基于分层强化学习的通用装配序列规划算法

赵铭慧[1,2], 张雪波[1,2†], 郭 宪[1,2], 欧勇盛[3]

(1. 南开大学 机器人与信息自动化研究所，天津 300350；2. 天津市智能机器人技术重点实验室，天津 300350；
3. 中国科学院 深圳先进技术研究院，广东 深圳 518055)

摘 要: 对于装配序列规划问题,现有算法大多聚焦于单一的目标构型. 对于多目标构型以及大规模问题,现有算法往往存在维数灾难及泛化能力差等问题. 为此,利用装配序列规划问题分层结构的特点,提出一种基于分层强化学习的适用于多构型装配任务的通用装配序列规划方法. 首先,将装配序列规划问题构建为一个分层的马尔科夫决策过程,其中,上层进行序列规划,下层进行零件的动作规划,符合装配过程层次化的结构,使规划方法更具灵活性,且可解释性更强;其次,针对分层马尔科夫决策过程,提出一种基于分层强化学习的通用装配序列规划算法,提高规划方法对多种目标构型任务的适应能力和泛化能力,以及对目标构型的信息利用率;最后,在搭建的仿真平台上进行验证,结果表明所提方法可以提取到关于装配问题的广义信息,对于不同零件初始位置以及其他多种构型装配任务均具有较好的决策能力,从而验证所提方法的有效性和通用性,表明该算法是适用于多目标构型的更加通用灵活的装配序列规划算法.

关键词: 智能装配；装配序列规划；深度强化学习；目标导向；分层强化学习；多构型任务

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2020.1289 **开放科学(资源服务)标识码(OSID):**

引用格式: 赵铭慧,张雪波,郭宪,等: 基于分层强化学习的通用装配序列规划算法[J]. 控制与决策, 2022, 37(4): 861-870.

A general assembly sequence planning algorithm based on hierarchical reinforcement learning

ZHAO Ming-hui[1,2], *ZHANG Xue-bo*[1,2†], *GUO Xian*[1,2], *OU Yong-sheng*[3]

(1. Institute of Robotics and Automatic Information System，Nankai University，Tianjin 300350，China；2. Key Laboratory of Intelligent Robotics，Tianjin 300350，China；3. Shenzhen Institutes of Advanced Technology，Chinese Academy of Sciences，Shenzhen 518055，China)

Abstract: For assembly sequence planning problems, most of the existing algorithms focus on a single target configuration. For multi-target configurations and large-scale problems, existing algorithms often have dimension disaster problems with poor generalization ability. Therefore, this paper uses the characteristics of the hierarchical structure of assembly sequence planning problems and conducts a general assembly sequence planning method based on hierarchical reinforcement learning, which is suitable for multi-configuration assembly tasks. First of all, this paper constructs the assembly sequence planning problem as a hierarchical Markov decision process, in which the upper layer performs sequence planning, and the lower layer carries out workpiece motion planning, which conforms to the hierarchical structure of the assembly process, making the planning method more flexible and interpretable. Then, in view of the hierarchical Markov decision process, this paper proposes a general assembly sequence planning algorithm based on hierarchical reinforcement learning, which improves the adaptability and generalization ability of the planning method to multi-target tasks and the information utilization of the target configuration. Finally, the proposed method is verified on the built simulation platform. The results show that the proposed method can extract general information about assembly problems, and it has good decision-making ability for any initial state and other various configurations assembly tasks, which verifies the effectiveness and flexibility of the method. Thus, a more general and flexible assembly sequence planning algorithm suitable for multiple configurations is realized.

Keywords: intelligent assembly; assembly sequence planning; deep reinforcement learning; target-oriented; hierarchical reinforcement learning; multi-configuration

收稿日期: 2020-09-15；**录用日期:** 2020-12-25.

基金项目: 国家自然科学基金项目(U1613210)；天津市杰出青年科学基金项目(19JCJQJC62100)；天津市自然科学基金项目(19JCYBJC18500)；中央高校基本科研业务费专项基金项目.

责任编委: 侯忠生.

†通讯作者. E-mail: zhangxuebo@nankai.edu.cn.

0 引 言

随着中国制造2025和工业4.0的提出,新一代信息技术与制造产业的创新融合与应用逐渐走入人们的视野,而作为中国制造2025五大工程之一的智能制造工程已成为近几年研究的热点. 在智能制造工程内容中,要求各大企业紧扣关键工序智能化、关键岗位机器人代替,建设重点领域的智能化工厂[1],可见,开展智能装配研究势在必行. 智能制造的根本目标是显著提升制造业领域的智能化水平,缩短产品生产周期,降低试点项目的运营成本. 伴随着对高精尖产品性能要求的不断提升,工业产品如船舶、汽车和航空航天产品等相较以往设计更加复杂,组装和生产过程更加困难[2-3],因此,对于工业产品生产过程中的装配序列规划研究迫在眉睫.

装配过程作为工业生产制造中的一个重要环节,被视为工业产品生产过程中的瓶颈阶段,对生产效率及产品质量有着较大影响[4]. 得到一组正确的装配工序是确保工业产品能够准确快速组装成功的关键.

进行装配序列规划需要在满足所有装配条件及要求的前提下,对装配过程中的装配工件选择和装配顺序进行规划,使复杂装配系统具有智能的决策能力,并根据优化指标等因素,得到满足条件的合理工序[5]. 进行装配序列规划可以从根本上提高装配效率,保证生产的可靠性和准确性,降低产品的开发成本.

为了提高规划算法的通用性以及对目标构型的信息利用率,针对复杂产品的多种构型装配序列规划问题,本文提出基于分层强化学习的目标导向型序列规划算法,将目标构型作为输入,使网络模型具备目标驱动的性质. 设计分层任务的框架,将序列规划问题转换为双层规划问题,实现灵活通用的装配序列规划算法,使其适用于任意初始装配状态与不同目标构型的装配任务.

1 研究现状

针对装配序列规划问题,现有的研究主要集中在以下3大类方法:基于图的序列规划方法[6-8]、基于知识的序列规划方法[9-10],以及一些启发式的规划算法[11-15]. 基于图的序列规划方法通过有向图来形式化编码可行的装配序列空间;基于知识的序列规划方法利用拓扑关系或其他关系描述装配对象的信息以及一些先验知识,以此建立多层模型结构以实现智能装配;而启发式规划方法则应用人工智能技术中的搜索算法,在线搜索优化,使用遗传算法或模拟退火算法解决零件的装配任务. 此3类规划方法中,基于图的序列规划方法提出较早、使用较多,现已发展成为一种比较成熟经典的装配序列规划算法.

基于图搜索的规划算法[7]如图1所示,其计算速度与关联图的稀疏性以及装配工件的数量有很大关系. 另外,此算法需要在每次启动时构建完整的装配关联图,当模型初始状态不同时必须重新构建装配图,不能适用于任意的装配模型状态. 解决不同类别的问题时,需要输入此模型对应的信息,构建所对应的装配关联图以及and/or图,每一个新的装配问题都需要重新规划,泛化能力较弱.

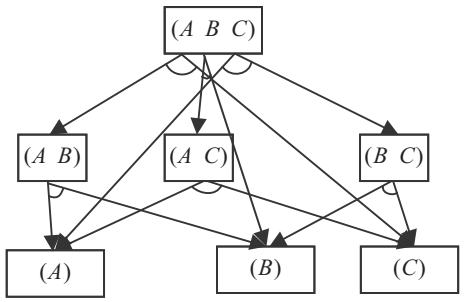

图1 基于图的序列规划方法[7]

基于知识的规划方法对先验知识或信息非常敏感,只适用于同类型的装配对象,规划结果会受到特定知识的严重限制. 对于启发式搜索规划算法, 计算量往往比较大且收敛速度受限,泛化能力不足.

随着计算资源和能力的不断提高,深度强化学习算法已经在各种游戏与博弈、机器人控制领域等序贯决策类问题中取得了十分优秀的成果. 但在智能装配方向上更多侧重于使用机械臂进行单个零件的抓取及安装的问题[16-17],在装配序列规划问题上还鲜有相关深入研究. 对于装配工序规划这类典型的序贯决策问题,利用无模型的深度强化学习方法可以通过与模拟器交互得到数据,并通过神经网络的强大知识表征能力处理大量数据[18-19]. 此类方法无需建模,并且可以通过学习获取关于所解决问题的广义信息,具有较高的泛化能力.

本文在之前的工作中[20],基于多零件、多可行序列的复杂模型进行初步的装配序列规划问题的研究,提出一种高效率的适用于任意初始装配状态的装配序列规划方法,但此方法只适用于单一目标构型的装配模型,通用性还有待提高. 因此,考虑到装配序列规划问题所具有的分层结构的特点,在本文中,提出一种新型的基于分层强化学习的通用序列规划方法,以此更好地解决复杂模型的装配序列规划问题,提高此规划算法对多构型任务的通用性和泛化能力.

2　问题构建与形式化

2.1　问题描述与构建

装配序列规划问题作为一种典型的序贯决策问题,目标是要在每个时间状态下,给出一个最佳的决策动作,从而形成一条最优的决策序列,即最优工序序列. 对于这种很难预先建立精确数学模型的问题,本文采用强化学习算法中无模型的方法,通过直接与环境交互获得所需的学习数据. 因此,本文首先给出装配序列规划问题的描述与构建,并在后续内容中对此问题进行形式化表示.

近些年来,随着人工智能方法的不断发展,研究者们也提出了一个针对序贯决策问题的标准模型,即马尔科夫决策过程(Markov decision process, MDP)[19],MDP可以解决大部分强化学习问题. 装配序列规划问题是一种典型的序贯决策问题,可以用马尔科夫决策过程描述.

强化学习的目标就是对于一个马尔科夫决策过程,找到最佳动作策略,实现智能化. 动作策略即在给定状态时动作集上的概率分布,一般用π表示动作策略,即

$$\pi(a|s)=p[A_t=a|S_t=s]. \tag{1}$$

在装配序列规划问题中,其目的是在每一个状态下确定是否应安装某个部件,以及应按特定顺序将其安装在何处. 在传统的强化学习算法中,智能体与环境交互,并根据所获得的回报奖励来学习更新自己的行为策略,所以此类方法往往通过提高累积回报值找到最优的序列,通过调整神经网络参数θ来最大化采样轨迹的累积回报J,即

$$J(\theta)=E_{s_0,a_0,\tau}\Big[\sum_{k=0}^{\lim}\gamma^k R_{t+k+1}\Big]. \tag{2}$$

其中:采样轨迹τ表示为从初始状态s_0选择动作a_0,直至k个时间步后终止状态的序列.

2.2　装配序列规划问题的形式化表示

状态空间设置S:为了满足问题描述的马尔科夫性,在本文的装配序列规划问题的形式化表示中,状态空间采用了图像的表示方法. 由于引入了目标导向的方法提高通用性,输入的状态表示中加入了目标构型. 对应的状态表示分为3部分,即

$$s_t=[x_{1t},x_{2t},x_{3t}]. \tag{3}$$

首先,状态表示(如图2所示)的第1部分x_{1t}代表当前装配区域的状态,是装配区域上方的相机实时返回的图像. 状态表示的第2部分x_{2t}是下一步即将进行决策的零件的二维图像,代表下一步即将产生的改变. 第3部分x_{3t}则是一张代表目标构型的图像. 这样的表示方法既提供了一些关于目标的直接信息,又可以使智能体通过当前状态与目标构型的差异,对所选择的零件进行运动规划,高效地学习动作策略.

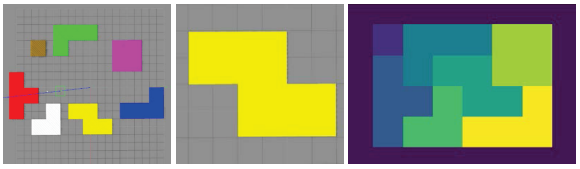

图2　状态空间设置

动作空间设置A:在通用装配序列规划方法研究中,为了使智能体的动作更加灵活通用,本文将动作设置为4个基本动作,相较于装配到某一位置的动作,具有较深一层的规划含义. 动作a_t表示为

$$a_t=\{0,1,2,3\}. \tag{4}$$

动作0、1、2和3分别代表对于所选择的零件,在当前所处位置的基础上向上、向下、向左以及向右移动一个单位长度.

回报函数设置R:如果装配网络能在预先设定的决策步数内正确安装所有零件,成功完成装配任务,则认为决策序列是正确的,会得到+1的奖励,若没有完成,则会得到−1的惩罚,即

$$r_t=[-1,1]. \tag{5}$$

目标函数:对于装配序列规划的序贯决策问题,目标函数为采样轨迹上的总累积回报. 使用深度Q网络(DQN)算法时,以θ为参数的损失函数$L_i(\theta_i)$可表示为

$$L_i(\theta_i)=E_{s,a\sim\rho}[(y_i-Q(s,a;\theta_i))^2]. \tag{6}$$

其中:y_i是目标值,通过对抽取出的数据进行小批量更新的训练,不断减小y_i与网络预测值$Q(s,a;\theta)$之间的差值,对策略进行更新改进.

3　基于分层强化学习的通用序列规划算法

对于复杂装配模型的装配序列规划这类大规模的问题,使用一般的强化学习方法直接学习难以取得好的效果,并且装配序列规划问题内部是具有一定层次结构的,包括零件的选择以及底层的规划问题. 因此,可以使用分层强化学习的方法解决. 本文实现了一种基于分层强化学习的通用序列规划算法,上层进行序列规划,下层进行零件的动作规划,提出了较新颖的状态空间和动作空间等表示方式,使其更具灵活性. 下面将从目标导向设计、分层结构各层功能、两层结构具体设置几方面对所提算法进行详细介绍.

3.1 目标导向的装配序列规划算法

对于一般的深度强化学习算法,在经过训练后可以很好地解决所构建的问题,但是对于这一组问题之外的新任务,往往较难胜任. 在本文之前的工作中[20]使用的深度网络结构,是与监督学习中标准的分类网络结构十分类似的深度网络结构. 文献[21]将使用这种结构的网络学习出的策略称作反应式策略,此反应式网络如图3所示: 在训练好此深度网络后,对于某个特定的输入,由于网络权值参数固定,其经过网络后的输出也是固定的,表现出一定固有的反应式,通用性较差.

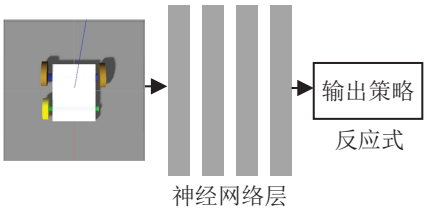

图 3 反应式网络结构

基于此,本文设计了一种目标导向的序列规划算法,使用目标驱动的模型结构[22],将目标构型作为额外的输入,显式地输入到网络中, 这样可以使网络模型直接接收目标所具有的信息,可以避免对每一组新的任务都重新训练一套网络模型. 对于目标导向的模型结构,策略的学习不仅仅依赖于当前状态,还取决于目标的构型. 这样的结构可以提高模型的灵活性和通用性,增强其泛化能力,适用于更多组不同的任务.

3.2 基于分层强化学习的通用装配序列规划方法

进行通用的装配序列规划意味着问题难度的增加,因此必将伴随着状态动作空间维度的增加,从而导致参数量的增加,传统的强化学习方法会遭遇维度灾难的打击,难以得到好的学习效果. 考虑到装配序列规划问题所具有的分层结构特点,本文在目标导向的基础上又设计了分层学习的通用算法,以此提高算法对大规模问题的解决效果.

文献[23]提出了MAXQ价值函数分解的思想,这种方法目前已经成为分层强化学习领域最重要的基础算法之一. 此方法将一个马尔科夫决策过程分解成多个子任务,使强化学习算法也可以应用于计算大规模的问题,并且随着近年来机器计算能力的提升,此方法已广泛应用于很多领域[24]. 本文设计的基于分层强化学习的通用装配序列规划算法也借鉴了这种方法的思想,设计了子任务,将一个复杂的强化学习问题分解成一些子问题并分别解决,以取得比直接解决整个问题更好的效果. 此分层结构的框架如图4所示.

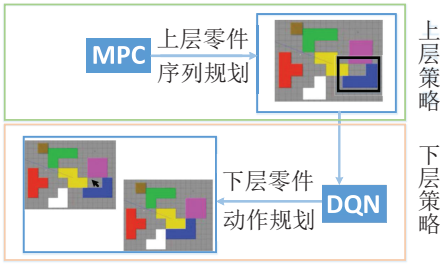

图 4 分层结构

针对装配序列规划问题的双层结构,本文也进行了分层的学习. 将装配序列规划转化为两层任务: 上层为序列规划,旨在选择正确的零件; 下层进行运动规划,对每一个上层选择的零件进行运动规划,学习在每个状态下应有的动作. 从而得到一整条满足条件且通用的装配序列.

为了实现装配序列规划任务,提高整体序列规划算法的效果,在算法训练过程中,首先对下层零件动作规划网络进行训练,对于不同的零件,采集其下层运动部分的数据,使用本文设计的基于孪生网络的目标导向深度网络架构进行训练,使其可以无碰撞地到达目标构型中对应的位置,为上层进行零件的序列规划提供基础. 然后进行上下层的分层规划学习,在进行上层策略选择的同时优化下层网络. 这样的分层结构回报设置会更加明确,也更符合装配序列中“序列”的概念.

下面将详细介绍此分层结构中各层的具体算法设计与实现.

3.2.1 基于MPC思想的上层序列规划方法

通过上一节的算法设计,本文将装配序列规划任务分解成为两个阶段的子任务. 其中,上层负责第1阶段的任务: 根据当前的状态选择此刻应当安装的零件. 对于这一部分任务,本文使用了模型预测控制(model predictive control, MPC)[25]的思想,提出一种对未来一段时间内的输出进行预测,然后滚动进行在线优化的上层规划方法. 目标函数J代表在k时刻内所得到的奖励分值的期望,即

$$J = E_{\pi}\left[\sum_{0}^{k}(R+\mathrm{sim})\right]. \tag{7}$$

因此,最佳策略就是选择合适的动作来最大化目标函数值

$$\pi : a = \arg\max J. \tag{8}$$

针对基于MPC思想的上层序列规划,在某一个状态下,要求智能体遍历每一个动作,在这里意味着尝试选择每一个零件,对每一个动作进行前向模拟,

按照下层模型的预测结果进行此零件的运动规划,直至到达下层任务的终点. 在此模拟过程中得到各个零件在进行下层模拟时收获的回报值 R,以及模拟后状态与目标构型的相似度值 sim,将两部分加和并做 Boltzmann 分布作为此动作的分值,并以此值来贪婪地选择策略.

$$S = R + \text{sim}, \tag{9}$$

$$B_i = \frac{\exp\left(-\dfrac{S_i}{kT}\right)}{\sum\limits_j \exp\left(-\dfrac{S_j}{kT}\right)}. \tag{10}$$

其中: S_i 是归一化前动作对应的分值, k 是波尔兹曼常数, T 是一个可以调节的参数, B_i 是归一化后对应的分值. 通过加入 Boltzmann 分布的变换,对每个动作的分值进行归一化,并提高探索率,使得不同动作之间的优劣评判更加直观. 上层规划流程如图 5 所示.

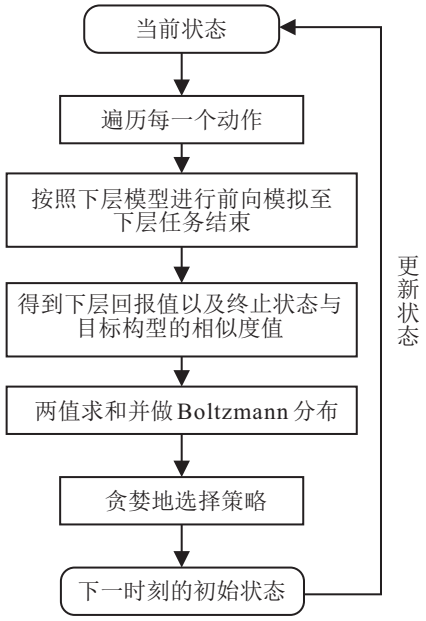

图 5　基于 MPC 思想的上层序列规划过程

3.2.2　基于 DQN 的下层动作规划方法

3.2.1 节介绍了基于分层强化学习的通用装配序列规划方法中的第 1 阶段,即上层序列规划部分,本节将介绍其中的第 2 阶段,也是下层子任务的规划部分.

在基于分层强化学习的通用装配序列规划方法的下层任务中,主要目标是对于上层所选择的零件进行运动规划,根据当前状态的图像表示,提取出与目标构型之间的差异,确保快速且无碰撞地到达目标构型中的对应位置,完成装配任务. 此任务的构建以及对应的马尔科夫决策过程的具体设置可见 2.2 节的形式化表示.

本文采用基于 DQN 的深度强化学习算法[26],学习在某一个状态下,每个动作所对应的行为——值函数 $Q(s,a)$. 在采集训练数据后,从经验回放池中抽取训练数据,然后使用小批量梯度下降的方法优化目标函数. 对于使用 TD 方法的 DQN 算法,损失函数 L_i 如前文式(6)所示.

$$y_i = E_{s'\sim\xi}[r + \gamma \max_{a'} Q(s', a'; \theta_{i-1})|s, a]. \tag{11}$$

其中: θ_i 是当前网络参数; y_i 是 TD 目标,使用当前回报和下一时刻的值函数估计近似代替当前状态下的累积回报,值 y_i 在这里可看作为 Q 的真值. 使用 DQN 算法可以在完成下层的动作规划任务的同时,提高所提规划算法的通用性和泛化能力,适用于更多同类的问题.

4　仿真设计与结果分析

4.1　仿真平台设置

本文使用机器人操作系统 (robot operating system, ROS) 下的 Gazebo 仿真软件,并且引入 Python 接口,搭建了 Gazebo+TensorFlow 的深度网络学习框架. 首先搭建用于装配序列规划问题的训练平台,如图 6 所示,以完成交互数据采集和装配过程可视化的任务.

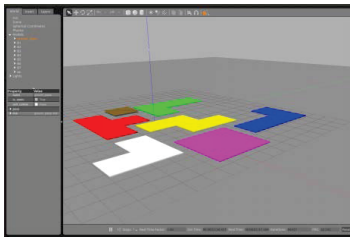

图 6　仿真平台

4.2　仿真设计与形式化

本文设计了一个类似于七巧板结构的 7 零件模型 (图 7) 作为本文仿真的装配对象. 此模型零件的形状比较特别,零件间的差异十分明显,不同零件之间会有多种组合关系,相互接触关系特别复杂,是一个典型的验证装配规划的测试案例.

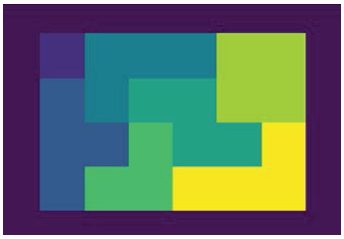

图 7　本文使用的 7 零件装配模型

状态设置: 在进行通用的问题形式化时,本文在状态表示部分加入了目标构型,将 3 部分灰度图像进

行堆叠,得到大小为80×80×3的最终状态表示.

动作设置:动作集使用了较为灵活的4个基本动作,不局限于装配类问题. 4个动作用索引{0, 1, 2, 3}表示,分别代表所选择的零件向上、向下、向左或是向右移动一个单位长度.

回报设置:在本文的下层动作规划中,主要研究零件如何无碰撞并快速到达目标构型中的对应位置,所以回报函数设计也将从这几个角度进行.

在下层的规划过程中,若零件与环境中的其他零件产生碰撞,则会立即终止,并获得−1的惩罚. 在每个零件进行下层规划之前,会使用SSIM(structural SIMilarity)结构相似度[27]评价方法对初始状态和目标构型进行相似度的衡量.

将初始的相似度值作为本次规划的基准值,之后每步的立即回报都是当前相似度值与基准相似度值的差值,差值为正代表所选动作是向装配完成的方向移动的. 使用相似度作为回报值,可以适用于很多同类的问题,更加通用. 此外,由于加入了立即回报,可以解决深度强化学习经常面临的延时回报问题,利于网络更快地学习到更优的策略.

网络结构设置:在引入目标导向思想后,为了能更高效地提取出当前状态与目标状态之间的差异,本文使用了孪生网络的网络结构,如图8所示. 针对本文设置的通用状态表示,设计了三支流的深度网络架构,对当前状态、目标构型以及待决策零件此3部分输入使用权值共享的神经网络层,映射到相同的特征空间,使得当前状态与目标状态之间的相似度度量更加直观.

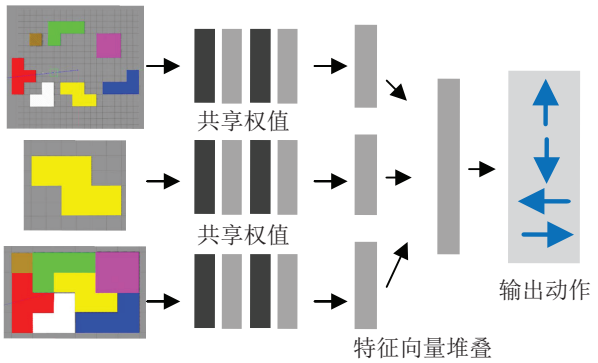

图 8 网络结构

在共享权值部分得到每一维输入在相同特征空间的特征向量表示后,将此3个特征向量堆叠,通过全连接层得到最后的输出,输出层的节点个数为4,分别代表4个动作所对应的行为——值函数,用作对各个动作的评估. 权值共享部分具体的网络设置如表1所示.

表 1 权值共享部分网络各层设置

层	核	特征图
input		80×80×1
conv1	[8, 8, 1, 32]	20×20×32
pool1	[1, 2, 2, 1]	10×10×32
conv2	[4, 4, 32, 64]	5×5×64
conv3	[3, 3, 64, 64]	5×5×64
fc1	[5×5×64, 1 024]	1 024
fc2	[1 024, 512]	512

课程学习:受到课程学习方法的启发,在本文的仿真中,也将设计不同难度的装配序列规划任务,以测试所学装配模型是否可以胜任于不同等级难度的问题. 任务的难度将从以下3个方面体现:

1)零件初始位置与正确装配位置之间的距离,此距离对应着分层结构中的下层,即基于DQN的下层动作规划部分的难度. 距离越大时,下层动作规划的过程中发生碰撞或难以到达的可能性就越大,装配任务的难度也越大.

2)零件的初始位置是否随机. 若零件初始位置随机,则每一次装配初始状态候选空间巨大,在训练时难以覆盖. 此时,对上、下层任务都具有较高的难度.

3)是否适用于多种目标构型. 为了验证本文所提方法的通用性,本文使用仿真模型所包含的7个零件设计了8种不同的目标构型,部分目标构型展示如图9所示,测试所学网络模型是否可以挖掘并学习到一些广义的装配知识并适用于多种构型的装配模型.

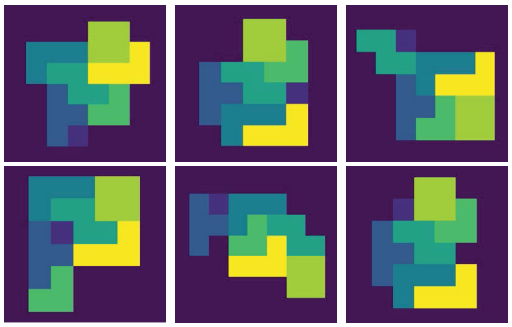

图 9 多种目标构型

4.3 仿真结果分析

对于4.2节所提出的不同难度体现,本节设计了如下3个不同级别难度的装配序列规划任务,以验证所提的基于双层结构的通用装配序列规划方法的有效性和通用性.

实验1(低难度案例) 首先对于零件初始位置与正确装配位置之间距离较近的情况,如图10左图所示,每个零件经过5步以内的运动规划即可到达正确装配位置.

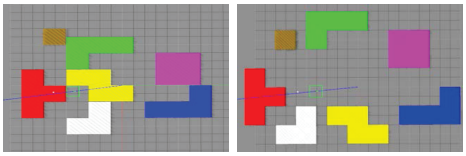

图 10　难度低、中高的初始状态

当使用端到端的DQN方法进行训练时,对装配网络模型的要求较高. 在训练结束后,生成100个场景进行测试,端到端的DQN方法完成装配任务的成功率较低,如表2所示,只有27 %.

表 2　低难度案例使用不同方法装配效果对比

	端到端DQN	分层MPC + DQN
装配成功率/%	27	100

由图11可以看出,使用端到端的DQN方法最终成功安装上的零件个数主要为4～7个,而本文所提的分层规划方法对于此类难度的装配序列规划任务可以100 %完成.

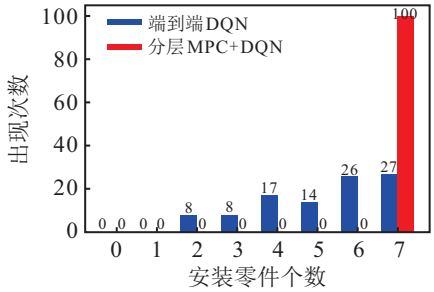

图 11　低难度案例使用不同方法成功安装零件个数统计

低难度案例的装配序列图如图12所示,验证了此方法的有效性. 此效果也验证了装配序列规划问题的分层特点,体现了本文所提的分层规划方法的良好决策能力和优越性.

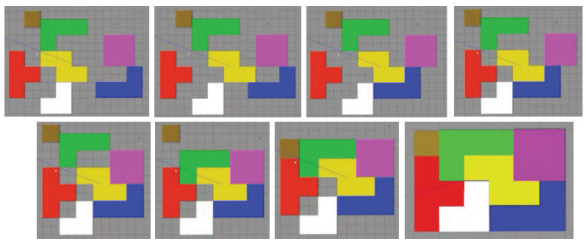

图 12　低难度装配序列实例

由图13损失函数下降曲线可以看出,在迭代80次后,网络的误差已经降到0.5以下,并且趋于收敛. 由于在装配过程中,当已安装的零件个数不同时,每安装一个新的零件后,就会变换为一个新的装配状态,对网络模型而言是一种全新的输入,而仿真结果表明此算法对于这些任意状态下的零件模型都可以完成后续的装配任务,也验证了此方法在此基础难度问题上的通用性.

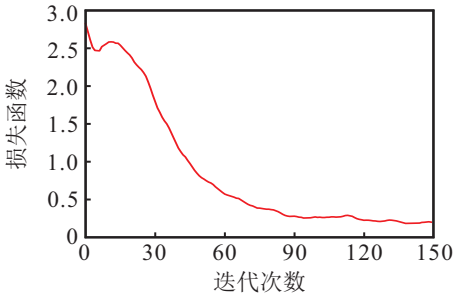

图 13　低难度案例损失函数下降曲线

实验2(中难度实例)　在低难度实例中,网络模型可以100 %完成装配任务与零件初始位置距离较近有一定的关系,因此,在中难度实例中,将此距离调大如图10右图所示,加大下层运动规划部分的难度,并且将每个零件的初始位置在一定范围内随机设置,保证其与目标位置的距离较远,更加提高了装配问题的难度. 综上,中难度实例的问题设置为: 当零件初始位置与正确装配位置距离较远时,需要15步～20步的规划才可达,且在每个零件初始位置随机的情况下进行装配序列规划.

由图14成功率曲线图可以看出,网络的测试成功率从接近零逐步上升,最后稳定于85 %左右,也验证了基于双层结构的通用装配序列规划算法对于此难度的有效性.

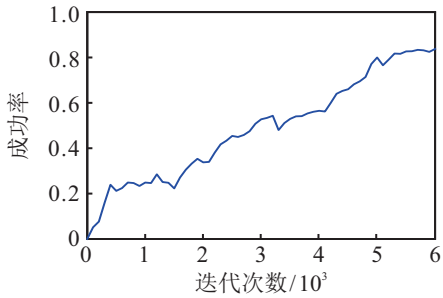

图 14　中难度实例装配成功率曲线

由图15的训练过程损失函数可以看出,由于问题难度较大,损失函数也有一定幅度的波动,但是整体的下降趋势还是十分明显的.

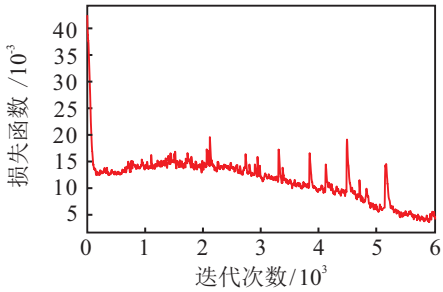

图 15　中难度实例损失函数曲线

由图16中难度装配序列实例可以看出,网络对于初始位置与正确装配位置较远的装配任务可以在较少步数内完成. 对于这一部分的难度设置,由于零

件的初始位置更加随机,初始给出的零件状态不同,并且其余未安装的零件也不在训练阶段所学习到的位置上,对于每一个未安装的零件都需要重新规划一条新的轨迹,输入状态与碰撞关系都和训练阶段的数据有较大差异,此结果也说明了所提方法对于不同随机初始位置的任务具有较好的通用性.

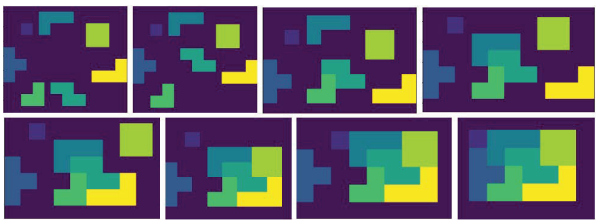

图 16 中难度装配序列实例

实验3(高难度实例) 为了体现本文所提出的基于分层强化学习的通用装配序列规划方法的通用性和泛化能力,使用在中难度问题上训练好的网络模型,进行对多种构型装配序列规划任务的仿真实验. 对于每一种新任务,输入的目标维度上都对应着当前任务下的目标构型. 设置训练环境为:零件初始位置与正确装配位置距离较远,且零件初始位置随机,每次训练的目标构型在8种其他构型中随机选择. 在解决了中难度装配任务的网络模型基础上进行初始装配时无已安装零件(零初始)条件下的多构型任务训练.

由图17成功率增长曲线可以看出,将由单一构型训练出的装配模型直接用于很多其他构型时,装配成功率是接近于零的,并且在迭代的过程中成功率增长十分缓慢. 这说明从单一构型转换到多种其他构型的巨大难度. 也从另一方面体现了装配序列规划问题的困难,当没有任何先验经验和决策能力时,很难完成对多种构型的装配任务.

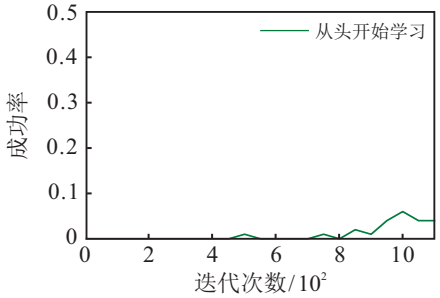

图 17 高难度零初始多构型任务直接训练成功率曲线

考虑到分层结构中上下层的互相影响,本文先对下层的零件动作规划网络进行训练,然后将此下层网络作为初始网络,继续对多种构型装配任务进行学习.

由图18可以看出,模型的装配成功率可以达到近60%,与图17相比,在相同的迭代次数下,成功率的增长更快. 训练结束后,随机生成100个初始环境来测试网络对多构型装配任务的效果,结果如表3所示,对于7零件的装配模型,有99%的情况下最终成功安装上了5个及以上零件.

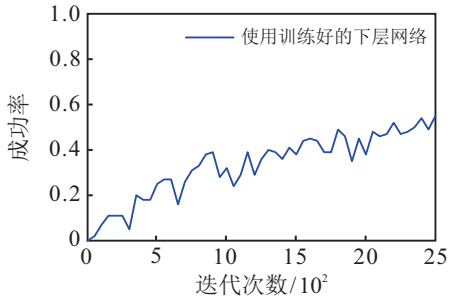

图 18 高难度案例零初始多构型任务初始使用下层网络训练成功率增长曲线

表 3 高难度案例零初始多构型任务成功安装的零件个数

安装零件个数	4	5	6	7
次数	1	16	28	55

由图19的序列实例可以看出,所提方法可以在较少步数内完成对其他目标构型任务的装配,验证了所提的基于分层强化学习的通用装配序列规划方法对多构型装配任务的有效性.

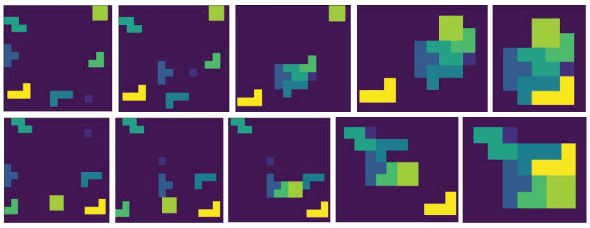

图 19 高难度案例零初始多种构型任务序列实例

在此基础上进行随机初始零件个数的多构型任务训练,在每50次迭代训练后进行一轮测试,每轮测试包含100个不同的装配任务,初始的零件个数在0～6中随机产生.

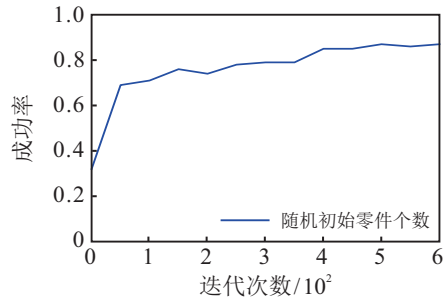

图 20 高难度案例随机初始多构型任务成功率变化曲线

由图20成功率增长曲线可以看出,对于随机初始零件个数的多构型装配任务,装配成功率最终可以达到近90%. 训练前后成功安装零件个数统计如

图21所示,在训练前,测试效果基本与随机初始化的情况持平,网络参数中没有适用于装配任务的有效信息,而在训练后,基本上全部任务都可以安装到6个及以上零件,表明网络模型中包含了很多对于装配序列规划问题的广义信息,不再局限于某一个特定的目标构型,对其他多种目标构型也具有一定的决策能力.

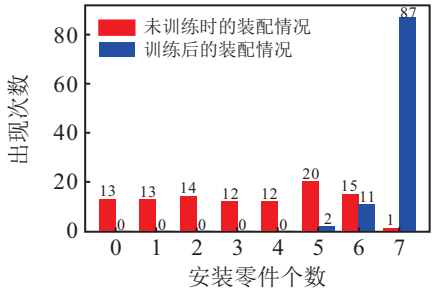

图 21　高难度案例随机初始多种构型任务成功安装零件个数直方图

本文分别给出了在初始4个以及3个零件的情况下对于其他构型任务的序列实例图,如图22所示. 可以看出,在初始位置随机且距离正确装配位置较远的情况下,所提方法可以成功完成不同目标构型的装配任务. 验证了本文所提出的基于分层强化学习的通用装配序列规划方法的通用性和适应能力.

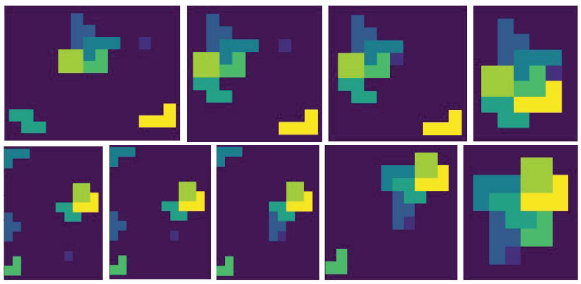

图 22　高难度案例随机初始多种构型任务序列实例

5　结　论

为了增强规划算法对于多种目标构型任务的适应能力和泛化能力,本文针对包含多个零件、多条可行序列的复杂装配模型,在进行目标驱动的基础上,考虑到装配序列规划的层次化结构特点,使用了分层强化学习的框架,提出了一种基于分层强化学习的目标导向型通用装配序列规划方法. 其上层进行装配序列规划,下层进行基于DQN的零件动作规划,能很好地解决规模较大带来的问题,使规划方法更具灵活性. 在问题建模方面,提出通用的形式化表示,设计了一种新颖的状态空间和动作空间表示方式,对解决类似问题具有一定借鉴意义;在网络架构设计及搭建方面,将目标构型作为一维输入,自主设计了基于孪生网络的目标导向深度网络架构;在强化学习训练过程中进行了探索性的初始化环境设置,并引入课程学习方法,提高了装配网络训练效率.

本文提出的装配序列规划算法不仅可以避免学习过程中没有目标构型的直接信息所带来的信息损失问题,还提高了算法的通用性和泛化能力,对于零件初始位置与目标位置较远且随机、初始零件个数以及初始零件均随机的装配任务有着准确的决策结果,并且对其他目标构型的装配任务也有很高的成功率. 仿真结果表明,所提方法适用于不同目标构型与任意初始状态的装配体,解决问题速度较快,具有一定的有效性、通用性和泛化能力.

参考文献(References)

[1] 王喜文. 中国制造2025解读[M]. 北京: 机械工业出版社, 2015: 5-10.
(Wang X W. Interpretation of made in China 2025[M]. Beijing: Machinery Industry Press, 2015: 5-10.)

[2] 刘炜, 刘峰, 倪阳咏, 等. 航天复杂产品智能化装配技术应用研究[J]. 宇航总体技术, 2018, 2(1): 33-36.
(Liu W, Liu F, Ni Y Y, et al. Application research of intelligent assembly technology for complicate products[J]. Astronautical Systems Engineering Technology, 2018, 2(1): 33-36.)

[3] 万晓琴, 严洪森, 汪峥. 知识化制造环境下航空发动机装配线调度及自重构[J]. 自动化学报, 2015, 41(1): 136-146.
(Wan X Q, Yan H S, Wang Z. Scheduling and self-reconfiguration of an aircraft engine assembly line in knowledgeable manufacturing[J]. Acta Automatica Sinica, 2015, 41(1): 136-146.)

[4] Xu L D, Wang C G, Bi Z M, et al. AutoAssem: An automated assembly planning system for complex products[J]. IEEE Transactions on Industrial Informatics, 2012, 8(3): 669-678.

[5] Jones R E, Wilson R H, Calton T L. On constraints in assembly planning[J]. IEEE Transactions on Robotics and Automation, 1998, 14(6): 849-863.

[6] Homem de Mello L S, Sanderson A C. Representations of mechanical assembly sequences[J]. IEEE Transactions on Robotics and Automation, 1991, 7(2): 211-227.

[7] Homem de Mello L S, Sanderson A C. A correct and complete algorithm for the generation of mechanical assembly sequences[J]. IEEE Transactions on Robotics and Automation, 1991, 7(2): 228-240.

[8] Karjalainen I, Xing Y, Chen G, et al. Assembly sequence planning of automobile body components based on liaison graph[J]. Assembly Automation, 2007, 27(2): 157-164.

[9] 李荣, 付宜利, 封海波. 基于连接结构知识的装配

序列规划[J]. 计算机集成制造系统, 2008, 14(6): 1130-1135.
(Li R, Fu Y L, Feng H B. Assembly sequence planning based on connecter-structure knowledge[J]. Computer Integrated Manufacturing Systems, 2008, 14(6): 1130-1135.)

[10] Kashkoush M, ElMaraghy H. Knowledge-based model for constructing master assembly sequence[J]. Journal of Manufacturing Systems, 2015, 34: 43-52.

[11] Cakir B, Altiparmak F, Dengiz B. Multi-objective optimization of a stochastic assembly line balancing: A hybrid simulated annealing algorithm[J]. Computers & Industrial Engineering, 2011, 60(3): 376-384.

[12] 宫华, 袁田, 张彪. 基于深度邻域搜索PSO算法的装配序列优化问题[J]. 控制与决策, 2016, 31(7): 1291-1295.
(Gong H, Yuan T, Zhang B. Assembly sequence planning problem based on particle swarm optimization algorithm with depth local search[J]. Control and Decision, 2016, 31(7): 1291-1295.)

[13] 周开俊, 李东波, 黄希. 基于遗传算法的装配序列规划研究[J]. 机械设计, 2006, 23(2): 30-33.
(Zhou K J, Li D B, Huang X. Research on assembly sequence planning based on genetic algorithm[J]. Journal of Machine Design, 2006, 23(2): 30-33.)

[14] Hong D S, Cho H S. A neural network-based computational scheme for generating optimized robotic assembly sequence[J]. Engineering Applications of Artificial Intelligence, 1995, 8(2): 129-145.

[15] Sinanoglu C, Borklu H R. An assembly sequence-planning system for mechanical parts using neural network[J]. Assembly Automation, 2005, 25(1): 38-52.

[16] Levine S, Pastor P, Krizhevsky A, et al. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection[J]. The International Journal of Robotics Research, 2018, 37(415): 421-436.

[17] Finn C, Levine S. Deep visual foresight for planning robot motion[C]. 2017 IEEE International Coference on Robotics and Automation. Singapore, 2017: 2786-2793.

[18] Arulkumaran K, Deisenroth M P, Brundage M, et al. Deep reinforcement learning: A brief survey[J]. IEEE Signal Processing Magazine, 2017, 34(6): 26-38.

[19] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. Cambrige: MIT Press, 1998: 47-68.

[20] Zhao M H, Guo X, Zhang X B, et al. ASPW-DRL: Assembly sequence planning for workpieces via a deep reinforcement learning approach[J]. Assembly Automation, 2019, 40(1): 65-75.

[21] Tamar A, Wu Y, Thomas G, et al. Value iteration networks[J/OL]. 2016, ArXiv: 1602.02867.

[22] Zhu Y K, Mottaghi R, Kolve E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning[C]. 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore, 2017: 3357-3364.

[23] Dietterich T G. Hierarchical reinforcement learning with the MAXQ value function decomposition[J]. Journal of Artificial Intelligence Research, 1999, 13(1): 227-303.

[24] Kulkarni T D, Narasimhan K R, Saeedi A, et al. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation[J/OL]. 2016, ArXiv: 1604.06057.

[25] Camacho E F, Bordons C. Model predictive control[M]. London: Springer, 2004: 13-19.

[26] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[27] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

作者简介

赵铭慧(1995−), 女, 助理实验师, 硕士, 从事深度强化学习的研究, E-mail: zmh@mail.nankai.edu.cn;

张雪波(1984−), 男, 教授, 博士生导师, 从事移动机器人视觉控制等研究, E-mail: zhangxuebo@nankai.edu.cn;

郭宪(1986−), 男, 副研究员, 博士, 从事强化学习的研究, E-mail: guoxian@nankai.edu.cn;

欧勇盛(1972−), 男, 研究员, 博士生导师, 从事低成本移动机器人导航研发与应用等研究, E-mail: ys.ou@siat.ac.cn.

(责任编辑：闫 妍)