

引用格式:李光涛,刘鲁齐,时鹏,等.考虑非齐次退化与资金折现的桥梁维修决策方法[J].工业建筑,2026,56(5):187-200. LI Guangtao, LIU Luqi, SHI Peng, et al. A Bridge Maintenance Decision-Making Method Considering Non-Homogeneous Deterioration and Cost Discounting[J]. Industrial Construction, 2026, 56(5): 187-200 (in Chinese). DOI:10.3724/j. gyjzG26031401

考虑非齐次退化与资金折现的桥梁维修决策方法*

李光涛¹ 刘鲁齐² 时鹏² 兰成明¹

(1. 北京科技大学未来城市学院城镇化与城市安全研究院,北京 100083;2. 北京科技大学国家材料服役安全科学中心,北京 102206)

摘 要 面向桥梁智慧运维中的长期维修决策优化需求,针对有限时域条件下桥梁状态退化具有时变性、维修成本存在资金折现且长期回报传递困难等特点,构建了一种考虑非齐次退化与折现效应的全生命周期维修决策模型。桥梁退化过程采用非齐次马尔科夫状态转移进行表征,以离散健康状态和维修动作为基础,将维护成本与风险成本统一纳入成本函数,并在决策过程中引入现金流折现。为克服常规强化学习方法难以有效处理有限时域阶段性决策需求、非齐次状态转移以及训练不稳定等问题,提出一种结合状态增广与反向课程学习的强化学习求解框架。其中,通过扩展状态表征增强策略对有限时域特征的识别能力;通过设置反向课程逐步扩展训练区间,提高策略学习的稳定性与收敛效率。算例结果表明,所提出方法能够有效适应桥梁非齐次退化下的有限时域维修决策问题,在策略性能及训练稳定性方面均表现出较好效果,可为桥梁全生命周期运维管理中的维修方案制定提供方法支撑。

关键词 桥梁智慧运维;维修决策;非齐次马尔科夫过程;强化学习;反向课程学习

A Bridge Maintenance Decision-Making Method Considering Non-Homogeneous Deterioration and Cost Discounting

LI Guangtao¹ LIU Luqi² SHI Peng² LAN Chengming¹

(1. Research Institute of Urbanization and Urban Safety, School of Future Cities, University of Science and Technology Beijing, Beijing 100083, China;2. National Center for Materials Service Safety, University of Science and Technology Beijing, Beijing 102206, China)

Abstract: To meet the demand for optimizing long-term maintenance decisions in intelligent bridge operation and maintenance, considering the characteristics of time-varying bridge deterioration under a finite horizon, financial discounting of maintenance costs, and difficulty in long-term reward propagation, a life-cycle maintenance decision model incorporating non-homogeneous deterioration and discounting effects was developed. The bridge deterioration process was characterized by non-homogeneous Markov state transitions. Based on discrete health states and maintenance actions, maintenance costs and risk costs were integrated into a unified cost function, while cash-flow discounting was introduced into the decision-making process. To address the limitations of conventional reinforcement learning methods in handling finite-horizon stage-wise decision tasks, non-homogeneous state transitions, and unstable training, a reinforcement learning framework combining state augmentation and backward curriculum learning was proposed. Expanded state representation enhanced the policy's capability to capture finite-horizon characteristics, while backward curriculum learning gradually extended the training

*国家重点研发计划(2023YFC3805700)。

第一作者:李光涛,博士研究生,主要从事桥梁智慧运维研究。

通信作者:兰成明,教授,主要从事桥梁智慧运维研究,lanchengming@ustb.edu.cn。

收稿日期:2026-03-14

interval to improve learning stability and convergence efficiency. Numerical results demonstrated that the proposed method effectively adapted to finite-horizon maintenance decision problems under non-homogeneous bridge deterioration and achieved favorable performance in both policy quality and training stability, thereby providing methodological support for maintenance planning in life-cycle bridge operation and maintenance management.

Keywords: intelligent bridge operation and maintenance; maintenance decision-making; non-homogeneous Markov process; reinforcement learning; backward curriculum learning

随着大量在役桥梁服役年限持续增长,结构材料劣化、环境侵蚀及重复荷载作用所引起的性能退化问题日益突出,桥梁运维管理已由单一缺陷处治逐步转向面向全寿命周期的系统性决策优化^[1-2]。在桥梁智慧运维背景下,如何依据结构状态演化规律,统筹维修投入、服役风险与长期养护效益,制定合理的阶段性维修策略,已成为保障桥梁服役安全、提升养护资源配置效率的重要问题^[3]。与一次性静态优化不同,桥梁维修管理具有显著的时序性和累积性特征^[4]。一方面,结构状态随服役时间不断退化,不同阶段的维修干预将对后续性能演化产生持续影响^[1];另一方面,维修资金投入、风险控制要求及时间价值共同作用,使得维修方案选择呈现出多阶段耦合的决策特征^[5]。因此,桥梁长期维修策略优化本质上是一个面向有限使用年限、兼顾退化演化与经济约束的序贯决策问题,亟需构建能够反映状态时变特征与运维经济性的分析方法。

现有桥梁维修决策研究已形成以随机退化建模和序贯决策优化为主的基本技术路线^[6]。其中,马尔可夫链及马尔可夫决策过程因能够描述桥梁状态在不同维修干预下的随机演化规律,较早被应用于桥梁养护管理与维修策略优化^[7]。相关研究表明,该类方法能够将结构状态预测、维修动作选择与全寿命周期成本控制纳入统一分析框架,为桥梁管理系统中的养护决策提供定量依据^[8]。在此基础上,部分研究进一步考虑了历史依赖、有限设计使用年限以及系统层级优化等因素,例如通过历史相关退化建模改进传统无记忆假设,或基于动态规划求解有限时域内的维修与重建策略,从而提升模型对实际桥梁服役过程的适应性^[9]。随着基础设施运维问题的状态维数和动作空间不断扩大,强化学习方法也逐步被引入桥梁及交通基础设施维修规划中,用于在复杂状态与动作映射关系下搜索近优策略^[10];已有研究

显示,深度强化学习、分层强化学习等方法在多构件桥梁和大规模设施维护场景中具有较好的扩展潜力^[11]。现有研究已形成由随机退化建模、动态规划和强化学习等方法构成的技术路径。

尽管现有桥梁维修决策研究在退化建模与策略优化方面已取得较多进展,但面向全寿命周期运维场景,仍存在若干关键问题有待进一步解决。首先,已有不少研究在建模时默认桥梁退化过程满足齐次马尔可夫假设^[12],即状态转移概率在不同时期保持不变,而实际桥梁构件在环境侵蚀、疲劳累积及材料老化等因素作用下,往往表现出随服役时间增长而加速退化的特征^[13],采用齐次转移描述难以充分反映长期服役条件下的时变退化规律。其次,在桥梁维修决策问题中,强化学习的规划期通常具有明确上限,具体表现形式为设计使用年限。不同剩余服役时间下的最优干预策略存在显著差异,因而该问题本质上属于有限时域序贯决策^[6]。然而,常规强化学习方法多面向平稳环境或无限时域回报近似设计,难以直接刻画阶段性决策需求对策略结构的影响。当目标函数同时纳入维护成本、风险成本及资金折现后,决策收益与代价在长时间尺度上的传递关系更加复杂,易导致训练过程中的长期信用分配困难,使策略学习出现收敛缓慢、稳定性不足等问题。因此,传统齐次退化建模与常规强化学习框架难以同时覆盖退化时变性、有限时域边界和运维经济性优化,需要针对问题结构重新设计建模与训练机制。

针对上述问题,本文以单桥系统下的构件级年度维修决策为研究对象,目标是在给定有限使用年限内,依据构件状态演化规律,在离散维修动作集合中确定各决策时刻的构件维修方案,以最小化全寿命周期累计总成本。为此,构建了考虑非齐次退化与资金折现的桥梁维修决策模型,并针对有限时域和时变退化条件下常规强化学习方法难以稳定求解的问题,提出了结合状态增广与

反向课程学习近端策略优化算法(SA-BCLPPO)。

1 桥梁非齐次马尔可夫维修决策模型

1.1 桥梁维修决策问题描述

桥梁维修决策问题是指在给定设计使用年限内,根据桥梁结构状态的演化情况,对不同构件在各决策时刻所采取的维修措施进行优化安排。由于构件退化具有随机性,且不同维修措施会对当前状态及后续退化路径产生持续影响,因此桥梁维修并不是孤立时刻上的单次选择问题,而是一个贯穿整个使用周期的连续决策过程。在这一过程中,每一时刻的维修行为既决定当期投入,又会影影响未来状态转移和后续成本累积,因而需要从全寿命周期角度对维修方案进行统一建模与分析。

本文研究对象来自加拿大魁北克省交通运输厅(MTQ)桥梁管理系统数据库中的典型梁桥桥面系统案例^[14-15]。该数据库是魁北克省公路结构管理综合系统的一部分。统计中考虑了环境和结构材料,包含截至2000年9678座省属桥梁的信息^[14]。并采用贝叶斯法则计算了中等环境下桥面系统构件的状态转移概率矩阵,相应的基准状态见表1^[15]。此外,其桥梁的维护在构件级别处理,因此状态和操作都在构件级别中考虑。本文桥面系统层面考虑由桥面铺装层、排水系统、两侧外侧面和中部及两端部的桥面板等共7个构件组成,如图1所示。

表1 各构件基准状态退化矩阵的主对角元素
Table 1 Main diagonal elements of the baseline deterioration matrix for each bridge component

构件	p_{00}	p_{11}	p_{22}	p_{33}	p_{44}	p_{55}
桥面铺装层	0.81	0.90	0.91	0.94	0.99	1
排水系统	0.91	0.90	0.86	0.94	0.93	1
外侧面1	0.86	0.97	0.94	0.89	0.99	1
外侧面2	0.86	0.96	0.95	0.92	0.91	1
端部桥面板1	0.87	0.95	0.94	0.92	0.99	1
端部桥面板2	0.87	0.96	0.94	0.94	0.99	1
中部桥面板	0.87	0.96	0.94	0.91	0.99	1

为便于建立桥面系统构件级维修决策模型,本文以典型梁桥的桥面系统为对象。在模型中,为控制状态空间规模并突出主要退化与维修特征,假定各构件在状态演化及维修效果上相互独立,暂不考虑构件间的耦合关系。在此基础上,系统状态由7个构件的组合状态表征,维修方案在构件层进行决策。桥梁维修过程具有典型的序贯决策特征^[16],

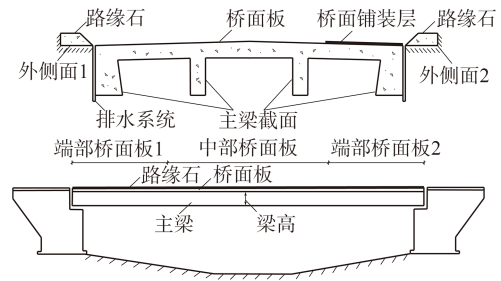


图1 桥梁横断面与纵断面示意

Fig. 1 Schematic diagram of bridge transverse and longitudinal sections

即桥梁在各离散决策时刻处于某一状态,决策者根据当前状态选择相应维修措施,而维修措施将通过影响后续状态演化进一步作用于未来成本。由于这一过程同时包含状态演化、动作干预和随机转移等要素,因此可采用式(1)对马尔可夫决策过程进行统一描述:

$$M = (S, A, P, R) \quad (1)$$

式中: S 为桥梁系统状态空间各构件状态,采用6级离散状态表示,记为 $\{0, 1, 2, 3, 4, 5\}$,其中状态值越大,表示构件劣化程度越高,0表示构件状态最好,5表示构件状态最差; A 为维修动作空间,单构件动作划分为5类,分别为不维修、小修、中修、大修和重建,桥梁系统动作则由7个构件动作组合形成联合动作; P 为状态转移概率矩阵,用于描述在给定当前状态和维修动作条件下,桥梁系统转移至下一时刻状态的概率; R 为奖励函数,用于衡量各时刻决策结果的优劣。

1.2 桥梁非齐次马尔可夫退化与状态转移建模

已有研究基于魁北克省交通运输厅桥梁检测数据,采用贝叶斯方法计算了中等环境条件下桥面系统构件的状态转移概率矩阵。矩阵的对角元素 p_{ii} 表示维持在状态 i 的概率,上一条对角线的元素为 $1-p_{ii}$ 表示在一年内由 i 退化到 $i+1$ 的概率。转移矩阵能够描述构件在统计意义下的平均状态演化规律,但尚未进一步显式刻画退化概率随服役时间变化的时变特征。为放宽传统齐次马尔可夫模型中状态转移概率在全寿命期内保持不变的假设,本文在既有基准退化矩阵基础上引入与构件服役时间相关的时变参数化形式。已有桥梁退化研究表明,现有桥梁管理系统中常用的状态型马尔可夫模型通常采用时间齐次假设,而这一假设难以充分反映桥梁构件在长期服役条件下随服役时间发展的时变退化特征^[17-19]。同时,桥梁状态或条件评分随

服役时间变化的经验退化曲线常可采用多项式函数进行近似描述^[18,20]。本文基于构件服役时间的二次经验函数构造了服役时间的衰减函数 $d(t)$ 以表征主对角元素 p_{ii} 随服役时间的变化,用以反映状态保持概率随服役时间增长而逐步衰减的特征,即式(2)。

$$p_{ii}(t) = p_{ii}(0)d(t) \quad (2)$$

式中: $p_{ii}(0)$ 为基准退化矩阵中状态 i 的初始保持概率; $d(t)$ 为随服役时间 t 的衰减函数,由式(3)确定。

$$d(t) = \max(d_{\min}, 1 - \alpha t - \beta t^2) \quad (3)$$

式中: $d_{\min}$ 为预设下限,用于避免在长服役时间退化下转移矩阵的变化过大而脱离实际退化情况,本文取 $d_{\min} = 0.5$; α 为对应构件在服役时间增长过程中由环境作用和长期服役引起的渐进性累积退化效应

参数; β 为刻画疲劳累积、材料老化、抗力衰减及其耦合作用所导致的加速损伤特征参数,使健康状态保持概率在寿命后期下降更快。各构件基准退化矩阵仅允许状态保持或向相邻劣化状态转移,其非对角元素由行概率守恒确定,即 $p_{i,i+1}(t) = 1 - p_{ii}(t)$ 。构件的基准状态退化矩阵的主对角元素见表1^[15],表中使用序号 $k=1\sim 7$ 对应各构件。

本文同时考虑小修、中修、大修和重建4类维修动作。对构件 k ,不同维修动作对应不同的状态转移矩阵,不维修对应的状态转移矩阵为自然退化的状态转移矩阵,其余维修动作对应的状态转移矩阵见表2^[15]。需要说明的是,模型方法本身并不依赖于该组具体数值,针对不同桥梁对象时可依据实际检测与维修数据重新标定。

表2 维修动作对应的状态转移矩阵
Table 2 State transition matrices under maintenance actions

状态	动作1小修状态转移矩阵						动作2中修的状态转移矩阵						动作3大修的状态转移矩阵						动作4重建的状态转移矩阵					
	0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3	4	5
0	0.99	0.01	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
1	0.03	0.95	0.02	0.00	0.00	0.00	0.50	0.45	0.05	0.00	0.00	0.00	0.90	0.10	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.03	0.95	0.02	0.00	0.00	0.00	0.50	0.45	0.05	0.00	0.00	0.70	0.20	0.10	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.03	0.95	0.02	0.00	0.00	0.00	0.50	0.45	0.05	0.00	0.50	0.30	0.20	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.03	0.95	0.02	0.00	0.00	0.00	0.50	0.45	0.05	0.40	0.30	0.20	0.10	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.90	0.10	0.40	0.30	0.20	0.10	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00

1.3 优化目标

构件层分别考虑维护成本和风险成本,其中维护成本与维修动作及构件状态相关,风险成本与构件当前状态相关。在时刻 t ,构件 k 的状态为 s_t^k ,采取的行动为 a_t^k ,其维护成本记为 $c_{t,\text{maint}}^k$,风险成本记为 $c_{t,\text{risk}}^k$,构件在时刻 t 的总成本为:

$$c_t^k = c_{t,\text{maint}}^k + c_{t,\text{risk}}^k \quad (4)$$

维护成本由构件基准成本、维修动作系数和状态系数共同确定,如式(5a)所示。

$$c_{t,\text{maint}}^k = C_k \cdot r_{\text{action}}(a_t^k) \cdot r_{\text{cond}}(s_t^k) \quad (5a)$$

构件的风险成本由构件基准成本和风险系数确定,如式(5b)所示。

$$c_{t,\text{risk}}^k = C_k \cdot r_{\text{risk}}(s_t^k) \quad (5b)$$

式中: C_k 为构件 k 的基准成本; $r_{\text{action}}(a_t^k)$ 为维修动作 a_t^k 对应的成本系数; $r_{\text{cond}}(s_t^k)$ 为状态 s_t^k 对应的维护成本调整系数; $r_{\text{risk}}(s_t^k)$ 为状态 s_t^k 对应的风险系数。相关参数取值见表3。

为综合反映维修支出与结构状态风险对决策的共同影响,本文将桥梁层的评价量表示为维护成

表3 维修成本与风险成本相关参数取值

Table 3 Parameter values related to maintenance and risk costs

符号	对应对象	取值
C_k	$k = 1, \dots, 7$	[100, 200, 400, 150, 300, 600, 800]
$r_{\text{action}}(a_t^k)$	$a = 0, \dots, 4$	[0.000, 0.003, 0.020, 0.060, 1.000]
$r_{\text{cond}}(s_t^k)$	$s = 0, \dots, 5$	[0.80, 0.85, 0.90, 0.95, 1.00, 1.00]
$r_{\text{risk}}(s_t^k)$	$s = 0, \dots, 5$	[0.0000, 0.0000, 0.0045, 0.0150, 0.0400, 0.1000]

本项与风险项之和。需要说明的是,二者在含义上并不完全相同;维护成本表示维修动作在当前决策时点产生的直接资金支出;风险项用于表征构件处于不同状态下的相对风险水平,作为桥梁维修决策中的风险代理量参与奖励和成本构造,而非严格意义上的财务损失现值。因此,桥梁总成本采用各构件评价量直接累加的方式表示,如式(6)所示。

$$C_t = \sum_{k=1}^7 c_t^k \quad (6)$$

考虑资金时间价值,本文仅对维护成本进行折现处理,风险成本保持原值。这是因为维护成本对

应维修活动在决策时刻发生的直接经济支出,具有明确的货币时间价值;而风险项主要用于表征结构状态风险对策略选择的影响,其本质上是风险量而非未来现金流损失的现值表达,因此未采用与维护支出相同的折现处理。设折现率为 r_d ,则构件 k 在时刻 t 的折现维护成本如式(7)所示^[21-22]。

$$c_{t, \text{maint}, \text{disc}}^k = \frac{c_{t, \text{maint}}^k}{(1 + r_d)^t} \quad (7)$$

桥梁系统在时刻 t 的折现总成本可表示为式(8)所示。

$$C_t^{\text{disc}} = \sum_{k=1}^7 (c_{t, \text{maint}, \text{disc}}^k + c_{t, \text{risk}}^k) \quad (8)$$

在给定桥梁构件的设计使用年限 T 内,本文的优化目标定义为最小化累计综合成本,如式(9)所示。

$$C_{t, \text{min}}^{\text{disc}} = \min \sum_{t=0}^{T-1} C_t^{\text{disc}} \quad (9)$$

该综合成本是面向维修决策优化构造的评价指标,其中维护成本反映现实支出,风险项反映状态风险偏好,二者共同服务于策略学习。

2 基于状态增广与反向课程学习的强化学习算法

强化学习是一类通过智能体与环境持续交互实现策略优化的序贯决策方法,如图2所示。其目标是在给定状态下学习动作选择策略,使累计回报达到最优。

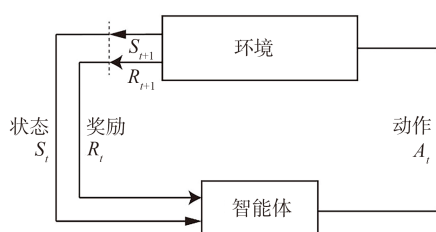


图2 强化学习示意

Fig. 2 Schematic diagram of reinforcement learning

对于桥梁维修决策问题,桥梁在各决策时刻的结构状态可视为环境状态,维修措施可视为动作选择,维修后的状态演化及其对应成本构成环境反馈,策略学习的目标则是在设计使用年限内获得更优的维修决策序列。对于本文所研究的问题,常规强化学习方法难以直接学习到有效策略,原因主要来自设计使用年限决定的有限规划期和结构的非齐次退化两个问题。对有限时域决策问题而言,最优策略不仅依赖于当前结构状态,还依赖于构件

剩余服役时间。设有限时域下时刻 t 的价值函数为 $V_t(s)$,则其递推形式为:

$$V_t(s) = \max \left[r_t(s, a) + \sum_{s'} P_t(s'|s, a) V_{t+1}(s') \right] \quad (10)$$

式(10)在设计使用年限 T 结束时应满足 $V_T(s) = 0$,若训练过程中未显式体现这一终端条件,则策略学习会隐含地假定桥梁在设计使用年限结束后仍存在后续价值,导致末期决策偏离真实有限时域问题。另外,桥梁退化过程属于非齐次马尔可夫过程,其状态转移概率随服役时间变化而变化。若仍采用不含时间相关信息的常规状态表述,则相同结构状态在不同时刻可能对应不同的最优决策,常规强化学习策略难以从原始状态中识别这种差异。

针对上述问题,本文构建了SA-BCLPPO算法。该算法包含两个核心环节:一是面向问题表述的状态增广方法,将与有限时域和退化演化相关的信息纳入策略输入,以增强策略对时变环境的识别能力;二是面向训练过程的反向课程学习机制,将训练任务由使用年限末端逐步扩展至完整时域,并通过阶段间参数继承改善训练稳定性。将两者共同嵌入强化学习训练流程,用于提高算法对桥梁有限时域非齐次维修决策问题的适应能力。所提算法框架如图3所示。

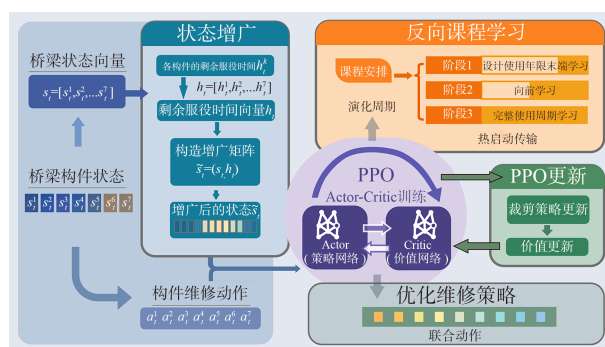


图3 基于状态增广与反向课程学习的PPO算法

Fig. 3 PPO algorithm based on state augmentation and backward curriculum learning

2.1 状态增广方法

状态增广用于表征桥梁维修决策中的有限时域特征与非齐次退化特征。相同构件状态在不同剩余服役时间和不同退化阶段下可能对应不同的最优动作,原始状态表述难以区分这种差异,因而在强化学习输入中引入时间相关信息,对状态进行扩展。

状态增广变量采用设计使用年限减去实际服

役时间得出的剩余服役时间表征。剩余服役时间更有利于刻画构件未来退化潜力及维修干预的时序影响。其反映当前状态距离规划终点的时间边界信息,其更新随决策自然推进决定,不随维修动作改变。构件状态与剩余服役时间共同构造增广状态作为强化学习的输入状态。

设在时刻 t 的第 k 个构件的状态为 s_t^k ,由各构件的状态构成构件状态向量 $\mathbf{s}_t=[s_t^1, s_t^2, \dots, s_t^k]$; 构件对应的设计使用年限为 T_k ,则该构件的剩余服役时间可表示为 $h_t^k=T_k-t$,由各构件的剩余服役时间构成向量 $\mathbf{h}_t=[h_t^1, h_t^2, \dots, h_t^k]$ 。因此强化学习输入的增广状态可表示为:

$$\tilde{\mathbf{s}}_t=(\mathbf{s}_t, \mathbf{h}_t) \quad (11)$$

式中: $\tilde{\mathbf{s}}_t$ 为时刻 t 的增广状态向量由构件状态向量 $\mathbf{s}_t$ 与剩余服役时间向量 $\mathbf{h}_t$ 共同构成,作为后续策略网络与价值网络的共同输入,用于强化学习训练过程中的动作选择与价值评估,从而使策略网络能够同时感知构件状态和构件所处的剩余服役时间,以区分相同状态但剩余服役时间不同的维修情景。

2.2 反向课程学习训练机制

反向课程学习用于改善有限时域条件下强化学习训练的稳定性。桥梁维修决策在设计使用年限末端不再存在后续累积影响,终端边界条件明确,相关子问题的学习难度相对较低;随着训练时域向前扩展,状态分布、动作影响范围和累计回报传播路径均随之增加,策略学习难度显著上升。若直接在完整设计使用周期内训练,策略需要同时处理长时域信用分配与终端价值截断问题,容易导致训练不稳定并陷入局部最优。反向课程学习的训练过程由设计使用年限的末端开始,并逐步扩展至完整使用周期,以降低早期训练阶段的学习难度。

反向课程学习按照由末期到全时域的方式组织训练任务。设桥梁构件的设计使用年限为 T ,第 m 个课程阶段对应的训练时间窗记为 $[T_m, T]$,其中 T_m 随课程阶段推进逐步减小。初始阶段仅在接近设计使用年限末端的短时域内进行训练,此时终端边界条件明确,策略学习主要关注末期状态下的动作选择;随后逐步向前扩展训练时间窗,将中后期决策过程纳入训练,最终覆盖完整使用周期,实现全时域策略学习。

相邻课程阶段之间采用参数继承方式进行衔

接。每一阶段训练结束后,将当前阶段得到的策略网络参数作为下一阶段的初始化参数,并在扩展后的训练时间窗内继续更新。该处理方式避免了各阶段训练从随机初始化重新开始,使后续阶段能够在已有策略基础上逐步适应更长时域下的状态分布与决策需求。随着训练窗口不断扩展,策略学习范围由设计使用年限末端逐步过渡至完整使用周期,训练过程与策略更新保持连续。

3 算例分析

算例分析在前文建立的桥梁非齐次马尔可夫维修决策模型基础上展开。为便于分析,假定桥梁系统每年发生一次状态转移,并同步进行一次维修决策。设计使用年限为 100 a,7 个构件的状态演化均服从前文给定的非齐次转移矩阵, α 和 β 分别为 0.005 和 0,资金折现率为 3%。基于上述设定,可进一步得到各构件在 100 a 设计使用年限内的状态演化概率分布。

在上述算例基础上,进一步设置对比实验、敏感性分析和消融实验三类实验。对比实验用于评估所提算法与其他强化学习方法、固定规则的维修策略方法以及动态规划方法在训练表现和维修决策结果方面的差异。在强化学习方法中,本文选择改进的参数化 DQN 算法^[7] (Parameterized-DQN) 和与 CNN 结合的 DQN 算法^[15] (CNN-DQN) 作为基准强化学习(RL)对比方法。在固定规则维修策略中,本文采用基于时间和基于构件状态的两类方法作为对比。基于时间的策略包括 Time-5、Time-10、Time-15 和 Time-20 四种方案,分别表示每隔 5, 10, 15, 20 a 实施一次周期性检修;检修时,对状态 1~3 及以上的构件分别采取小修、中修和大修。基于构件状态的策略包括 Condition-1~3 三种方案,均按年度检修,但对应不同的维修触发阈值:Condition-1 在状态 1~3 及以上时分别实施小修、中修和大修;Condition-2 在状态 2~4 及以上时分别实施小修、中修和大修;Condition-3 在状态 3~5 时分别实施小修、中修和大修。动态规划方法采用逆向归纳进行求解。由于本构件的算例属于有限时域下的低维决策问题,动态规划能够得到数学上的精确解,因此可作为本算例的强基准,用于评估各算法对精确解的逼近程度。敏感性分析用于讨论退化因子和折现率变化时维修策略的响应特征。实验中主要训练参数设置如表 4 所示,其

中,Parameterized-DQN 将构件状态和构件服役时间作为算法输入;CNN-DQN 将构件状态进行独热编码,构件服役时间进行二进制编码之后通过卷积特征提取后进行 Q 值估计。评估阶段每 2000 时间步进行一次策略评估,每次评估 5 个规划期,取平均值作为最终结果。算例分析的结果是在给定 100 a 设计使用年限下,不同构件在不同年份和不

同状态条件下应采取何种维修动作,以及由此形成的全寿命周期时序维修策略。因而,后文中的策略热图不仅表示维修措施类型,也反映了维修时机和动作切换边界在设计使用年限内的变化特征。本文设置消融实验来分析不同改进点的具体贡献,对状态增广和反向课程学习对算法的训练稳定性和求解精度的提升进行了研究。

表 4 PPO 与 DQN 类方法的超参数设置

Table 4 Hyperparameter settings for PPO and DQN-based methods

方法	学习率	折扣因子 γ	GAE 系数 λ	采样步数	批量大小	训练轮数	裁剪阈值	策略网络结构	经验回放池容量
PPO	1×10^{-4}	0.99	0.95	128	512	10	0.2	两层全连接 (128, 128)	—
DQN 类	1×10^{-4}	0.99	—	—	—	—	—	—	50000

3.1 对比实验

从训练过程看,SA-BCLPPO 算法在收敛稳定性方面表现出更明显优势。由图 4 可见,CNN-DQN 在训练初期累计奖励较低,随后随训练推进逐步提升,但曲线波动较为明显,说明其策略学习过程稳定性相对不足。参数化 DQN 的累计奖励变化相对平缓,但整体提升幅度有限,训练后期仍未达到较高的奖励水平。相比之下,SA-BCLPPO 在训练前期的改进过程相对平缓,但随着训练推进,其平均奖励持续上升、平均成本持续下降,并在后期逐步收敛到更稳定的水平。

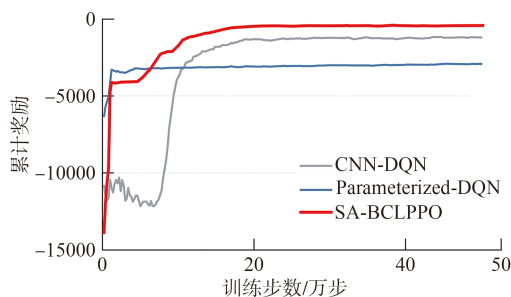


图 4 SA-BCLPPO 与基准算法的累计奖励收敛曲线

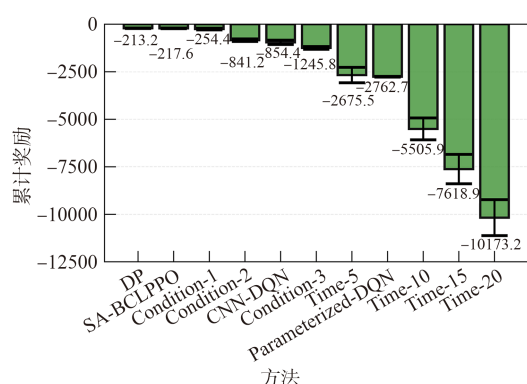
Fig. 4 Cumulative reward convergence curves of SA-BCLPPO and baseline algorithms

为进一步比较各方法在训练收敛后的决策性能,本文将训练完成后的各强化学习模型与其他基准方法应用于当前桥面系统,并进行 100 次蒙特卡洛模拟。每次模拟均覆盖 100 a 规划周期,最终取各项指标的平均值,用于对比不同方法下的累计奖励、累计维修成本和累计风险成本,结果如图 5 所示。

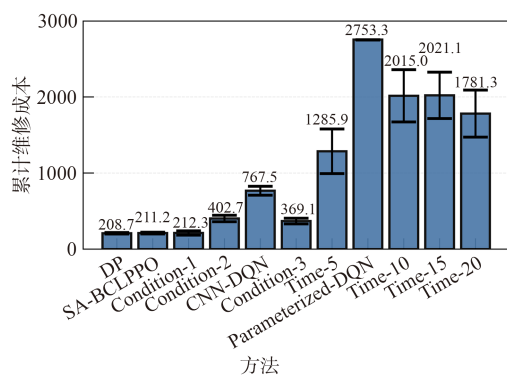
由图 5 可见,不同方法在 100 a 设计使用年限内

的累计奖励、维修成本和风险成本方面存在明显差异。DP 作为当前场景下的理论参照,取得了最高的累计奖励和最低的总成本;SA-BCLPPO 的累计奖励与 DP 接近,且明显优于 Parameterized-DQN、CNN-DQN 以及规则基线,说明所提方法在长期维修决策中具有较好的近似最优性。进一步比较成本构成可以发现,SA-BCLPPO 在维修成本与风险成本之间实现了较好的平衡:相较规则策略,其能够显著降低长期风险积累;相较 DQN 类方法,其避免了过高的维修成本或不稳定的风险控制结果。基于时间规则类维修策略总体表现较差,表明单纯依赖固定周期维修难以适应桥面构件状态的动态演化。为进一步揭示不同方法的决策机制,表 5 给出了各方法在 100 次模拟中的详细量化指标。可以看出;DP 在当前小规模场景下取得最优累计奖励,为 -213.21,为理论精确基准;SA-BCLPPO 的累计奖励为 -217.64,与 DP 的差距仅为 4.43,明显小于其他基准方法,表明所提方法能够较好地逼近有限时域逆向归纳得到的最优策略。

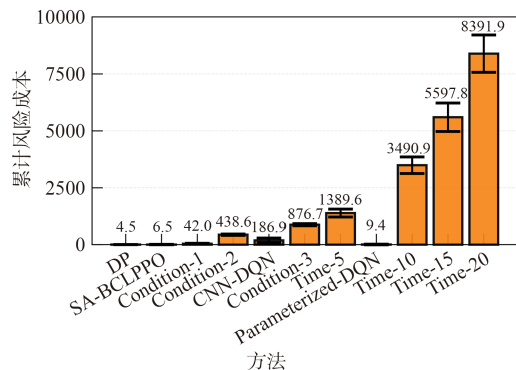
从成本构成看,SA-BCLPPO 的累计维修成本为 211.15,与 DP 的 208.67 非常接近;累计风险成本为 6.49,也维持在较低水平。这表明 SA-BCLPPO 并不是单纯通过增加维修投入来降低风险,而是在维修成本和结构风险之间取得了较好的平衡。相比之下,Parameterized-DQN 虽然平均状态最低,仅为 0.0507,但其累计维修成本高达 2753.35,说明该方法倾向于采取过度维修策略,虽然能维持较好的构件状态,却显著牺牲了经济性。CNN-DQN 的累计维修成本为 767.53,风险成本为 186.91,均明显高于 SA-BCLPPO,表明其在长期维修决策中未能形成



(a) 累计奖励对比



(b) 累计维修成本对比



(c) 累计风险成本对比

图5 100 a设计使用年限下不同维修决策方法的
累计奖励与成本对比

Fig. 5 Comparison of cumulative rewards and costs under
different maintenance decision-making methods over a
100-year design service life

稳定且经济的风险控制策略。规则基线中,Condition-1的表现相对最好,累计奖励为-254.36,但仍与DP相差41.15,且风险成本达到42.04,明显高于SA-BCLPPO。这说明简单的状态阈值策略虽然具有一定有效性,但难以充分利用长时域动态决策信息。Condition-2和Condition-3的累计奖励进一步下降,风险成本分别达到438.58和876.73,构件处

于状态4或5的时间占比也逐步提高,说明较宽松的状态触发条件会导致维修干预滞后,从而带来更高的结构风险积累。Time类固定周期策略整体表现最差。随着维修间隔从5 a增加到20 a,累计奖励从-2675.55下降到-10173.17,风险成本从1389.63增加到8391.87,构件处于状态4或5的时间占比也由17.944%升至53.410%。

这表明单纯依赖固定时间间隔进行维修难以适应构件退化过程的随机性和非齐次性,尤其在较长维修间隔下,构件会长期处于较差状态,导致风险成本快速累积。综合来看,SA-BCLPPO在累计奖励、维修成本、风险成本和状态保持指标上均与DP最优基准最为接近,且明显优于DQN类方法和传统规则策略。

本文模型实际输出的是有限设计使用年限内桥面系统7个构件的联合维修策略,即在每个决策时刻分别确定各构件的维修动作,并共同形成面向桥面系统的年度维修方案。考虑到7个构件在100 a设计使用年限内的完整策略信息量较大,若逐一展示,不利于突出不同算法在决策结构上的差异。为此,选取桥面构件作为代表对象,对策略热图进行可视化分析,以说明所提方法在维修时机与采取维修等级两方面的决策特征。

由图6可见,不同算法在桥面构件上的维修策略存在明显差异。总体来看,SA-BCLPPO算法得到的策略具有更清晰的阶段性和更合理的时序特征。在设计年限前中段,随着构件状态由较好逐步向较差演化,策略能够根据状态等级变化采取由小修、中修到大修的分层干预;而在设计使用年限末段,维修动作的频率和强度明显减弱,更多表现为低强度维修或不再实施高强度干预。这一特征与有限时域维修决策问题的基本规律一致,即在接近设计使用年限终点时,后续可获得的长期收益有限,继续实施大修或重建等高成本动作的必要性下降。SA-BCLPPO算法能够学习到这一末期维修收缩特征,表明其更能反映有限时域条件下末期干预收缩的决策规律。相比之下,CNN-DQN算法得到的策略虽在不同状态下表现出一定差异,但整体上仍以较高频率的维修动作为主,该策略虽然能够在一定程度上响应构件状态变化,但维修动作分布较为单一,末期维修强度收缩特征不够明显。这表明该算法在有限时域条件下对终端阶段决策特征的识别仍然不足。Parameterized-DQN算法的策略则表现

表 5 不同维修决策方法的多指标性能对比

Table 5 Multi-metric performance comparison of different maintenance decision-making methods

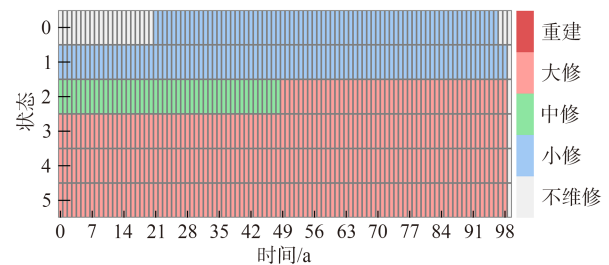
方法	累计奖励	与 DP 差距	维修成本	风险成本	平均状态	构件处于状态 4 或 5 的时间占比/%
DP	-213.21 (13.82)	—	208.67 (9.69)	4.54 (4.31)	0.1724 (0.0661)	0.013 (0.041)
SA-BCLPPO	-217.64 (17.57)	4.43	211.15 (11.75)	6.49 (8.25)	0.1573 (0.0684)	0.013 (0.046)
Parameterized-DQN	-2762.73 (19.20)	2549.52	2753.35 (5.05)	9.38 (14.87)	0.0507 (0.0470)	0.327 (0.896)
CNN-DQN	-954.43 (112.99)	741.22	767.53 (58.67)	186.91 (104.63)	0.2802 (0.0362)	0.950 (0.661)
Condition-1	-254.36 (45.46)	41.15	212.32 (28.31)	42.04 (20.33)	0.9303 (0.0282)	0.143 (0.155)
Condition-2	-841.23 (68.75)	628.02	402.66 (42.67)	438.58 (30.74)	1.3056 (0.0300)	1.646 (0.426)
Condition-3	-1245.79 (73.04)	1032.58	369.06 (37.55)	876.73 (47.82)	1.4355 (0.0478)	11.990 (0.696)
Time-5	-2675.55 (406.28)	2462.33	1285.92 (293.19)	1389.63 (173.52)	1.5511 (0.0614)	17.944 (1.747)
Time-10	-5505.91 (575.68)	5292.70	2014.99 (343.49)	3490.92 (364.76)	1.9932 (0.0697)	33.514 (2.435)
Time-15	-7618.90 (773.37)	7405.69	2021.06 (304.97)	5597.84 (624.94)	2.3335 (0.0989)	43.141 (2.885)
Time-20	-10173.17 (945.23)	9959.96	1781.30 (309.15)	8391.87 (819.38)	2.7635 (0.1048)	53.410 (2.957)

注:表中左侧数字为平均值,右侧括号内的数字为标准差。

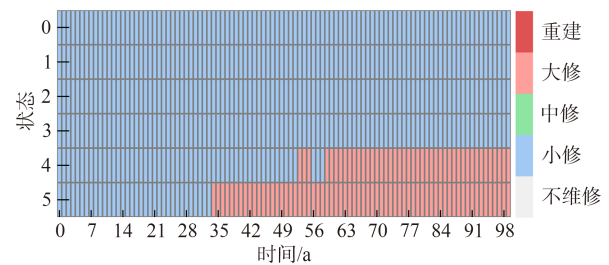
出更强的局部波动性,在设计使用年限前段,其策略以较高强度维修动作为主;进入中后期后,维修动作在部分状态上快速转向低强度维修或不维修,策略边界相对突兀,策略结构的平滑性和阶段一致性相对较弱。与这两类对照算法相比,SA-BCLPPO 在不同状态等级和不同时间位置上表现出更稳定的决策边界,能够更合理地平衡规划前期的状态维持需求与规划末期的维修投入收缩。

为进一步从时间维度分析不同策略在 100 a 设计使用年限内的状态控制效果和维修投入差异,本文选取 DP 策略、两类 DQN 基线以及三类状态阈值规则策略进行对比,并基于 100 次蒙特卡洛模拟结果绘制桥面构件平均状态与累计维修成本的演化曲线,如图 7 所示。

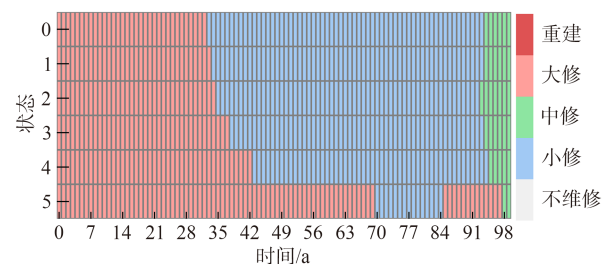
可以看出:SA-BCLPPO 与 DP 参照策略在 100 a 设计使用周期内的平均状态演化趋势较为接近,二者均能将桥面构件平均状态控制在较低水平,并且在设计使用周期末期时,能够在控制状态的前提下避免过度维修,表明所提方法能够较好地逼近 DP 参照策略;从累计维修成本演化看,SA-BCLPPO 的成本增长趋势也与 DP 保持一致,未出现明显的过度维修或维修滞后现象,表明其在长期状态保持与维修投入之间形成了较为合理的权衡;与 DQN 类基准相比,SA-BCLPPO 的优势更加明显;Parameterized-DQN 虽然能够在设计使用周期内维持更低的平均状态,但其累计维修成本增长显著快于 SA-BCLPPO 和 DP,说明该方法主要依赖更高频率或更高强度的维修来改善构件状态,存在过



(a) SA-BCLPPO 算法



(b) CNN-DQN 算法

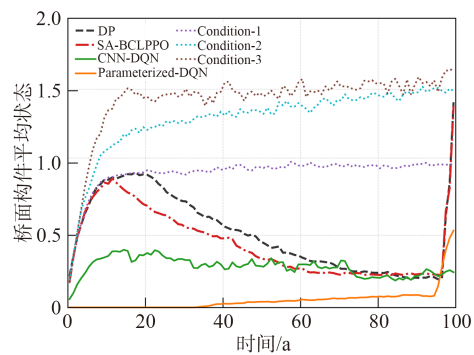


(c) Parameterized-DQN 算法

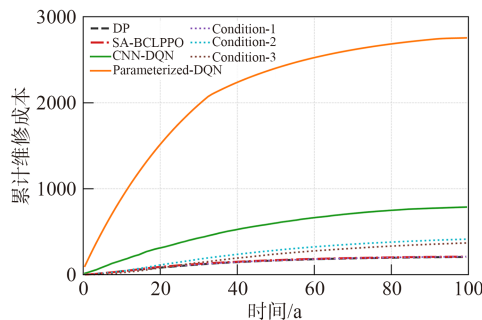
图 6 不同算法下桥面构件维修策略热图对比

Fig. 6 Comparison of maintenance policy heatmaps for the deck component under different algorithms

度维修倾向;CNN-DQN 的累计维修成本低于 Parameterized-DQN,但其平均状态整体高于 SA-BCLPPO,说明其状态控制效果不足,难以在维修成



(a) 平均状态线



(b) 累计维修成本

图7 100a设计使用年限下不同算法的状态演化与累计维修成本对比

Fig. 7 Comparison of state evolution and cumulative maintenance costs under different algorithms over a 100-year design service life

本和结构状态之间取得稳定平衡。

与 Condition 类规则策略相比,SA-BCLPPO 能够在更低状态水平下保持相对合理的维修成本增长。Condition-1 的维修成本与 SA-BCLPPO 较为接近,但其平均状态明显更高,说明简单阈值策略虽然能够控制维修投入,却难以实现精细化状态管理;Condition-2 和 Condition-3 随着触发阈值放宽,维修成本进一步降低,但平均状态明显劣化,反映出维修干预滞后带来的状态控制不足。综合来看,SA-BCLPPO 在状态保持效果和维修成本累积方面均更接近 DP 参照策略,并明显优于 DQN 类方法和规则基线,进一步验证了所提方法在长期桥面维修决策中的近似最优性和工程适用性。

综上所述,SA-BCLPPO 在训练稳定性、累计奖励、成本控制和状态保持方面均表现出较好的综合性能。其结果与 DP 精确基准最为接近,说明所提方法能够较好地逼近有限时域最优维修策略;同时,相较于 DQN 类方法和传统规则策略,SA-BCLPPO 能够更合理地平衡维修成本与结构风险,并避免规划末期过度维修,从而形成更优的长期桥

面维修决策方案。

3.2 敏感性分析

退化速度和资金时间价值均会影响桥梁维修决策,因此分别围绕两种退化因子和折现率开展敏感性分析。在敏感性分析中,考虑不同的退化因子和折现率变量(表6),比较不同条件下 SA-BCLPPO 智能体对桥面构件的维修策略热图。

表6 敏感性分析

Table 6 Sensitivity analysis

组别	α	β	折现率/%
基准组	0.005	0	3
对比组 I	0.010	0	3
	0.001	0	3
对比组 II	0.005	1×10^{-5}	3
	0.005	2×10^{-5}	3
对比组 III	0.005	0	2
	0.005	0	5

在对比组 I ($\beta=0, r_d=3\%$) 中,SA-BCLPPO 智能体在不同 α 取值下的最优维修策略整体保持了较稳定的分层特征,但优良状态下的干预时机随 α 变化较为明显。由图8可见,当 α 分别取 0.001, 0.005, 0.01 时,状态 1 在整个设计使用年限内始终对应小修,状态 3~5 始终稳定对应大修,状态 2 均表现为前期采用中修、后期转为大修的两阶段模式,说明 α 的变化并未改变轻度劣化及较差状态下的主导维修等级。

相比之下,优良状态即状态 0 对 α 更为敏感:当 $\alpha=0.001$ 时,状态 0 在规划前中期主要采取不维修策略,至第 40 年后才转为小修;当 $\alpha=0.005$ 时,该切换时刻提前至第 20 年;当 $\alpha=0.01$ 时,小修进一步提前至规划初期即开始出现。该结果表明,随着 α 增大,构件在长期服役过程中呈现更快的渐进性退化特征,SA-BCLPPO 智能体因而倾向于前移优良状态下的预防性干预时刻。总体来看, α 的变化并未改变 SA-BCLPPO 智能体最优维修策略的基本等级结构,但会显著影响优良状态下小修启动的时间位置。

在对比组 II ($\alpha=0.005, r_d=3\%$) 中,SA-BCLPPO 智能体在不同 β 取值下的维修策略同样表现出较强稳定性,仅在优良状态下出现有限调整。由图9可见,当 β 分别取 1×10^{-5} 和 2×10^{-5} 时,状态 1 在全寿命期内始终对应小修,状态 3~5 始终对应大修,状态 2 均维持前期中修、后期大修的分段特征,表明在当

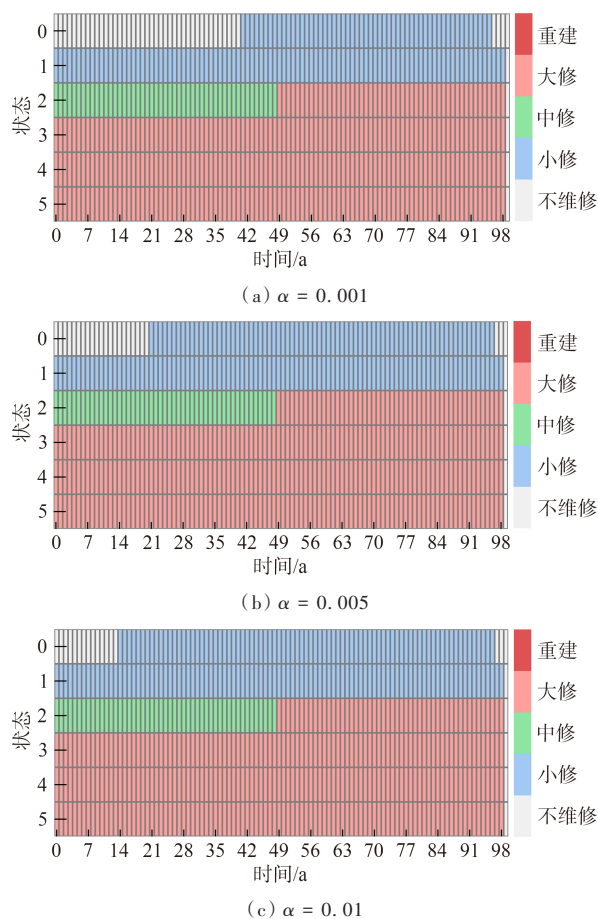


图8 不同退化因子 α 下桥面构件维修策略热图($\beta = 0, r_d = 3\%$)

Fig. 8 Maintenance policy heatmaps of the deck component under different deterioration factors α ($\beta = 0, r_d = 3\%$)

前参数范围内, β 的变化并未改变中、差状态下的优势维修动作。相较之下,状态0的不维修向小修的切换时刻随 β 增大而略有提前,说明当寿命后期的加速损伤效应增强时,SA-BCLPPO智能体会适当前移优良状态下的干预时刻,以缓解后期退化加快所带来的维修压力。不过,与 α 敏感性结果相比, β 变化引起的策略调整幅度明显较小,表明在本文设定范围内, β 对维修策略的影响主要集中在寿命后段退化过程的修正,对整体维修等级结构的扰动相对有限。总体而言, β 的增大不会改变不同劣化状态下的基本维修模式,但会使优良状态下的预防性干预略有前移。

在对比组Ⅲ($\alpha=0.005, \beta=0$)中,SA-BCLPPO智能体在不同折现率下的最优维修策略呈现出明显差异,表明资金时间价值对长期维修决策具有较强影响。由图8(b)和图10对比可知,当折现率分别取2%、3%和5%时,状态3~5下的策略基本保持稳定,

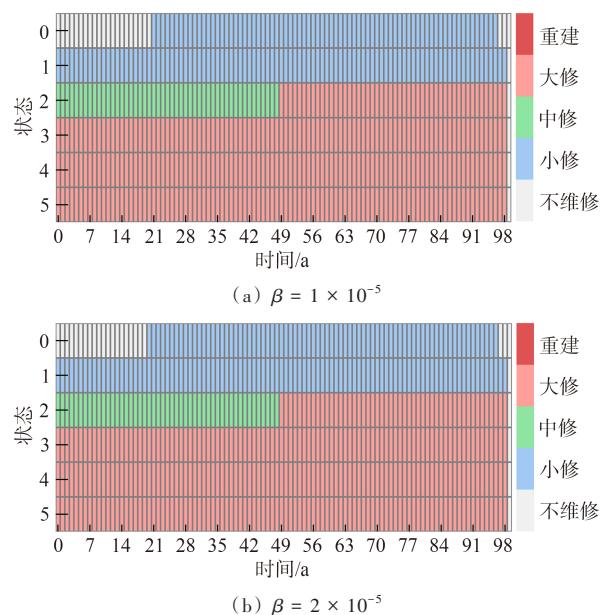


图9 不同退化因子 β 下桥面构件维修策略热图
($\alpha = 0.005, r_d = 3\%$)

Fig. 9 Maintenance policy heatmaps of the deck component under different deterioration factor β ($\alpha = 0.005, r_d = 3\%$)

始终以大修为主,说明对于已处于较差状态的构件,智能体更倾向于采取较强干预措施,且该决策不易受折现率变化影响;状态1在三种情形下也总体保持为小修,仅在高折现率条件下局部时段出现短暂的大修插入,表明轻度劣化状态下的维修等级整体较为稳定。相比之下,状态0和状态2对折现率变化更为敏感。当折现率为2%和3%时,状态0均在规划前期保持不维修、随后转为小修,且状态2均表现为前期中修、后期大修的分段特征,说明在较低折现水平下,SA-BCLPPO智能体仍会保留一定的中后期强化干预策略。但当折现率提高至5%后,状态0转为小修的時刻并未明显提前,反而整体策略更加偏向于在较长时期内维持较低即时支出。与此同时,状态2对应的最优动作由中修至大修转变为较早进入大修,且状态1在寿命后段也出现局部大修。

该结果表明,随着折现率升高,未来维修支出在当前决策中的权重被进一步削弱,智能体因而更倾向于压缩分阶段、渐进式的养护安排,而在状态恶化至一定程度后直接选择更强维修动作,以减少长期连续投入所带来的折现后效益减弱问题。总体来看,折现率变化并未改变SA-BCLPPO智能体在严重劣化状态下偏向大修的基本策略,但会显著影响优良和中等状态下的维修时序与维修强度,其

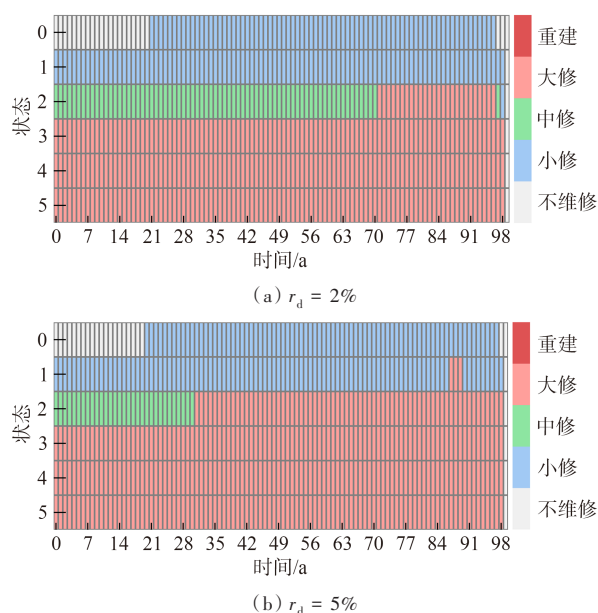


图 10 不同折现率 r_d 下桥面构件维修策略热图
($\alpha = 0.005, \beta = 0$)

Fig. 10 Maintenance policy heatmaps of the deck component under different discount rates r_d ($\alpha = 0.005, \beta = 0$)

中高折现率条件下策略的阶段特征更弱、维修动作更趋集中,反映出资金时间价值对长期养护节奏具有较强调节作用。

3.3 消融实验

为进一步分析状态增广和反向课程学习在所提算法中的具体作用,设置消融实验对不同设计进行对比。由于服役时间增广状态已应用在 Parameterized-DQN 算法中,因此状态增广部分分别采用无状态增广和增广状态两种输入形式;训练机制部分分别设置采用和不采用反向课程学习两种情形,据此设置 4 组对照:原始 PPO、状态增广+PPO、反向课程学习+PPO 以及完整的 SA-BCLPPO。基于上述组合,可进一步比较不同状态表征方式和训练机制对策略学习结果的影响。

图 11 给出了不同算法在消融实验中的训练表现。可知:原始 PPO 算法在训练过程中波动最大,平均奖励曲线长期处于较大幅度振荡状态,平均成本下降过程也不稳定,说明在不进行状态增广和课程学习的条件下,策略难以有效适应有限时域与非齐次退化共同作用下的决策环境。

引入状态增广后,训练曲线波动明显减小,平均奖励提升和平均成本下降过程均更为平滑,表明增强状态表征有助于提高策略对时变决策特征的识别能力。相比之下,仅采用反向课程学习时,训

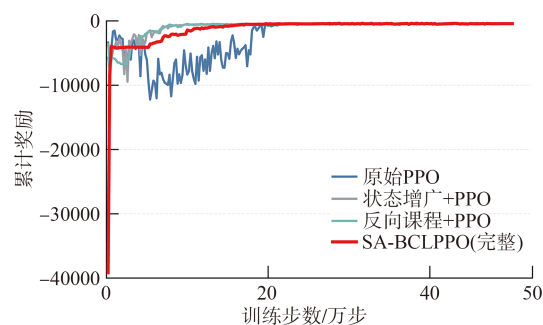


图 11 消融实验中不同算法的训练表现对比

Fig. 11 Comparison of training performance of different algorithms in the ablation study

练过程较原始 PPO 有所改善,但整体波动仍然较大,未表现出与状态增广相当的稳定收敛特征。完整的 SA-BCLPPO 算法在两项机制共同作用下表现出最优的训练特征,其平均奖励曲线提升更连续,平均成本曲线下降更稳定,后期波动幅度也最小。

图 12 进一步给出了不同消融算法在训练完成后的最终累计奖励对比结果。为保证评估结果的稳定性,本文对各算法分别进行了 1000 次蒙特卡洛法模拟,每次模拟均覆盖一个 100 a 设计使用年限,并统计累计奖励;图中的误差棒表示累计奖励的 95% 置信区间。

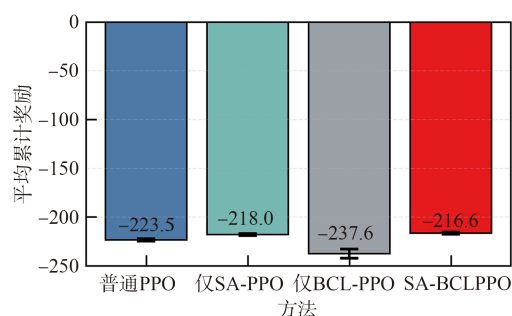


图 12 消融实验中不同算法的最终累计奖励对比

Fig. 12 Comparison of final cumulative rewards of different algorithms in the ablation study

结果表明:SA-BCLPPO 取得了最高的累计奖励,为-216.6,说明状态增广与反向课程学习联合使用时能够获得更优的最终决策性能;仅 SA-PPO 的累计奖励与完整方法较为接近,表明状态增广对策略性能提升具有重要贡献。由表 7 可见,SA-BCLPPO 的累计维修成本为 209.90,低于普通 PPO、仅 SA-PPO 和仅 BCL-PPO,说明完整方法在获得最高累计奖励的同时,也能有效控制维修投入。

仅 SA-PPO 的风险成本和平均状态略优于完整

表 7 消融实验中不同算法的多指标性能对比
Table 7 Multi-metric performance comparison of different algorithms in the ablation study

方法	平均累计奖励	维修成本	风险成本	平均状态
普通 PPO	-223.49	212.85	10.64	0.2120
仅 SA-PPO	-218.00	213.06	4.94	0.1307
仅 BCL-PPO	-237.59	232.36	5.23	0.1314
SA-BCLPPO	-216.56	209.90	6.66	0.1577

方法,但其维修成本较高,表明仅依靠状态增广可能会带来更偏保守的维修策略。仅 BCL-PPO 的维修成本最高,达到 232.36,导致其累计奖励明显下降,说明在缺乏有效状态表征的条件下,单独引入反向课程学习难以取得理想效果。综合来看,状态增广和反向课程学习的联合使用能够在维修成本、风险成本和状态保持之间形成更合理的平衡。

4 结 论

本文围绕桥梁在有限设计寿命条件下的长期维修决策问题展开研究,针对退化过程时变、维修投入具有时间价值以及传统方法难以兼顾非齐次退化与有限时域决策特征等现实需求,构建了相应的建模与求解框架。通过理论建模、算法设计与算例验证,本文较系统地讨论了桥梁非齐次退化条件下维修决策的关键问题,并验证了所提方法在策略学习效果与决策合理性方面的可行性与有效性。

1)针对桥梁维修决策中构件退化过程随服役时间变化、设计使用年限有限以及资金时间价值并存的特点,建立了考虑非齐次退化与维护成本折现的桥梁非齐次马尔可夫维修决策模型。该模型在构件层刻画了状态退化、维修动作与成本演化之间的关系,可用于桥梁 100 a 设计使用年限内的长期维修决策分析。

2)针对常规强化学习方法难以同时处理有限时域边界条件和非齐次退化时变特征的问题,提出了基于状态增广与反向课程学习的强化学习算法 SA-BCLPPO。该方法以剩余决策时域构造增广状态输入,并通过由设计使用周期末端向全时域逐步扩展的课程训练方式改善策略学习过程,形成了适用于桥梁有限时域非齐次维修决策问题的算法框架。

3)基于魁北克桥面系统算例的对比实验结果表明,SA-BCLPPO 在训练稳定性、收敛表现和最终成本结果方面均优于 Parameterized-DQN 算法和

CNN-DQN 算法。策略热图分析进一步表明,所提方法能够学习到更符合有限时域维修决策规律的动作结构,尤其能够在设计使用年限末段减少不必要的高强度维修投入。

4)敏感性分析表明,退化因子变化主要影响维修时机的前移或后移,折现率变化则主要影响不同状态下维修强度及动作切换边界,所提方法能够对关键参数变化作出一致性的策略响应。消融实验表明,状态增广和反向课程学习在所提算法中具有不同作用,其中状态增广对提升策略质量更为直接,反向课程学习主要用于改善训练稳定性,两者联合使用时能够获得更优结果。

5)本文算例分析基于给定桥面系统和示例退化参数条件展开,尚未进一步考虑构件间相互作用、检测误差、预算约束及退化参数的工程标定问题,尤其是非齐次退化函数中的经验参数,当前主要用于说明所提建模与求解框架,其取值仍需结合长期检测数据、状态转移统计结果或可靠度分析进一步校准。后续研究可在此基础上拓展至更复杂桥梁系统、多约束运维场景及不确定性条件下的维修决策问题。

参 考 文 献

- [1] 秦权. 桥梁结构的健康监测[J]. 中国公路学报, 2000, 13(2): 38-43.
- [2] 中华人民共和国交通运输部. 公路桥涵养护规范: JTG 5120—2021[S]. 北京: 人民交通出版社股份有限公司, 2021.
- [3] 周绪红, 张喜刚. 关于中国桥梁技术发展的思考[J]. Engineering, 2019, 5(6): 1120-1130.
- [4] 贺桥, 雷家艳, 高婧, 等. 可持续发展的项目级桥梁多目标维修决策优化[J]. 厦门大学学报(自然科学版), 2025, 64(6): 1026-1033.
- [5] ABU DABOUS S, ALKASS S. Managing bridge infrastructure under budget constraints: a decision support methodology [J]. Canadian Journal of Civil Engineering, 2011, 38(11): 1227-1237.
- [6] 刘纲, 孙瑞卿, 李琦, 等. 基于序贯决策的桥梁多阶段维修加固策略优化方法[J]. 重庆大学学报, 2026, 49(1): 60-69.
- [7] DU A, GHAVIDEL A. Parameterized deep reinforcement learning-enabled maintenance decision-support and life-cycle risk assessment for highway bridge portfolios[J]. Structural Safety, 2022, 97: 102221.
- [8] GENG Z, ZHANG C, JIANG Y, et al. Integrating multi-source data for life-cycle risk assessment of bridge networks: a system digital twin framework [J]. Journal of Infrastructure Preservation and Resilience, 2025, 6(1): 1-20.
- [9] 戴理朝, 康哲, 陈瑞, 等. 基于 NSGA-III 的桥梁网络多目标维修决策优化研究[J]. 土木工程学报, 2024, 57(5): 41-52.

- [10] SUTTON R S, BARTO A G. Reinforcement Learning: An Introduction[M]. Massachusetts: The MIT Press, 2018.
- [11] 吕颖钊, 贺拴海. 在役桥梁承载力模糊可靠性的马尔科夫预测[J]. 长安大学学报(自然科学版), 2005(4): 39-43.
- [12] TAO Z, COROTIS R B, ELLIS J H. Reliability-based bridge design and life cycle management with markov decision processes[J]. Structural Safety, 1994, 16(1/2): 111-132.
- [13] 刘渝. 基于部分可观测马尔可夫决策过程的钢桥腐蚀维护策略优化研究[D]. 重庆:重庆大学, 2025.
- [14] MORCOUS G. Performance prediction of bridge deck systems using markov chains[J]. Journal of Performance of Constructed Facilities, 2006, 20(2): 146-155.
- [15] WEI S, BAO Y, LI H. Optimal policy for structure maintenance: A deep reinforcement learning framework[J]. Structural Safety, 2020, 83: 101906.
- [16] 王超. 马尔科夫链在桥梁状态预测中的研究与应用[D]. 北京:北京交通大学, 2009.
- [17] SRIKANTH I, AROCKIASAMY M. Deterioration models for prediction of remaining useful life of timber and concrete bridges: a review[J]. Journal of Traffic and Transportation Engineering (English Edition), 2020, 7(2): 152-173.
- [18] BOLUKBASI M, MOHAMMADI J, ARDITI D. Estimating the future condition of highway bridge components using national bridge inventory data[J]. Practice Periodical on Structural Design and Construction, 2004, 9(1): 16-25.
- [19] TOLLIVER D, LU P. Analysis of bridge deterioration rates: a case study of the northern plains region[J]. Journal of the Transportation Research Forum, 2012, 50(2): 87-100.
- [20] AGRAWAL A K, KAWAGUCHI A. Bridge element deterioration rates: project: C-01-51[R]. New York: Transportation Infrastructure Research Consortium New York State, Department of Transportation, 2009.
- [21] ARROW K, CROPPER M, Gollier C, et al. Determining benefits and costs for future generations[J]. Science, 2013, 341: 349-350.
- [22] TREASURY H M. The green book: central government guidance on appraisal and evaluation[M]. Updated ed. London: OGL Press, 2022.