

基于POMDP模型的检修与备件库存联合优化

汪凯¹ 顾刘栋² 周一帆^{1*}

1. 东南大学机械工程学院, 南京, 211189

2. 东南大学数学学院, 南京, 211189

摘要:在维修与备件库存联合优化时,已有的研究大多假设系统状态监测是完美的,忽略了实际应用中的误差。为了解决该问题,以包含不完美状态监测和固定检修周期的单部件系统为研究对象,考虑该系统的视情维修(CBM)与备件库存管理问题,采用部分可观测的马尔可夫决策过程(POMDP)对系统进行建模,并推导系统状态转移概率。为了处理复杂的信念状态空间,提高算法求解效率,采用了一种改进的Perseus算法。在数值案例部分验证了该算法的有效性,并对最优策略结构进行分析,结果表明:信念状态相比于观测值能相对合理地表示状态信息,同时也对比了有无备件库存的情况,证明了备件库存的有效性。

关键词:不完美状态监测;部分可观测的马尔可夫决策过程;信念状态;维修与备件库存联合优化

中图分类号:TH17;O224

DOI:10.3969/j.issn.1004-132X.2025.08.020

开放科学(资源服务)标识码(OSID):



Joint Optimization of Inspection, Maintenance and Spare Parts Inventory Based on POMDP Model

WANG Kai¹ GU Liudong² ZHOU Yifan^{1*}

1.School of Mechanical Engineering, Southeast University, Nanjing, 211189

2.School of Mathematics, Southeast University, Nanjing, 211189

Abstract: In the research direction of joint optimization of maintenance and spare parts inventory, most of the existing research assumed that the system condition monitoring was perfect and ignored the detection errors in practical applications. In order to solve the above problems, a single-component system containing imperfect condition monitoring and fixed inspection and maintenance intervals was studied, and the condition-based maintenance (CBM) and spare parts inventory management problems of the systems were considered. POMDP were used to model the system and derive the system state transfer probabilities. In order to deal with the complex belief state spaces and improve the algorithm solution efficiency, an improved Perseus algorithm was used. In the numerical case section, the effectiveness of the algorithm was verified, and then the optimal policy structure was analyzed to show that the belief states may represent the state information relatively reasonably compared with the observations, and the effectiveness of the spare parts inventory is also demonstrated by comparing the cases without spare parts inventory.

Key words: imperfect condition monitoring; partially observable Markov decision processes (POMDP); belief state; joint optimization of maintenance and spare parts inventory

0 引言

维修与备件库存的联合优化问题是目前设备健康管理领域的研究热点^[1]。合理的备件库存策略会减少因备件短缺导致维修更换延迟或设备停机的风险,但在实际应用中,设备状态信息监测存在误差,因此,需要在维修与备件库存联合优化的基础上考虑状态监测误差导致的不确定性。部分

可观测的马尔可夫决策过程(partially observable Markov decision processes, POMDP)模型为描述不确定的系统状态提供了合适的模型框架。但是,对于维修与备件联合优化问题,需要同时考虑不确定的部件状态和确定的库存状态,构建合理的POMDP模型框架。此外,考虑备件库存因素后,也需要用合适的算法应对快速增长的系统状态空间。本文基于POMDP模型,采用改进的Perseus算法,研究不完美监测下的视情维修(condition-based maintenance, CBM)与备件库存的联合优化问题。

收稿日期:2024-05-16

基金项目:国家自然科学基金(72071044);江苏省研究生科研与实践创新计划(KYCX24_0358)

目前,针对视情维修和备件库存的联合优化问题已有许多研究。LI等^[2]采用马尔可夫决策过程(Markov decision processes, MDP)对两部件系统视情维修与备件库存策略的联合优化问题进行建模,并使用策略迭代算法进行求解。WANG等^[3]采用半马尔可夫决策过程研究了两部件串联系统的视情维修与备件库存策略联合优化问题。ZHENG等^[4]将检测间隔内发生故障的K-out-of-N系统的CBM与备件订购问题建模为MDP。但是这些联合优化的问题都假设状态监测是完美的,没有考虑状态监测误差。状态信息在收集和處理等过程中难免会受到噪声的影响,如传感器收集到有噪声的信息、对振动信号等信息处理的过程会产生误差^[5-6]。

针对状态监测存在误差的情况,有学者将有误差的状态监测与维修联合优化。张红旗等^[7]针对机电装备运行维护问题,考虑不完全观测的设备运行信息,提出了基于POMDP理论的动态可靠性评价方法。刘大玲等^[8]通过对高速铁路无砟轨道系统的监测数据进行分析,建立了无砟轨道状态BP神经网络预测模型,以指导运维部门合理的养护维修。ROUX等^[9]将一个随机退化的单部件系统建模为POMDP模型。MORATO等^[10]通过将动态贝叶斯网络和POMDP相结合,对海上风机支架结构进行退化建模,并与基于启发式的参数化策略进行对比,发现POMDP方案显著地降低了成本。NGUYEN等^[11]使用高斯噪声描述检测误差,提出了一种基于POMDP模型的视情维修和动态监测的策略,将不同监测质量的动作作为决策变量,相比静态策略取得了更好的效果。上述研究同时考虑状态监测与维修,大部分将问题建模为POMDP模型,但未同时考虑状态监测、维修和备件库存等因素。若同时考虑这些因素,会增加POMDP模型下信念空间的维度和复杂性。

本文以钢厂的炼炉系统为研究对象,考虑了不完美的状态监测和备件库存管理。实际中一般需要考虑生产安全等因素,设置固定周期的检修来对系统进行整体的检测和维护。同时,为降低检修周期内系统故障停机的风险,在检修周期内设置多个监测点,根据监测的不完美状态信息采取额外的维修措施。在检修或监测的同时可以进行备件库存盘点,以满足维修更换的需求。

本文的主要贡献如下:①研究不完美的状态监测下的CBM与备件库存管理联合优化问题,

并将该问题建模为POMDP模型;②采用QMDP方法改进Perseus算法,以提高求解复杂信念状态空间下POMDP模型的效率;③以某钢厂转炉系统为案例,验证改进的Perseus算法,并分析CBM和备件库存联合优化的必要性。

1 问题假设

现实中有许多设备的状态监测存在误差,如钢铁厂高炉的耐火材料^[12]、高铁的牵引变压器设备^[13]、滚动轴承^[14]等。系统部件随时间退化,在 t 时刻的退化状态表示为 $x(t)$ 。由于Gamma过程对单调退化的随机过程建模最为合适,它对腐蚀、磨损、疲劳裂纹扩展等退化具有很好的拟合效果^[15],故假设部件的退化状态 $x(t)$ 服从Gamma过程,其初始状态 $x(0)=0$,失效阈值为 F ,即当 $x(t) \geq F$ 时部件失效。根据Gamma过程具有独立增量的性质,部件的退化增量可以表示为

$$x(t + \Delta t) - x(t) \sim G(\alpha \Delta t, \beta) \quad (1)$$

式中, α 为形状参数; β 为尺度参数; $G(\cdot)$ 为Gamma过程的累积概率密度函数。

系统状态检测在检修点和监测点进行,图1所示为检修点和监测点之间的关系。

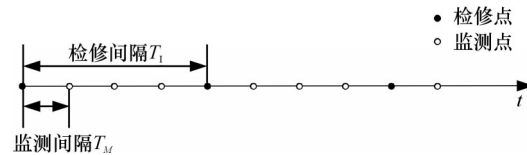


图1 检修点与监测点的关系

Fig.1 The relationship between inspection and monitoring points

考虑生产安全等因素,设置固定周期的检修,但检修间隔时间较长,因此,检修周期内通过额外的状态监测以获得状态信息,状态监测一般通过人工监测或传感器等方式获得状态数据。假设监测周期为 T_M ,检修周期为 T_I (T_I 为 T_M 的整数倍)。在检修点时能检测部件的真实退化状态,而在监测点时得到的部件退化状态存在噪声误差,部件退化状态的观测值可以表示为

$$x_c(t) = \begin{cases} x(t) & t \bmod T_I = 0 \\ x(t) + \epsilon & \text{其他} \end{cases} \quad (2)$$

其中, ϵ 为监测点观测到部件退化状态的误差。由于现有研究大部分采用正态分布描述状态观测误差^[11,16],故本文同样选取正态分布作为误差噪声,假设误差值服从正态分布 $\epsilon \sim N(0, \sigma)$ 。此外,正态分布并非本研究所提方法的必要假设,它只影响状态观测函数,如果噪声误差符合其他分布,只需修改相应的状态观测函数。

部件的维修活动主要在检修点或监测点进行,主要包括两种方式:事后维修和预防性维修。当部件退化状态超过失效阈值 F 时,需要立即进行事后维修,更换故障部件以防止造成更多的损失。事后维修又可分为两种情况:①无库存备件时发生事后维修,需要紧急订购备件并进行事后维修,假设紧急订购备件时间忽略不计,此时维修成本为 C_E ;②有库存备件时发生事后维修,则直接使用备件维修替换,此时维修成本为 C_C 。在仓库库存无可备用件的情况下无法提供预防性维修更换。由于检修点的检测维修活动是定期进行的,故不需要专门派人或设备进行维修,相对于监测点有更低的预防性维修更换成本。预防性维修更换在检修点和监测点的成本分别为 C_P 、 C'_P 。显然,不同维修方式下的成本关系为 $C_E > C_C > C'_P > C_P$ 。由于检测与维修的时间远小于监测间隔时间,故假设检测与维修的时间忽略不计。

仓库的库存盘点周期和监测周期一致,即在检修点或监测点获得部件状态信息后判断是否进行预防性维修,之后盘点备件库存。根据备件库存和部件状态判断是否进行备件订购,备件订购的提前期为 T_1 。备件订购的成本为 C_O ,同时备件运输和储存也都会有相应的成本,假设运输和储存的单位时间内成本相同,为 C_H 。库存容量与系统部件数相关。

在监测点或检修点时各个事件的执行顺序如图 2 所示。首先订购备件到货,其次检测或监测部件退化状态,然后判断是否进行预防性维修更换,最后判断是否进行备件订购。

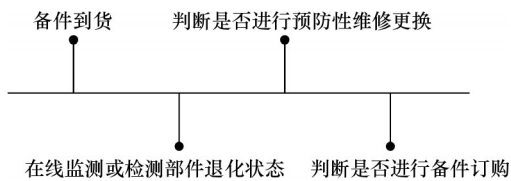


图2 事件发生顺序
Fig.2 Sequence of events

2 POMDP 模型

由于不完美监测导致系统状态信息损失,故采用 POMDP 模型对系统进行建模。与 MDP 模型类似,POMDP 模型也是一套通过交互式学习实现目标的理论框架,其中智能体表示进行学习和实施决策的抽象机器^[17]。相比 MDP 模型,POMDP 模型中的智能体只能观测到有误差的状态或局部状态。为了更加准确地描述系统状态,此时通过智能体的观测值计算系统状态的概率分

布——信念状态(belief state)。一般将离散时基于信念空间的 POMDP 模型描述为 8 元组 $(S, B, A, \tau, R, \Omega, O, \gamma)$ 。其中, S 为系统状态空间, B 为系统信念状态空间, A 为动作空间, τ 为信念状态的状态转移函数, R 为期望奖励函数, Ω 为观测状态空间, O 为状态观测函数, γ 为折扣因子。

POMDP 模型中智能体与环境交互的过程如图 3 所示,智能体将当前环境下的观测值通过信念更新模型得到当前系统的信念状态,然后采取动作,环境随之响应更新状态,并获得奖励。

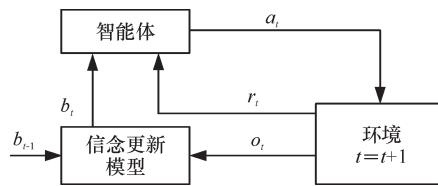


图3 智能体与环境交互示意图
Fig.3 Schematic diagram of the interaction between agent and environment

2.1 状态空间和信念状态空间

系统状态由监测点状态、部件的退化状态和库存状态组成,可表示为

$$s = [I, y, z] \quad (3)$$

其中, $I \in \{0, 1\}$ 表示系统是否处于检修点; z 表示系统部件的离散退化状态; y 表示备件库存状态。当离散状态数为 N 时,离散状态 z 和连续状态 x 之间的关系为

$$z = \begin{cases} n & x \in [nL, (n+1)L], x < F \\ N & x \geq F \end{cases} \quad (4)$$

其中, $L = F/N$ 为离散状态的长度,离散状态的范围为 $z \in [0, N)$ 。库存状态可表示为

$$y = \begin{cases} 0 & \text{无库存} \\ 1 & \text{有库存} \\ m & 1 < m \leq T_L \text{ 在途库存} \end{cases} \quad (5)$$

其中, $y = m$ 表示在途库存需 $m - 1$ 个单位时间到货。

为了描述部件状态的不确定性,采用部件状态的概率分布即信念状态来表示部件状态。部件的信念状态可以表示为 $b_{\text{comp}} = [p(s_0), p(s_1), \dots, p(s_n)]$,在此基础上考虑库存状态, b_{comp}^m 表示库存状态为 m 时的部件状态,系统信念状态可进一步表示为 $b_{\text{inv}} = [b_{\text{comp}}^0, b_{\text{comp}}^1, \dots, b_{\text{comp}}^m]$,最后考虑检测点状态,系统信念状态最终表示为 $b = [b_{\text{inv}}^0, b_{\text{inv}}^1]$,上标 0 和 1 分别表示检修点和监测点两种检测点状态。

2.2 动作空间

动作主要包括预防性维修和订购备件,所以

将动作空间表示为

$$A = \{DN, OA, MA\} \quad (6)$$

其中, $a = DN$ 表示不采取任何动作, $a = OA$ 表示订购备件, $a = MA$ 表示预防性维修更换。当仓库有备件时, 不进行订购备件的动作; 当仓库无库存时, 不进行预防性维修更换的动作。为了后续方便表示, 用 $a = 0$ 表示 DN, $a = 1$ 表示 OA, $a = 2$ 表示 MA。

2.3 奖励函数

奖励函数在 MDP 和 POMDP 问题中都至关重要。在 POMDP 问题中, 奖励函数表示为

$$R_E(b, a) = \sum_{s \in S} b(s) R(s, a) \quad (7)$$

其中, b 表示系统信念状态; $b(s)$ 表示处于状态 s 下的概率值; $R(s, a)$ 表示在状态 s 下采取动作 a 后单位时间内产生的奖励。奖励函数 $R(s, a)$ 可具体表示如下:

$$R(s, a) = -E(C_p(s, a) + C_c(s, a, s') + C_o(s, a) + C_h(s, a, s'))$$

其中, $E(\cdot)$ 表示取一个时间周期内的期望值。预防性维修更换的成本为

$$C_p(s, a) = (C_p \cdot \mathbb{I}(t \% T_1 = 0) + C_p' \cdot \mathbb{I}(t \% T_1 \neq 0)) \cdot \mathbb{I}(a = 2)$$

其中, $\mathbb{I}(\cdot)$ 表示指示函数。事后维修成本为

$$C_c(s, a, s') = (C_c \cdot \mathbb{I}(y = 1) + C_e \cdot \mathbb{I}(y \neq 1)) \cdot \mathbb{I}(x' \geq F)$$

备件订购成本为

$$C_o(s, a) = C_o \cdot \mathbb{I}(a = 1)$$

备件运输和持有成本为

$$C_h(s, a, s') = C_h \cdot (\mathbb{I}(y > 0) + \mathbb{I}(y' > 0)) / 2$$

2.4 信念状态转移函数

信念状态转移函数是 POMDP 模型中的关键部分。在已知部件退化状态和系统环境的情况下, 采取动作 a 使信念状态 b 更新到下一刻的状态 b' , 即表示为

$$\tau(b, a, b') = \frac{O(a, s', o) \sum_{s \in S} T(s, a, s') b(s)}{\Pr(o|b, a)}$$

其中, $O(a, s', o)$ 为系统状态观测函数; $T(s, a, s')$ 为系统状态转移函数; $\Pr(o|b, a)$ 起到归一化作用, 可表示为

$$\Pr(o|b, a) = \sum_{s \in S} b(s) \sum_{s' \in S} T(s, a, s') O(a, s', o)$$

从信念状态转移概率的计算公式中可以看出, 需要获得系统状态的转移函数 $T(s, a, s') = \Pr(s_{t+1} = s' | s_t = s, a_t = a)$ 和状态观测函数 $O(a, s, o) = \Pr(o | s, a)$ 。由于系统状态 s 包含部件状态、备件库存状态和检测点的信息, 故需要先计算部件的状态转移概率, 然后在此基础上考虑库存

状态因素, 最后考虑检测点信息, 最终得到系统状态的转移概率。

2.4.1 状态转移概率

状态转移概率受事后维修的影响。在两个时刻间部件的退化可分为两种情况: ① 部件发生失效; ② 部件未发生失效。如图 4 所示, t_i 为部件故障的时刻, T 为监测时间间隔。当 $t_i > T$ 时, 部件在当前监测时间间隔内未发生失效, 部件正常退化; 当 $t_i \leq T$ 时, 部件在当前监测时间间隔内发生失效, 在 t_i 时刻进行事后维修。

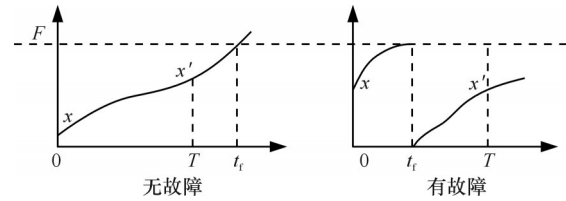


图4 部件失效曲线示意图

Fig.4 Diagram of component failure curves

部件状态转移概率的计算过程如下:

$$\Pr(x(t+1) = x' | x(t) = x) =$$

$$\Pr(x(t+1) = x' | x(t) = x, t_i \geq T) \Pr(t_i \geq T | x(t) = x) + \Pr(x(t+1) = x' | x(t) = x, t_i < T) \Pr(t_i < T | x(t) = x)$$

1) 对于未发生失效的情况, 有

$$\Pr(x(t+1) = x' | x(t) = x, t_i \geq T) \Pr(t_i \geq T | x(t) = x) = \int_T^{+\infty} \Pr(x(T) = x' | x(0) = x, x(t_i) = F) f(t_i | x(0) = x) dt_i = \int_T^{+\infty} B\left(\frac{x' - x}{F - x}, \alpha T, \alpha(t_i - T)\right) f(t_i | x(0) = x) dt_i$$

其中, 对于 $\Pr(x(T) = x' | x(0) = x, x(t_i) = F)$ 部分概率的求解需用到 Gamma Bridge 过程^[18]; $B(\cdot)$ 为不完全 Beta 函数。

基于 Gamma 过程的部件寿命分布函数表示为^[15]

$$F(t) = \Pr(t_i \leq t) = \Pr(x(t) \geq F) = \int_F^{+\infty} g(x | \alpha t, \beta) dx = \frac{\Gamma(\alpha t, \beta F)}{\Gamma(\alpha t)}$$

为了减少计算复杂性, 部件寿命分布的概率密度函数可用 B-S 分布来近似^[19]:

$$f(t|x) = \frac{dF(t)}{dt} \approx \Phi\left(\frac{1}{2\sqrt{2\pi}uv} \left[\left(\frac{u}{t}\right)^{1/2} + \left(\frac{u}{t}\right)^{3/2}\right], \exp\left[-\frac{1}{2v^2} \left(\frac{t}{u} - 2 + \frac{u}{t}\right)\right]\right)$$

$$u = x\beta/\alpha \quad v = 1/\sqrt{x\beta}$$

其中, $\Phi(\cdot)$ 为标准正态分布的累积分布函数。

2) 对于发生失效的情况, 有

$$\Pr(x(t+1) = x' | x(t) = x, t_i < T) \Pr(t_i < T | x(t) = x) = \int_0^T \Pr(x(T) = x' | x(t_i) = 0) f(t_i | x(0) = x) dt_i = \int_0^T G(x' | \alpha(T - t_i), \beta) f(t_i | x(0) = x) dt_i$$

得到部件状态转移概率 $\Pr(x(t+1)=x'|x(t)=x)$ 后,可以计算部件的离散状态转移概率矩阵。不采取维修动作时,部件的离散状态转移概率为

$$\Pr(z(t+1)=j|z(t)=i)=\int_T^{+\infty} (B(\frac{(j+1)L-(i+0.5)L}{F-(i+0.5)L}, \alpha T, \alpha(t_i-T)) - B(\frac{\max\{j, i+0.5\}L-(i+0.5)L}{F-(i+0.5)L}, \alpha T, \alpha(t_i-T))) \cdot f(t_i|(i+0.5)L) dt_i \cdot \mathbb{I}(j \geq i) + \int_0^T (G((j+1)L|\alpha(T-t_i), \beta) - G(jL|\alpha(T-t_i), \beta)) \cdot f(t_i|(i+0.5)L) dt_i$$

此时无维修动作下部件的状态转移概率矩阵记号表示为 T_{comp}^0 , 即 $T_{\text{comp}}^0[i, j] = \Pr(z(t+1)=j|z(t)=i)$, 有维修动作下的部件状态转移矩阵表示为

$$T_{\text{comp}}^1 = ET_{\text{comp}}^0$$

其中, E 为部件维修状态转移矩阵, 它是与 T_{comp}^0 维度一致但第一列全为 1、其他位置全为 0 的矩阵。此时用上标 0 和 1 区分有无维修动作。

由于事后维修会对库存备件状态产生影响, 当发生事后维修时无备件库存, 则进行紧急订购, 当发生事后维修时有备件库存, 则使用备件进行事后维修, 故用上标 cm、ucm 区分监测周期内是否发生事后维修。

进一步加入备件因素后, 使用分块矩阵表示系统状态转移矩阵, T_{inv}^0 表示采取动作 a 时的状态转移矩阵。其中, 非零块表示部件的状态转移矩阵。

1) 不采取任何动作 ($a=0$) 时:

$$T_{\text{inv}}^0 = \begin{bmatrix} T_{\text{comp}}^0 & 0 & \dots & 0 \\ T_{\text{comp}}^{0, \text{cm}} & T_{\text{comp}}^{0, \text{ucm}} & 0 & \dots & 0 \\ 0 & T_{\text{comp}}^0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & T_{\text{comp}}^0 & 0 \end{bmatrix}$$

2) 采取订购动作 ($a=1$) 时:

$$T_{\text{inv}}^1 = \begin{bmatrix} 0 & 0 & \dots & 0 & T_{\text{comp}}^0 \\ T_{\text{comp}}^{0, \text{cm}} & T_{\text{comp}}^{0, \text{ucm}} & 0 & \dots & 0 \\ 0 & T_{\text{comp}}^0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & T_{\text{comp}}^0 & 0 \end{bmatrix}$$

3) 采取维修替换动作 ($a=2$) 时:

$$T_{\text{inv}}^2 = \begin{bmatrix} T_{\text{comp}}^0 & 0 & \dots & 0 \\ T_{\text{comp}}^1 & 0 & \dots & 0 \\ 0 & T_{\text{comp}}^0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & T_{\text{comp}}^0 & 0 \end{bmatrix}$$

不同动作下库存状态转移的差别在于前两行。对于第一行库存状态为 0 即无备件库存情

况: 当采取订购动作时, 库存状态转移到最后一列; 而采取预防性维修动作或无动作时, 由于没有库存, 库存状态保持为 0。对于第二行库存状态为 1 即有备件库存的情况: 采取订购动作或无动作时, 库存状态随事后维修而产生相应的变化, 发生事后维修则库存状态变为 0, 不发生事后维修则库存状态保持不变; 而采取预防性维修动作时, 库存状态变为 0。

最后考虑检测点因素, 此时只需要考虑检修点转移到监测点的情况。不同动作下检测点状态转移情况相同, 都是从状态 1(检修点)转移到状态 0(监测点)。此时, 得到最终的信念状态转移矩阵:

$$T_0 = \begin{bmatrix} T_{\text{inv}}^0 & 0 \\ T_{\text{inv}}^0 & 0 \end{bmatrix} \quad T_1 = \begin{bmatrix} T_{\text{inv}}^1 & 0 \\ T_{\text{inv}}^1 & 0 \end{bmatrix} \quad T_2 = \begin{bmatrix} T_{\text{inv}}^2 & 0 \\ T_{\text{inv}}^2 & 0 \end{bmatrix}$$

最终 $T(s, a, s') = T_a[i, j]$, 其中 i, j 分别表示 s, s' 在信念状态 b 中的位置序号。

2.4.2 状态观测概率

状态观测概率可以通过多种方式获得, 包括假设误差的概率分布、仿真和根据实际误差数据拟合等。本文假设监测误差服从正态分布, 由于部件的状态监测与动作无关, 故部件状态观测概率为

$$\Pr(x_e(t)=x_e|x(t)=x, a) = \varphi\left(\frac{x_e-x}{\sigma}\right)$$

式中: $\varphi(\cdot)$ 为标准正态分布的概率密度函数; x_e 为部件退化状态的观测值。

同理, 对应的离散状态观测概率为

$$\Pr(z_e(t)=j|z(t)=i, a) = \Phi\left(\frac{(j+1)L-(i+0.5)L}{\sigma}\right) - \Phi\left(\frac{jL-(i+0.5)L}{\sigma}\right)$$

所以部件的状态观测概率矩阵 $O_{\text{comp}}[i, j] = \Pr(z_e(t)=j|z(t)=i)$ 。由于状态观测矩阵与库存状态和检测点状态无关, 故系统的状态观测矩阵为

$$O = \begin{bmatrix} O_{\text{comp}} & & & & \\ & O_{\text{comp}} & & & \\ & & \ddots & & \\ & & & O_{\text{comp}} & \end{bmatrix}$$

其中, 矩阵 O 中只有对角线处为部件状态观测概率矩阵, 其余位置均为 0。系统状态观测矩阵维度与系统状态转移矩阵相同。系统状态观测函数 $O(a, s, o) = O[i, j]$, i, j 分别表示状态 s, o 在信念状态 b 中的位置序号。

3 改进的基于点的 Perseus 算法

基于点的算法在求解 POMDP 问题中被广泛

运用。首先定义 POMDP 模型下的值函数：

$$V(b) = \max_{a \in A} R_B(b, a) + \gamma \sum_{b' \in B} T(b, a, b') V(b')$$

POMDP 的值函数可以建模为有限线性函数集上的包络^[20]，并用一系列向量来表示值函数 $V = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 。因此，给定信念状态 b 下的值函数为

$$V(b) = \max_{\alpha \in V} b \cdot \alpha$$

对于固定的检修周期，POMDP 模型下的信念空间有限，故采用基于点的 Perseus 算法来求解。原始的 Perseus 算法分为两部分：首先随机探索信念空间得到一个足够大的信念点集合；然后对信念点集中的所有信念点计算最优值函数集合。

由于原始算法中选取信念点部分过于随机，

不能有效地选择有利于更新值函数的信念点，故本文根据问题的特性对原始的 Perseus 算法进行改进。将信念点按监测时刻分层，第 n 层的信念点由第 $n-1$ 层更新而来。而每层的信念点数量呈指数级增长，所以需要每层的信念点进行筛选。由于值函数的计算取决于相邻且有较高状态价值的信念点，故本文根据 QMDP 算法获得每层信念点的近似状态价值 $\hat{V}(b)$ ，然后收集 $\hat{V}(b)$ 值较高的一些信念点。收集信念点部分的算法如图 5 所示。虽然 QMDP 算法需要计算 MDP 模型下的最优策略，但相比求解 POMDP 模型，计算 MDP 模型下的最优策略所花费的时间可以忽略不计^[21]。所以，根据近似状态价值 $\hat{V}(b)$ 可以更好地选择有利于更新值函数的信念点。

```

算法 1 探索收集信念点
1: function EXPLORE( $B_0, Q$ ) ▷ 输入参数为初始信念点集合和 MDP 解的  $Q$  表
2:    $B \leftarrow B_0, B_{old} \leftarrow B_0, B_{new} \leftarrow \phi, \hat{V} \leftarrow \phi$ 
3:   repeat
4:     for all  $b \in B_{old}$  do
5:       for all  $a \in A, o \in O \& \Pr(o | b, a) > 0$  do
6:          $B_{new} \leftarrow B_{new} \cup \{b^{a,o}\}$ 
7:          $\hat{V} \leftarrow \hat{V} \cup \{\arg \max_a \sum_{s \in S} Q(s, a) b(s)\}$  ▷ 计算新增信念点的近似状态价值
8:         根据  $\hat{V}$  保留  $B_{new}$  中价值较高的信念点 ▷ 需设置合适的数量阈值
9:          $B \leftarrow B \cup B_{new}$ 
10:         $B_{old} \leftarrow B_{new}, B_{new} \leftarrow \phi, \hat{V} \leftarrow \phi$ 
11:   until 到达下一检修点 ▷ 探索两检修点之间的信念点
    
```

图 5 算法 1——探索收集信念点

Fig.5 Algorithm 1——explore to collect belief points

基于点的 Perseus 算法中，计算最优值函数集合的步骤如图 6 所示。初始化集合 V 和 Backup 操作是最为关键的两个步骤。选择合适的初始值

函数可以有效减少算法的迭代次数，一般选取尽可能接近最优值函数 V^* 的初始值。对于 Perseus 算法，集合 V 的初始值通常设置如下：

```

算法 2 计算值函数集合
1: function PERSEUSUPDATE( $B, V$ ) ▷ 输入参数为信念点集合和初始值函数集合
2:   repeat
3:      $B' \leftarrow B, V' \leftarrow \phi$ 
4:     while  $B' \neq \phi$  do
5:       随机选取  $b \in B'$ 
6:        $\alpha \leftarrow \text{BACKUP}(b, V)$ 
7:       if  $\alpha \cdot b \geq V(b)$  then
8:          $B' \leftarrow \{b \in B' : \alpha \cdot b < V(b)\}$ 
9:          $\alpha_b \leftarrow \alpha$ 
10:      else
11:         $B' \leftarrow B' - \{b\}$ 
12:         $\alpha_b \leftarrow \arg \max_{\alpha \in V} \alpha \cdot b$ 
13:       $V' \leftarrow V' \cup \{\alpha_b\}$ 
14:      $V \leftarrow V'$ 
15:   until  $V$  收敛 ▷  $V$  集合内  $\alpha$  数量恒定时收敛
    
```

图 6 算法 2——计算值函数集合

Fig.6 Algorithm 2——compute the set of value functions

$$R_{\min} = \min_{s \in S, a \in A} R(s, a)$$

$$V_0 = \{\alpha_0\} \quad \alpha_0(s) = \frac{R_{\min}}{1 - \gamma}$$

Backup 操作作为值函数更新的核心，其表达式如下：

$$\text{backup}(b, V) = \arg \max_{a^b: a \in V, a \in A} b \cdot \alpha_a^b$$

$$\alpha_a^b = r_a + \gamma \sum_{o \in \Omega} \arg \max_{a^o: a \in V} b \cdot \alpha^{a,o}$$

$$\alpha^{a,o} = \sum_{s' \in S} \alpha(s') O(a, s', o) T(s, a, s')$$

最后, 经过 Perseus 算法得到值函数集合 V 后, 最优策略可表示为

$$\pi_V(b) = \arg \max_{a \in V} b \cdot a$$

4 数值案例

本节以某钢厂中的钢铁转炉系统为研究对象^[22], 该系统功能为去除铁水中的杂质并将其转换为钢材。系统会随着生产而不断退化, 产生裂纹或磨损等缺陷。该工厂每 3 个月会进行定期的检修活动, 确保设备正常运转, 同时在此期间可以进行额外的监测, 每 2 周进行一次状态监测, 根据状态监测结果判断是否进行维修活动。监测和维修导致平均停机时间为 2 h 和 4.5 h, 相比监测周期或检修周期可忽略不计。通过 Gamma 过程对该系统的退化过程进行建模, 相关参数见表 1, 其中, 成本单位为百元。

表 1 系统及成本参数

Tab.1 System and cost parameters

| | | | | | | | |
|----------|---------|----------|----------|-------|-------|-------|-------|
| α | β | σ | γ | F | N | C_O | C_H |
| 0.4 | 1.6 | 0.3 | 0.99 | 5 | 10 | 25 | 5 |
| C_C | C_E | C_P | C_P' | T_I | T_M | T_L | |
| 90 | 300 | 30 | 50 | 6 | 1 | 2 | |

4.1 算法评估

对算法的求解效率和值函数集合的数量进行对比。为了减少算法中随机性的影响, 两种算法都重复 5 次实验, 并取平均值作为最终结果。由表 2 可以看出, 在同一离散状态数下, 改进后的 Perseus 算法相比原始的算法求解时间大大缩短, 并且最终的值函数集合的数量 $|V|$ 也比原始算法有所减少。较小的 $|V|$ 可以缩短实际使用中从值函数集合 V 中寻找最优策略所花费的时间。当离散状态数增加时, 信念空间规模变大, 改进后的 Perseus 算法相比原始算法提升效果更明显。从整体上看, 在相同空间规模下, 改进的 Perseus 算法在保证求解所得最优策略相同的同时提高了算法求解和策略使用的效率。

表 2 不同规模下算法性能对比

Tab.2 Comparison of algorithm performance at different scales

| 状态数 | 算法 | 时间/s | $ V $ | 期望成本 |
|-----|----------------|------|-------|------|
| 5 | 原始 Perseus 算法 | 65 | 18 | 6.18 |
| | 改进的 Perseus 算法 | 41 | 12 | |
| 10 | 原始 Perseus 算法 | 2492 | 38 | 6.15 |
| | 改进的 Perseus 算法 | 856 | 24 | |

4.2 对比实验

对比的实验如下: ①不考虑备件库存的维修

与订购的联合优化; ②基于观测值的维修与备件库存管理的联合优化。首先, 对不考虑备件库存的维修与订购问题进行对比实验。备件订购到货后会进行维修替换或者退货, 维修更换则产生预防性维修更换的成本, 退货则产生退货成本 $C_B = 30$ 。由于该实验也采用信念状态, 故同样使用 Perseus 算法求解。其次, 对基于观测值的维修与备件库存管理问题进行对比实验, 采用基于门限值的参数策略进行对比分析。设置备件订购门限值和预防性维修更换门限值 $[L_O, L'_O, L_P, L'_P]$, 其中, L_O 是检修点处的订购门限值, L'_O 是监测点处的订购门限值, L_P 是检修点处的预防性维修更换门限值, L'_P 是监测点处的预防性维修更换的门限值。由于部件离散状态数有限, 故此时采用穷举法寻找最优的参数化策略。

3 次实验的平均成本见表 3。可以看出, 考虑库存因素和采用信念状态的效果是最优的, 相比无备件库存的情况, 成本降低了 9.25%; 相比基于观测值的情况, 成本降低了 25.63%。这表明考虑检修与备件库存联合优化是必要的, 它能有效地降低长期期望成本。并且信念状态更加合理地表示系统的状态信息, 所以得到了更好的策略。

表 3 不同实验下的成本与策略

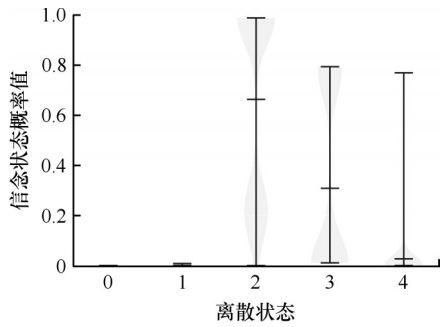
Tab.3 Costs and strategies under different experiments

| 实验 | 长期期望成本 | 参数化策略 |
|--------|--------|--------------|
| 本文实验 | 6.18 | / |
| 对比实验 1 | 6.81 | / |
| 对比实验 2 | 8.31 | [1, 4, 3, 4] |

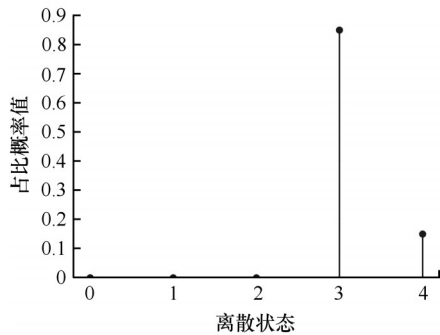
4.3 策略分析

基于点的 Perseus 算法得到的订购与维修策略如图 7 和图 8 所示, 图 7 表示有备件库存下的订购与维修策略, 图 8 表示无备件库存下的维修与订购的联合策略。在图 7 和图 8 的基础上分析有无库存下策略的差异, 并讨论检修与备件库存联合优化的合理性。本文将策略分成两种情况分析: 检修点和监测点, 检修点处统计订购或维修的次数, 并计算不同状态下的占比情况; 监测点处统计订购或维修的信念状态。

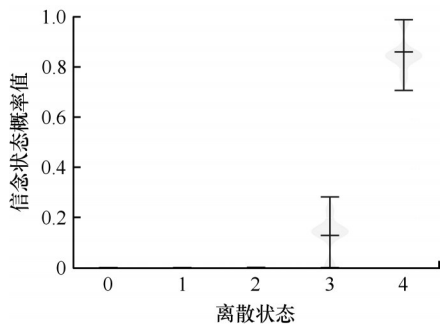
首先, 对于有备件库存下的订购策略, 由图 7a 可以看出, 监测点处智能体采取订购动作时信念状态集中于状态 2 和 3。图中显示状态 2 和 3 的分布集中于上下两端, 表示状态 2 可能性高时其他状态可能性低, 状态 3 可能性高时其他状态可能性低, 这是符合实际概率分布的。状态 4 的可能性大部分都较低, 但也有可能性高的情况, 这表示部件状态突然变差并进行订购备件的小概率情



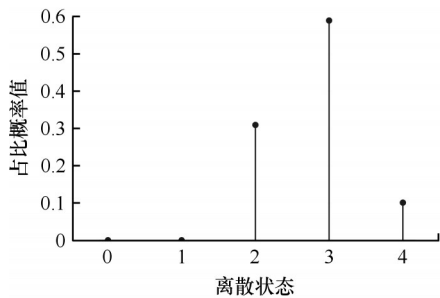
(a)监测点订购时的信念状态



(b)检修点时不同状态下订购的占比情况



(c)监测点维修时的信念状态



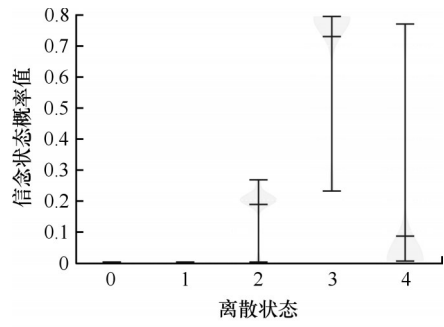
(d)检修点时不同状态下维修的占比情况

图7 有备件库存时的订购和维修策略

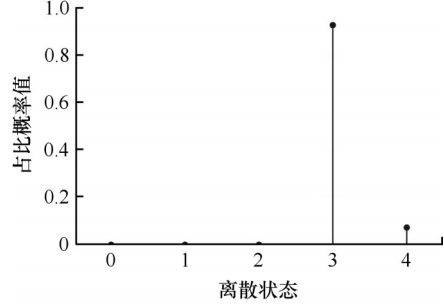
Fig.7 Ordering and maintenance strategy when spare parts are in stock

况。图7b显示检修点处智能体采取订购动作时部件状态大部分位于状态3,小部分位于状态4。由于监测处获得部件状态信息有误差,故在监测点订购时的部件状态相比检修点略低,策略也略微保守。而与参数化策略相比,参数化策略在监测点处的订购更加保守,在状态1处就采取订购动作。

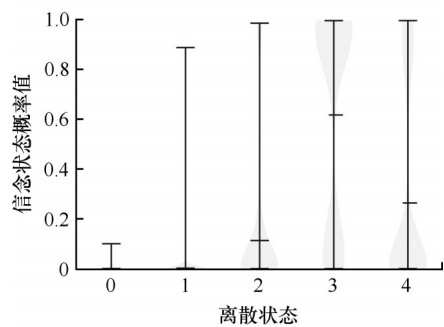
其次,对于维修策略,由图7c可以看出,监测



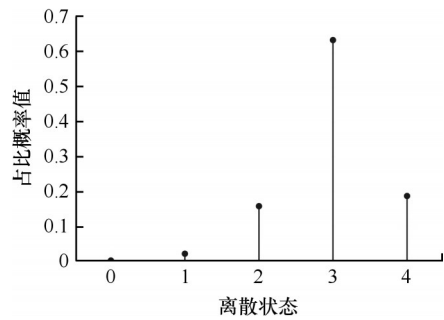
(a)监测点订购时的信念状态



(b)检修点时不同状态下订购的占比情况



(c)监测点维修时的信念状态



(d)检修点时不同状态下维修的占比情况

图8 无备件库存时的订购和维修策略

Fig.8 Ordering and maintenance strategy when no spare parts are in stock

点处智能体采取预防性维修更换动作的信念状态集中于状态3和4,并且状态4的可能性较高,这是由于监测点处预防性维修替换成本较高,部件退化状态接近故障阈值时才进行预防性维修更换。图7d显示检修点处智能体采取预防性维修更换动作时的状态大部分位于状态2和3,小部分位于状态4。很显然这是因为检修点处维修更换成本较低,所以更早地进行预防性维修更换,而监测点处

维修成本较高,需要等待状态接近故障阈值时才进行维修更换,降低了检修间隔内系统故障的风险。此时最优的维修策略和参数化策略相似。

图 8 所示为无备件库存下的策略,它和图 7 所示的有备件库存下的策略在检修点处较为相似,但在监测点处有所不同。由图 8a 和图 8c 可以看出,无备件库存时的订购相比有备件库存时有所滞后,而维修则有所提前,这符合订购到货后立即维修的特点。由此看出,有备件库存不仅增加了原本部件的使用时长,还可以起到预防性维修的作用,降低了检修间隔内系统故障的风险。

对有无库存下策略进行分析(图 7 和图 8)后可以发现,考虑维修与备件库存的联合优化得到了更加合理的策略。在联合优化下,监测点处可以更早期地订购备件以提供潜在维修更换的需求,降低维修时无备件的风险,同时更晚地进行预防

性维修以提高部件的使用时长。检修与备件库存的联合优化综合影响了订购和维修策略,使得策略更加合理,进而降低长期期望成本。

最后对有备件库存下仿真过程中的系统状态与智能体采取的动作进行分析。使用最优策略仿真 60 个时间步,该过程中系统状态和智能体所采取的动作如图 9 所示,从上往下依次表示智能体在每个决策点采取的动作、库存状态、部件的观测状态与真实状态、部件的信念状态。其中,部件状态图中深蓝色(1)表示部件的真实状态,深红色(-1)表示在监测点监测到的有误差的状态。部件信念状态图表示部件当前时刻的信念状态,虚线部分表示检修点处的准确状态。由图 9 可以看出,在监测点订购时信念状态大部分集中于状态 2,在监测点维修时信念状态集中于状态 3 和 4,与图 7 的分析是吻合的。

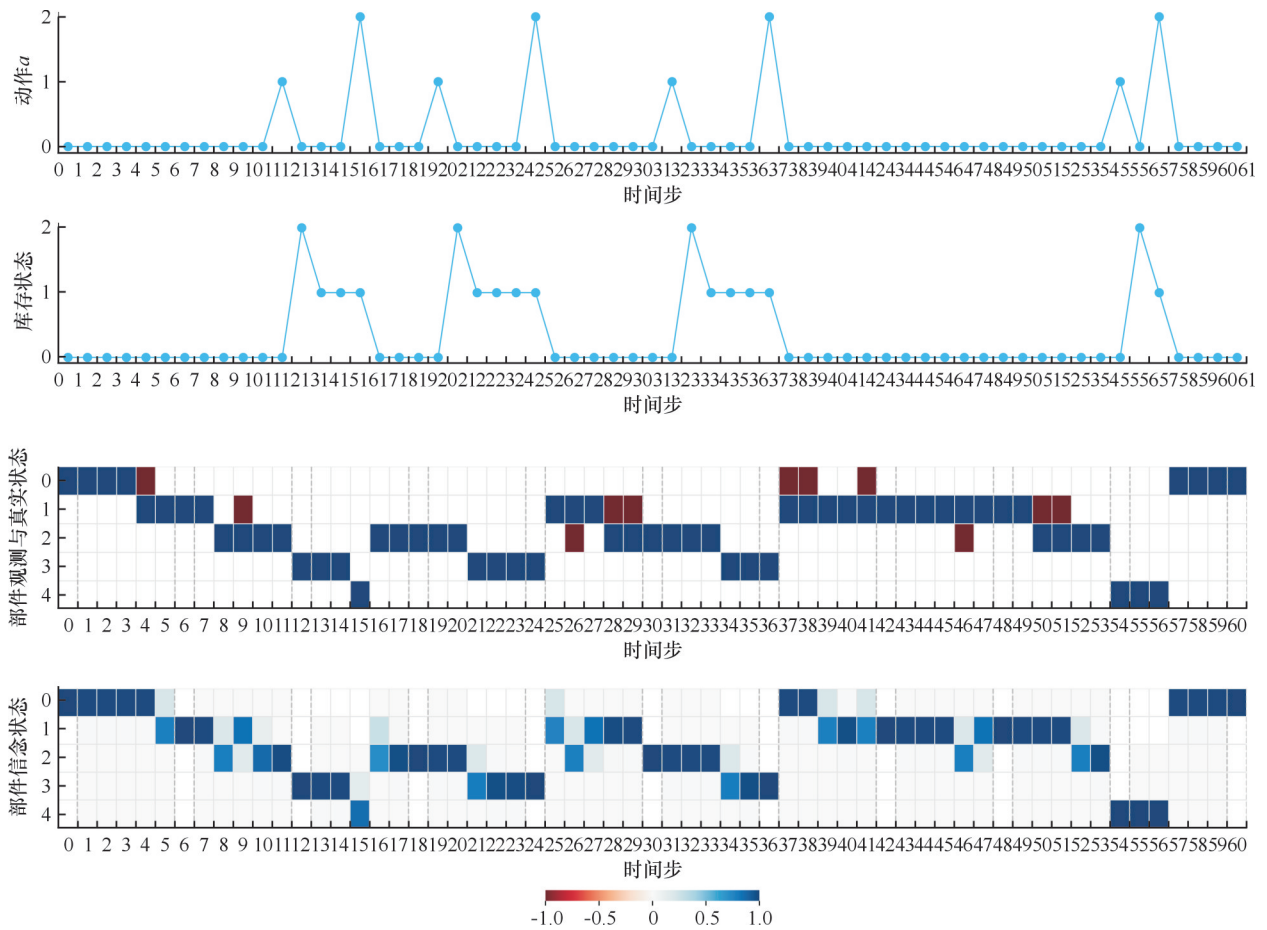


图 9 仿真过程中状态和动作分析

Fig.9 Analyzing states and actions during simulation

4.4 敏感性分析

用敏感性分析相关参数对长期期望成本的影响。选择以下关键参数进行分析:监测点处预防性维修成本 C'_p , 备件运输和存储成本 C_H , 检修周期 T_1 , 备件订购提前期 T_L 和离散状态数 N 。

根据不同参数下的长期期望成本(表 4)对不同参数进行分析。随着监测点处的预防性维修更换成本的降低,长期期望成本也随之下落,同时增加了监测点处预防性维修更换的机会,降低部件发生故障的风险。运输和库存成本与长期期望成

表4 不同参数下的期望成本

Tab.4 Expected costs under different parameters

| 参数 | C'_P | | | C_H | | | T_1 | | | T_L | | | N | | |
|----|--------|------|------|-------|------|------|-------|------|------|-------|------|------|------|------|------|
| | 30 | 40 | 50 | 0 | 5 | 10 | 4 | 6 | 8 | 1 | 2 | 3 | 3 | 5 | 10 |
| 成本 | 5.92 | 6.05 | 6.18 | 4.23 | 6.18 | 6.95 | 6.05 | 6.18 | 6.31 | 5.57 | 6.18 | 6.71 | 6.95 | 6.18 | 6.15 |

本正相关,当库存成本为0时,可随意储存备件,使得长期期望成本大大减少。检修间隔的大小也会影响长期期望成本,当检修间隔减小时,监测点处的预防性维修更换减少,进而降低长期期望成本。备件订购提前期减少,使得备件订购更快响应,不仅减少运输成本,也降低运输期间发生故障的风险。最后,对于离散状态数 N 的取值情况, N 越大,计算所得的策略越准确,长期期望成本越小,但过大的 N 不仅会导致信念空间维度爆炸,算法的效率大大减少,并且在实际应用中也难以划分过多的离散状态。当 $N=10$ 时,相比 $N=5$,长期期望成本提升微小,但计算时间大大增加,从41 s增加到850 s。

5 结语

本文考虑在不完美监测下单部件系统的视情维修与备件库存管理的联合优化问题,并以钢厂中的钢铁转炉系统为对象进行数值案例分析。该系统具有固定检修周期的特性,并在检修周期内可进行不完美的状态监测,在检修点和监测点同时进行备件库存盘点。部件退化过程建模为Gamma过程,并推导考虑检测点状态、库存状态和部件状态的系统状态转移概率矩阵,系统建模为POMDP模型,采用QMDP算法改进基于点的Perseus算法以求解近似最优策略。在数值案例中验证了改进Perseus算法的性能,分析信念状态相比观测值能更加合理地表示状态信息,以得到更好的策略,并且发现备件库存可以增加部件使用时间,以及降低检修间隔内故障的风险。

未来的研究可以有以下拓展方向:①将单部件系统推广到多部件系统^[23],研究基于不完美监测下的多部件系统的维修与库存联合优化问题;②针对上述问题,设计开发高状态维度下的POMDP模型以更好地求解算法。

参考文献:

- [1] 徐兆平,郭波.复杂装备故障预测方法研究综述[J].长沙理工大学学报(自然科学版),2023,20(2):10-26.
XU Zhaoping, GUO Bo. A Research Review on Fault Prognostic Techniques for Complex Equip-

ments[J]. Journal of Changsha University of Science & Technology (Natural Science), 2023, 20(2): 10-26.

- [2] LI H, ZHOU Y. Joint Optimization of Condition-based Maintenance and Spare Parts Orders for a Two-unit System[C]//2022 Global Reliability and Prognostics and Health Management. Yantai, 2022: 1-6.
- [3] WANG Jingjing, QIU Qingan, WANG Huanhuan. Joint Optimization of Condition-based and Age-based Replacement Policy and Inventory Policy for a Two-unit Series System[J]. Reliability Engineering & System Safety, 2021, 205: 107251.
- [4] ZHENG Meimei, LIN Jie, XIA Tangbin, et al. Joint Condition-based Maintenance and Spare Provisioning Policy for a K-out-of-N System with Failures during Inspection Intervals[J]. European Journal of Operational Research, 2023, 308(3): 1220-1232.
- [5] TCHAKOUA P, WAMKEUE R, TAMEGHE T A, et al. A Review of Concepts and Methods for Wind Turbines Condition Monitoring [C] //2013 World Congress on Computer and Information Technology (WCCIT). Sousse, Tunisia, 2013: 1-9.
- [6] LI Yanrong, PENG Shizhe, LI Yanting, et al. A Review of Condition-based Maintenance: Its Prognostic and Operational Aspects[J]. Frontiers of Engineering Management, 2020, 7(3): 323-334.
- [7] 张红旗,邵晓东,胡祥涛.基于部分可观察马尔可夫决策过程的机电装备动态可靠性评价方法[J].中国机械工程,2016,27(18):2482-2486.
ZHANG Hongqi, SHAO Xiaodong, HU Xiangtao. Dynamic Reliability Assessment Method Based on POMDP for Electromechanical Equipment[J]. China Mechanical Engineering, 2016, 27(18): 2482-2486.
- [8] 刘大玲,黄小钢.高速铁路无砟轨道系统状态监测及预防性维修[J].中国机械工程,2019,30(3): 349-353.
LIU Daling, HUANG Xiaogang. Condition Monitoring and Preventive Maintenance of Ballastless Track Systems for High-speed Railways[J]. China Mechanical Engineering, 2019, 30(3): 349-353.
- [9] ROUX M, FANG Y P, BARROS A. Maintenance Planning under Imperfect Monitoring: an Efficient

- POMDP Model Using Interpolated Value Function [J]. IFAC-PapersOnLine, 2022, 55(16): 128-135.
- [10] MORATO P G, NIELSEN J S, MAI A Q, et al. POMDP Based Maintenance Optimization of Off-shore Wind Substructures Including Monitoring[C]// 13th International Conference on Applications of Statistics and Probability in Civil Engineering. Seoul, 2019:1-8.
- [11] NGUYEN K T, DO P, HUYNH K T, et al. Joint Optimization of Monitoring Quality and Replacement Decisions in Condition-based Maintenance[J]. Reliability Engineering & System Safety, 2019, 189: 177-195.
- [12] ZHAO Fei, LIU Xuejuan, PENG Rui, et al. Joint Optimization of Inspection and Spare Ordering Policy with Multi-level Defect Information[J]. Computers & Industrial Engineering, 2020, 139: 106205.
- [13] 张奥. 考虑状态演变过程的高速铁路牵引供电设备维修策略[D]. 成都: 西南交通大学, 2017.
ZHANG Ao. High Speed Railway Traction Power Supply Equipment Maintenance Strategy Considering State Evolution Process[D]. Chengdu: Southwest Jiaotong University, 2017.
- [14] 张龙, 黄文艺, 熊国良, 等. 基于多域特征与高斯混合模型的滚动轴承性能退化评估[J]. 中国机械工程, 2014, 25(22): 3066-3072.
ZHANG Long, HUANG Wenyi, XIONG Guoliang, et al. Assessment of Rolling Bearing Performance Degradation Using Gauss Mixture Model and Multi-domain Features[J]. China Mechanical Engineering, 2014, 25(22): 3066-3072.
- [15] van NOORTWIJK J M. A Survey of the Application of Gamma Processes in Maintenance[J]. Reliability Engineering & System Safety, 2009, 94(1): 2-21.
- [16] NEWBY M J, BARKER C T. A Bivariate Process Model for Maintenance and Inspection Planning[J]. International Journal of Pressure Vessels and Piping, 2006, 83(4): 270-275.
- [17] SUTTON R S, BARTO A G. Reinforcement Learning: an Introduction [M]. Cambridge: The MIT Press, 2018.
- [18] BRODY D C, HUGHSTON L P, MACRINA A. Dam Rain and Cumulative Gain[J]. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2008, 464(2095): 1801-1822.
- [19] PARK C, PADGETT W J. Accelerated Degradation Models for Failure Based on Geometric Brownian Motion and Gamma Processes[J]. Lifetime Data Analysis, 2005, 11(4): 511-527.
- [20] SHANI G, PINEAU J, KAPLOW R. A Survey of Point-based POMDP Solvers [J]. Autonomous Agents and Multi-Agent Systems, 2013, 27(1): 1-51.
- [21] VIRIN Y, SHANI G, SHIMONY S E, et al. Scaling Up: Solving POMDPs through Value Based Clustering[C]// AAAI Conference on Artificial Intelligence. Vancouver, 2007:1290-1295.
- [22] YANG Li, YE Zhisheng, LEE C G, et al. A Two-phase Preventive Maintenance Policy Considering Imperfect Repair and Postponed Replacement [J]. European Journal of Operational Research, 2019, 274(3): 966-977.
- [23] 周一帆, 郭凯, 李帮诚. 基于多智能体强化学习的多部件系统维修优化[J]. 长沙理工大学学报(自然科学版), 2023, 20(2): 27-34.
ZHOU Yifan, GUO Kai, LI Bangcheng. Maintenance Optimization of Multi-component System Based on Multi-agent Reinforcement Learning [J]. Journal of Changsha University of Science & Technology (Natural Science), 2023, 20(2): 27-34.

(编辑 陈 勇)

作者简介:汪 凯,男,2000年生,硕士研究生。研究方向维修优化等。周一帆(通信作者),男,1981年生,教授、博士研究生导师。研究方向为可靠性、维修优化、库存优化等。E-mail: yifan.zhou@seu.edu.cn。

本文引用格式:

汪凯,顾刘栋,周一帆. 基于 POMDP 模型的检修与备件库存联合优化[J]. 中国机械工程,2025,36(8):1853-1863.

WANG Kai, GU Liudong, ZHOU Yifan. Joint Optimization of Inspection, Maintenance and Spare Parts Inventory Based on POMDP Model[J]. China Mechanical Engineering, 2025, 36(8): 1853-1863.