DOI: 10.7641/CTA.2015.40217

可变服务率模式下基于需求驱动的传送带给料加工站系统的优化控制

唐 昊[1,2†], 许玲玲[2], 周 雷[2], 谭 琦[1]

(1. 合肥工业大学 电气与自动化工程学院, 安徽 合肥 230009; 2. 合肥工业大学 计算机与信息学院, 安徽 合肥 230009)

摘要: 本文主要研究可变服务率模式下基于需求驱动的传送带给料加工站(CSPS)系统的优化控制问题, 主要目标是对系统的随机优化控制问题进行建模和提供解决方案. 论文以缓冲库和成品库剩余容量为联合状态, 以站点前视距离和工件服务率为控制变量, 将其最优控制问题描述为半马尔科夫决策过程(SMDP)模型. 该模型为利用策略迭代等方法求解系统在平均准则或折扣准则下的最优控制策略提供了理论基础, 特别地, 据此可引入基于模拟退火思想的Q学习算法等优化方法来寻求近似解, 以克服理论求解过程中的维数灾和建模难等困难. 仿真结果说明了本文建立的数学模型及给出的优化方法的有效性.

关键词: 传送带给料加工站; 可变服务率; 半马尔科夫决策过程; Q学习

中图分类号: TP202 **文献标识码**: A

Optimization control of demand-driven conveyor-serviced production station with changeable service rate

TANG Hao[1,2†], XU Ling-ling[2], ZHOU Lei[2], TAN Qi[1]

(1. School of Electrical Engineering and Automation, Hefei University of Technology, Hefei Anhui 230009, China;
2. School of Computer and Information, Hefei University of Technology, Hefei Anhui 230009, China)

Abstract: The optimal control of demand-driven conveyor-serviced production station with changeable service rate is concerned in this paper. We focus on modeling the stochastic control problem and providing solutions. First, the vacancies of the buffer and the bank are jointed to be viewed as the system state, and the look-ahead range and service rate are viewed as the control variable. Then we set up in detail a semi-Markov decision process for the optimal control problem. As a result, policy iteration can be used to obtain the optimal look-ahead range and service rate under either average or discounted-cost criteria. Furthermore, to avoid the disaster of dimensionality and the difficulties of modeling in numerical optimization methods, we also propose a Q-learning algorithm combined with simulated annealing technique to derive the approximate solutions. Simulation results are finally used to validate the effectiveness of our established model and proposed optimization methods.

Key words: conveyor-serviced production station; changeable service rate; semi-Markov decision process; Q-learning

1 引言(Introduction)

在现实生产制造业中存在一类由生产中心和销售中心组成的生产系统, 前者配有有限容量的缓冲库, 用于存放从传送带上卸载的未加工工件, 销售中心设有有限容量的成品库用于存放生产中心加工的成品, 并为顾客的需求服务. 统称此类生产系统为基于需求驱动的传送带给料加工站(conveyor-serviced production station, CSPS)[1–2]. 它源于以福特生产线为代表的工业生产自动化过程, 被广泛用在世界著名的先进企业中[3]. 对于传统的仅具有单个加工站的CSPS的优化控制, 松井正之教授等建立了一种半马尔科夫决策过程(semi-Markov decision process, SMDP)模型, 并总结了不同控制模式下的CSPS系统物理模型[1,3]. 结合这些工作, 文献[4]提出了基于性能势的在线策略迭代算法, 以解决单站点CSPS系统基于仿真或观测数据的最优控制问题. 但这些研究中, 一般皆假设成

收稿日期: 2014−03−18; 录用日期: 2015−03−19.

†通信作者. E-mail: htang@hfut.edu.cn; Tel.: +86 551-62901416.

国家自然科学基金项目(61174186, 61374158, 71231004), 国家国际科技合作项目(2011FA10440), 教育部新世纪优秀人才计划项目(NCET–11–0626), 高等学校博士学科点专项科研基金项目(20130111110007)资助.

Supported by National Natural Science Foundation of China (61174186, 61374158, 71231004), International S & T Cooperation Program of China (2011FA10440), Program for New Century Excellent Talents in University (NCET–11–0626) and Specialized Research Fund for the Doctoral Program of Higher Education (20130111110007)

品库容量无限大, 即最优生产控制并未考虑销售情况. 随着经济的发展, 在生产供应链管理(supply chain management, SCM)时代, 实时生产可能会极大地受到市场的影响[5], 这意味着在不确定顾客需求的情况下, 生产中心的生产调度应该考虑销售中心的库存管理. 所以, 需求驱动优化控制在SCM时代变得必要且更具有吸引力[6–8].

在现代生产制造业中, 基于需求驱动的CSPS系统被视为连接销售中心的生产中心, 尽管该生产模式在SCM背景下越来越普及, 但是目前为止, 关于该模型的研究很少. 松井正之教授在作业调度模型中第一次讨论了该模型, 后来将其发展为管理博弈模型[1,9]. 然而, 前述研究都是假设生产中心的工件服务率是固定的, 而采用固定服务率的生产模式不利于发挥资源的最佳使用效率, 例如, 高服务率生产将导致工作站操作机械(或工人)劳动强度大, 机器损耗高, 造成设备维护成本增加, 低服务率生产不利于提高生产设备的利用率, 并影响销售中心的供货. 因此, 根据实时生产销售情况动态采取可变的服务率进行生产, 在现代生产制造业中已显示出其必要性, 相关研究近年来也引起了工业工程或控制领域学者的关注[10–11]. 文献[12]研究了服务率可变的面向订单装配(assemble-to-order, ATO)系统, 可根据实际情况动态地选择服务率, 使系统的整体性能最优. 文献[13]将服务率作为系统的一个控制变量, 研究了服务率可变的单站点CSPS系统的优化控制. 结合文献[13]和文献[14]等前期工作, 本文主要综合考虑顾客需求随机到达和服务率可动态实时变化情况下的CSPS系统, 根据系统工作模式和决策机制的特点, 对其最优控制问题进行半马尔科夫决策过程建模, 进而给出相关优化模型和算法.

2 系统组成与决策过程(Composition and decision-making process of the system)

基于需求驱动的CSPS系统可看作是由生产中心和销售中心组成的两中心模型, 其示意图见文献[14]中图1所示. 生产中心的加工主体agent可以动态地检测传送带上游一定范围内是否有工件到达, 并记录工件的一些位姿信息, 如工件位置和角度. 假设传送带是匀速运动的, 则传送带上的前视距离可折算成时间来等价表示. 该系统的典型工作过程如下: 在决策时刻, agent根据缓冲库和成品库的剩余容量选择一段前视距离, 如果前视距离内有工件, agent将等待直到第一个工件到达捡取点, 并将其卸载到缓冲库中. 否则, agent将从缓冲库中取出一个工件并以一个可变的服务率$u \in [u_{\min}, u_{\max}]$进行加工. 加工时, 任何流经捡取点的工件都将丢失, 加工后的成品被放入成品库并提供给顾客. 在等待卸载或加工的过程中, 顾客随机到达销售中心. 假设每个顾客只从成品库中取走一个成品, 如果成品库为空, 顾客将流失. 这里, 生产中心的前视距离和服务率被视为控制或者决策变量.

在本文中, 加工、卸载操作常伴随多种类型生产代价, 如等待代价、缓冲库和成品库的存储代价、加工代价等. 优化控制目标是找到一个最优策略使系统在平均性能准则或者折扣性能准则下的长远期望代价最小. 为了便于说明, 系统的一些重要参数变量如表1所示.

表 1　参数说明

Table 1　Parameter specification

变量	物理意义
λ_a	工件的泊松流到达率
λ_c	顾客的泊松流到达率
u	生产中心服务率
$S(t,u)$	单个工件的加工时间分布函数
$u_{\min}$	最小服务率
$u_{\max}$	最大服务率
$\hat{u}$	额定服务率设为$(u_{\min}+u_{\max})/2$
N	缓冲库的容量
K	成品库的容量
n	缓冲库的剩余容量
k	成品库的剩余容量
$v_{S_{k,n}}$	前视距离, 用等价时间表示
$l_{\min}$	最小前视距离
$l_{\max}$	最大前视距离
k_1	单个待加工工件的单位时间库存代价
k_2	单位时间加工代价
k_3	单位时间等待代价
k_4	加工完一个工件的即时报酬, 用负数表示
k_5	单位长度前视距离的即时代价
k_6	单个成品的单位时间库存代价
k_7	售出一个成品的即时报酬, 用负数表示
k_8	加工时服务率偏离额定值的单位时间代价

定义$k \in \Phi_1 = \{0, 1, \cdots, K\}$为成品库状态, $n \in \Phi_2 = \{0, 1, \cdots, N\}$为缓冲库状态, $S_{k,n}$为系统的联合状态, 状态空间为$\Phi = \Phi_1 \times \Phi_2$. 在状态$S_{k,n}$下, 系统采取的行动表示为二元组$(v_{S_{k,n}}, u_{S_{k,n}})$, 由状态相关的前视距离$v_{S_{k,n}}$和服务率$u_{S_{k,n}}$组成.

系统运行时, 如果缓冲库和成品库都为满, agent将一直等待, 直到有顾客到达销售中心并从成品库中取走一个成品, 然后才开始新的决策过程. 此情况下, 不失一般性, 可令$v_{S_{0,0}} \equiv 0$, $u_{S_{0,0}} \equiv \hat{u}$. 如果只有成品库为满且缓冲库不为空, 假设agent不取工件进行加工, 要么等到工件到达捡取点这一事件发生并进行卸载, 要么等到有顾客到达销售中心取走一个成品这一事件发生, 才启动下个决策周期. 此情况下, 可令$v_{S_{0,n}} \equiv \infty$, $n = 1, 2, \cdots, N-1$, 且$u_{S_{0,n}} \equiv \hat{u}$. 如果只有缓冲库为满, 则$v_{S_{k,0}} \equiv 0$, $k = 1, 2, \cdots, K$. 当缓冲库为空时, agent将一直等待到有工件到达捡取点并卸载至缓冲库中, 所以$v_{S_{k,N}} \equiv \infty$, $k \in \Phi_1$. 对于其他的

一些典型工作过程, 有$v_{S_{k,n}}\in[l_{\min},l_{\max}]$, $u_{S_{k,n}}\in[u_{\min},u_{\max}]$. 于是, 系统所有前视距离的集合为$D=[l_{\min},l_{\max}]\cup\{0,\infty\}$, 平稳策略为$V=\{(v_{S_{k,n}},u_{S_{k,n}})|\ v_{S_{k,n}}\in D,u_{S_{k,n}}\in[u_{\min},u_{\max}],\ S_{k,n}\in\Phi\}$.

假设在控制策略V作用下, 系统由初始决策时刻$T_0=0$开始进行状态演化, 其状态过程记为X_t, $t\geqslant 0$. 在决策时刻T_m, 若系统状态为$S_{k,n}$, 记$X_{T_m}=X_m=S_{k,n}$. 根据策略V, 采取行动$(v_{S_{k,n}},u_{S_{k,n}})$, 如果agent实际采取等待动作, 记等待时间记为ϖ_m, 则下一个决策时刻$T_{m+1}=T_m+\varpi_m$. 典型情况下, T_{m+1}时系统转移到状态$X_{m+1}=S_{k',n-1}$, $k'\in\{k,k+1,\cdots,K\}$. 作为一个特例, 当$X_m=S_{0,n}(n\neq N)$时, 以一定概率转移到$X_{m+1}=S_{1,n}$.

相反, 如果$v_{S_{k,n}}$范围内没有工件, agent将以服务率$u_{S_{k,n}}$加工从缓冲库中取出的工件. 为保证决策过程的无记忆性, 定义$T_{m+1}=T_m+\tau_m+\varpi_m$, 这里, τ_m为服务时间, $\varpi_m=\max\{v_{S_{k,n}}-\tau_m,0\}$是等待时间, 并且$X_{m+1}=S_{k',n+1},k'\in\{k-1,k,\cdots,K\}$.

3 系统优化的半马尔科夫决策过程模型 (Semi-Markov decision process model for system optimization)

由上节分析可知, 系统状态过程X_t是连续时间半马尔科夫过程, $\{X_0,X_1,\cdots,X_m,\cdots\}$是一个嵌入的马氏链[15]. 记策略$V$下的逗留时间分布矩阵

$$F^V(t)=[F_{S_{k,n}S_{k',n'}}(t,(v_{S_{k,n}},u_{S_{k,n}}))]_{S_{k,n},S_{k',n'}\in\Phi},$$

其中$F_{S_{k,n},S_{k',n'}}(t,(v_{S_{k,n}},u_{S_{k,n}}))$表示系统在状态$S_{k,n}$下采取行动$(v_{S_{k,n}},u_{S_{k,n}})$并转移到下一个状态$S_{k',n'}$的逗留时间, 在服务率可变的系统中, 其公式在形式上与服务率固定的情形类似[14], 受篇幅限制, 此处不再赘述. 记系统的嵌入链转移矩阵为

$$P^V=[P_{S_{k,n}S_{k'n'}}(v_{S_{k,n}},u_{S_{k,n}})]_{S_{k,n},S_{k',n'}\in\Phi},$$

其中$P_{S_{k,n}S_{k',n'}}(v_{S_{k,n}},u_{S_{k,n}})$表示系统在状态$S_{k,n}$下采取行动$(v_{S_{k,n}},u_{S_{k,n}})$转移到下一状态$S_{k'n'}$的概率. 首先, 有

$$\begin{cases}P_{S_{0,0}S_{1,0}}(v_{S_{0,0}},u_{S_{0,0}})\equiv 1,\\ P_{S_{0,n}S_{0,n-1}}(v_{S_{0,n}},u_{S_{0,n}})=\dfrac{\lambda_a}{\lambda_a+\lambda_c},\\ P_{S_{0,n}S_{1,n}}(v_{S_{0,n}},u_{S_{0,n}})=\dfrac{\lambda_c}{\lambda_a+\lambda_c},\\ n=1,2,\cdots,N-1.\end{cases}\tag{1}$$

对于其他情况, 在一个给定的行动$(v_{S_{k,n}},u_{S_{k,n}})$作用下, 缓冲库和成品库的状态转移相互独立, 前者由工件的随机到达决定, 后者由顾客的随机到达决定. 记行动$(v_{S_{k,n}},u_{S_{k,n}})$下缓冲库由状态$n$转移到$n'$的概率为$\hat{p}_{n,n'}(v_{S_{k,n}},u_{S_{k,n}})$, 成品库由状态$k$转移到下一状态$k'$的概率为$\bar{p}_{k,k'}(v_{S_{k,n}},u_{S_{k,n}})$, 两者在形式上亦与固定服务率的情况类似[14]. 因此, 在行动$(v_{S_{k,n}},u_{S_{k,n}})$下, 如果agent采取卸载操作, 则系统状态转移概率为

$$\begin{cases}P_{S_{k,n}S_{k',n-1}}(v_{S_{k,n}},u_{S_{k,n}})=\\ \hat{p}_{n,n-1}(v_{S_{k,n}},u_{S_{k,n}})\times\bar{p}_{k,k'}(v_{S_{k,n}},u_{S_{k,n}}),\\ k=1,2,\cdots,K-1;\ n=1,2,\cdots,N-1.\end{cases}\tag{2}$$

若agent采取加工操作, 则系统状态转移概率为

$$\begin{cases}P_{S_{k,n}S_{k',n+1}}(v_{S_{k,n}},u_{S_{k,n}})=\\ \hat{p}_{n,n+1}(v_{S_{k,n}},u_{S_{k,n}})\times\bar{p}_{k,k'}(v_{S_{k,n}},u_{S_{k,n}}),\\ n=1,2,\cdots,N-1.\end{cases}\tag{3}$$

于是, 计算半马尔科夫核[15]

$$Q^V(t)=[Q(S_{k,n},S_{k',n'},(v_{S_{k,n}},u_{S_{k,n}}),t)]_{S_{k,n},S_{k',n'}\in\Phi},$$

且

$$Q(S_{k,n},S_{k',n'},(v_{S_{k,n}},u_{S_{k,n}}),t)=P_{S_{k,n}S_{k',n'}}(v_{S_{k,n}},u_{S_{k,n}})\times F_{S_{k,n}S_{k',n'}}(t,(v_{S_{k,n}},u_{S_{k,n}})).$$

显然, 由于系统的状态空间由缓冲库和成品库的剩余容量组成, 行动由前视距离和服务率组成, 状态–行动对很大, 很容易给系统造成维数灾. 同时由于系统状态过多, 状态转移比较复杂, 系统的状态转移矩阵推导和后期的系统矩阵等价处理面临较大的计算困难[15] , 因此给系统带来一定的建模难问题.

令$f^V=[f(S_{k,n},(v_{S_{k,n}},u_{S_{k,n}}),S_{k',n'})]$为系统性能矩阵, 其中$f(S_{k,n},(v_{S_{k,n}},u_{S_{k,n}}),S_{k',n'})$表示系统在状态$S_{k,n}$采取行动$(v_{S_{k,n}},u_{S_{k,n}})$转移到下一状态$S_{k',n'}$之前系统单位时间内获得的期望代价. 假设系统在状态X_m下, 采取行动(v_{X_m},u_{X_m}), 并转移到下一个状态X_{m+1}, 系统得到一个样本转移$<X_m,(v_{X_m},u_{X_m}),X_{m+1},\omega_m,\tau_m>$, 其中: τ_m是服务时间, ω_m是T_m到T_{m+1}的时间间隔, 即$\omega_m=T_{m+1}-T_m$. 系统在某个时间点或时间间隔内将付出各种各样形式的代价, 如前视距离代价, 分别在$[T_m,T_m+\tau_m]$和$[T_m+\tau_m,T_{m+1}]$时间区间内积累的服务代价和等待代价等. 另外, 在服务过程中, 如果服务率偏离额定值, 还将会产生相应代价. 所有这些代价可等价折算成$[T_m,T_{m+1}]$整个时间段内的单位时间代价$f(S_{k,n},(v_{S_{k,n}},u_{S_{k,n}}),S_{k',n'})$, 详细的转换过程可参考文献[14].

于是, 系统的优化目标是对任意折扣因子$\alpha>0$, 寻找一个最优策略使得下列折扣性能准则

$$\begin{cases}\eta_\alpha^V(S_{k,n})=\\ \mathrm{E}[\lim\limits_{M\to\infty}\sum\limits_{m=0}^{M-1}\int_{T_m}^{Tm+1}\alpha\mathrm{e}^{-\alpha t}f_{X_m}\mathrm{d}t|X_0=S_{k,n}],\\ S_{k,n}\in\Phi.\end{cases}\tag{4}$$

的值最小. 这里$f_{X_m}=f(X_m,(v_{X_m},u_{X_m}),X_{m+1})$. 当$\alpha$趋于零时, 性能向量$\eta_\alpha^V$任一分量的极限都等于平均代价性能准则值$\eta^V$.

通过计算等价无穷小因子A_α^V和等价性能矩阵f_α^V, 前述半马尔科夫决策过程可转换为连续时间马尔科夫决策过程模型$(X_t,\Phi,D,A_\alpha^V,f_\alpha^V)$[15–17]. 则优化问题可由策略迭代方法求解, 即在一个典型迭代过程中, 对于当前策略V_k, 可计算性能势向量$g_\alpha^{V_k}$, 并进行策略更新$V_{k+1}\in\arg\min\limits_V\{f_\alpha^V+A_\alpha^V g_\alpha^{V_k}\}$.

由于理论优化算法依赖于模型, 可引入Q学习算法来求解最优或者次优控制策略. 前面建立的半马尔科夫决策过程模型为Q学习算法提供了理论基础, 与文献[4]类似, 对于观测得到或仿真得到的一个样本转移$<X_m,(v_{X_m},u_{X_m}),X_{m+1},\omega_m,\tau_m>$, 平均准则和折扣准则下的统一差分公式为

$$\begin{aligned}c_m=\ & f'(X_m,(v_{X_m},u_{X_m}),X_{m+1})-\\ & T_\alpha(\omega_m)\eta_m+\mathrm{e}^{-\alpha\omega_m}\min_{d\in D}Q_\alpha(X_{m+1},d)-\\ & Q_\alpha(X_m,(v_{X_m},u_{X_m})).\end{aligned}\tag{5}$$

这里$Q_\alpha(\cdot,\cdot)$是在折扣情况下状态–行动对的Q值. 令$f'(X_m,(v_{X_m},u_{X_m}),X_{m+1})$表示从$T_m$到$T_{m+1}$时间内累积的折扣代价, 则有

$$\begin{aligned}&f'(S_{0,0},(v_{S_{0,0}},u_{S_{0,0}}),S_{1,0})=\\ &T_\alpha(\omega_m)\times(k_1N+k_3+k_6K)+\mathrm{e}^{-\alpha\tau_m}k_7,\end{aligned}\tag{6}$$

$$\begin{aligned}&f'(S_{0,n},(v_{S_{0,n}},u_{S_{0,n}}),S_{1,n})=\\ &T_\alpha(\omega_m)\times t[k_1(N-n)+k_3+k_6K]+\mathrm{e}^{-\alpha\tau_m}k_7,\end{aligned}\tag{7}$$

$$\begin{aligned}&f'(S_{0,n},(v_{S_{0,n}},u_{S_{0,n}}),S_{0,n-1})=\\ &T_\alpha(\omega_m)\times[k_1(N-n)+k_3+k_6K],\end{aligned}\tag{8}$$

$$\begin{aligned}&f'(S_{k,n},(v_{S_{k,n}},u_{S_{k,n}}),S_{k',n-1})=\\ &T_\alpha(\omega_m)\times[k(N-n)+k_3+k_6(K-k)]+\\ &k_5v_{S_{k,n}}\times\chi_N(v_{S_{k,n}})+\mathrm{e}^{-\alpha\tau_m}k_7(k'-k),\end{aligned}\tag{9}$$

$$\begin{aligned}&f'(S_{k,n},(v_{S_{k,n}},u_{S_{k,n}}),S_{k',n+1})=\\ &T_\alpha(\omega_m)\times[k_1(N-n-1)+k_6(K-k)]+\\ &k_2\mathrm{T}_\alpha(\tau_m)+k_3[T_\alpha(\omega_m)-T_\alpha(\tau_m)]+\\ &k_4\mathrm{e}^{-\alpha\tau_m}+k_5v_{S_{k,n}}+\mathrm{e}^{-\alpha\tau_m}k_7(k'-k+1)+\\ &k_8|u_{S_{k,n}}-\hat{u}|\mathrm{e}^{-\alpha\tau_m}.\end{aligned}\tag{10}$$

上述式子中, $T_\alpha(t)=\int_0^t\mathrm{e}^{-\alpha t}\mathrm{d}t$. 另外, η_m是平均代价η^V的当前估计值, 可根据样本轨道直接计算. 因此, Q值更新公式为

$$\begin{aligned}&Q_\alpha(X_m,(v_{X_m},u_{X_m})):=\\ &Q_\alpha(X_m,(v_{X_m},u_{X_m}))+\gamma(X_m,(v_{X_m},u_{X_m}))c_m,\end{aligned}\tag{11}$$

这里$\gamma(X_m,(v_{X_m},u_{X_m}))$是学习步长.

传统的Q学习通常采用ε-greedy算法, 但若ε取值不当, 则有可能让算法进入局部最优, 为了克服这种困难, 可引入基于模拟退火的Q学习算法(simulated annealing Q–learning, SA–Q), 其详细过程可参考文献[14], 本文不再赘述.

4　仿真结果(Simulation results)

仿真中, 设工件服务时间服从(L,u)的Erlang分布[4], 系统物理参数和算法参数的设置如表2所示.

表 2　参数设置表

Table 2　The value of parameters

λ_a	λ_c	$u_{\min}$	$u_{\max}$	N	K	$l_{\min}$	$l_{\max}$	L	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	$\hat{u}$
1	0.8	3.5	4.5	5	6	0	1	4	0.1	5	1	–10	0.1	0.01	0	1	4

图1是策略迭代算法的优化曲线, 其中用三角形图标标注的曲线表示平均准则下的优化曲线, 其余是折扣因子$\alpha=0.01$时以下3个特殊状态的优化曲线: 状态3, 缓冲库的空余量为2, 成品库为满; 状态6, 缓冲库为满, 成品库的空余量为1; 状态8, 缓冲库的空余量为2, 成品库的空余量为1.

通过观察图1可以看出策略迭代算法收敛速度很快, 并且通过策略迭代算法系统可以得到最优前视距离和服务率, 在折扣准则下的优化曲线是相互分离的且分布在平均准则优化曲线的两侧. 在状态3时, 成品库一开始是满的, 系统的生产将被暂停直到有一个顾客到达销售中心并从成品库中取走一个成品, 这一过程将导致缓冲库的存储代价和等待代价增大, 所以状态3的优化曲线在图1的最上面. 此外, 在状态7时, 缓冲库一开始是满的, 这时agent不用向前看, 以一定的服务率服务一个工件, 产品报酬将会立即产生, 因此总的累积代价减少, 这就解释了为什么状态6的优化曲线在图1的最下面. 在状态6时, 系统学习的行动为(0, 4.5), 这表明, 在缓冲库为满时, agent不需要前视距离, 并且以最大的服务率进行加工, 与实际生产情况符合.

图2是平均准则下, 服务率固定和服务率可变下的优化曲线. 从图中可以看出, 对不同的k_8, 服务率变化下系统的平均代价均小于服务率固定时系统的

平均代价, 本实验中, 当系统服务率变化时, 系统会产生一定的代价, 所以随着k_8的增大, 系统的平均代价逐渐增大. 算法开始时系统的平均代价较大, 说明此时系统的服务率与额定服务率相差较大, 但是随着学习的进行, 系统根据缓冲库和成品库的库存情况合理地选择服务率, 使系统的最终代价减少并且收敛到最优值. 进一步说明, 通过策略迭代算法, 可变服务率模式下基于需求驱动的CSPS系统取得了较好的优化效果, 通过动态地控制生产中心的服务率可以有效的降低系统的平均代价.

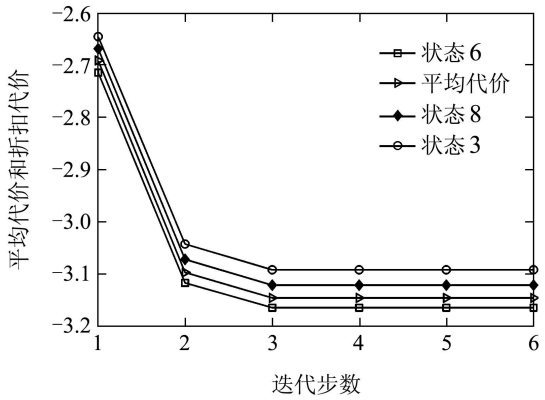

图 1 平均准则和折扣准则下的策略迭代优化

Fig. 1 Policy iteration under average- and discounted-criterion

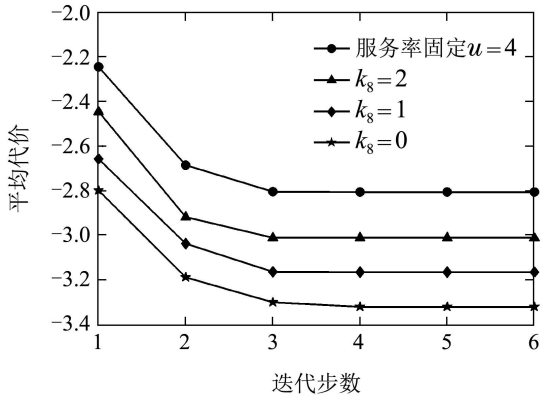

图 2 服务率固定与服务率变化的优化曲线

Fig. 2 The optimization plots under fixed and changeable service rate

图3和图4分别表示平均准则和折扣准则下SA–Q算法优化曲线. 其中每个曲线学习1000个片段, 每个片段包括100步, 学习过程中根据当前的Q值表选择一个贪婪策略, 然后通过理论方法评估这个策略. 图中可以看到, 刚开始曲线波动非常大, 最后趋于稳定. 通过SA–Q算法, 可以得到次优策略和性能值, 且与策略迭代算法得到的最优值非常接近. 因此, SA–Q算法可以有效的解决可变服务率模式下基于需求驱动的CSPS系统的优化控制问题. 事实上, 它比使用“ε-greedy”算法进行探索的传统Q学习算法在学习速度和性能上更有优势.

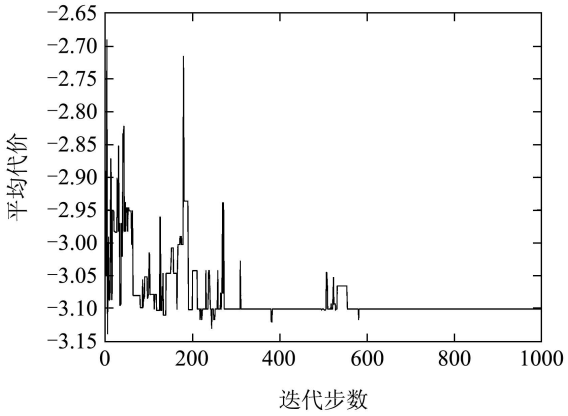

图 3 平均准则下SA–Q算法优化曲线

Fig. 3 The optimization plot of SA–Q under average criterion

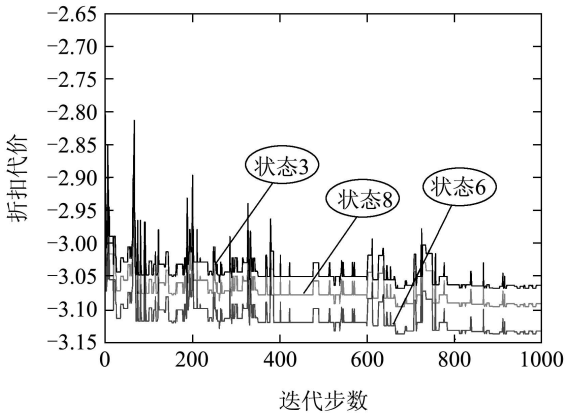

图 4 折扣准则下SA–Q算法优化曲线

Fig. 4 The optimization plots of SA–Q under discounted criterion

图5表示服务率固定和服务率可变的基于需求驱动的CSPS系统的单位时间生产率(后面简称生产率), 服务率固定时令$u=4$. 从图中可以看出, 可变服务率模式下基于需求驱动的CSPS系统的生产率比服务率固定情况下的生产率有明显的提高. 进一步说明通过调节生产中心的服务率, 让系统根据缓冲库和成品库中工件的个数动态实时地选择服务率可以有效减少系统的等待时间, 增加系统总的服务时间, 从而提高系统的生产率. 从图6中还可以看出, 在服务率固定的情况下, 随着服务率的增大, 系统等待时间逐渐减少最终收敛到固定值. 图6中菱形图标标注的曲线表示服务率可变时$u\in[3.5,4.5]$系统等待时间优化曲线, 此时系统的等待时间为0.2388, 虽然略大于$u=4.5$时系统的等待时间, 但是却远远小于服务率固定为$u=4$和$u=3.5$时系统的等待时间, 而且结合表3可以看出, 在这种情况下

的系统平均代价小于服务率固定为$u = 4.5$时的平均代价, 所以根据系统的实际情况合理地选择生产中心的服务率, 可以使系统的整体性能最优.

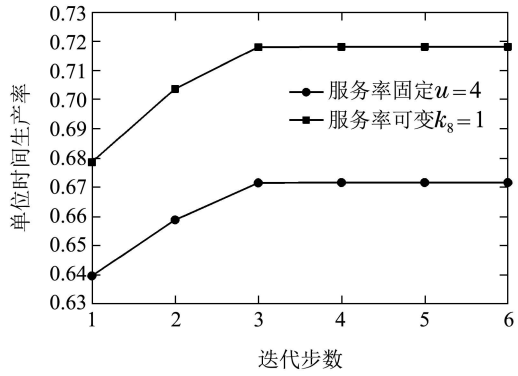

图 5 服务率固定与服务率变化的单位时间生产率

Fig. 5 The production rate per unit time under fixed or changeable service rate

表3统计了服务率固定与服务率可变情况下系统的平均代价、等待时间和单位时间生产率. 可见, 在服务率固定的情况下, 随着服务率的增加系统的平均代价和等待时间减少, 单位时间生产率有所提高. 但服务率并不是越大越好, 如果服务率过大, 将导致生产中心劳动量过大, 机器损耗过大, 导致系统成本增加. 系统根据缓冲库和成品库的库存情况合理地选择服务率, 不仅可以减少系统的平均代价和等待时间, 还可以有效地提高系统单位时间的生产率, 如表3的最后两列所示. 从表3的最后3行可以看出, k_8增加对系统的等待时间的影响非常小, 结合图6, 进一步说明了可变服务率模式下基于需求驱动的CSPS系统可以有效减少系统的等待时间, 增加系统生产率. 从表3还可以看出, k_8对系统的平均代价有一定的影响, 对系统的单位时间生产率影响并不大, 说明了该模型与研究方法的合理性, 通过动态控制生产中心的服务率可以有效地改善CSPS系统的性能, 使系统整体性能达到最优.

表 3 服务率固定与变化情况下系统的性能值

Table 3 Performance values under fixed or changeable service rate

服务率	策略迭代算法	基于模拟退火Q学习算法	等待时间	单位时间生产率
服务率固定$u = 3.5$	–2.1656	–2.0154	1.6431	0.6080
服务率固定$u = 4$	–2.8041	–2.7669	1.4892	0.6715
服务率固定$u = 4.5$	–3.3635	–2.2582	1.3481	0.7418
服务率变化$k_8 = 0$	–3.3198	–3.1899	1.3545	0.7383
服务率变化$k_8 = 1$	–3.1640	–3.0442	1.3829	0.7231
服务率变化$k_8 = 2$	–3.0102	–2.8913	1.3924	0.7182

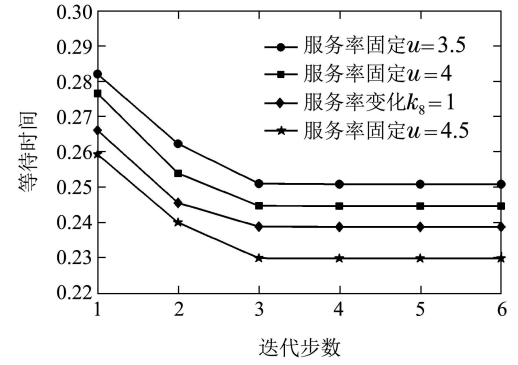

图 6 服务率固定与服务率变化的系统等待时间

Fig. 6 The waiting time under fixed or changeable service rate

5 总结(Conclusions)

本文将顾客需求随机到达和服务率可实时控制的情况同时考虑到CSPS系统中, 给出了详细的半马尔科夫决策过程模型建模过程和理论求解算法. 由于本系统将前视距离和服务率作为系统的控制变量, 并将销售中心的库存情况作为影响系统前视距离控制和服务率控制的一个重要因素, 故该系统的优化问题比传统的CSPS系统更复杂. 为了避免理论算法带来的"维数灾"和"建模难"的问题, 本文给出的基于观测样本转移构建的基于模拟退火的Q学习算法, 可有效解决系统的优化控制问题. 本文目前只考虑了一个生产中心、一个销售中心和单品种产品加工的情况, 而具有多个生产中心、多个销售中心和多品种工件的可变服务率CSPS系统的优化控制问题研究将是一件具有挑战性的工作.

参考文献(References):

[1] MATSUI M. CSPS model: look-ahead controls and physics [J]. *International Journal of Production Research*, 2005, 43(10): 2001 – 2025.

[2] CHEN Y J, TANG H, ZHOU L, MA X S. Look-ahead control of conveyor-serviced production station under stochastic demand [C] //*Proceedings of the 8th World Congress on Intelligent Control and Automation*. Newyork: IEEE, 2010: 6 – 9.

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

[4] TANG H, ARAI T. Look-ahead control of conveyor-serviced produc-

tion station by using potential-based online policy iteration [J]. *International Journal of Control*, 2009, 82(10): 1917 – 1928.

[5] MIKURIYA H, NAKADE K. The optimal production-instruction policy for the production and sales model of two-type products [C] //*Presentation Conference of Chubu Branch*. Japan: Japan Industrial Management Association, 2013: 5 – 6.

[6] YAMADA T, SATOMI K, MATSUI M. Strategic selection of assembly systems under viable demands [R]. *Assembly Automation*, 2006, 26(4):335 – 342.

[7] SHEN W J, DUENYAS I, KAPUSCINSKI R. Optimal pricing, production, and inventory for new Product diffusion under supply constraints [J]. *Manufacturing & Service Operations Management*, 2014, 16(1): 1523 – 4614.

[8] JIANG Y H, MARIA A R, IIRO H, et al. Optimal supply chain design and management over a multi-period horizon under demand uncertainty. Part II: A Lagrangean decomposition algorithm [J]. *Computers and Chemical Engineering*, 2014，62: 211 – 224.

[9] MATSUI M. On a joint policy of order-selection and switch-over [J]. *Journal of Japan Industrial Management Association*, 1988, 39(2): 83 – 87.

[10] BEN-SALEM A, GHARBI A, HAJJI A. An environmental hedging point policy to control production rate and emissions in unreliable manufacturing systems [J]. *International Journal of Production Research*, 2015, 53(2): 435 – 450.

[11] ZANONI S, BETTONI L, GLOCK C H. Energy implications in a two-stage production system with controllable production rates [J]. *International Journal of Production Economics*, 2014, 149(special issue): 164 – 171.

[12] DI Y, BAO Y Y, XIN ZH. Performance analysis of ATO system with changeable service rates [C] //*International Conference on E-Business and E-Government (ICEE)*. New York: IEEE, 2011: 1 – 3.

[13] QING Q C, TANG H, ZHOU L, et al. The optimization control of single conveyor-serviced production station with variable service rate [C] //*Proceedings of the 32nd Chinese Control Conference*. New York: IEEE, 2013: 2180 – 2184.

[14] TANG H, XU L L, SUN J, et al. Modeling and optimization control of a demand-driven, conveyor-serviced production station [J]. *European Journal of Operational Research*. doi:10.1016/j.ejor.2015.01.009.

[15] CAO X R. *Stochastic Learning and Optimization: A Sensitivity-Based View* [M]. New York: Springer, 2007.

[16] 殷保群, 奚宏生, 周亚平. 排队系统性能分析与Markov控制过程 [M]. 合肥：中国科学技术大学出版社, 2004.
(YIN Baoqun, XI Hongsheng, ZHOU Yaping. *Queueing System Performance Analysis and Markov Control Processes* [M]. Hefei: Press of University of Science and Technology of China, 2004.)

[17] CAO X R. Semi-Markov decision problems and performance sensitivity analysis [J]. *IEEE Transactions on Automatic Control*, 2003, 48(5): 758 – 769.

作者简介:

唐 昊 (1972−), 男, 教授, 博士生导师, 主要研究方向为离散事件动态系统、智能生产系统、强化学习和神经元动态规划, E-mail: htang@hfut.edu.cn;

许玲玲 (1989−), 女, 硕士, 主要研究方向为离散事件动态系统和强化学习, E-mail: llxu891213 @163.com;

周 雷 (1981−), 男, 讲师, 主要研究方向为离散时间动态系统和强化学习, E-mail: zhoulei @hfut.edu.cn;

谭 琦 (1985−), 男, 讲师, 主要研究方向为智能优化方法, E-mail: tanqi@hfut.edu.cn.