形式:正序分段交叉算子(positive segmented crossover, PSX)、逆序分段交叉算子(reverse segmented crossover, RSX)和错位分段交叉算子(intersect segmented crossover, ISX).同时,3个交叉算子即为智能体(种群)在任何状态下都可以选择的3个动作.

2.2.1 正序分段交叉算子(PSX)

在RLSGA中,交叉定义为:一个染色体在每段 path_j 上不同位置随机选取一定长度的基因片段与另一染色体的相同位置相同长度的基因片段进行交换.交叉包括多片段交换、重复置空、完备性检测与互补、空位补全4个步骤.染色体的正序分段交叉操作如图2所示.其中: path_1 上的固定孔集合 $m_1: \{1, 2\}$,编码右上角以“ Δ ”符号标记; path_2 上的固定孔集合 $m_2: \{3, 4\}$,编码右上角以“ \circ ”符号标记;未标记的编码是可决策孔.

步骤1:多片段交换.在RLSGA中,每个染色体都由多条路径组成,不同染色体每条路径上孔的数量不尽相等.当对两个染色体执行正序分段交叉操作时,在每个路径上随机截取片段进行交换,如图2(a)所示.为保证片段等长、等位,在对某路径进行片段截

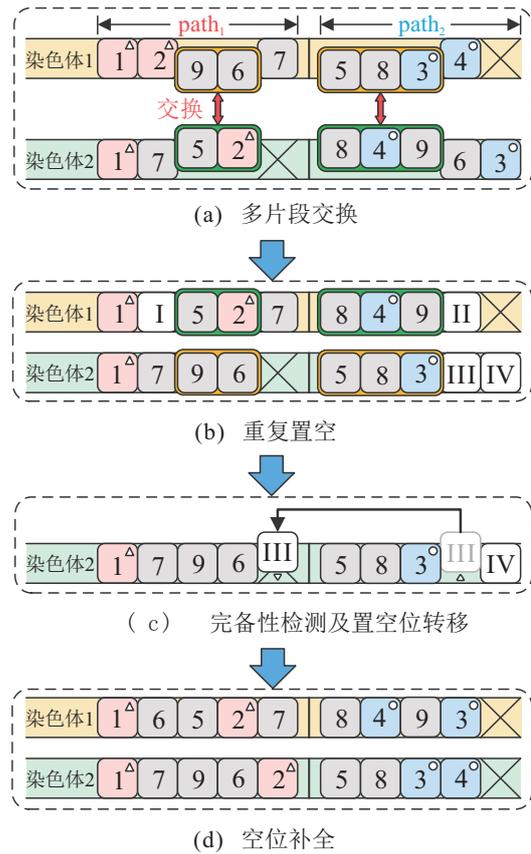


图2 正序分段交叉算子步骤

取时,起、终点从两染色体中该段路径的较短者中随机选取.

步骤2:重复置空.多个片段交换后,交换片段内的编码将固定不动,片段外若有与片段内存在相同的编号,则置空等待补全.在图2(b)中,对于置空的位置以白底罗马数字标记,染色体1在接收来自染色体2的片段 $\{5, 2\}$ 后,因为出现了两个孔2,所以片段外的孔2被置空,标记为“ I ”.同理标记“ II ”“ III ”“ IV ”.

步骤3:完备性检测与置空位转移. m_j 是路径 j 上固定孔的集合,即对于 path_j , $|m_j|$ 是恒定的.基于这一性质,路径的完备性定义为:在置空状态下,若某条路径上尚存的固定孔数量加上置空位数量小于 $|m_j|$,则其为非完备路径;若两者相等,则为完备路径;若前者大于后者,则为过完备路径.非完备路径无法执行步骤4的“空位补全”,因为其空位数量小于需要补入的固定孔的数量.此时,需要不断将随机一条过完备路径上的随机置空位转移至非完备路径,直到变为完备路径.对所有非完备路径进行置空位转移,直到染色体中不存在非完备路径,见图2(c).

步骤4:空位补全.所有未在染色体上的固定孔先补入合理置空位后,对于剩下的可决策孔采用增量最小原则以随机顺序依次补入剩余置空位.图2(d)中,右下角以“ $*$ ”符号标记的编码是经过补全的编码.

2.2.2 逆序分段交叉算子(RSX)

逆序分段交叉与正序分段交叉的区别在于,逆序分段交叉对截取的片段前后倒转后再进行交换. 其余步骤与正序分段交叉相同.

2.2.3 错位分段交叉算子(ISX)

错位分段交叉用于增加种群的多样性. 错位分段交叉在截取片段时,只要求片段等长,不要求片段等位. 在执行错位分段交叉时,在路径较短者上确定片段起、终点后,在较长者上以随机满足条件的起点截取相同长度的片段后进行交换. 其余步骤与正序分段交叉相同.

2.3 即时奖励

环境反馈给智能体的即时奖励能够反映某状态下其执行的动作在本时刻的合理性. RLSGA中,智能体的即时奖励 R 分为3个部分,分别是多样性奖励 R^1 、最佳适应度值奖励 R^2 和平均适应度值奖励 R^3 .

2.3.1 多样性奖励

在GA算法的迭代中后期,有极大的概率使种群中的所有染色体全部同化为同一个染色体,这种现象容易让GA陷入局部最优. 为了在一定程度上抑制种群同化的速度,保持种群的多样性,在RLSGA中加入了多样性奖励. 多样性奖励表达为种群多样性系数 div 的变化, t 时刻多样性奖励由下式计算:

$$R_t^1 = div(pop_t) - div(pop_{t-1}). \quad (6)$$

2.3.2 最佳适应度值奖励

RLSGA的目标是提高种群中最佳适应度值 $best[fit]$,从而在迭代结束时,将其输出作为本次优化的最优解. 因此,若最佳适应度值提高,则需要给予种群正向反馈;否则,将给予种群负向反馈. 在 t 时刻,最佳适应度值奖励为

$$R_t^2 = \begin{cases} +1.5, & best[fit_t] > best[fit_{t-1}]; \\ -0.5, & best[fit_t] = best[fit_{t-1}]; \\ -1.0, & best[fit_t] < best[fit_{t-1}]. \end{cases} \quad (7)$$

2.3.3 平均适应度值奖励

种群的平均适应度值 $mean[fit]$ 反映了整个种群的进化状态和趋势. 在 t 时刻,平均适应度值奖励由下式计算:

$$R_t^3 = \begin{cases} +1.0, & mean[fit_t] > best[fit_{t-1}]; \\ -2.0, & mean[fit_t] = best[fit_{t-1}]; \\ -0.5, & mean[fit_t] < best[fit_{t-1}]. \end{cases} \quad (8)$$

需要强调的是,若种群 t 时刻的平均适应度值与 $t-1$ 时刻的相等,则说明种群极大概率没有发生任何变化,此时将给予较大的惩罚.

智能体在 t 时刻的即时奖励为3部分之和,即

$$R_t = R_t^1 + R_t^2 + R_t^3. \quad (9)$$

2.4 RLSGA学习与迭代流程

综上,在RLSGA中,智能体的学习与迭代流程如图3所示.

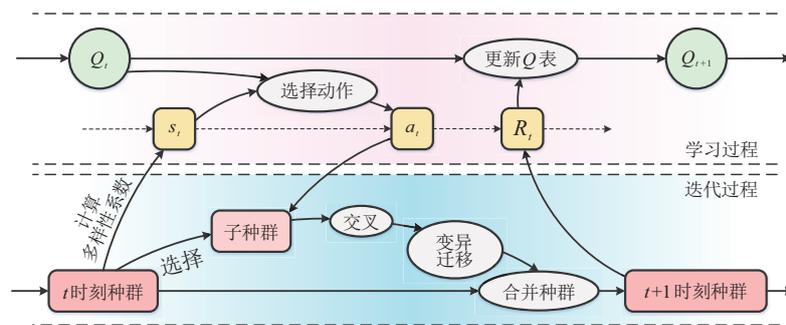


图3 RLSGA算法学习与迭代流程

图3中, Q_t 是 t 时刻的 Q 表,用于存储每个状态下不同动作的累积奖励. Q 表的更新使用 Q -learning 算法,更新时单步采样的动作 a_{t+1} 与 $t+1$ 时刻选取的 a_{t+1} 是相互独立的. Q -learning 算法中, Q 表由下式进行更新:

$$Q_{t+1}(s_t, a_t) = (1 - \alpha)Q_t(s_t, a_t) + \alpha[R_t + \gamma Q_t(s_{t+1}, a_{t+1})]. \quad (10)$$

其中: $\alpha \in [0, 1)$ 是学习率, $\gamma \in [0, 1]$ 是折扣率. Q -learning 算法允许智能体在迭代过程中增量式地学

习,更新累积奖励,从而使得智能体随着迭代的进行,能够更加精准地判断在当前状态下应采取何种动作才能谋取最大的收益.

RLSGA的流程可以概括如下.

- 1) 选择. 采用轮盘赌的方式,从大小为 n 的种群中选出 n 个染色体(可重复)组成子种群.
- 2) 选择动作. 种群根据自身的多样性系数获取状态 s_t 后,根据状态和累积奖励 Q_t 选择一个动作(交叉算子),动作的选择方式采用 ϵ -贪心法,即以 ϵ ($\epsilon \in$

[0, 1]) 的概率随机选择一个动作, 以 $1 - \epsilon$ 的概率选择累积奖励最大的动作。

3) 交叉. 子种群中染色体以 p_C 概率进行交叉。

4) 变异. 子种群中每个染色体以 p_{Mu} 概率发生变异, 变异定义为: 随机路径上的随机两个编码互换。

5) 迁移. 子种群中每个染色体以 p_{Mi} 概率进行迁移, 迁移定义为: 随机路径上某个可决策孔转移至另一路径上的随机位置。

6) 合并种群. 子、父种群共 $2n$ 个染色体, 两者合并后取前 n 个适应度值大的个体作为下一代种群。

7) 获取即时奖励. 根据种群前后代适应度值、多样性系数等关系获取即时奖励 R_t 。

8) 更新 Q 表为 Q_{t+1} 。

3 实验设计与结果分析

对于本文所提出的MTdDPO模型, 设计5个测试问题. 在测试问题上检测RLSGA的有效性。

3.1 测试问题设计

本文设计3种规模共5个测试问题, 包括: 14孔2路径(14-2), 29孔2、3路径(29-2, 29-3), 51孔3、4路径(51-3, 51-4)问题. 14孔问题是根据文献[3, 5]所给单刀具DPO基准测试问题改编而成, 其孔位如图1所示. 所有测试问题的孔类型和位置参见3.3节的图4。

3.2 对比算法及算法参数

粒子群算法PSO^[14]、最优觅食算法OFA^[15]、鲸鱼优化算法WOA^[16]是优秀的启发式算法, 本文使用文献[5]提出的离散算子对3种算法改进以应用于MTdDPO. 特别地, 为了验证强化学习框架对GA的影响, 在对比算法中加入不带强化学习框架的分段遗传算法SGA. 该算法的种群没有状态、即时奖励等属性, 染色体在执行交叉时, 随机选择一个交叉算子执行。

算法的参数设置如下: 在PSO中, 设置 $\omega = 0.3$, $c_1 = c_2 = 0.5$; 在OFA中, $k = 0.9 - 0.5 \frac{t}{t_{\max}}$;

RLSGA中, 设定迭代过程中 $p_C = 0.6$, $p_{Mu} = 0.1$, $p_{Mi} = 0.1$, 设定学习过程中 $\alpha = 0.9$, $\gamma = 0.9$, $\epsilon = 0.1$; SGA的参数与RLSGA迭代过程的参数相同。

5种算法使用同一个随机初始化后的种群作为初始种群进行优化. 5种算法在测试问题14-2、29-2、29-3、51-3、51-4上的种群大小 n 均分别设置为50、80、100、150、200, 最大迭代次数均设置为 $t_{\max} = 2000$. 每种算法在每个测试问题上独立运行30次。

3.3 实验与分析

表1是5种算法在测试问题上独立运行30次所得结果的所有路径总长度的平均值和最优值(最佳值以粗体标记). RLSGA在每个测试问题的平均值和最优值上都取得了最好的效果, 且RLSGA所得的平均值与最优值之间的差距明显小于其他算法, 说明RLSGA的随机性较小, 性能稳定. 在复杂度较低的问题14-2上, 所有算法的每次运行都能成功得到最优解. 当测试问题孔数或路径增加时, 问题的复杂度也随之增加. 在29孔的两个测试问题29-2和29-3上, RLSGA的平均值和最优值都没有表现出明显的优势, 所有算法表现仍差距不大. RLSGA在51孔的两个测试问题上则明显领先其他算法. 总体来看, RLSGA的表现最好, 其次是SGA和PSO, 而OFA和WOA的表现并不理想. 结果表明, RLSGA能够合理且有效地优化MTdDPO模型。

图4是RLSGA在5个测试问题上最优解所对应的刀具路径, 图例中“固定孔1”指的是 $path_1$ 上的固定孔集合 m_1 , 其余路径同理。

5种算法在29-2和29-3上获得最优解的迭代曲线如图5所示. 从图5中可以看出, 在这两个问题上, 所有算法的收敛值差距很小. 在处理29-2时, RLSGA、SGA和PSO的收敛速度较快, 且RLSGA和SGA几乎同时在迭代达到100时完成了收敛. OFA和WOA的收敛速度较为接近. 在29-3的求解上, 所有算

表1 算法运行结果对比

问题	总长度平均值					总长度最优值				
	RLSGA	PSO	OFA	WOA	SGA	RLSGA	PSO	OFA	WOA	SGA
14-2	436.73	436.73	436.73	436.73	436.73	436.73	436.73	436.73	436.73	436.73
29-2	1142.21	1172.65	1157.19	1153.49	1174.30	1135.14	1140.51	1135.42	1137.38	1143.94
29-3	1157.79	1191.48	1169.29	1167.91	1193.15	1152.36	1153.17	1152.36	1152.36	1168.33
51-3	2391.05	2722.75	3172.52	3221.55	2576.97	2357.55	2570.89	2906.38	2989.91	2446.75
51-4	2698.22	2981.39	3280.44	3278.75	2928.78	2659.97	2829.62	3099.60	3156.89	2812.12

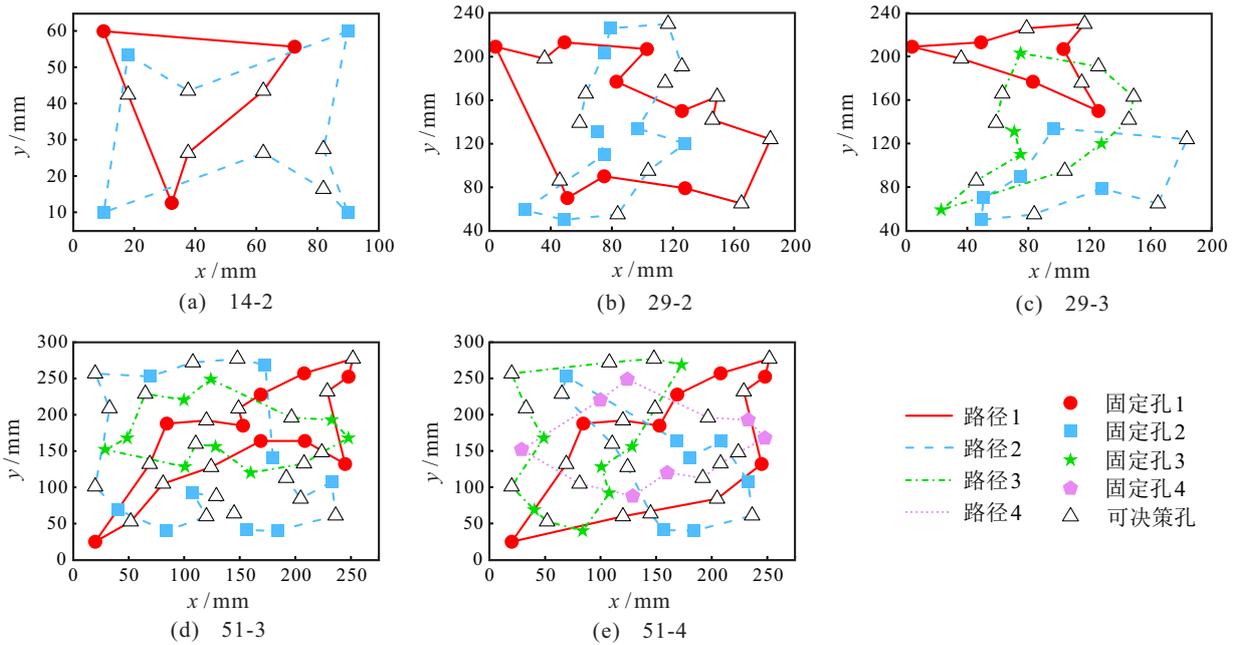


图4 RLSGA在测试问题上所得的最优路径图

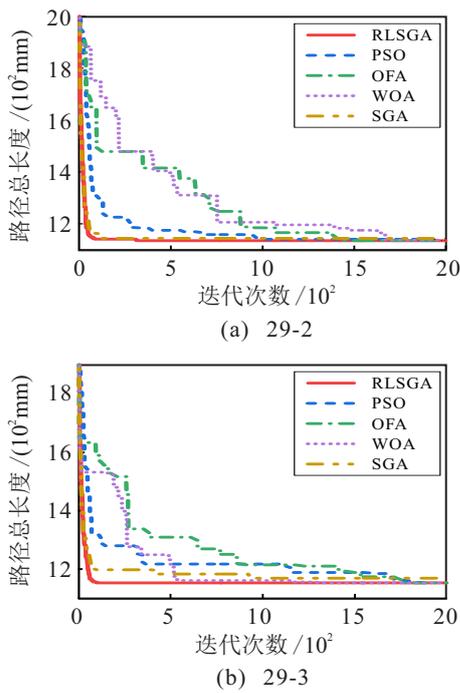


图5 29孔问题迭代曲线

法的收敛速度的差距不大.

51-3和51-4的迭代曲线如图6所示,在51孔的两个测试问题上,相较于其他3种算法,RLSGA表现出了明显的优势. RLSGA和PSO在51-3上经过迭代1000次左右收敛到最优值. SGA在51-3上明显陷入局部最优,在迭代末期才跳出局部最优,找到更优解. 而在51-4上,SGA虽然收敛速度快于PSO,且收敛曲线更为平滑,但最终收敛结果并没有明显优于PSO. 从表1的总长度平均值和最优值可以看出,SGA和PSO在51-4上的表现基本一致. 在两个测

试问题上, OFA和WOA在迭代2000次时的曲线表明,种群仍然处于搜索状态,并没有收敛.

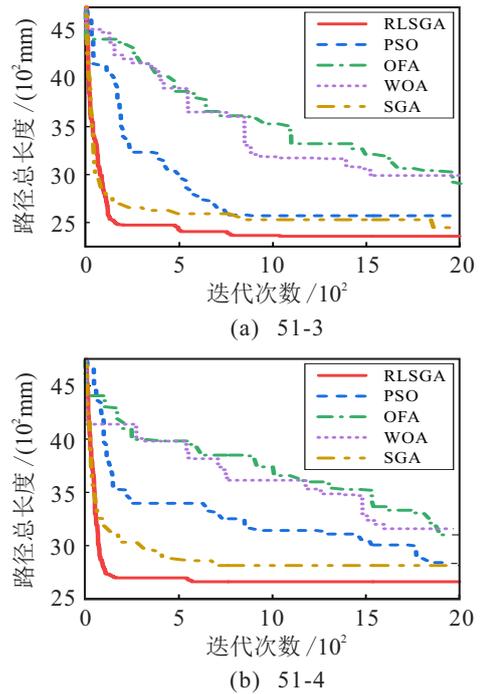


图6 51孔问题迭代曲线

根据RLSGA和SGA在29孔问题和51孔问题上的表现可以总结出以下结论:

1) RLSGA算法的性能好于所有比较算法, SGA的表现总体上优于除RLSGA以外的其他启发式算法,说明遗传算法在优化MTdDPO时表现良好,与处理其他组合优化问题时的表现相当.

2) 从图5和图6可以看出,迭代前期,RLSGA的收敛速度略微逊色于SGA,尤其在51-3问题上最为

明显,这是在强化学习框架的即时奖励中的多样性奖励 R^1 设计时所期望的结果——牺牲部分收敛速度以增强种群的多样性。可以看出,RLSGA 在各个问题上的收敛曲线都相对平滑,不会出现明显的断层式曲线,说明算法较少陷入局部最优或陷入局部最优后能及时跳出。

4 结论

针对现实生产活动中的孔加工过程,本文提出了一种带决策的孔加工多刀具路径优化问题模型(MTdDPO),该模型将工件上的孔分为固定孔与可决策孔。为了实现MTdDPO的优化,本文提出了基于强化学习的分段遗传算法(RLSGA),该算法使得种群能够通过尝试与积累,在不同的多样性系数状态下执行不同的分段交叉算子以获得最大的收益。将RLSGA与其他4种算法在所设计的MTdDPO的5个测试问题上进行了比较,结果表明,RLSGA的表现明显优于其他算法,RLSGA能够有效解决MTdDPO问题。

在未来的研究中,应对MTdDPO模型进一步完善,考虑更多现实生产活动中存在的因素,并设计更多规模的测试问题。

参考文献(References)

- [1] Dewil R, Küçükolu İ, Luteyn C, et al. A critical review of multi-hole drilling path optimization[J]. Archives of Computational Methods in Engineering, 2019, 26(2): 449-459.
- [2] Abidin N W Z, Rashid M F F A, Mohamed N M Z N. A review of multi-holes drilling path optimization using soft computing approaches[J]. Archives of Computational Methods in Engineering, 2019, 26(1): 107-118.
- [3] Zhang W B, Zhu G Y. Drilling path optimization by optimal foraging algorithm[J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 2847-2856.
- [4] Hatem N, Yusof Y, Kadir A Z A, et al. A novel integrating between tool path optimization using an ACO algorithm and interpreter for open architecture CNC system[J]. Expert Systems with Applications, 2021, 178: 114988.
- [5] Zhu G Y, Zhang W B. Drilling path optimization by the particle swarm optimization algorithm with global convergence characteristics[J]. International Journal of Production Research, 2008, 46(8): 2299-2311.
- [6] Ilker K, Tulin G, Fatma B, et al. Application of precedence constrained travelling salesman problem model for tool path optimization in CNC milling machines[J]. An International Journal of Optimization and Control: Theories & Applications: IJOCTA, 2019, 9(3): 59.
- [7] Sarin S C, Sherali H D, Judd J D, et al. Multiple asymmetric traveling salesmen problem with and without precedence constraints: Performance comparison of alternative formulations[J]. Computers & Operations Research, 2014, 51: 64-89.
- [8] Kolahan F, Liang M. Optimization of hole-making operations: A tabu-search approach[J]. International Journal of Machine Tools and Manufacture, 2000, 40(12): 1735-1753.
- [9] Aamir M, Giasin K, Tolouei-Rad M, et al. A review: Drilling performance and hole quality of aluminium alloys for aerospace applications[J]. Journal of Materials Research and Technology, 2020, 9(6): 12484-12500.
- [10] Zhu G Y, Chen L F. Holes machining process optimization with genetic algorithm[J]. Key Engineering Materials, 2011, 460/461: 117-122.
- [11] Khalkar S, Yadav D, Singh A. Optimization of hole making operations for sequence precedence constraint[J]. International Journal of Innovative and Emerging Research in Engineering, 2015, 2(7): 26-31.
- [12] Mazyavkina N, Sviridov S, Ivanov S, et al. Reinforcement learning for combinatorial optimization: A survey[J]. Computers & Operations Research, 2021, 134: 105400.
- [13] 李润佛, 张新宇, 李俊杰, 等. 基于强化学习的自学习遗传算法在船舶调度中的应用[J]. 大连海事大学学报, 2022, 48(3): 20-30.
(Li R F, Zhang X Y, Li J J, et al. Application of self-learning genetic algorithm based on reinforcement learning in ship scheduling[J]. Journal of Dalian Maritime University, 2022, 48(3): 20-30.)
- [14] Poli R, Kennedy J, Blackwell T. Particle swarm optimization[J]. Swarm Intelligence, 2007, 1(1): 33-57.
- [15] Zhu G Y, Zhang W B. Optimal foraging algorithm for global optimization[J]. Applied Soft Computing, 2017, 51: 294-313.
- [16] Mirjalili S, Lewis A. The whale optimization algorithm[J]. Advances in Engineering Software, 2016, 95: 51-67.

作者简介

朱光宇(1970—),男,教授,博士,从事车间调度、智能制造、优化设计等研究, E-mail: zhugy@fzu.edu.cn;

张德颂(1998—),男,硕士生,从事智能制造、优化设计等研究, E-mail: z.desong@foxmail.com.