DOI: 10.3901/JME.2019.14.197

多优先级与离开机制并存的 M/M/c/N 排队系统

印明昂 田刚 孙志礼 孙尧
(东北大学机械工程与自动化学院 沈阳 110004)

摘要: 生产中重要设备发生故障后需要及时维修, 普通部分设备会因等待时间过长而离开维修系统, 因此需要建立设备多优先级, 并存且包含离开机制的排队系统模型。在马尔科夫过程理论基础上结合分块理论, 得出分块的系统状态转移率矩阵。针对状态转移率矩阵为块三对角矩阵的特点, 利用分块矩阵理论得到稳态下的状态概率以及状态转移频度。在此基础上获得关键状态转移频度、待修时间等系统运行指标, 并据此提出系统盈利计算方法。最后以某汽车维修厂为实例, 验证该模型在实际生产中的有效性。所提的排队系统建模方法通过引入稳态状态转移频度, 建立了完备的系统指标体系, 为系统配置优化提供了理论基础。

关键字: 马尔科夫过程; 排队理论; 分块矩阵; 状态转移率矩阵; 状态转移频度

中文分类号: TB114

M/M/c/N Queue with Multiple-priority and Abandonment

YIN Mingang TIAN Gang SUN Zhili SUN Yao
(College of Mechanical and Automation, Northeastern University, Shenyang 110004)

Abstract: In practice production, when failures occurs to the important equipment, it needs to be repaired in time, and some equipment will leave the system because of impatience. Therefore, model with multi-priority and abandonment should be established in the queuing system. The system state revolution matrix is deduced by the Markov process method and block matrix method. Because For the state transfer rate matrix is block tridiagonal matrix, the steady-state probability and steady-state transfer frequency of each state are obtained by block matrix analysis. The system indexes such as the transfer frequency between key states and the total waiting time are calculated, and the relations between profit and the parameters in the model are deduced. Finally, an example is given to further explain effectiveness of the model. The queuing system modeling method establishes a complete system index by introducing the steady state transfer frequency, which provides a theoretical basis for optimization of system configuration.

Key words: Markov process; queuing theory; partitioned matrix; state transfer rate matrix; state transfer frequency

0 前言

随着工业现代化、集成化、智能化的不断推进, 设备维修成本与人力成本在总成本中所占比重逐渐上升, 因此对该部分成本的控制成为了设备维修管理领域研究的新热点。在众多理论中排队理论和马尔科夫过程为系统的成本分析提供了坚实的基础, 受到了广泛关注。而实际生产中, 较为传统的排队模型因其缺乏灵活性难以广泛应用在复杂的实际生产过程中。因此, 考虑优先级和离开机制的排队模型研究更具有现实意义。

截至目前关于优先级与离开机制的研究成果

如下。

LIU 等^[1]首先提出优先级这一概念, 并将其应用在处理器的多任务调度中。SAIFULLAH 等^[2]在此基础上将优先级细分为两类, 一种为强占性优先级(Preemptive priorities)^[3-4], 指一旦具有优先级的设备进入系统, 维修人员会停下手中其他一切工作立刻开始对其进行维修; 另一种为非强占性优先级^[5], 指具有优先级的设备进入系统, 会排自动排到队列最前边, 一旦有维修工空闲, 便立刻对其进行维修。

HALFIN 等^[6]首先提出离开机制的概念, 并将其应用在交通运输系统的研究中。WANG 等^[7]将离开机制分成止步、等待离开、重新排队三种情况。关于止步机制, HAIGHT 等^[8]和 GARNET 等^[9]分别根据队长和系统负载(系统负载指当前时刻维修工

* 国家自然科学基金资助项目(51775097)。20181102 收到初稿, 20190401 收到修改稿

将系统中所有故障设备均修理完毕的时间)对故障设备是否进入系统进行判断。假设存在一个阈值,当设备故障时只有少于 N 个设备待修或者系统负载小于 b 时,该设备会进入系统,否则离开。关于等待离开机制,由于设备进入系统之后,便会产生一个等待时间上限 T ,当待修时间超过阈值 T 后,会因缺乏耐心而离开系统。对此,WEI 等^[10]将 T 定义为一个常值;WANG 等^[11]考虑了个体差异,令 T 服从指数分布;WARD 等^[12]将 T 的分布形式进一步扩展到任意分布。KE 等^[13]提出了重新排队机制的概念,即离开系统后会再次进入系统重新排队。

截至目前,对大多数排队系统都是将优先级和离开机制分割开来进行研究。然而随着对排队理论研究的不断深入,二者相结合的研究逐渐增多。在多优先级与离开机制并存的排队模型研究中,IRAVANI 等^[14]探讨了单一维修工且其中一种设备具有强占性优先级的情况;JOUNI 等^[15]探讨了多个维修工且设备的优先级为非强占性的情况。

在模型求解方面,韩超等^[16]指出基本上所有分析多优先级排队系统的研究都利用马尔科夫过程建模,并通过一些改进的算法来求解系统状态概率。例如:HEIJDEN 等^[17]通过聚合设备种类和减少维修工方式降低系统状态维度,从而对系统状态概率进行估计;WANG^[18]通过将二维马尔科夫过程压缩为一维的方式对系统状态概率进行了精确计算。多优先级和离开机制并存的模型也继承了多优先级模型的求解思路。然而,以上方法均是在利用马尔科夫过程求解系统状态概率。系统状态概率可理解为一段时间内系统在各状态上停留时间的相对比例,其容易忽略一些停留时间较短的状态,而造成模型指标体系构建的不完备,使得模型建立脱离实际。例如等待离开系统的次数便无法利用系统状态概率来计算。

综上所述,目前的研究中,多优先级模型、离开机制模型等单个模型的研究结构相对简单,面对实际生产过程中产生的复杂多变的问题上,缺乏灵活性与准确性。虽然有二者相结合的模型研究,但在指标体系构建方面较为初级,脱离实际。因此,本文将多优先级和离开机制相结合建立一个复合模型,根据系统状态转移频度建立系统指标体系,以此提高分析的灵活度与精确度;并在此基础上,以排队系统收益为目标函数,优化排队系统的组成方式,进而为维修方的决策提供更加完善的理论依据。

1 模型描述

为了更为贴近实际生产过程,本文以设备维修工厂为例建立模型,将待修设备分为普通设备和特殊设备。特殊设备享有强占性优先级,并且由多个维修工服务。为体现协同维修对于维修速度的提升,假设在理想情况下维修同一台设备的维修工数量越多,维修速度越快。以维修工数量 c 、特殊设备容量 G ,特殊设备所需的维修工数量 k 、普通设备容量 N 为变量,结合马尔科夫过程计算状态转移频度、待修设备等待时间、系统运行费用等系统指标。

该模型可以表述为 $M/M/c/N$ 排队系统,普通设备的维修方式为先到先服务,令 $n(t)$ 代表 t 时刻普通设备数目,令 $g(t)$ 代表 t 时刻特殊设备数目, $v(t)$ 表示 t 时刻因等待即将离开系统的普通设备数,那么系统的状态便可以表示为 $(n(t), g(t), v(t))$, 后简记为 (n, g, v) 。具体内容如下所述。

(1) 该系统具有普通、特殊两类设备,普通、特殊设备在单位时间内到达次数服从参数为 λ_1 、 λ_2 的泊松分布,即两种设备的到达时间均服从 λ_1 、 λ_2 的指数分布。

(2) 系统中共有 c 个维修工,1 个普通设备由 1 名维修工维修,1 个特殊设备由 k 名维修工维修。特殊设备到达系统时, k 个维修工立即对其进行维修,当维修工数量不足时,会抢占普通设备的维修工优先对其进行维修。普通、特殊设备维修时间分别服从 μ_1 、 μ_2 的指数分布。为保障特殊设备的优先性,系统中最多存在 $\lfloor c/k \rfloor$ 个特殊设备。

(3) 等待离开表示:普通设备待修时间超过等待时间上限时,普通设备便会离开维修厂,等待时间上限服从参数为 ρ 的指数分布。 $v=0$ 表示没有设备会因为等待而准备离开系统,但不排除有设备正在等待; $v=1$ 表示有一个普通设备因等待而准备离开系统,当此普通设备离开系统后,系统会立即进入 $v=0$ 的对应状态。普通设备等待离开时间服从参数为 μ_3 的指数分布,其中 $\mu_3 \rightarrow \infty$,表示两个状态转移是无延时的。因此,也不会存在多个普通设备同时离开系统的现象,所以 v 最大为 1。

(4) 止步表示:当普通设备数到达 N 时,修理厂将不再接受普通设备修理任务,直到普通设备个数小于 N 。当普通设备数量到达 N 后,普通设备仍可到达维修厂,但是到达之后立即离开,即离开时间服从参数为 μ_4 的指数分布,其中 $\mu_4 \rightarrow \infty$ 。

(5) 普通、特殊设备的到达间隔、维修时间相互独立,普通设备服务规则为 FCFS(first come first

served)。

如此描述系统状态的指标构成了一个三维马尔科夫过程 (n, g, v) , 其状态空间为

$$E = \{[n, g, v] | n=0, 1, 2, \dots, N+1; g=0, 1, \dots, G; v=0, 1\} \quad (1)$$

马尔科夫过程的状态转移机制如图 1 所示。结合分块化思想, 将系统按照普通设备个数 n 进行划分。当 $n=0$ 时, 把 $(0,0,0)$ 称为 1 状态, $(0,0,1)$ 为 2 状态, $(0,1,0)$ 为 3 状态, $\dots$, $(0,G,1)$ 为 $2(G+1)$ 状态。以此类推, 在第 n 个子系统中, $(n,0,0)$ 为 $2(G+1)n+1$ 状态, $\dots$, $(n,G,1)$ 为 $2(G+1)(n+1)$ 状态。当 $n=N$ 时, 由于转移机制的特殊性, $(N,0,0)$ 为 $2(G+1)N+1$ 状态, $(N+1,0,0)$ 为 $2(G+1)N+2$ 状态, $(N,1,0)$ 为 $2(G+1)N+3$ 状态, $\dots$, $(N+1,G,0)$ 为 $2(G+1)(N+1)$ 状态, 令 $GT=2(G+1)(N+1)$ 。全集 $S=\{s|s=1, 2, \dots, GT\}$ 为包含所有状态的集合。全集 S 、 E 表示状态相同, 只不过状态的表示形式不同。用三个参数表示状态会使得有些状态数学中可以表达, 实际中是不存在的, 如 $(0,0,1)$,

即普通设备和特殊设备均未进入系统, 但却有一个普通设备要离开。系统不可能处于此类状态, 此类状态到达率为 0, 只为简化模型构建过程而设置。

2 模型分析

将空间状态按字典排序, 状态转移机制如图 1 所示。将 n 和 v 相同的状态集合称为一个状态序列, 图中有三个关键状态序列: ① 为等待离开的序列, 此时系统中普通设备数量为 $l+1$, $l=c-kG$, 在特殊设备全部到齐的情况下普通设备会出现等待离开; ② 维修工缺乏序列, 此时普通设备数量为 c , 只要特殊设备加入队列便会出现普通设备等待离开; ③ 为止步序列, 此时普通设备数量为 N , 队列中普通设备数量达到容量, 后续普通设备一旦进入系统便会马上离开。为满足特殊顾客都能够得到及时维修, 则假设 $N \geq c \geq kG$ 。

图 1 模型状态转移机制

系统的状态转移率矩阵可以写成

$$Q = \{q_{i,j}\}_{GT \times GT} = \begin{pmatrix} A_0 & C_0 & & & & \\ B_1 & A_1 & C_1 & & & \\ & B_2 & A_2 & C_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & B_n & A_n & C_n \\ & & & & \ddots & \ddots & \ddots \\ & & & & & B_N & A_N \end{pmatrix} \quad (2)$$

矩阵 B_n 表示普通设备减少的转移率矩阵。这里需要注意前两个关键状态序列发生的位置

$$B_n = \{B_{i,j}\}_{4(G+1) \times (G+1)}, \quad n=1,2,\dots,N \quad (3)$$

$$\begin{cases} B_{2i+1,2i+1} = \min\{(c-ki)\mu_1, n\mu_1\} \\ B_{2i+1,2i+2} = \max\{0, (n-(c-ki))\rho\} \end{cases} \quad i=0,1,\dots,G \quad (4)$$

矩阵 C_n 表示普通设备增多的转移率矩阵

$$C_n = \{C_{i,j}\}_{4(G+1) \times (G+1)} \quad n=0,1,\dots,N-1 \quad (5)$$

$$C_{2i+1,2i+1} = \lambda_1 \quad i=0,1,\dots,G, \quad (6)$$

当 $n=0,1,\dots,N-1$ 时, A_n 表示系统中普通设备数量均为 n 的两个状态序列中状态内部之间的转移率矩阵; 当 $n=N$ 时, A_n 表示 $n=N, N+1, v=0$ 的两个状态序列中状态之间的转移率矩阵。这里包含了最后一个关键状态序列发生的位置

$$A_n = \{A_{i,j}\}_{4(G+1) \times (G+1)}, \quad n=0,1,\dots,N \quad (7)$$

其实 A 中只有非对角线元素才表示这 $2(G+1)$ 个状态内部之间的转移, 对角线元素 $a_{i,i}$ 为马尔科夫过程的形式元素

$$a_{i,i} = q_{i,i} = - \sum_{j \in S, j \neq i} q_{i,j} \quad (8)$$

式中, 当 $n=0,1,\dots,N-1$ 时

$$\begin{cases} A_{2i+2,2i+1} = \mu_3 & i=0,1,\dots,G \\ A_{2i+3,2i+1} = (i+1)\mu_2 & i=0,1,\dots,G-1 \\ A_{2i+1,2i+3} = \lambda_2 & i=0,1,\dots,G-1 \\ A_{i,i} = a_{i,i} & i=0,1,\dots,2(G+1) \end{cases} \quad (9)$$

当 $n=N$ 时

$$\begin{cases} A_{2i+2,2i+1} = \mu_4 & i=0,1,\dots,G \\ A_{2i+3,2i+1} = (i+1)\mu_2 & i=0,1,\dots,G-1 \\ A_{2i+1,2i+3} = \lambda_2 & i=0,1,\dots,G-1 \\ A_{2i+1,2i+2} = \lambda_1 & i=0,1,\dots,G \\ A_{i,i} = a_{i,i} & i=0,1,\dots,2(G+1) \end{cases} \quad (10)$$

3 系统稳态求解

系统稳态求解包括稳态状态概率和稳态状态转移频度。

3.1 稳态状态概率求解

令 $P(t) = (P_1(t), P_2(t), \dots, P_{GT}(t))$ 表示在时间 t 时系统处于各状态的概率

$$\begin{cases} P'(t) = P(t) \cdot Q \\ P(0) = \left(1, \underbrace{0, 0, \dots, 0}_{GT-1}\right) \end{cases} \quad (11)$$

两端做拉普拉斯变换, 可得

$$P^*(s) = P(0)(sI - Q)^{-1}, \quad s > 0 \quad (12)$$

式中, I 为单位阵。根据托贝尔定理可知

$$P = \lim_{t \rightarrow \infty} P(t) = \lim_{s \rightarrow 0} sP^*(s) \quad (13)$$

式中, $P = (P_1, P_2, \dots, P_{GT})$ 表示稳态时, 处于各状态的概率, 也可以理解为稳态时各状态停留时间的比例。

3.2 稳态状态转移频度

本文状态转移频度指状态间的转移频度, 区别以往的仅仅是进入状态的频度。设 $M_{l,z}^0(t+\Delta t)$ 表示从 i 状态出发在 $t+\Delta t$ 时间内从 l 状态向 z 状态转移的次数, $l, z \in S$ 。

当 $l \neq z$ 时

$$M_{l,z}^{(i)}(t+\Delta t) = \begin{cases} \sum_{j=1, j \neq z}^{GT} (P_{l,j}(\Delta t) \cdot M_{l,z}^{(j)}(t)) + \\ P_{l,z}(\Delta t) (1 + M_{l,z}^{(z)}(t)) + o(\Delta t) & i=l \\ \sum_{j=1}^{GT} P_{i,j}(\Delta t) \cdot M_{l,z}^{(j)}(t) + o(\Delta t) & i \neq l \end{cases} \quad (14)$$

式中, 第一式中的 1 表示系统状态在 Δt 内从 l 状态向 z 状态转移了 1 次, 即 $M_{l,z}^0(\Delta t)=1$, $P_{i,j}(\Delta t)$ 表示在 Δt 时间内由 i 状态向 j 状态转移概率, 由指数分布性质可得

$$P_{i,j}(\Delta t) = \begin{cases} q_{i,j} \Delta t & j \neq i \\ 1 - \sum_{j=1, j \neq i}^{GT} q_{i,j} \Delta t & j = i \end{cases} \quad (15)$$

式中, $q_{i,j}$ 即为矩阵 Q 中第 i 行第 j 列的元素。式(15)可化简为

$$P_{i,j}(\Delta t) = \begin{cases} q_{i,j} \Delta t & j \neq i \\ 1 + q_{i,i} \Delta t & j = i \end{cases} \quad (16)$$

代入式(14)可得到

$$\frac{M_{l,z}^{(i)}(t+\Delta t) - M_{l,z}^{(i)}(t)}{\Delta t} = \begin{cases} \sum_{j=1}^{GT} q_{l,j} \cdot M_{l,z}^{(j)}(t) + q_{l,z} + \frac{o(\Delta t)}{\Delta t} & i=l \\ \sum_{j=1}^{GT} q_{i,j} \cdot M_{l,z}^{(j)}(t) + \frac{o(\Delta t)}{\Delta t} & i \neq l \end{cases} \quad (17)$$

令 $\Delta t \rightarrow 0$, 式(17)右端的极限存在, 因此左端的极限也存在, 故 $M_{l,z}^{(i)}(t)$ 可微。由 $M_{l,z}^{(i)}(t)$ 的定义, 显然有 $M_{l,z}^{(i)}(0)=0, i \in S$ 。

初始状态从 1 到 GT 排序, 可以写成矩阵形式

$$\begin{cases} \mathbf{M}'_{l,z}(t) = \mathbf{Q}\mathbf{M}_{l,z}(t) + \mathbf{W}_{l,z} \\ \mathbf{M}_{l,z}(0) = \mathbf{0} \end{cases} \quad (18)$$

式中, $\mathbf{M}_{l,z}(t)$ 为列矢量 $(M_{l,z}^{(1)}(t), M_{l,z}^{(2)}(t), \dots, M_{l,z}^{(GT)}(t))^T$, $\mathbf{W}_{l,z}$ 为列矢量 $(W_{l,z}^{(1)}, W_{l,z}^{(2)}, \dots, W_{l,z}^{(GT)})^T$, 只有 $W_{l,z}^{(i)} = q_{l,z}$, 其余位置均为零, $\mathbf{0}$ 为元素全为零的 GT 维列矢量。

前面已证 $\mathbf{M}_{l,z}(t)$ 是可微的, 令其导数为

$$m_{l,z}^{(i)}(t) = dM_{l,z}^{(i)}(t) / dt, i \in S \quad (19)$$

称 $m_{l,z}^{(i)}(t)$ 为系统从 i 状态出发, 在时刻 t 从 l 状态转移到 z 状态的瞬时频度。为解方程组(19), 对该式的两端乘以 $\exp(-st)$, 并将 t 从 0 到 ∞ 积分, 并且

$$m_{l,z}^{(i)*}(s) = \int_0^\infty \exp(-st) m_{l,z}^{(i)}(t) dt = \int_0^\infty \exp(-st) dM_{l,z}^{(i)}(t), \quad s > 0 \quad (20)$$

可得方程组

$$s\mathbf{m}_{l,z}^*(s) = \mathbf{Q}\mathbf{m}_{l,z}^*(s) + \mathbf{W}_{l,z} \quad (21)$$

式中, $\mathbf{m}_{l,z}^*(s) = (m_{l,z}^{(1)*}(s), m_{l,z}^{(2)*}(s), \dots, m_{l,z}^{(GT)*}(s))^T$, 则

$$\mathbf{m}_{l,z}^*(s) = (s\mathbf{I} - \mathbf{Q})^{-1} \mathbf{W}_{l,z} \quad (22)$$

反演式(22)可求得 $\mathbf{m}_{l,z}(t)$ 。同样根据托贝尔定理, 可以得出稳态状态转移频度 $\mathbf{m}_{l,z}$ 。

$$\mathbf{m}_{l,z} = \lim_{t \rightarrow \infty} \mathbf{m}_{l,z}(t) = \lim_{s \rightarrow 0} s\mathbf{m}_{l,z}^*(s) \quad (23)$$

由于 $\mathbf{m}_{l,z}$ 为稳态时从 l 状态向 z 状态转移的频度, 故与初始状态无关, 因此 $\mathbf{m}_{l,z}$ 中个元素均相同。后文均建立在系统稳态的基础上, 为书写方便令

$$m_{l,z} = m_{l,z}^{(i)}, i \in S \quad (24)$$

当 $l=z$ 时, 根据 Markov 过程性质, 在 Δt 时间内不能发生两次状态转移, 因此 $m_{z,z}=0$ 。

4 系统运行指标构建

结合 DUDIN 等^[19]对系统队列长度、等待时间

等参数的计算分析, 以及张兰芳等^[20]提出的系统容量 N 对系统指标的影响, 提出下列系统运行指标。

4.1 关键状态转移频度

系统中关键状态间的频度是系统运行指标的重要组成部分, 其主要包括了设备维修完成的频度和由于离开机制离开的设备频度。

(1) 维修普通设备频度 L_n

$$L_n = \sum_{g=0}^G \sum_{n=1}^N m_{2(G+1)n+2g+1, 2(G+1)(n-1)+2g+1} \quad (25)$$

(2) 维修特殊设备频度 L_g

$$L_g = \sum_{g=0}^{G-1} \sum_{n=0}^N m_{2(G+1)n+2g+3, 2(G+1)n+2g+1} \quad (26)$$

从式(25)、(26)中可以看出部分转入状态是相同的, 如果不对转出状态加以区分, 则不能判断出是普通设备维修完毕还是特殊设备维修完毕。因此, 状态间的转移频度, 要比单纯的状态转入频度更具价值。

(3) 等待离开的设备频度 L_a

当 $n > c - kg$ 时, 即维修工人手不足, 普通设备不能得到及时维修, 会出现等待离开现象。从 $(n, g, 0)$ 状态转移到 $(n-1, g, 1)$ 状态, 因为 $\mu_3 \rightarrow \infty$, 而后 $(n-1, g, 1)$ 状态瞬间转移到 $(n-1, g, 0)$ 状态。计算 $(n-1, g, 1)$ 状态转移到 $(n-1, g, 0)$ 状态的频度即可得到 L_a

$$L_a = \sum_{g=0}^G \sum_{n=c-kg+1}^N m_{2(G+1)n+2g+2, 2(G+1)(n-1)+2g+1} \quad (27)$$

从 $(n-1, g, 1)$ 状态转移到 $(n-1, g, 0)$ 状态是瞬间完成的, 故在 $(n-1, g, 1)$ 状态上停留的时间为零, 即任何时刻系统处于该状态的概率为零。而 $(n-1, g, 1)$ 状态是判断普通设备是否等待离开系统的决定性状态, 因此, 无法利用状态概率判断是否有普通设备等待离开系统。从式(27)可以看出, 状态转移频度可以妥善解决以上问题。所以状态转移频度比状态转移概率在系统指标求解方面表现的更加灵活。这种优势同样体现在止步离开的设备频度求解上。

(4) 止步的设备频度 L_e

若系统的普通设备容量为 N , 那么当再有普通设备到达系统后, 便会马上离开。系统从 $(N, g, 0)$ 状态转移到 $(N+1, g, 0)$ 状态。因为 $\mu_3 \rightarrow \infty$, 而后 $(N+1, g, 0)$ 瞬间转移到 $(N, g, 0)$ 。只需计算 $(N+1, g, 0)$ 状态转移到 $(N, g, 0)$ 状态的频度即可得到 L_e

$$L_e = \sum_{g=0}^G m_{2(G+1)N+2g+2, 2(G+1)N+2g+1} \quad (28)$$

4.2 稳态时状态转移频度与状态转移概率的关系

稳态时状态转移概率矩阵 P_{ij}

$$P_{i,j} = t_i \cdot m_{i,j} \quad (29)$$

稳态时状态概率 P_i

$$P_i = t_i \sum_{j \in S} m_{i,j} \quad (30)$$

式中, $m_{i,j}$ 表示系统从状态 i 向状态 j 转移的频度。 t_i 表示从 i 状态向其他可直达状态转移的时间, 由于是竞争关系, 所以 t_i 为从 i 状态向各状态转移时间的最小值。由多个独立分布的最小值期望与指数分布的性质, 可以推导出

$$t_i = \frac{1}{\sum_{j \in S, j \neq i} q_{i,j}} \quad (31)$$

4.3 普通设备待修时间

在系统稳态下普通设备待修时间 T_d , 指系统运行区间 T_s 内, 所有普通设备待修时间总和。不仅需要考虑状态停留时间, 同时还需考虑对应状态待修普通设备数量。状态转移之后, 可以根据转移过程得出状态改变原因, 因此根据状态改变的原因将 T_d 划分为 4 个部分组成。

$$T_d = T_{d1} + T_{d2} + T_{d3} + T_{d4} \quad (32)$$

普通设备维修完毕前的等待时间 T_{d1} , 由 $(n, g, 0)$ 状态转移到 $(n-1, g, 0)$ 状态。

$$T_{d1} = T_s \sum_{g=0}^G \sum_{n=c-kg+1}^N (n-c+kg) P_{i,j} \quad (33)$$

式中, $i=2(G+1)n+2g+1$, $j=2(G+1)(n-1)+2g+1$; $n=c-kg+1$ 表示有普通设备待修;

特殊设备维修完毕前的等待时间 T_{d2} , 由 $(n, g, 0)$ 状态转移到 $(n, g-1, 0)$ 状态。

$$T_{d2} = T_s \sum_{g=0}^{G-1} \sum_{n=c-kg}^N (n-c+kg) P_{i,j} \quad (34)$$

式中, $i=2(G+1)n+2g+3$, $j=2(G+1)n+2g+1$ 。

普通设备等待离开前的等待时间 T_{d3} , 由 $(n, g, 0)$ 状态先转移到 $(n-1, g, 1)$ 状态, 再由 $(n-1, g, 1)$ 状态转移到 $(n-1, g, 0)$ 状态, 由于 $\mu_3 \rightarrow \infty$, $(n-1, g, 1)$ 状态转移到 $(n-1, g, 0)$ 状态时间忽略不计。

$$T_{d3} = T_s \sum_{g=0}^G \sum_{n=c-kg+1}^N (n-c+kg) P_{i,j} \quad (35)$$

式中, $i=2(G+1)n+2g+1$, $j=2(G+1)(n-1)+2g+2$ 。

普通设备止步前的等待时间 T_{d4} , 由 $(N, g, 0)$ 状态转移到 $(N+1, g, 0)$ 状态, 再由 $(N+1, g, 0)$ 状态转移到 $(N, g, 0)$ 状态, 由于 $\mu_4 \rightarrow \infty$, $(N+1, g, 0)$ 状态转移到 $(N, g, 0)$ 状态时间忽略不计。

$$T_{d4} = T_s \sum_{g=0}^G (N-c+kg) P_{i,j} \quad (36)$$

式中, $i=2(G+1)N+2g+1$, $j=2(G+1)N+2g+2$ 。

对式(33)~(36)的求解过程的解释如下: $T_s \cdot P_{i,j}$ 表示系统运行时间内, 向 j 状态转移之前在 i 状态停留的时间。再乘以此段时间内待修设备的数量 $(n-c+kg)$, 得出该状态转移下普通设备等待的总时间。最后对相应状态转移进行求和, 得出对应原因下普通设备待修时间总和。

4.4 系统运行费用

参考将马尔科夫过程在网络上的应用^[21]和排队理论在医院方面的应用^[22], 给出本文所提模型在维修厂方面的应用。

应用模型时可以根据系统参数 c 、 k 、 N 来计算系统的盈利, 由此以系统盈利最大为目标函数来匹配系统参数, 影响系统盈利的因素。

(1) 单位时间内系统因维修设备的盈利, 包括两部分: 维修普通设备频度 $\times$ 单个普通设备工厂盈利, 即 $L_n C_n$; 维修特殊设备频度 $\times$ 单个特殊设备工厂盈利, 即 $L_g C_g$ 。

(2) 单位时间内系统因普通设备离开而产生的费用, 分为两部分, 包括单位时间内等待离开系统设备个数 $\times$ 单个设备等待离开的费用 $L_a C_a$ 和因为工厂待修设备饱和而不能服务的设备 $\times$ 单个设备止步的费用 $L_e C_e$ 。

(3) 系统因待修设备等待时间而产生的费用, 普通设备等待维修时间乘以单位时间排队设备的花费, 即 $T_d C_d$ 。

(4) 系统雇佣维修工的费用, 维修工数量乘以单个维修工工资, 即 $c C_c$ 。

(5) 由于维修车辆会抢占保养车辆的位置, 因此无需额外增设, 所以场地费用只需考虑保养车辆数量。维修厂场地费用, 保养车辆上限乘以单个工位费用, 即 $N C_N$ 。

系统总盈利 C 为

$$C = T_s (L_n C_n + L_g C_g - L_a C_a - L_e C_e) - T_d C_d - c C_c - N C_N \quad (37)$$

5 系统算例

本文以某维修厂为例。该维修厂主要负责汽车的保养和维修。维修厂内有 c 个维修工, 每个维修工负责一辆汽车的保养。有汽车需要维修时, 会有 k 个维修工立即对其进行维修, k 越大维修效率越高, 其余维修工仍负责汽车的保养。如果维修工人数不足时, 会中断当前的保养工作, 优先服务于维修车辆。为保证维修的及时性以及维修厂场地的限

制,维修厂最多只能同时接纳 G 辆汽车维修及 N 辆汽车保养。为控制系统规模,又不失一般性,这里取 $G=2$ 。

需要保养及维修的汽车到达时间服从指数分布,保养汽车平均到达时间 40 min/辆,维修汽车平均到达时间 200 min/辆,汽车的维修及保养时间服从指数分布,汽车平均保养时间 240 min/辆,汽车平均维修时间 $360/1.2^{k-1}$ min/辆。若有保养汽车在队列中等待,排队中的车辆会以参数为 120 min/辆的指数分布离开维修厂。

保养汽车维修厂盈利 1 200 元/辆,维修汽车维修厂盈利 7 000 元/辆,汽车等待成本为 10 元/(min·辆),汽车等待离开时维修厂会损失 1 200 元/辆,汽车止步时维修厂会损失 3 000 元/辆,维修工雇佣费 3 万元/(人·月),场地金为 1 万元/(辆·月)。那么,可以通过确定系统参数 k 、 c 、 N 来实现工厂盈利最大化。

5.1 分块矩阵与全矩阵求逆对比

在马尔科夫过程的运算中,最耗费时间的是式(12)、(22)中 $(sI-A)$ 矩阵的求逆过程,本文采用分块矩阵法对该部分进行了计算简化,并于全矩阵求逆的方法进行对比。分块矩阵的求解方法可参照文献[23]。计算机性能为 i5-4590 的中央处理器,16 GB 内存,64 位 Win7 操作系统,计算软件采用 Matlab2018a。对比结果如图 2 所示。

图2 矩阵求逆时间对比

图 2a 为采用分块矩阵求逆法所用的, b 为采用 Matlab 自带的全矩阵求逆函数所需时间,两种方法计算时间差距随矩阵规模的增大而拉大。当 $k=2$, $c=15$, $N=18$ 时,采用分块矩阵求逆只需 83 s,而采用全矩阵求逆法则需要 1 794 s,两者相差近 20 倍。因此对于大规模矩阵求逆来说分块矩阵法是十分有利的。矩阵规模跟 N 有直接关系,其具体数据如表 1 所示。

表 1 矩阵规模

保养汽车数量 N /辆	9	12	14	18
矩阵规模	60^2	78^2	90^2	114^2

5.2 状态转移频度算法的模拟与理论计算对比

为证明本文所提状态转移频度算法的正确性,采用 MCMC (Markov Chain Monte Carlo)方法对其进行验证,采用 $k=2$, $c=5$, $N=5$ 的系统,选取了出现次数较多的几个状态转移过程进行对比,如图 3 所示。

图3 模拟与理论计算结果对比

从图 3 中可以看出,随着模拟运行时间的增加,模拟结果与理论计算结果越来越接近。当运行时间为 10^6 min 时(7.6 年)时已基本接近理论计算结果,运行到 10^7 min 时,各状态转移频度的模拟与理论计算结果几乎重合,表明此时系统已处于稳态。

5.3 优化分析

以一月为限,每月工作 30 d,每天工作 8 h,则全月工作时间 $T_s=30 \times 8 \times 60=14\,400(\text{min})$,利用式(37)计算系统盈利,对于 k 取不同值时,系统最大盈利情况如表 2 所示。

表 2 不同 k 下系统的最大盈利情况

k	1	2	3	4
系统盈利/(万元)	27.4	30.1	32.0	33.0
$[c, N]$	$[10, 12]$	$[11, 12]$	$[12, 12]$	$[12, 12]$
k	5	6	7	
系统盈利/(万元)	34.0	35.1	33.8	
$[c, N]$	$[12, 12]$	$[12, 12]$	$[14, 14]$	

从表 2 可以看出,随着 k 不断增大,系统的最大盈利呈现先增后减的趋势,在 $k=6$, $c=12$, $N=12$ 时取得最大值 $C=35.1$ 万元。其中各部收支占比如图 4 所示。

图 4 系统各部分收支占比

图 4 中外圈表示系统收入，共 92.2 万元，内圈表示系统支出，共 57.1 万元。系统的收入主要为汽车维修，占总收入的 56%；系统中人员和场地支出占比较高，分别占比支出的 63% 和 25%，而由于未能及时维修造成的赔偿金额则仅占比支出的 12%，说明在此配置下车辆基本能够得到妥善保养，到厂后，不保养便离开的情况发生较少。由式(27)、(28)计算可得每月等待离开的保养车辆为 32.7 辆，止步的保养车辆为 3.7 辆，供需赔偿 5 万元。

6 结论

以多种优先级和离开机制并存的 $M/M/c/N$ 排队系统模型为研究对象，结合马尔科夫过程与分块矩阵理论对其求解，并在此基础上对该模型进行了优化分析，所得结论如下。

(1) 改变了基于马尔科夫过程求解排队模型的计算目标，从计算状态转移概率转为计算状态转移频度，并通过分块矩阵理论提升了模型求解效率。

(2) 建立了稳态时状态转移频度和状态转移概率之间的关系。利用状态转移频度得出了维修设备的频度、等待离开和止步设备的频度、设备待修总时间、系统运行费用等系统运行指标，进而完备了多优先级和离开机制相结合的排队模型的指标体系。

(3) 通过模拟验证了状态转移频度的理论计算方法的准确性，并在完备的指标体系基础上，对维修厂的配置进行了优化分析，进而指导实际生产。

参 考 文 献

- [1] LIU C L, LAYLAND J W. Scheduling algorithms for multiprogramming in a hard-real-time environment[J]. Readings in Hardware/software Co-design, 2002, 20(1): 179-194.
- [2] SAIFULLAH A, FERRY D, LI J, et al. Parallel real-time scheduling of dags[J]. IEEE Transactions on Parallel & Distributed Systems, 2014, 25(12): 3242-3252.
- [3] HEIJDEN M V D, HARTEN A V, SLEPTCHENKO A. Approximations for Markovian multi-class queues with preemptive priorities[J]. Operations Research Letters, 2004, 32(3): 273-282.
- [4] SLEPTCHENKO A, HARTEN A V, HEIJDEN M V D. An exact solution for the state probabilities of the multi-class, multi-server queue with preemptive priorities[J]. Queueing Systems, 2005, 50(1): 81-107.
- [5] 周杰. 考虑急诊非强占优先权的 CT 室容量管理研究[D]. 成都: 西南交通大学, 2016.
- ZHOU Jie. Study on capacity management of CT department with emergency non-preemptive priority[D]. Chengdu: Southwest Jiaotong University, 2016.
- [6] HALFIN S, WHITT W. Heavy-traffic limits for queues with many exponential servers[J]. Operations Research, 1981, 29(3): 567-588.
- [7] WANG K, LI N, JIANG Z. Queueing system with impatient customers: A review[C]//IEEE International Conference on Service Operations and Logistics and Informatics. Qing Dao: IEEE, 2010: 82-87.
- [8] HAIGHT F A. Queueing with balking. II[J]. Biometrika, 1960, 47(3): 285-296.
- [9] GARNET O, MANDELBAUM A, REIMAN M. Designing a call center with impatient customers[J]. Manufacturing & Service Operations Management, 2002, 4(3): 208-227.
- [10] WEI X, ALTIOK T. An approximation for multi-server queues with deterministic reneging times[J]. Annals of Operations Research, 2009, 172(1): 143-151.
- [11] WANG K H, CHANG Y C. Cost analysis of a finite $M/M/R$ queueing system with balking, reneging, and server breakdowns[J]. Mathematical Methods of Operations Research, 2002, 56(2): 169-180.
- [12] WARD A R, Glynn P. A diffusion approximation for a $gi/gi/1$ queue with balking or reneging[J]. Queueing Systems, 2005, 50(4): 371-400.
- [13] KE J C, CHANG F M. Modified vacation policy for $M/G/1$ retrial queue with balking and feedback[J]. Computers & Industrial Engineering, 2009, 57(1): 433-443.
- [14] IRAVANI F, BALCIOGLUB. On priority queues with impatient customers[J]. Queueing Systems, 2008, 58(4): 239-260.
- [15] JOUINI O, ROUBOS A. On multiple priority multi-server queues with impatience[J]. Journal of the Operational Research Society, 2014, 65(5): 616-632.
- [16] 韩超. 基于马尔科夫的服务能力分享策略排队模型研究[D]. 上海: 上海交通大学, 2012.
- HAN Chao. Service capacity sharing strategy of queueing

- model based on Markov theory.[D]. Shanghai: Shanghai Jiao Tong University, 2012.
- [17] HEIJDEN M V, HARTEN A V, SLEPTCHENKO A. Approximations for Markovian multi-class queues with preemptive priorities[J]. *Operations Research Letters*, 2004, 32(3): 273-282.
- [18] WANG J, BARON O, SCHELLERWOLF A. M/M/c queue with two priority classes[J]. *Operations Research*, 2015, 63(3): 1-16.
- [19] DUDIN A, KLIMENOK V, LEE M H. Recursive formulas for the moments of queue length in the BMAP/G/1 queue[J]. *IEEE Communications Letters*, 2009, 13(5): 351-353.
- [20] 张兰芳, 王知, 方守恩. 机场航站楼路边交通容量需求分析[J]. *同济大学学报*, 2007, 35(4): 486-490.
ZHNAG Lanfang, WANG Zhi, FANG Shouen. Study on traffic capacity and demand of airport terminal curbside[J]. *Journal of Tongji University(Natural Science)*, 2007, 35(4): 486-490.
- [21] SHEIKH S M, WOLHUTER R, ENGELBRECHT H A. A model for analyzing the performance of wireless multi-hop networks using a contention-based CSMA/CA strategy[J]. *KSII Transactions on Internet & Information Systems*, 2017, 11(5): 2499-2522.
- [22] ANDERSEN A R, BO F N, REINHARDT L B. Optimization of hospital ward resources with patient relocation using Markov chain modeling[J]. *European Journal of Operational Research*, 2017, 260(3): 1152-1163.
- [23] LIU X Q, HUANG T Z, FU Y D. Estimates for the inverse elements of tridiagonal matrices[J]. *Applied Mathematics Letters*, 2006, 19(6): 590-598.
-
- 作者简介: 印明昂(通信作者), 男, 1985 年出生, 博士, 讲师。主要研究方向为系统可靠性。
E-mail: yinma@mail.neu.edu.cn