文章编号: 1001-0920(2017)09-1614-07 DOI: 10.13195/j.kzyjc.2016.1037

两类品种工件混流的多站点CSPS系统优化控制

唐 昊†, 李博川, 王 彬, 谭 琦

(合肥工业大学 电气与自动化工程学院，合肥 230009)

摘 要: 研究一种两类品种工件混流的多站点传送带给料加工站系统的优化控制问题. 系统中的站点如何协同工作完成工件加工任务,是提高系统生产率的重要课题. 将前视距离作为各站点的决策变量,通过站点间的局部信息交互,提出一种品种均衡工作模式,并运用一种模型无关的串行反馈式多agent强化学习算法求解系统的最优策略. 实验结果验证了该工作模式的合理性和算法的有效性,并分析了部分参数变化对系统性能的影响.

关键词: 多站点CSPS系统；两类品种工件；多agent强化学习；前视距离控制

中图分类号: TP278 **文献标志码:** A

Optimal control of multiple CSPS system with two-type product mixed flow

TANG Hao†, LI Bo-chuan, WANG Bin, TAN Qi

(School of Electrical Engineering and Automation，Hefei University of Technology，Hefei 230009，China)

Abstract: This paper is mainly concerned with the optimal control of multiple CSPS system with two-type product mixed flow. How the stations in the system work cooperatively to complete the machining task is one of the important subjects to improve productivity. An operating mode based on equilibrium of products is proposed by using look-ahead range as the control variable of each station and through the local information interaction between stations. A model-free multi-agent reinforcement learning algorithm is used to derive the optimal control policy. Simulation results show the rationality of the proposed operating mode and the effectiveness of the proposed algorithm, and the impact of different parameters on system performance is also analyzed.

Keywords: multiple CSPS system；two-type product；multi-agent reinforcement learning；look-ahead control

0 引 言

在现代化企业生产中,存在着一类由生产加工站作为加工主体的自动化生产线,其中生产加工站用于传输工件进行加工,这样一类系统称为传送带给料加工站(CSPS)[1-4]. 例如,一些著名IT品牌的手机或硬盘等智能装配线,以及实际生产中广泛存在的食品包装和产品分拣系统等,都可抽象化为CSPS生产模型. 因此,对于这类系统的优化控制问题的研究具有十分重要的现实意义.

CSPS研究中,日本经营工学领域著名教授松井正之做了大量基础性工作,例如单站点CSPS系统的半Markov决策过程模型的建模和理论方程建立、CSPS的各种控制模式和物理原理分析等[1-2]. 在此工作基础上,文献[3]提出了一种基于性能势的在线策略迭代学习优化算法,用于解决单站点CSPS系统的前视距离控制问题. 关于多站点生产线的协同控制问题亦有诸多研究,例如,文献[5]研究了自动化冲压线多个加工站的协同控制优化问题,文献[6-7]给出了多站点CSPS基于学习优化的协同控制方法. 以上这些研究都是考虑单一品种工件加工的情况,在实际生产过程中,往往存在多种产品混线生产模式[8]. 文献[9]将多站点CSPS系统作为一个整体,使用遗传算法及其改进算法求解最优前视距离控制问题. 文献[10]使用人工蜂群算法和禁忌搜索法优化多品种混流装配线平衡问题.

本文考虑两类品种工件混合到达的多站点

收稿日期: 2016-08-14；**修回日期:** 2016-11-16.

基金项目: 国家自然科学基金面上项目(61174186, 61573126, 71231004)；教育部高等学校博士学科点专项科研基金项目(20130111110007)；教育部新世纪优秀人才计划项目(NCET-11-0626)；合肥工业大学应用科技成果培育计划项目(JZ2016YYPY0052).

作者简介: 唐昊(1972–), 男, 教授, 博士, 从事离散事件动态系统、强化学习、神经元动态规划等研究；李博川(1990–), 男, 硕士生, 从事离散事件动态系统、强化学习的研究.

†通讯作者. E-mail: htang@hfut.edu.cn

CSPS 系统. 同时在文献[11]的前期研究基础上,提出一种品种均衡工作模式,采用模型无关的串行反馈式多agent强化学习算法来求解系统的最优前视距离协同控制问题. 实验结果验证了该工作模式的合理性和算法的有效性,并分析了部分参数变化对系统性能的影响.

1　系统模型和工作模式

图1为两类品种工件混流的多站点CSPS系统物理模型图. 传送带匀速运行,两个品种的工件分别按照独立的泊松过程随机到达,传送带一侧串行分布着多个功能相同的加工站点,每个加工站点都是一个决策主体,故可视作一个agent,可以执行工件卸载或者工件加工的操作. 每个站点都配备前视距离传感器用以感知传送带上一定范围内的工件信息,并分别配备两个容量有限的缓存库(buffer),用来存放两个品种的待加工工件. 此外,系统还为每个品种的工件配备了无限容量的成品库(bank)用来存放加工好的工件.

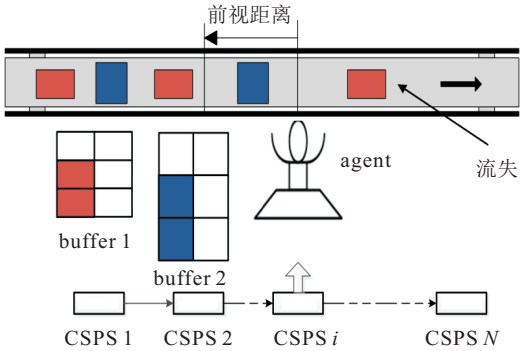


图 1　系统物理模型

在传统的CSPS系统中,站点根据缓存库状态信息选取合适的前视距离,根据该范围内有无工件情况选择工件卸载或者工件加工操作. 在本文研究的两类品种工件混线的多站点CSPS系统中,由于两个品种工件混流到达,如果不考虑不同品种工件的到达和缓存情况而直接按单品种CSPS情况进行卸载或者加工操作,则会造成各品种生产明显不均衡的情况. 因此,本文提出一种品种均衡工作模式,其工作过程概括为: 在决策时刻,站点根据自身两个缓存库的状态信息选取一段前视距离,若前视距离内有工件,则等待其中缓存库相对剩余量最大的品种的第1个工件到达本站捡取点,并卸载放入对应的缓存库中;否则,从相对剩余量最小的缓存库中取一个工件进行加工. 此种情况下,若两个品种的缓存库相对剩余量相等,则随机选取一个品种工件进行加工. 这里,缓存库相对剩余量定义为该缓存库剩余量与该缓存库容量的比值.

不失一般性,作以下假设:

1) 传送带匀速运行,前视距离可以等效为前视时间进行处理;

2) 各站点沿传送带间距相等;

3) 每个品种工件的加工时间服从不同参数的Erlang分布;

4) 每个品种工件独立存放在相应的缓存库中,工件的卸载时间忽略不计,工件加工完毕放入成品库的时间也忽略不计.

2　优化模型和方法

2.1　决策过程

假设系统中站点个数为N. 记品种j的到达率为λ_j,品种m在各站点的缓存库容量皆为C_m,站点i品种m的缓存库剩余量为s_m^i,则有$0 \leqslant s_m^i \leqslant C_m, i = 1,2,\cdots,N, m = 1,2$. 站点$i$的状态$s^i$由该站点所有品种缓存库剩余量组成,即$s^i = \{s_1^i, s_2^i\}$,其前视距离记为$a_i(s^i)$.

假设在控制策略的作用下,各加工站点都由其初始决策时刻$T_0 = 0$开始进行状态演化. 为了简单起见,记站点i的第n个决策时刻为T_n,状态为$s^i(T_n) = \{s_1^i(T_n), s_2^i(T_n)\}$,行动为$a_i(s^i(T_n))$. 若决策时刻存在一个品种的缓存库为满,则agent i直接从中取一个工件进行加工,即$a_i(s^i(T_n)) = 0$;若所有品种缓存库都为空,则agent i一直等待传送带上有工件到达并捡取,即等价于$a_i(s^i(T_n)) = \infty$. 一般情况下,有$a_i(s^i) \in [l_{\min}, l_{\max}]$,其中$l_{\min}$为最小前视距离,$l_{\max}$为最大前视距离. 根据前述品种均衡工作模式,若前视距离内有工件品种m待捡取,且其到达捡取点的时间为ϕ_n^i,则下一决策时刻为$T_{n+1} = T_n + \phi_n^i$. 如果$m = 1$,则下一决策时刻状态为$\{s_1^i(T_n) - 1, s_2^i(T_n)\}$;如果$m = 2$,则下一决策时刻状态为$\{s_1^i(T_n), s_2^i(T_n) - 1\}$. 若前视距离内没有工件到达,则根据前面所提到的品种均衡工作规则,选取工件品种m进行加工,记其服务时间为$u_m^i(T_n)$,则下一个决策时刻T_{n+1}为$T_{n+1} = T_n + \max(u_m^i(T_n), a_i(s^i(T_n)))$. 如果$m = 1$,则下一决策时刻状态为$\{s_1^i(T_n) + 1, s_2^i(T_n)\}$;如果$m = 2$,则下一决策时刻状态为$\{s_1^i(T_n), s_2^i(T_n) + 1\}$.

两品种混线生产的多站点CSPS系统的优化控制目标是在兼顾各站点间负载平衡性的情况下最大化系统在无穷时间段内的各品种工件相对处理率及系统总处理率,因此定义T时段内工件到达总数为$Q(T)$,系统加工的工件总数为$P(T)$,其中品种M的工件到达数和加工数分别为$Q_m(T)$和$P_m(T)$,同时agent i加工品种m的工件个数为$P_m^i(T)$,则系

统总的工件处理率为$P(T)/Q(T)$,品种m的相对处理率为$P_m(T)/Q_m(T)$,站点i对品种m的负载率为$P_m^i(T)/Q_m(T)$.

2.2 代价函数定义

在多品种多站点CSPS系统中,每个加工站点都是一个决策主体. 由于上游站点优先捡取工件,其决策会对下游所有站点的决策产生影响,容易造成上游站点负荷过重,下游站点负荷过轻的情况. 为了平衡各站点的负载率,可以借鉴文献[6]提出的站点局部信息交互思想,将站点$i(i \neq N)$与其相邻的下游站点进行信息交互. 具体地,将相邻下游站点各品种缓存库的状态信息通过代价函数反馈到其相邻上游站点,作为其决策参考因素. 通过逐级信息反馈,最终提高系统各站点间的相互协作水平.

另外,当站点既不进行卸载又不进行加工操作时只能等待,而每个品种工件的期望加工时间以及单位时间流入生产线的总数皆为定值,于是从系统的长期运行角度看,单位时间内一个站点的等待时间越短,其加工时间越长,该站点的工件处理率也越高. 因此在CSPS系统的研究中,一个站点的单位时间等待代价将是系统优化控制目标的一个重要部分.

最后,每个站点对每个品种的工件处理率越高,则单位时间内该品种缓存库中临时存放的工件越少,因此系统优化中也可将各品种缓存库的存储代价作为一个性能进行考虑.

综上所述,定义以下代价系数.

K_1^m:相邻两站点间品种m的单位缓存库剩余量差值的单位时间反馈代价;

K_2:站点的单位时间等待代价;

K_3^m:单位时间品种m的单个工件存储代价.

站点i在决策时刻T_n时,根据状态$s^i(T_n)$采取行动$a_i(s^i(T_n))$,在转移到下一决策时刻状态$s^i(T_{n+1})$前的单位时间代价函数记为$f^i(s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1}),t)$,其计算分为以下两种情况:

1) 若站点i执行加工品种j工件的操作,记样本转移为$\langle s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1}),u_j^i(T_n)\rangle$,则

$$\begin{aligned}&f^i(s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1}),t)=\\&\sum_{m=1}^{2}[K_1^m\cdot(\theta_i\cdot s_m^{i+1}(T_n)-s_m^i(T_n))]+\\&K_2\cdot\gamma_t+\sum_{m=1,m\neq j}^{2}[K_3^m\cdot(C_m-s_m^i(T_n))]+\\&K_3^j\cdot(C_j-s_j^i(T_{n+1})),\ T_n\leqslant t<T_{n+1}.\end{aligned}\tag{1}$$

其中

$$\theta_i=\begin{cases}1,\ i\neq N;\\0,\ i=N.\end{cases}\tag{2}$$

$$\gamma_t=\begin{cases}0,\ T_n\leqslant t<T_n+u_j^i(T_n);\\1,\ T_n+u_j^i(T_n)\leqslant t<T_{n+1}.\end{cases}\tag{3}$$

2) 若站点i执行等待品种j工件到达的操作,记样本转移为$\langle s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1})\rangle$,则

$$\begin{aligned}&f^i(s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1}),t)=\\&\sum_{m=1}^{2}[K_1^m\cdot(\theta_i\cdot s_m^{i+1}(T_n)-s_m^i(T_n))]+K_2+\\&\sum_{m=1}^{2}[K_3^m\cdot(C_m-s_m^i(T_n))],\ T_n\leqslant t<T_{n+1}.\end{aligned}\tag{4}$$

式(1)和(4)中,$K_1^m\cdot(\theta_i\cdot s_m^{i+1}(T_n)-s_m^i(T_n))$即为具有控制作用反馈功能的上下游站点信息交互项,它与相邻的下游站点的状态有关. 若代价函数中有信息交互项,则称为协同工作模式,否则称为独立工作模式.

2.3 基于性势能的多agent Q学习算法

利用式(1)和(4)构造基于局部反馈的多agent学习算法. 系统中,每个站点都可视为一个agent. 根据前述决策时刻的定义,统计意义上,任意决策时刻两个或两个以上agent同时决策的概率为零. 故每个决策时刻将N个agent分成两组[12],其中进行决策的agent单为一组,其余不参与决策的$(N-1)$个agent为另一组. 学习过程中,决策agent根据自身状态选择对应行动,并根据局部反馈信息更新其状态行动值,即Q值. 决策agent i每次决策后至其下一决策时刻获得的观测样本统一记为$\langle s^i(T_n),s^{i+1}(T_n),a_i(s^i(T_n)),s^i(T_{n+1}),w^i(T_n),u_j^i(T_n)\rangle$,其中$w^i(T_n)=T_{n+1}-T_n$,表示决策间隔时间. 为表达间接起见,若$i=N$,则约定该样本数据中$s^{i+1}(T_n)$为空. 若站点$i$执行的是等待品种$j$工件到达的操作,则假设$u_j^i(T_n)=0$. 那么根据该样本,可以计算$T_n\sim T_{n+1}$之间的折扣累积代价$f_\alpha^i(s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1}))$,此式中的$\alpha(\alpha\geqslant 0)$为折扣因子[15]. 与式(1)和(2)相对应,其计算也分为两种情况:

1) 若站点i执行加工品种j工件的操作,则

$$\begin{aligned}&f_\alpha^i(s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1}))=\\&\sum_{m=1}^{2}[K_1^m\cdot(\theta_i\cdot s_m^{i+1}(T_n)-s_m^i(T_n))]\cdot T_\alpha(w^i(T_n))+\\&K_2\cdot(T_\alpha(w^i(T_n))-T_\alpha(u_j^i(T_n)))+\end{aligned}$$

$$T_\alpha(w^i(T_n))\cdot\sum_{m=1,m\neq j}^{2}[K_3^m\cdot(C_m-s_m^i(T_n))]+$$
$$T_\alpha(w^i(T_n))\cdot K_3^j\cdot(C_j-s_j^i(T_n)-1). \quad (5)$$

2) 若执行等待品种j工件到达的操作,则

$$f_\alpha^i(s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1}))=$$
$$\sum_{m=1}^{2}[K_1^m\cdot(\theta_i\cdot s_m^{i+1}(T_n)-s_m^i(T_n))]\cdot T_\alpha(w^i(T_n))+$$
$$K_2\cdot T_\alpha(w^i(T_n))+$$
$$\sum_{m=1}^{2}[K_3^m\cdot(C_m-s_m^i(T_n))]\cdot T_\alpha(w^i(T_n)). \quad (6)$$

记$T_\alpha(t)=\int_0^t \mathrm{e}^{-\alpha t}\mathrm{d}t$,且$T_\alpha(t)$在$\alpha\to 0$时的极限为$T_0(t)=t$.

与文献[6]类似,根据公式可以构造串行反馈式Wolf-PHC学习算法[13-14]. 该算法是一种模型无关的多agent强化学习算法,决策agent利用样本观测值,根据式(5)或(6)计算平均和折扣性能准则下统一的即时差分公式,然后再更新其Q值$Q_i(s^i(T_n),a_i(s^i(T_n)))$,并更新其随机策略$\pi_i$和平均策略$\bar{\pi}_i$,详细过程此处不再赘述.

文献[6]采取的是ε-greedy探索策略,但ε的取值规则往往依赖于经验知识或反复试凑,其取值好坏对学习效果影响较大. 为更好地平衡学习过程中探索与利用之间的矛盾,引入模拟退火方法解决行动探索问题,即决策agent在每步学习时根据Metropolis准则确定是执行探索行动还是利用行动. 具体过程是,每次学习时,随机选取一个探索行动a_r,并根据当前的随机策略π_i选取一个利用行动a_p,再产生一个随机数$\delta\in(0,1)$. 若$\delta>\mathrm{e}^{(Q_i(s^i(T_n),a_r)-Q_i(s^i(T_n),a_p))/\mathrm{Temp}}$,则接受探索行动;否则接受利用行动. 这里,初始温度Temp和温度衰减因子τ由算法初始化时设定为恒值.

综上,给出agent i的串行反馈式Wolf-PHC算法流程.

Step 1: 令所有状态行动对的Q值为0,初始化随机策略π_i,为每个状态行动对设置相同的概率,初始化温度Temp、温度衰退因子τ、学习步数. 令$n=0$, $T_0=0$.

Step 2: 在当前状态$s^i(T_n)$下随机产生一个行动a_r,根据随机策略π_i产生行动a_p,并根据Metropolis准则决定$a_i(s^i(T_n))$是采取探索行动a_r还是采取利用行动a_p.

Step 3: 执行行动$a_i(s^i(T_n))$,记录样本观测数据

$$\langle s^i(T_n),s^{i+1}(T_n),a_i(s^i(T_n)),s^i(T_{n+1}),$$
$$w^i(T_n),u_j^i(T_n)\rangle,$$

并计算$T_n\sim T_{n+1}$之间的折扣累积代价$f_\alpha^i(s^i(T_n),a_i(s^i(T_n)),s^i(T_{n+1}))$,进而更新状态行动值$Q_i(s^i(T_n),a_i(s^i(T_n)))$.

Step 4: 更新随机策略π_i和平均策略$\bar{\pi}_i$在状态$s^i(T_n)$时的行动选择概率.

Step 5: 若算法终止条件满足,则学习结束;否则$n:=n+1$, $\mathrm{Temp}:=\mathrm{Temp}\cdot\tau$,转至Step 2.

图2为系统在均衡工作模式下一个站点的在线学习流程.

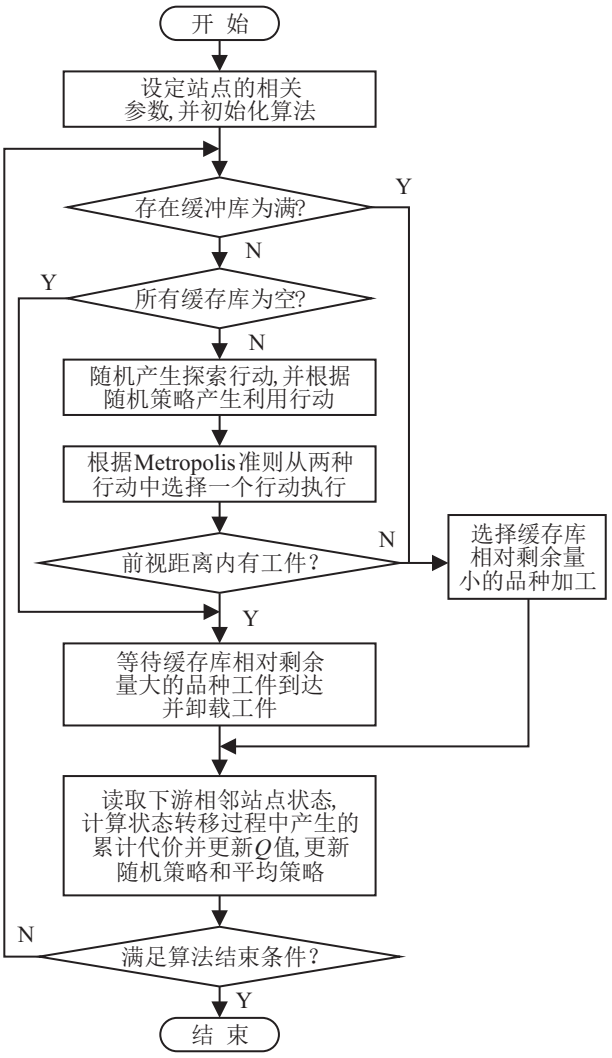


图 2　品种均衡工作模式下站点在线学习流程

3　实验结果及分析

在两类品种工件混线的多站点CSPS系统仿真实验中,取$N=4$, j品种工件在各站点的工件服务时间服从L阶的Erlang分布,每相的加工时间服从参数为u_j的指数分布. 表1为相关参数设置.

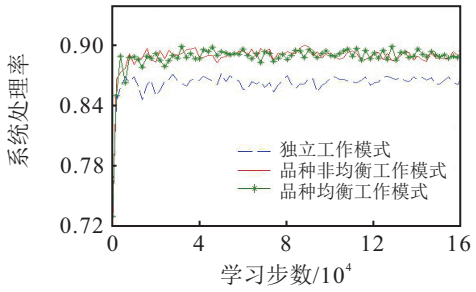
表1 相关参数设置

	C_1	C_2	$l_{\min}$	$l_{\max}$	u_1	u_2	L	λ_1	λ_2	K_1^1	K_1^2	K_2	K_3^1	K_3^2
参数值	4	5	0	4	4	5	4	0.4	0.6	0.6	0.9	2.0	0.2	0.3

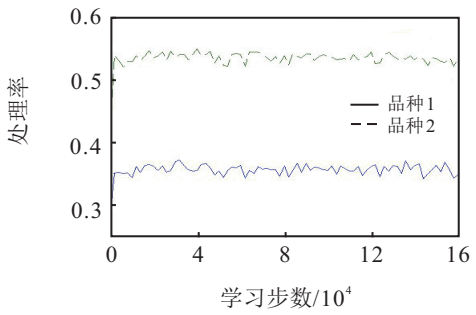
算法性能评估时每次进行10次独立实验,每次独立实验时系统仿真运行20万步,然后取相关性能值的统计平均值. 在学习优化过程中每运行2 000步对学习效果进行一次取样评估. 每次学习中,每个站点在各状态下采取各行动的初始概率都相等. 这里,非均衡工作模式的工作过程与传统的单一品种CSPS的工作过程类似:在决策时刻,前视距离内有工件,则卸载其中第1个到达的工件;前视距离内无工件,则从站点的所有缓存库中随机选取1个工件进行加工.

实验首先将本文提出的品种均衡工作模式与非均衡工作模式进行比较,分析这两种工作模式的优劣. 图3(a)为系统分别在独立工作模式和协同工作模式下的总体处理率优化曲线图. 其中,协同工作模式又分为品种均衡工作模式和非均衡工作模式两种情况. 由图3(a)可见,独立工作模式和协同工作模式都能显著提升系统的工件总处理率,但是协同工作模式下系统的总处理率更高,其原因在于协同工作模式中相邻agent间进行了局域信息交互,改善了各站点的出力均衡性,故间接提高了系统的总处理率. 此外可知,品种均衡工作模式和非均衡工作模式下系统的总处理率几乎一致,并无明显优劣.

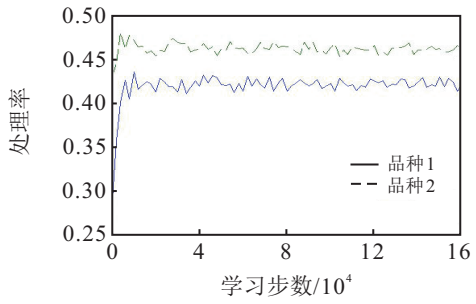
图3(b)与图3(c)分别为与之对应的不同品种工件处理率的优化曲线. 显然,品种均衡工作模式显著缩小了不同品种工件的相对处理率之差,这是由于品种均衡工作模式下agent在执行卸载或者加工操作时需要适当平衡不同品种的缓存库剩余容量,从而显著缓解两品种工件混线生产不均衡的问题. 但是,在仿真实验中,由于两品种工件的到达率和服务率存在差异,无法完全实现均衡生产. 随后,实验从系统的负载率和代价性能两个方面分别对品种均衡工作模式与独立工作模式进行比较. 图4为系统分别在独立工作模式和品种均衡工作模式下的总平均代价优化曲线,其变化趋势与工件处理率正好相反. 随着学习的进行,两种情况下系统的平均代价都逐渐减小、趋于水平并作微小波动. 其中,品种均衡工作模式下需将下游站点各品种的缓存库剩余容量与上游站点对应品种的缓存库剩余容量相减反馈到上游站点的学习中,由于代价函数中该反馈项的存在,在学习开始时,其总平均代价比独立工作模式情况下要大. 随着学习步数的增多,各站点间对应品种缓存库剩余容量差值



(a) 3种工作模式下系统总处理率



(b) 非均衡工作模式下两品种工件处理率



(c) 品种均衡工作模式下两品种工件处理率

图3 各工作模式下系统处理率

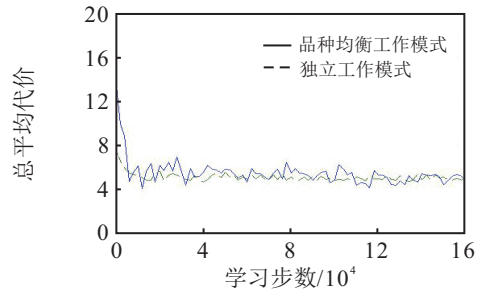
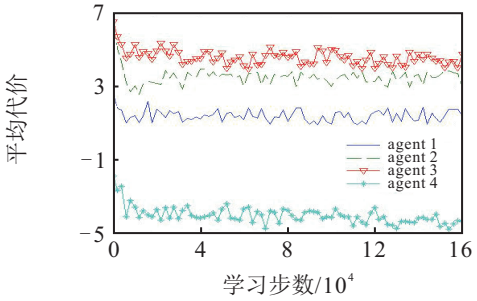


图4 两种工作模式下系统总平均代价

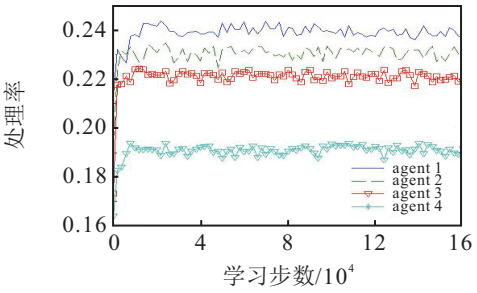
趋于减小,平均代价也迅速减小,最终两者的代价学习值基本接近.

图5给出了品种均衡工作模式下系统中各站点的平均代价优化曲线和工件处理率优化曲线. 可见,各个站点的平均代价随着学习的进行逐渐减小、趋于水平并作微小震荡,而各站点的工件处理率随着学习的进行也逐渐增大、趋于水平并作微小震荡.

为了直观地分析各个站点的出力平衡性,本文定义各站点总体生产均衡度和品种m的生产均衡度,用以表示各站点实际负载率与理想负载率的吻合程



(a)　品种均衡工作模式下各agent平均代价



(b)　品种均衡工作模式下各agent工件处理率

图5　各agent的平均代价与工件处理率

度,计算公式分别为

$$\sigma = \left(1 - \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \left(\frac{P_m^1(T) + P_m^2(T)}{P(T)} - \bar{p}\right)^2}\right) \times 100\%, \tag{7}$$

$$\sigma^m = \left(1 - \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (p_i^m - \bar{p})^2}\right) \times 100\%. \tag{8}$$

其中:σ为各站点的总体生产均衡度,σ^m为各站点中品种m的生产均衡度,p_i^m为站点i品种m的实际负载率,$\bar{p}$为站点的理想负载率. 本文所述系统中,生产线共有4个加工站点,每个站点的理想负载率皆为25%. 表2比较了独立工作模式和品种均衡生产模式下的生产均衡度(单位为%). 从仿真结果中可以看出,品种均衡工作模式改善了站点间的负载均衡性,提高了站点间的协作能力.

表2　两种工作模式下的生产均衡度

	独立工作模式	品种均衡工作模式
品种1生产均衡度	96.32	98.07
品种2生产均衡度	96.37	97.99
总体生产均衡度	96.35	98.02

实验同时分析了各品种工件缓存库容量配比和系统工件处理率之间的关系. 表3给出了缓存库容量配比不同情况下系统总的工件处理率仿真实验结果. 表3中每行数据表示品种1的缓存库容量固定时,品种2缓存库容量变化对系统总处理率的影响. 可以看出,在品种1缓存库容量固定的情况下,系统总处理率随着品种2缓存库容量的增大而增大,但增幅逐渐放缓. 同时,表中也给出了缓存库容量之和固定时,不同缓存库容量配比对系统工件处理率的影响. 由于工件到达率反映各品种工件到达的密集程度,而服务率反映的是系统加工相应品种工件的能力,故$(\lambda_1/\mu_1):(\lambda_2/\mu_2)$反映系统在长期运行过程中,各品种工件在缓存库中的累积程度. 由表3可知,当品种1和品种2的缓存库容量配比越接近$(\lambda_1/\mu_1):(\lambda_2/\mu_2)$时,系统的工件处理率就越大. 其中黑体部分表示缓存库容量之和为不同固定值时最大的系统工件处理率,对应的是最优缓存库容量配比. 在当前的参数配置中,$(\lambda_1/\mu_1):(\lambda_2/\mu_2)$为5:6,表3中黑体部分对应的都是在缓存库容量之和为固定值时最接近5:6的缓存库容量配比. 例如,当两个品种缓存库容量之和固定为9时,显然配比4:5对应的系统总的处理率最大与5:6也最接近.

表3　各品种缓存库容量配比不同时系统总的工件处理率

品种1容量	品种2容量						
	2	3	4	5	6	7	8
2	0.8336	**0.8573**	0.8691	0.8713	0.8770	0.8808	0.8912
3	0.8465	0.8672	**0.8749**	0.8842	0.8893	0.8939	0.8964
4	0.8535	0.8722	0.8853	**0.8927**	0.8949	0.8976	0.9006
5	0.8660	0.8811	0.8919	0.8968	**0.8987**	0.9037	0.9040
6	0.8699	0.8915	0.8964	0.8979	0.8998	**0.9055**	0.9085
7	0.8728	0.8961	0.8964	0.8995	0.9043	0.9071	**0.9084**
8	0.8771	0.8957	0.8979	0.8997	0.9057	0.9068	0.9093

下面研究服务率与到达率之比变化时的情况. 图6给出了两个品种总的到达率不变,各品种到达率之比不同时系统的总处理率. 当$u_1 = 4, u_2 = 5$和$u_1 = 3, u_2 = 6$时,由于品种1工件在各站点的服务率比品种2工件在各站点的服务率低,随着品种2工件到达率的提高,每个站点加工品种2工件的概率便提高,单位时间内期望的加工工件数会随之增加,系统总的处理率也逐步提高. 反之,当$u_1 = 5, u_2 = 4$时,由于品种1工件在各站点的服务率比品种2工件在各站点的服务率高,随着品种2工件到达率的提高,每个站点加工品种1工件的概率便降低,单位时间内期望的加工工件数会随之减少,系统总的处理率也逐步降低. 此外,由图6可见,$u_1 = 4, u_2 = 5$时,各品种到达率之比变化,系统工件处理率始终比$u_1 = 3, u_2 = 6$时高,这是因为当$u_1 = 4, u_2 = 5$时,$(\lambda_1/\mu_1):(\lambda_2/\mu_2)$始终比$u_1 = 3, u_2 = 6$更接近

于两个品种的缓存库容量之比.

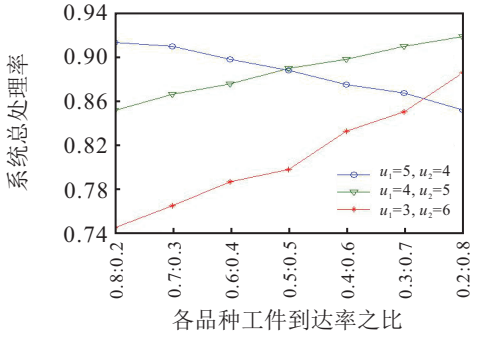


图 6 系统总的处理率

4 结 论

本文研究了两类品种工件混流到达的多站点CSPS系统优化问题,提出了一种品种均衡工作模式,采取模型无关的串行反馈式多agent强化学习算法,得到了较好的优化效果,并对系统参数性能进行了比较分析. 本文仅考虑了工件捡取点固定、工件服务率固定的情况,在实际生产过程中,为了提高生产效率,机械臂可移动捡取传送带上的工件,并且其路径规划和加工速度可以适当调节. 因此,对于多类品种工件混线的可移动卸载式CSPS系统的优化控制将是一个有意义的待拓展的研究课题.

参考文献(References)

[1] Matsui M. A generalized model of convey-serviced production station(CSPS)[J]. J of Japan Industrial Management Association, 1993, 44(1): 25-32.

[2] Matsui M. CSPS model: Look-ahead controls and physics[J]. Int J of Production Research, 2005, 43(10): 2001-2025.

[3] Tang H, Arai T. Look-ahead control of conveyor-serviced production station by using potential-based online policy iteration[J]. Int J of Control, 2009, 82(10): 1917-1928.

[4] Tang H, Xu L L, Sun J, et al. Modeling and optimization control of a demand-driven conveyor-serviced production station[J]. European J of Operational Research, 2015, 243(3): 839-851.

[5] Glorieux E, Danielsson F, Svensson B, et al. Constructive cooperative coevolutionary optimisation for interacting production stations[J]. Int J of Advanced Manufacturing Technology, 2015, 80(1): 1-16.

[6] 唐昊, 万海峰, 韩江洪, 等. 基于多agent强化学习的多站点CSPS系统的协作look-ahead控制[J]. 自动化学报, 2010, 36(2): 289-296.
(Tang H, Wan H F, Han J H, et al. Coordinated look-ahead control of multiple CSPS system by multi-agent reinforcement learning[J]. Acta Automatica Sinica, 2010, 36(2): 289-296.)

[7] Tang H, Wang C, Matsui M, et al. Hierarchical coordinated control of a multi-procedure CSPS system by learning optimisation methods[J]. Int J of Production Research, 2015, 53(7): 2055-2072.

[8] Caputo A C. A decision model for selecting parts feeding policies in assembly lines[J]. Industrial Management & Data Systems, 2015, 115(6): 974-1003.

[9] Feyzbakhsh S A, Matsui M. Adam-eve-like genetic algorithm: a methodology for optimal design of a simple flexible assembly system[J]. Computers & Industrial Engineering, 1999, 36(2): 233-258.

[10] Buyukozkan K, Kucukkoc I, Satoglu S I, et al. Lexicographic bottleneck mixed-model assembly line balancing problem: Artificial bee colony and tabu search approaches with optimised parameters[J]. Expert Systems with Applications, 2016, 50(2): 151-166.

[11] 周旸旻. 面向多品种部件的单站点CSPS系统优化控制模型和方法[D]. 合肥: 合肥工业大学计算机学院, 2015.
(Zhou Y M. Optimal Control model and method of single conveyor-serviced production station with multi-type products[D]. Hefei: College of Computer, Hefei University Technology, 2015.)

[12] Scherrer B, Charpillet F. Cooperative co-learning: A model-based approach for solving multi agent reinforcement problems[C]. The 24th IEEE Int Conf on Tools with Artificial Intelligence. Washington: IEEE Computer Society, 2002: 463-468.

[13] Bowling M, Veloso M. Rational and convergent learning in stochastic games[C]. Proc of the 17th Int Joint Conf on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers Inc, 2001: 1021-1026.

[14] Bowling M, Veloso M. Multiagent learning using a variable learning rate[J]. Artificial Intelligence, 2002, 136(2): 215-250.

[15] Cao X R. Semi-Markov decision problems and performance sensitivity analysis[J]. IEEE Trans on Automatic Control, 2003, 48(5): 758-769.

（责任编辑：郑晓蕾）