文章编号: 1000−8152(2012)03−0323−08

一类排队服务系统的最优控制策略研究

刘名武[1], 杨迎春[2], 马永开[3]

(1. 重庆交通大学 管理学院, 重庆 400074;

2. 上海财经大学 国际工商管理学院, 上海 200433; 3. 电子科技大学 经济与管理学院, 四川 成都 610054)

摘要: 建立一个双阈值排队服务模型, 用来协调顾客的等待时间和服务系统的运行成本. 采用一种精细的概率分解方法获得瞬态和稳态队长的概率分布性质; 再根据系统稳态性能指标, 建立系统营运利润函数, 并设计一种针对双离散变量函数的全局优化收索算法来研究最优控制策略; 进一步的数值实验揭示出双重阈值策略的优越性.

关键词: 排队服务; 控制策略; $M/G/1$排队; 双重阈值

中图分类号: O213.2; O226 **文献标识码**: A

Optimal control policy for a queuing service system

LIU Ming-wu[1], YANG Ying-chun[2], MA Yong-kai[3]

(1. School of Management, Chongqing Jiaotong University, Chongqing 400074, China;

2. School of International Business Administration, Shanghai University of Finance and Economics, Shanghai 200433, China;

3. School of Management and Economics, University of Electronic Science and Technology of China, Chengdu Sichuan 610054, China)

Abstract: This paper builds a double-threshold queuing service model for coordinating customer's waiting time and operating costs. Using an elaborated probability decomposition technique, we derive the transient and stationary queue-length probability distribution. Next, according to the system steady-state performance index, the system net profit function is modeled. A global algorithmic approach to optimize the dual discrete variable function is proposed, which is for learning the optimal control policy. Furthermore, numerical experiments show the advantages of the double threshold control policy.

Key words: queueing service; control policy; $M/G/1$ queueing; double threshold

1 引言(Introduction)

考虑一类生产管理中如来料生产/加工、机器设备维修等排队问题. 当等待服务的顾客数(顾客对应于等待加工的原材料或者等待维修的机器)很少时, 启动服务系统(服务系统对应于生产设备或者机器维修设备)来服务往往不经济. 管理员常常采用等待顾客累计达到一定量(整数N)时才启动服务系统的策略来节约系统的运行费用. 一般来说, 系统工作前需要预热一段时间, 这就会导致顾客的等待时间过长, 不仅造成服务水平的下降也会增加系统容纳顾客的存贮成本. 因此需要深化运作策略, 提高系统服务效率.

大量文献研究了排队系统的最优控制策略, Tadj和Choudhury[1]对排队系统的最优设计和控制作了详细的综述, 控制策略如: N–策略、早启动的N–策略等广泛应用于连续时间排队系统[2–6]和离散时间排队系统[7]中. 这些文献推导出稳态队长分布的概率母函数, 并考察了系统的最优控制策略. 但是, 系统的稳态队长概率分布却少有研究. 此外, 对于排队网控制策略的研究也受到学者的关注[8].

本文借鉴上述研究成果, 通过引入一双阈值(m, N且$m < N$)策略建立一类$M/G/1$型服务排队模型. 所谓双阈值策略是指当系统中等待服务的顾客数达到m时就启动服务系统, 启动预热完成后如果系统中等待服务的顾客数累计达到N就立即开始服务, 否则系统处于休眠的等待状态直到要求服务的顾客数达到N. 这种双阈值策略用阈值m来控制服务系统的启动对冲服务系统预热的副作用, 用阈值N控制服务系统的服务启动以降低系统频繁启动/关闭的频率, 起到节约系统运行成本的目的. 不同于上述文献, 本文研究系统队长的概率分布, 为系统的容量(对应于原材料或者等待修理的机器设备存储空间)设计提供依据; 并且从利润最大化角度研究的系统最优控制策略, 为管理者提供决策支持.

2 模型描述(Model description)

本文建立双重阈值(m, N)−控制的$M/G/1$排队系统, 刻画一类如来料生产/加工、机器设备维修等生产管理中遇到排队问题. 基本假设如下:

收稿日期: 2010−10−21; 收修改稿日期: 2011−03−24.
基金项目: 国家自然科学基金资助项目(70672104, 71171138); 重庆市科委自然科学基金资助项目(cstcjjA30014).

1) 顾客到达服务台形成参数为λ的Poisson过程, 并且根据到达顺序排成一队列;

2) 服务台一次只能服务一个顾客, 服务时间(记作G)是独立同分布的随机变量且有一般的分布函数$G(t)$, $t \geqslant 0$;

3) 系统实施(m, N)–控制策略: 当系统空闲时服务台立即关闭, 系统进入闲期. 当闲期中到达的顾客数累计到m, 服务台立即启动预热且随机的预热时间为Y且Y的分布函数为$Y(t)$, $t \geqslant 0$. 启动预热完成后, 若系统中的等待服务的顾客不少于N, 服务台马上为顾客服务进入忙期, 否则服务台处于休眠状态进入等待期, 直到系统中的顾客数达到N时才开始为顾客服务.

为了便于分析, 把服务台闲期分成3个子期, 分别称为建立期、启动期和等待期. 记随机变量Y的期望为$\mathrm{E}[Y]$, 其Laplace-Stieltjes变换(LST)为

$$y(s) = \int_0^\infty \mathrm{e}^{-st}\mathrm{d}Y(t).$$

3 系统队长的概率分布(Probability distribution of the queue-length)

采用更新过程理论和全概率分解技术对系统结构进行分解, 把处于任一时刻队长的概率分布表达出来, 然后对其采取Laplace变换导出队长的瞬态分布迭代关系式, 再利用阿贝尔引理推导队长的稳态分布递推关系式. 首先从系统忙期开始来研究忙期中的队长. 记随机变量B表示从一个顾客开始的忙期长度; $N(t)$代表时刻t系统中的顾客数; $Q_j(t)$则表示忙期中时刻t系统中顾客数为j的概率, 其Laplace变换为$q_j^*(s)$. 忙期中的队长性质同经典的$M/G/1$排队系统.

3.1 经典的$M/G/1$排队系统忙期结果(Results of the busy periods for classical $M/G/1$ queue)

对于$M/G/1$排队系统, 当有顾客到达系统时服务台开始服务, 忙期开始. 根据文献[9]中的结论, 有

$$\lim_{t\to\infty} B(t) = \lim_{s\to 0^+} b(s) = 1,\ \mathrm{E}[b] = \frac{\rho}{\lambda(1-\rho)},\ \rho < 1, \tag{1}$$

$$q_j^*(s) = \frac{1}{g(s+\lambda)}\{b(s)\int_0^\infty \mathrm{e}^{-(s+\lambda)t}\bar{G}(t)\frac{(\lambda t)^{j-1}}{(j-1)!}\mathrm{d}t + \sum_{k=1}^{j-1}\frac{q_{j-k}^*(s)}{b^k(s)}[b(s) - \sum_{i=0}^{k}\int_0^\infty \mathrm{e}^{-(s+\lambda)t}\frac{[\lambda b(s)]^i}{i!}\mathrm{d}G(t)]\}, \tag{2}$$

其中:

$$q_j^*(s) = \int_0^\infty \mathrm{e}^{-st}Q_j(t)\mathrm{d}t,$$

$$g(s+\lambda) = \int_0^\infty \mathrm{e}^{-(s+\lambda)t}\mathrm{d}G(t),$$

$$\rho = \lambda \mathrm{E}[G],\ \bar{G}(t) = 1 - G(t).$$

3.2 队长的瞬态分布和稳态分布(Transient and stationary probability distribution of the queue-length)

假设$p_{i,j}(t)$表示初始状态队长为i而时刻t的队长为j的概率, 即$p_{i,j}(t) = p\{N(t) = j | N(0) = i\}$, 其Laplace变换记为$p_{i,j}^*(s) = \int_0^\infty \mathrm{e}^{-st}p_{i,j}(t)\mathrm{d}t$. 下面来分析系统在时刻$t$的瞬态队长分布.

定理 1 当$j = 0$, 如果$R(s) > 0$, 那么$p_{0,0}^*(s)$和$p_{i,0}^*(s)$由下面的式子确定:

$$p_{0,0}^*(s) = \frac{1-f(s)}{s}[1 + \frac{b(s)f(s)}{\varOmega(s)}], \tag{3}$$

$$p_{i,0}^*(s) = \frac{1-f(s)}{s}\frac{b^i(s)}{\varOmega(s)}, \tag{4}$$

其中:

$$\varOmega(s) = 1 - b^m(s)f^m(s)y^*(\varLambda) + b^m(s)f^m(s)\varPhi_2(s) - b^N(s)f^N(s)\varPhi_1(s),$$

$$\varPhi_1(s) = \sum_{n=0}^{N-m-1}\int_0^\infty \mathrm{e}^{-(s+\lambda)t}\frac{[\lambda t/f(s)]^n}{n!}\mathrm{d}Y(t),$$

$$\varPhi_2(s) = \sum_{n=0}^{N-m-1}\int_0^\infty \mathrm{e}^{-(s+\lambda)t}\frac{[\lambda b(s)t]^n}{n!}\mathrm{d}Y(t),$$

$$y^*(\varLambda) = \int_0^\infty \mathrm{e}^{-\varLambda t}\mathrm{d}Y(t),\ \varLambda = s + \lambda - \lambda b(s),$$

$$f(s) = \int_0^\infty \mathrm{e}^{-st}\mathrm{d}F(t) = \frac{\lambda}{s+\lambda},$$

$$b(s) = \int_0^\infty \mathrm{e}^{-st}\mathrm{d}B(t).$$

证 记$L_k = \sum\limits_{i=1}^{k}\tau_i (k \geqslant 1)$且$L_0 = 0$, 其中$\tau_i (i = 1, 2, \cdots)$顾客相继到达的间隔时间序列. 由系统假设和示意图可知服务台服务完成系统中所有的顾客后就立即进入闲期. 如果初始时刻$t = 0$时的队长为0, 那么时刻t队长为0的概率为

$$\begin{aligned} &p_{0,0}(t) = \\ &p\{t < \tau_1\} + p\{\tau_1 + b_1 \leqslant t < \tau_1 + b_1 + \hat{\tau}_2\} + \\ &p\{Y \geqslant L_{N-m}, \tau_1 + b_1 + L_m + Y \leqslant t, N(t) = 0\} + \\ &p\{Y \leqslant L_{N-m-1}, \tau_1 + b_1 + L_N \leqslant t, N(t) = 0\}. \end{aligned} \tag{5}$$

同理, 如果初始时刻$t = 0$时的队长为0, 那么时刻t队长为i的概率为

$$\begin{aligned} &p_{i,0}(t) = \\ &p\{b^{<i>} \leqslant t < b^{<i>} + \tau_1\} + \\ &p\{Y \geqslant L_{N-m}, b^{<i>} + L_m + Y \leqslant t, N(t) = 0\} + \\ &p\{Y \leqslant L_{N-m-1}, b^{<i>} + L_N \leqslant t, N(t) = 0\}, \end{aligned} \tag{6}$$

其中: 事件$Y \geqslant L_{N-m}$表示在启动期Y内到达的至少$N - m$顾客; 事件$Y < L_{N-m}$是说在启动期Y内

到达的顾客数至多$N-m-1$个. 根据Poisson到达过程的无后效性, 有

$$
\begin{aligned}
&p_{0,0}(t)=\\
&1-F(t)+\int_0^t F(t-x)\mathrm{d}\,[F(x)\times B(x)]+\\
&\sum_{n=0}^{N-m-1}\int_0^{\infty}\mathrm{e}^{-yt}\frac{(\lambda y)^n}{n!}p\,\{\tau_1+b_1+L_{N-n}+y\leqslant t,\\
&N(t)=0\}\,\mathrm{d}Y(t)+\sum_{n=N-m}^{\infty}\int_0^{\infty}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p\{\tau_1+\\
&b_1+L_m+y\leqslant t,N(t)=0\}\mathrm{d}Y(t),
\end{aligned}\tag{7}
$$

$$
\begin{aligned}
&p_{i,0}(t)=\\
&\int_0^t 1-F(t-x)\mathrm{d}B^i(x)+\\
&\sum_{n=0}^{N-m-1}\int_0^{\infty}\mathrm{e}^{-yt}\frac{(\lambda y)^n}{n!}p\{b^{<i>}+L_{N-n}+y\leqslant t,\\
&N(t)=0\}\mathrm{d}Y(t)+\sum_{n=N-m}^{\infty}\int_0^{\infty}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p\{b^{<i>}+\\
&L_m+y\leqslant t,N(t)=0\}\mathrm{d}Y(t).
\end{aligned}\tag{8}
$$

由于Poisson到达过程, 每个忙期开始时刻都是一个更新点, 所以有

$$
\begin{aligned}
p_{0,0}(t)=\;&1-F(t)+\int_0^t F(t-x)\mathrm{d}[F(x)\times B(x)]+\\
&\sum_{n=0}^{N-m-1}\int_0^t\int_0^{t-y}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p^*_{N,0}(t-x-y)\mathrm{d}[F^{(N-n+1)}(x)\times B(x)]\mathrm{d}Y(t)+\\
&\sum_{n=N-m}^{\infty}\int_0^t\int_0^{t-y}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p^*_{m+n,0}(t-x-y)\mathrm{d}[F^{(m+1)}(x)\times B(x)]\mathrm{d}Y(t),
\end{aligned}\tag{9}
$$

$$
\begin{aligned}
p_{i,0}(t)=\;&\int_0^t 1-F(t-x)\mathrm{d}B^i(x)+\\
&\sum_{n=0}^{N-m-1}\int_0^t\int_0^{t-y}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p^*_{N,0}(t-x-y)\mathrm{d}[F^{(N-n)}(x)\times B^{(i)}(x)]\mathrm{d}Y(t)+\\
&\sum_{n=N-m}^{\infty}\int_0^t\int_0^{t-y}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p^*_{m+n,0}(t-x-y)\mathrm{d}[F^{(m)}(x)\times B^{(i)}(x)]\mathrm{d}Y(t).
\end{aligned}\tag{10}
$$

对式(9)和式(10)取Laplace变换:

$$
\begin{aligned}
p^*_{0,0}(s)=\;&\frac{1-f(s)}{s}(1+f(s)b(s))+\\
&f^{N+1}(s)b(s)\sum_{n=0}^{N-m-1}p^*_{N,0}(s)\cdot\\
&\int_0^{\infty}\mathrm{e}^{-(s+\lambda)t}\frac{[\lambda t/f(s)]^n}{n!}\mathrm{d}Y(t)+\\
&f^{m+1}(s)b(s)\sum_{n=N-m}^{\infty}p^*_{m+n,0}(s)\cdot\\
&\int_0^{\infty}\mathrm{e}^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t),
\end{aligned}\tag{11}
$$

$$
\begin{aligned}
&p^*_{i,0}(s)=\\
&\frac{1-f(s)}{s}b^i(s)+f^N(s)b^i(s)\sum_{n=0}^{N-m-1}p^*_{N,0}(s)\cdot\\
&\int_0^{\infty}\mathrm{e}^{-(s+\lambda)t}\frac{[\lambda t/f(s)]^n}{n!}\mathrm{d}Y(t)+f^m(s)b^i(s)\cdot\\
&\sum_{n=N-m}^{\infty}p^*_{m+n,0}(s)\int_0^{\infty}\mathrm{e}^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t).
\end{aligned}\tag{12}
$$

联立式(11)和式(12), 得到$p^*_{0,0}(s)$和$p^*_{i,0}(s)$之间的关系:

$$
p^*_{i,0}(s)=\frac{b^{i-1}(s)}{f(s)}[p^*_{0,0}(s)-\frac{1-f(s)}{s}].\tag{13}
$$

下面把式(13)式代入式(11)和式(12), 经过不太复杂的计算就可得到式(3)和式(4).　　证毕.

定理 2　对$j=1,\cdots,m-1$且$R(s)>0$, 那么$p^*_{0,j}(s)$和$p^*_{i,j}(s)$有下面的表达式:

$$
\begin{aligned}
&p^*_{0,j}(s)=f(s)q^*_j(s)+\\
&\qquad\frac{[1-f(s)]f^{j+1}(s)b(s)+s[\delta_1(s)+\delta_2(s)]}{s\varOmega(s)},
\end{aligned}\tag{14}
$$

$$
\begin{aligned}
&p^*_{i,j}(s)=\\
&\sum_{k=1}^{i}q^*_{j-i+k}(s)b^{k+1}(s)+\\
&\frac{[1-f(s)]f^j(s)b^i(s)+sb^{i-1}(s)\dfrac{\delta_1(s)+\delta_2(s)}{f(s)}}{s\varOmega(s)},
\end{aligned}\tag{15}
$$

其中:

$$
\begin{aligned}
&\delta_1(s)=\\
&\sum_{n=N-m}^{\infty}f^{m+1}(s)b(s)\sum_{k=1}^{m+n}q^*_{j-(m+n)+k}(s)\cdot\\
&b^{k-1}(s)\int_0^{\infty}\mathrm{e}^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t),\\
&\delta_2(s)=\\
&\sum_{n=0}^{N-m-1}f^{N+1}(s)b(s)\sum_{k=1}^{N}q^*_{j-N+k}(s)b^{k-1}(s)\cdot\\
&\int_0^{\infty}\mathrm{e}^{-(s+\lambda)t}\frac{[\lambda t/f(s)]^n}{n!}\mathrm{d}Y(t).
\end{aligned}
$$

证　当$j=1,\cdots,m-1$时, 时刻t系统队长等于j, 那么时刻t必须处于建立期或者忙期. 采用全

概率公式$p_{0,j}(t)$和$p_{i,j}(t)$可以分解如下:

$$p_{0,j}(t)=$$
$$p\{\tau_1\leqslant t<\tau_1+b_1,N(t)=j\}+$$
$$p\{\tau_1+b_1+L_j\leqslant t<\tau_1+b_1+L_{j+1}\}+$$
$$p\{Y\geqslant L_{N-m},\tau_1+b_1+L_m+Y\leqslant t,N(t)=j\}+$$
$$p\{Y\leqslant L_{N-m-1},\tau_1+b_1+L_N\leqslant t,N(t)=j\},\tag{16}$$
$$p_{i,0}(t)=$$
$$p\left\{0\leqslant t<b^{<i>},N(t)=j\right\}+$$
$$p\left\{Y\geqslant L_{N-m},b^{<i>}+L_m+Y\leqslant t,N(t)=j\right\}+$$
$$p\left\{Y\leqslant L_{N-m-1},b^{<i>}+L_N\leqslant t,N(t)=j\right\},\tag{17}$$

其中: $b^{<i>}$定义为有i个顾客开始的忙期. 由于Poisson到达过程, 忙期$b^{<i>}$等于$b_1+b_2+\cdots+b_i$, 每个b_i独立且同分布. 忙期$b^{<i>}$中时刻t系统中的顾客数等于j的概率等于$\sum\limits_{k=1}^{i}Q_{j-i+k}(t)\times B^*(t)$. 因此, 有

$$\begin{aligned}p_{0,j}(t)=&\int_0^t Q_j(t-x)\mathrm{d}F(t)+F^{(j+1)}(t)\times B(t)-F^{(j+2)}(t)\times B(t)+\\&\sum_{n=N-m}^{\infty}\int_0^t\int_0^{t-y}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p^*_{m+n,j}(t-x-y)\mathrm{d}\left[F^{(m+1)}(x)\times B(x)\right]\mathrm{d}Y(t)+\\&\sum_{n=0}^{N-m-1}\int_0^t\int_0^{t-y}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p^*_{N,0}(t-x-y)\mathrm{d}\left[F^{(N-n+1)}(x)\times B(x)\right]\mathrm{d}Y(t),\end{aligned}\tag{18}$$

$$\begin{aligned}p^*_{i,j}(s)=&\sum_{k=1}^{i}Q_{j-i+k}(t)\times B^*(t)+\sum_{n=N-m}^{\infty}\int_0^t\int_0^{t-y}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p^*_{m+n,j}(t-x-y)\mathrm{d}[F^{(m)}(x)\times B^{(i)}(x)]\mathrm{d}Y(t)+\\&\sum_{n=0}^{N-m-1}\int_0^t\int_0^{t-y}\mathrm{e}^{-\lambda y}\frac{(\lambda y)^n}{n!}p^*_{N,0}(t-x-y)\mathrm{d}[F^{(N-n)}(x)\times B^{(i)}(x)]\mathrm{d}Y(t).\end{aligned}\tag{19}$$

同理, 也对上述两式取Laplace变换, 得

$$\begin{aligned}&p^*_{0,j}(s)=\\&f(s)q^*_j(s)+f^{j+1}(s)b(s)\frac{1-f(s)}{s}+f^{m+1}(s)b(s)\cdot\\&\sum_{n=N-m}^{\infty}p^*_{m+n,j}(s)\int_0^\infty\mathrm{e}^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t)+\\&f^{N+1}(s)b(s)\sum_{n=0}^{N-m-1}p^*_{N,j}(s)\cdot\\&\int_0^\infty\mathrm{e}^{-(s+\lambda)t}\frac{[\lambda t/f(s)]^n}{n!}\mathrm{d}Y(t),\end{aligned}\tag{20}$$

$$\begin{aligned}&p^*_{i,0}(s)=\\&\sum_{k=1}^{i}q^*_{j-i+k}(s)b^{k-1}(s)+b^i(s)f^{j+1}(s)b(s)\frac{1-f(s)}{s}+\\&f^m(s)b^i(s)\sum_{n=N-m}^{\infty}p^*_{m+n,j}(s)\cdot\\&\int_0^\infty\mathrm{e}^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t)+f^N(s)b^i(s)\cdot\\&\sum_{n=0}^{N-m-1}p^*_{N,j}(s)\int_0^\infty\mathrm{e}^{-(s+\lambda)t}\frac{[\lambda t/f(s)]^n}{n!}\mathrm{d}Y(t).\end{aligned}\tag{21}$$

同定理1, 首先得到$p^*_{0,j}(s)$和$p^*_{i,j}(s)$的关系式, 把它代入式(20)和式(21)化简即可. 类似于定理1和2, 下面的两个定理容易证明.

定理 3 对$j=m,\cdots,N-1$且$R(s)>0$, 那么$p^*_{0,j}(s)$和$p^*_{i,j}(s)$有下面的表达式:

$$p^*_{0,j}(s)=f(s)q^*_j(s)+\frac{f^m(s)b(s)\varGamma(s)+\sum\limits_{k=1}^{j}q^*_j(s)}{\varOmega(s)},\tag{22}$$

$$p^*_{i,j}(s)=\sum_{k=1}^{i}q^*_{j-i+k}(s)b^{k+1}(s)+\frac{b^{i-1}(s)[f^m(s)b(s)\varGamma(s)+\sum\limits_{k=1}^{j}q^*_j(s)]}{\varOmega(s)f(s)},\tag{23}$$

其中$\varGamma(s)=\int_0^\infty\mathrm{e}^{-(s+\lambda)t}\frac{(\lambda t)^{j-m}}{(j-m)!}\mathrm{d}F(t)$.

定理 4 对$j=N,N+1,\cdots$且$R(s)>0$, 那么$p^*_{0,j}(s)$和$p^*_{i,j}(s)$有下面的表达式:

$$p^*_{0,j}(s)=f(s)q^*_j(s)+\frac{f^m(s)b(s)\varPsi(s)+\delta_3(s)+\delta_4(s)}{\varOmega(s)},\tag{24}$$

$$p^*_{i,j}(s)=\sum_{k=1}^{i}q^*_{j-i+k}(s)b^{k+1}(s)+\frac{b^{i-1}(s)\left[f^m(s)b(s)\varPsi(s)+\delta_3(s)+\delta_4(s)\right]}{\varOmega(s)f(s)},\tag{25}$$

其中:

$$\delta_3(s)=$$
$$\sum_{n=N-m}^{\infty}f^{m+1}(s)b(s)\sum_{k=1}^{m+n}q^*_{j-(m+n)+k}(s)b^{k-1}(s)\cdot$$

$$\int_0^\infty \mathrm{e}^{-(s+\lambda)t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t),$$

$$\delta_4(s)=\sum_{n=0}^{N-m-1} f^{N+1}(s)b(s)\sum_{k=j-N+1}^{j} q_k^*(s)b^{k-1}(s)\cdot \int_0^\infty \mathrm{e}^{-(s+\lambda)t}\frac{[\lambda t/f(s)]^n}{n!}\mathrm{d}Y(t),$$

$$\Psi(s)=\int_0^\infty \mathrm{e}^{-(s+\lambda)t}\frac{(\lambda t)^{j-m}}{(j-m)!}\bar{Y}(t)\mathrm{d}t.$$

定理1–4给出系统瞬态队长分布的递推关系式. 当$\rho<1$, 系统能达到长期均衡, 使用L'hospital法则和式(1), 系统队长稳态的分布有下面的定理给出.

定理 5　稳态队长分布为$p_j=\lim\limits_{s\to 0^+} sp_{i,j}^*(s)$. 当$\rho<1$时, 平稳状态下的稳态队长的概率分布为

$$p_0=\frac{1-\rho}{m+\lambda\mathrm{E}[Y]+\sum\limits_{n=0}^{N-m-1}(N-m-n)\int_0^\infty \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t)},\tag{26}$$

$$p_j=p_0[1+\lambda\sum_{k=1}^{j}\theta_k],\ j=1,2,\cdots,m-1,\tag{27}$$

$$p_j=p_0\lambda[\Gamma(0)+\sum_{k=1}^{j}\theta_k],\ j=m,m+1,\cdots,N-1,\tag{28}$$

$$p_j=p_0\lambda[\Psi(0)+\delta_3(0)+\delta_4(0)],\ j=N,N+1,\cdots,\tag{29}$$

其中:

$$\theta_j=\frac{1}{g(\lambda)}\{\int_0^\infty \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-1}}{(j-1)!}\bar{G}(t)\mathrm{d}t+\sum_{k=1}^{j-1}\theta_{j-k}[1-\sum_{i=0}^{k}\int_0^\infty \mathrm{e}^{-\lambda t}\frac{(\lambda t)^i}{i!}\mathrm{dG}(t)]\},\ j\geqslant 1,$$

$$\delta_3(0)=\sum_{n=N-m}^{\infty}\sum_{k=1}^{m+n}\theta_{j-(m+n)+k}\cdot\int_0^\infty \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t),$$

$$\delta_4(0)=\sum_{n=0}^{N-m-1}\sum_{k=1}^{N}\theta_{j-N+k}\int_0^\infty \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathrm{d}Y(t),$$

$$\Gamma(0)=\int_0^\infty \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-m}}{(j-m)!}\mathrm{d}F(t),$$

$$\Psi(0)=\int_0^\infty \mathrm{e}^{-\lambda t}\frac{(\lambda t)^{j-m}}{(j-m)!}\bar{Y}(t)\mathrm{d}t.$$

注 1　假设$m=N$且$\mathrm{P}\{Y=0\}=1$, 本文研究的模型就是常见的$N-$策略$M/G/1$系统. 注意到: $\delta_3(0)=\sum\limits_{k=1}^{N}\theta_{j-N+k}$, $\Psi(0)=0$且$\delta_4(0)=0$, 定理5就同文献[9]中对应的结论一致.

4 系统稳态性能指标(Steady-state performance indexes for the system)

下面考虑非奇异的双阈值控制策略, 即排除$m=N$这种特殊的退化情况. 采用文献[2]中的结论, 容易给出系统平均队长、平均闲期、平均忙期以及平均启动期.

从稳态队长分布的概率母函数出发容易得出系统的平均队长(E[L])为

$$\mathrm{E}[L]=\frac{\lambda^2\mathrm{E}[G^2]}{2(1-\rho)}+\rho+\frac{m(m-1)}{2(m+\lambda\mathrm{E}[Y]+\sum\limits_{n=0}^{N-m-1}\beta_n)}+\frac{\sum\limits_{n=0}^{N-m-1}(m+n)\beta_n+\lambda m\mathrm{E}[Y]+\frac{\lambda^2\mathrm{E}[Y^2]}{2}}{m+\lambda\mathrm{E}[Y]+\sum\limits_{n=0}^{N-m-1}\beta_n},\tag{30}$$

其中β_n等于等待过程中系统状态达到$m+n$的概率. 下面先定义一个概率: $\gamma_i=$在启动期Y内到达i个顾客的概率. 由于启动期开始时刻系统中就有m个顾客, 只有启动期内到达的顾客数n满足条件: $n\leqslant N-m-1$时, 等待过程系统状态才能达到水平$m+n$. 因此有$\beta_n=\sum\limits_{i=0}^{n}\gamma_i$. 接着, 定义指标随机变量

$$\delta_{m,n}=\begin{cases}1, & \text{等待期中系统状态达到}m+n,\\ 0, & \text{其他}.\end{cases}\tag{31}$$

因此, 有

$$\beta_n=\mathrm{P}(\delta_{m,n}=1)=\mathrm{E}(\delta_{m,n}).\tag{32}$$

采用更新理论, 易得系统的平均闲期长度($T_{m,N}$)、平均忙期长度($I_{m,N}$)以及系统的平均忙期循环$C_{m,N}$为

$$\mathrm{E}[I_{m,N}]=\frac{m}{\lambda}+\mathrm{E}[Y]+\frac{1}{\lambda}\sum_{n=0}^{N-m-1}\beta_n,\tag{33}$$

$$\mathrm{E}[B_{m,N}]=\frac{\mathrm{E}[G](m+\lambda\mathrm{E}[Y]+\sum\limits_{n=0}^{N-m-1}\beta_n)}{1-\rho},\tag{34}$$

$$\mathrm{E}[C_{m,N}]=\frac{m+\lambda\mathrm{E}[Y]+\sum\limits_{n=0}^{N-m-1}\beta_n}{\lambda(1-\rho)}.\tag{35}$$

5 系统的最优设计(Optimal design for the system)

本节研究系统的最优控制策略(m^*,N^*). 先构建系统的净收益函数, 接着设计有效的搜索算法,

最后, 通过数值例子来考察不同情况下的最优控制策略.

5.1 净收益函数(Net profit function)

从系统单位时间平均净收益最大化角度考察系统的最优控制策略, 并且比较其与单阈值N–策略的优劣性. 定义如下的成本–收益因子:

Cr: 单位服务时间的收益;

Cs: 每个忙期循环的系统启动成本;

Co: 单位时间系统的运行成本;

Cst: 系统启动后服务前单位时间的等待成本;

Ch: 单位时间系统容纳单个顾客的存储成本.

根据定义的成本–收益因子, 构建系统单位时间平均净收益函数为

$$\begin{aligned}&\mathrm{Profit}(m,N)=\\&\frac{\mathrm{Cr}\cdot\mathrm{E}[B_{m,N}]}{\mathrm{E}[C_{m,N}]}-\{\frac{\mathrm{Cs}}{\mathrm{E}[C_{m,N}]}+\mathrm{Co}\frac{\mathrm{E}[B_{m,N}]}{\mathrm{E}[C_{m,N}]}+\\&\frac{\mathrm{Cst}(1-\rho)\sum\limits_{n=0}^{N-m-1}\beta_n}{\mathrm{E}[C_{m,N}]}+\mathrm{Ch}\cdot\mathrm{E}[L]\}=\\&\rho\mathrm{Cr}-\frac{\lambda(1-\rho)\mathrm{Cs}}{m+\lambda\mathrm{E}[Y]+\sum\limits_{n=0}^{N-m-1}\beta_n}-\rho\mathrm{Co}-\\&\frac{(1-\rho)\sum\limits_{n=0}^{N-m-1}\beta_n\mathrm{Cst}}{m+\lambda\mathrm{E}[Y]+\sum\limits_{n=0}^{N-m-1}\beta_n}-\mathrm{Ch}\{\frac{\lambda^2\mathrm{E}[G^2]}{2(1-\rho)}+\rho+\\&\frac{\frac{m^2-m}{2}\sum\limits_{n=0}^{N-m-1}(m+n)\beta_n+\lambda m\mathrm{E}[Y]+\frac{\lambda^2\mathrm{E}[Y^2]}{2}}{m+\lambda\mathrm{E}[Y]+\sum\limits_{n=0}^{N-m-1}\beta_n}\}.\end{aligned}\tag{36}$$

5.2 算法设计(Algorithm design)

上述净收益函数的最优解的存在性以及唯一性是否存在? 对这个具有两个离散变量的净收益函数来说, 从分析的角度去证明上述问题, 数学上推导较难. 下面来导出这个函数具有的某些性质, 依此来设计可行的算法. 便于后续推导, 记

$$\varphi_{k,m}=\sum_{n=0}^{k-m}\beta_n,\ \delta_{k,m}=\sum_{n=0}^{k-m}n\beta_n,$$

那么

$$\begin{aligned}&\mathrm{Profit}(m,k+1)-\mathrm{Profit}(m,k)=\\&\frac{\lambda(1-\rho)\mathrm{Cs}}{(m+\lambda\mathrm{E}[Y]+\varphi_{k,m})(m+\lambda\mathrm{E}[Y]+\varphi_{k-1,m})}\beta_{k-m}-\\&\frac{(1-\rho)\mathrm{Cst}(m+\lambda\mathrm{E}[Y])}{(m+\lambda\mathrm{E}[Y]+\varphi_{k,m})(m+\lambda\mathrm{E}[Y]+\varphi_{k-1,m})}\beta_{k-m}+\\&\frac{\mathrm{Ch}\frac{m^2-m}{2}}{(m+\lambda\mathrm{E}[Y]+\varphi_{k,m})(m+\lambda\mathrm{E}[Y]+\varphi_{k-1,m})}\beta_{k-m}+\\&\frac{\mathrm{Ch}(\lambda m\mathrm{E}[Y]+\frac{\lambda^2\mathrm{E}[Y^2]}{2})}{(m+\lambda\mathrm{E}[Y]+\varphi_{k,m})(m+\lambda\mathrm{E}[Y]+\varphi_{k-1,m})}\beta_{k-m}+\\&\frac{\mathrm{Ch}\,[-k(m+\lambda\mathrm{E}[Y])+\delta_{k-1,m}-(k-m)\varphi_{k-1,m}]}{(m+\lambda\mathrm{E}[Y]+\varphi_{k,m})(m+\lambda\mathrm{E}[Y]+\varphi_{k-1,m})}\cdot\\&\beta_{k-m}=\\&\frac{\beta_{k-m}M(m,k)}{(m+\lambda\mathrm{E}[Y]+\varphi_{k,m})(m+\lambda\mathrm{E}[Y]+\varphi_{k-1,m})},\end{aligned}\tag{37}$$

其中

$$\begin{aligned}&M(m,k)=\\&\lambda(1-\rho)\mathrm{Cs}+\mathrm{Ch}[\frac{\lambda^2\mathrm{E}[Y^2]}{2}+\frac{m^2-m}{2}+\\&\lambda m\mathrm{E}[Y]-k(m+\lambda\mathrm{E}[Y])+\delta_{k-1,m}-\\&(k-m)\varphi_{k-1,m}]-(1-\rho)(m+\lambda\mathrm{E}[Y])\mathrm{Cst}.\end{aligned}$$

由于

$$\frac{\beta_{k-m}}{(m+\lambda\mathrm{E}[Y]+\varphi_{k,m})(m+\lambda\mathrm{E}[Y]+\varphi_{k-1,m})}>0,$$

因此净收益函数Profit(m,k)的单调性由$M(m,k)$的正负号确定. 对于一个给定的m, 有

$$\begin{aligned}&M(m,k+1)-M(m,k)=\\&-\mathrm{Ch}(m+\lambda\mathrm{E}[Y]+\varphi_{k,m})<0.\end{aligned}\tag{38}$$

函数$M(m,k)$是关于k的减函数. 因此, 如果$n=\min\{k|M(m,k)<0\}$, 那么净收益函数具有下面的性质:

$$\mathrm{Profit}(m,k)<\mathrm{Profit}(m,n),\ k>n.\tag{39}$$

净收益函数的上述性质表示: 对于一个给定的m, 最优的阈值$N^*(m)$是第一个使得$M(m,k)<0$的k. 所以对于每一个m, 都可以得到最优的阈值以及相应的最大净收益Profit$(m,N^*(m))$. 依据上述性质, 设计如下算法来寻找全局最优的控制阈值, 具体步骤为:

Step 1 置$m=1$, 由

$$N^*(m)=\min\{k|M(m,k)<0\}$$

得到阈值$N^*(m)$以及相应的净收Profit$(m,N^*(m))$;

Step 2 继续计算$N^*(m+1)$以及对应的Profit$(m+1,N^*(m+1))$;

Step 3 如果Profit$(m+1,N^*(m+1))<$ Profit$(m,N^*(m))$, 则停止搜索, 全局最优的控制阈值为$(m^*,N^*(m^*))$; 否则, 重复Step 2.

5.3 数值试验(Numerical experimentation)

接下来通过数值实验来考察不同经济系统参

数下的最优控制策略. 假设到达过程的参数$\lambda = 2$; 服务时间和启动时间都服从负指数分布, 且平均服务时间和平均启动时间分别为$\mathrm{E}[G] = 0.25$和$\mathrm{E}[Y] = 1$. 由控制阈值的搜索方式可知单位时间收益因子不影响系统的最优控制阈值.

表1给出不同的启动成本下最优控制阈值并且对比了单个阈值的最优控制策略. 结果显示双重控制策略优于单个阈值控制策略, 系统净收益相对于单个阈值策略的情况提高了.

表2给出不同运行成本下的最优控制阈值并且对比了单个阈值的最优控制策略. 结果显示双重控制策略优于单个阈值控制策略; 但是系统的最优控制策略在两种情况下都不随运行成本的变化而变化.

表3给出不同等待成本下的最优控制阈值并且对比了单个阈值的最优控制策略. 结果显示双重控制策略等价于单个阈值控制策略, 最优控制策略固定为(3, 3).

表4给出不同存储成本下的最优控制阈值并且对比了单个阈值的最优控制策略. 结果显示双重控制策略优于单个阈值控制策略.

表 1 最优控制阈值vs系统启动成本

Table 1 Optimal control thresholds vs system's setup cost

Cs	$(m^*, M^*(m))$	Profit$(m, M^*(m))$	N^*	Profit(N^*)
10	(1, 2)	83.38	1	83.04
30	(2, 3)	77.91	2	77.62
50	(3, 4)	73.56	3	73.37
100	(5, 6)	65.01	5	64.94
150	(6, 7)	58.37	6	58.25
200	(7, 8)	52.62	7	52.49

注 2 本表对应的排队系统参数为$\lambda = 2$, $\mathrm{E}[G] = 0.25$, $\mathrm{E}[Y] = 1$; 经济系统参数为Cr = 200, Co = 1, Cst = 10, Ch = 5.

表 2 最优控制阈值vs系统运行成本

Table 2 Optimal control thresholds vs system's operation cost

Co	$(m^*, M^*(m))$	Profit$(m, M^*(m))$	N^*	Profit(N^*)
0	(3, 4)	74.06	3	73.83
5	(3, 4)	71.56	3	71.37
10	(3, 4)	69.06	3	68.87
15	(3, 4)	66.56	3	66.37
20	(3, 4)	64.06	3	63.87
100	(3, 4)	24.06	3	23.88

注 3 本表对应的排队系统参数为$\lambda = 2$, $\mathrm{E}[G] = 0.25$, $\mathrm{E}[Y] = 1$; 经济系统参数为Cr = 200, Cs = 50, Cst = 10, Ch = 5.

表 3 最优控制阈值vs系统等待成本

Table 3 Optimal control thresholds vs system's waiting cost

Cst	$(m^*, M^*(m))$	Profit$(m, M^*(m))$	N^*	Profit(N^*)
20	(3, 3)	73.38	3	73.38
40	(3, 3)	73.38	3	73.38
60	(3, 3)	73.38	3	73.38
80	(3, 3)	73.38	3	73.38
100	(3, 3)	73.38	3	73.38
120	(3, 3)	73.38	3	73.38

注 4 本表对应的排队系统参数为$\lambda = 2$, $\mathrm{E}[G] = 0.25$, $\mathrm{E}[Y] = 1$; 经济系统参数为Cr = 200, Co = 1, Cs = 50, Ch = 5.

表 4 最优控制阈值vs系统存储成本

Table 4 Optimal control thresholds vs system's storage cost

Ch	$(m^*, M^*(m))$	Profit$(m, M^*(m))$	N^*	Profit(N^*)
10	(2, 3)	59.02	2	58.25
20	(1, 3)	33	2	29.5
30	(1, 3)	6.75	1	4.08
40	(1, 3)	−19.5	1	−2.17
50	(1, 3)	−45.75	1	−48.4
100	(1, 3)	−177	1	−179.7

注 5 本表对应的排队系统参数为$\lambda = 2$, $\mathrm{E}[G] = 0.25$, $\mathrm{E}[Y] = 1$; 经济系统参数为Cr = 200, Co = 1, Cs = 50, Cst = 10.

6 结束语(Conclusion)

本文针对生产管理中出现的排队服务问题, 建立双阈值控制策略的排队模型. 通过考察系统队长概率分布的研究, 为系统的容量设计提供依据. 考虑系统营运效率, 建立了系统利润模型并设计有效的全局最优算法, 通过数值方法揭示本文设计的双阈值模型的性能优于单个阈值控制的系统. 这一策略的优越性本质上在于通过设置系统启动预热阈值, 节约了顾客的等待时间, 从而提高了系统效益, 对于生产管理有着重要的意义. 进一步的研究可以考虑稳定队长分布对于有限容量的系统设计以及顾客批到达的情况.

参考文献(References):

[1] TADJ L, CHOUDHURY G. Optimal design and control of queues[J]. *Top*, 2005, 13(2): 359 – 412.

[2] LEE H W, PARK J O. Optimal strategy in N-policy production system with early set-up[J]. *Journal of the Operational Research Society*, 1997, 48(3): 306 – 313.

[3] LEE H W, PARK J G, KIM B K, et al. Queue length and waiting time analysis of a batch arrival queue with bilevel control[J]. *Computers and Operations Research*, 1998, 25(3): 191 – 205.

[4] LEE H W, PARK N I, JEON J. Queue length analysis of batch arrival queues under bilevel threshold control with early set-up[J]. *International Journal of Systems Science*, 2003, 34(3): 195 – 204.

[5] KE J C. Bi-level control for batch arrival queues with an early startup and un-reliable server[J]. *Applied Mathematical Modelling*, 2004, 28(5): 469 – 485.

[6] KE J C. An M/G/1 queue under hysteretic vacation policy with an early startup and un-reliable server[J]. *Mathematical Methods of Operations Research*, 2006, 63(2): 357 – 369.

[7] HERNÁNDEZ-DÍAZ A G, MORENO P. A discrete-time single-server queueing system with an N-policy, an early setup and a generalization of the Bernoulli feedback[J]. *Mathematical and Computer Modelling*, 2009, 49(5): 977 – 990.

[8] 周亚平, 奚宏生, 殷保群, 等. 一类受控闭排队网络基于性能势的最优性方程[J]. 控制理论与应用, 2002, 19(4): 521 – 526.
(ZHOU Yaping, XI Hongsheng, YIN Baoqun, et al. Optimality equations based performance potentials for a class of controlled closed queueing networks[J]. *Control Theory & Applications*, 2002, 19(4): 521 – 526.)

[9] 唐应辉. 延迟N–策略M/G/1排队系统队长的瞬态和稳态分布[J]. 系统工程理论与实践, 2007, 27(11): 130 – 134.
(TANG Yinghui. The transient and equilibrium distributions of the queue-length for M /G/1 queue with delayed N-policy[J]. *Systems Engineering—Theory & Practice*, 2007, 27(11): 130 – 134.)

作者简介:

刘名武 (1979—), 男, 博士, 讲师, 目前研究方向为系统优化与决策, E-mail: liumingwu2007@yahoo.cn;

杨迎春 (1980—), 男, 讲师, 博士研究生, 目前研究方向为国际贸易理论与政策;

马永开 (1963—), 男, 教授, 博士, 博士生导师, 目前研究方向为资本市场和公司金融、管理运筹学.