中文引用格式:赵长啸,李道俊,孙亦轩,等. 基于深度强化学习的综合航电系统安全性优化方法[J]. 中国安全科学学报,2024,34(7):123-131.

英文引用格式:ZHAO Changxiao, LI Daojun, SUN Yixuan, et al. Integrated avionics system safety optimization method based on deep reinforcement learning[J]. China Safety Science Journal, 2024, 34(7): 123-131.

基于深度强化学习的综合航电系统安全性优化方法*

赵长啸[1,2]副教授, 李道俊[1], 孙亦轩[1], 景 鹏[1], 田 毅**[1,2]副研究员

(1 中国民航大学 安全工程与科学学院,天津 300300;
2 中国民航大学 民航航空器适航审定技术重点实验室,天津 300300)

中图分类号:X949 **文献标志码**:A **DOI**: 10.16265/j.cnki.issn1003-3033.2024.07.0228

资助项目:国家重点研发计划项目(2021YFB1600601);天津市高等学校研究生教育改革研究计划项目(TJYG135);中国民航大学研究生科研创新资助项目(2023YJSKC09015)。

【摘 要】 为解决传统基于人工检查的安全性设计方法难以应对航电系统大规模集成带来的可选驻留方案爆炸问题,构建航电系统分区模型、任务模型以及安全关键等级量化模型,将考虑安全性的综合化设计优化问题模型化为马尔可夫决策过程(MDP)问题,并提出一种基于 Actor-Critic 框架的柔性动作-评价(SAC)算法的优化方法;为得到 SAC 算法的参数选择和训练结果之间的相关性,针对算法参数灵敏度开展研究;同时,为验证基于 SAC 算法的优化方法在优化考虑安全性的综合化设计方面的优越性,以深度确定性策略梯度(DDPG)算法和传统分配算法为对象,开展优化对比试验。结果表明:在最佳的参数组合下,使用的 SAC 算法收敛后的最大奖励相较于其他参数组合提升近8%,同时,收敛时间缩短近16.6%;相较于 DDPG 算法和传统分配算法,基于 SAC 算法的优化方法在相同的参数设置下获得的最大奖励、约束累计违背率、分区均衡风险效果、分区资源利用以及求解时间方面最大提升分别为62%、7464%、8370%、2123%和775%。

【关键词】 深度强化学习; 综合航电系统; 安全性; 优化方法; 马尔可夫决策过程(MDP); 综合化设计

Integrated avionics system safety optimization method based on deep reinforcement learning

ZHAO Changxiao[1,2], **LI Daojun**[1], **SUN Yixuan**[1], **JING Peng**[1], **TIAN Yi**[1,2]

(1 School of Safety Engineering and Science, Civil Aviation University of China, Tianjin 300300, China;
2 Key Laboratory of Civil Aviation Airworthiness Certification Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: To solve the problem that traditional safety design methods based on manual inspection were difficult to cope with the explosion of optional residence solutions caused by the large-scale integration of avionics systems, an avionics system partition model, task model and safety criticality level quantification model were constructed, and the comprehensive design optimization considering safety was modeled as an

* 文章编号:1003-3033(2024)07-0123-09; 收稿日期:2024-01-18; 修稿日期:2024-04-21

** 通信作者:田 毅(1983—),男,陕西汉中人,硕士,副研究员,主要从事机载电子硬件适航审定、航空专用集成电路设计、计算机体系结构方面的研究。E-mail: ytian@cauc.edu.cn。

MDP problem. An optimization method of Soft Action-Critic (SAC) algorithm based on Actor-Critic framework was proposed. In order to obtain the correlation between the parameter selection and training results of SAC algorithm, the sensitivity of the algorithm parameters was studied. At the same time, to verify the superiority of the optimization method based on the SAC algorithm in optimizing the comprehensive design considering safety, optimization comparison experiments were carried out with the Deep Deterministic Policy Gradient (DDPG) algorithm and the traditional allocation algorithm as the objects. The results show that under the optimal parameter combination, the maximum reward after using convergence of SAC algorithm increases by nearly 8% compared with other parameter combinations, and the convergence time is shortened by nearly 16.6%. Compared with the DDPG algorithm and the traditional allocation algorithm, the optimization method based on SAC algorithm has improved approximately 62%, 7464%, 8370%, 2123% and 775% in terms of the maximum reward, cumulative constraint violation rate, partition balance risk effect, partition resource utilization and solution time

Keywords: deep reinforcement learning; integrated modular avionics; safety; Markov decision process (MDP); integrated design

0 引 言

安全是民机产业的生命线[1],航电系统作为飞机的大脑和中枢神经[2],对飞机整机的安全性水平起着至关重要的作用。相较于已在波音 737、空客 A320 等系列飞机上大规模应用的联合式航电系统,综合航电系统在共享的高性能平台上通过时空分区机制实现了多个航电功能集成,通过时空分区的机制保证不同航电任务的综合执行,而不同航电功能失效对飞机安全性的影响程度不同,按照文献[3],飞机功能失效状态可划分为灾难性 I 类—无影响 V 类,如飞行控制系统失效在最严酷的条件下可能导致机毁人亡,该失效状态类别即为 I 类失效状态,而客舱娱乐系统的某类失效不会对飞机的运行安全带来影响,该失效状态类别即为 V 类无影响。当基于综合航电架构将不同安全关键等级的航电功能进行集成设计时,如何在不同分区中规划合理的航电功能任务,是综合航电系统安全性设计的关键。

近年来,学者们针对多安全关键功能的规划、调度开展了相关研究。KHAMVILAI 等[4]受航空电子系统中多核架构的启发,将并行计算架构上的任务分配问题表述为整数线性规划形式的优化问题,提出一种任务的在线分配方法。LU Hui 等[5]重点关注严格周期性和抢占式分区调度策略,提出一种基于粒子群优化(Particle Swarm Optimization, PSO)算法的优化方法,以增强系统的可重构性和可调度性。ZHOU Tianran 等[6]针对多层资源分配,提出一种基于遗传模拟退火(Genetic Simulated Annealing, GSA)的启发式调度方法,以通信代价和工作负载为优化目标,将预定义任务有效分配到处理节点。为优化航电系统性能并增强系统的自适应性,ZHOU Xuan 等[7]建立了协调分区处理和时间触发通信的通用分布式综合航电系统模型,基于构建的模型提出一种混合整数规划(Mixed Integer Programming, MIP)的混合调度算法。然而,随着技术的发展,航电系统中综合的功能数量不断提高,以波音 787 飞机航电系统为例,有 36 项功能驻留在综合模块化航空电子(Integrated Modular Avionics, IMA)系统,其中,涉及 16 个硬件通用处理模块(General Processor Module, GPM),在 1 个 GPM 仅 2 个分区假设下,可能的综合化方案达到上百万种,传统的优化方法从时间成本和人力成本上都难以接受。而相较于传统的优化算法,深度强化学习具有更广泛的应用范围、更强的适应性和自主学习能力,能够有效处理不确定环境和连续空间问题,在自动驾驶、机器人控制和能源管理等领域已有广泛应用[8-11]。

鉴于此,笔者拟引入深度强化学习方法综合化设计多主流航电系统。基于综合航电系统分区模型、任务模型以及安全关键等级量化模型的构建,将考虑安全性的综合化设计问题制定为马尔可夫决策过程(Markov Decision Process, MDP)问题,并应用提出的深度强化学习算法求解分析,以期为航电系统的综合化设计提供安全有效的新方法。

1 综合航电系统建模

依靠具体任务落实航电功能,航电系统综合化设计过程如图 1 所示。在设计层面,通过将不同航电任务分配到独占性资源的分区来处理任务。

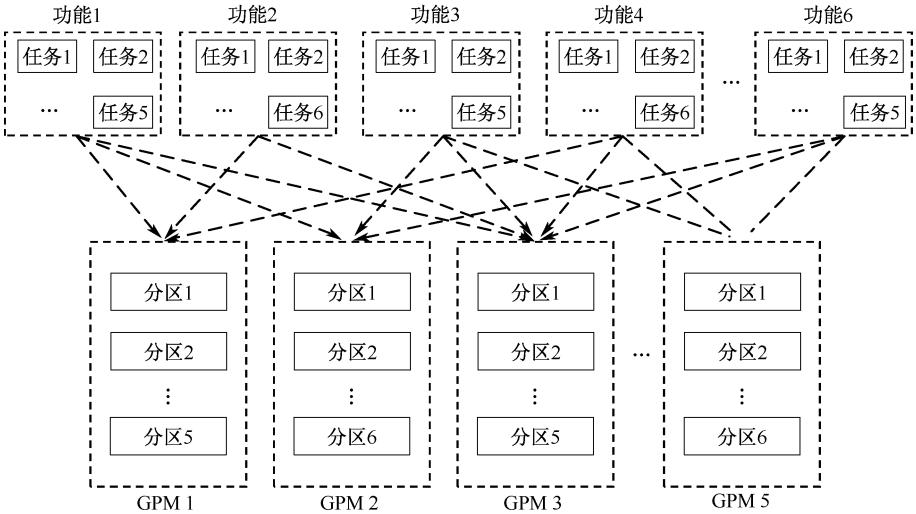

图 1　航电系统综合化设计过程

Fig. 1　Avionics system integrated design process

1.1　分区模型构建

IMA 系统可具体化为由共享的软硬件资源组成的开放式平台,系统通过时空分区机制实现资源的共享和重用。通用处理模块作为航电系统中的核心组件,负责通过航空电子全双工交换式以太网(Avionics Full-Duplex Switched Ethernet, AFDX)接收、处理和传输各种航电数据和指令。根据任务需求,将处理后的数据分发到适当分区。

分区是模块划分的不同独立部分或区域,每个部分都拥有独立的系统资源并互相隔离,使系统灵活配置能力得到最大程度的开发。每个分区可根据当前的系统配置方案分配特定任务,从而提高系统处理性能和效率。分区集合 P 表示为:

$$P = \{P_1, P_2, \cdots, P_{N_P}\} \tag{1}$$

式中 N_P 为系统分区集合 P 内所有分区数量之和。

结合具体的综合航电系统架构和相关行业标准,每个分区拥有的资源被定义为:核心处理能力和内存。因此,使用二元数组 (C_i, M_i) 表示第 i 个分区 P_i 占有的系统独立资源, C_i 表示第 i 个分区核心处理能力, M_i 表示第 i 个分区内存资源能力。

1.2　任务模型构建

航电功能涵盖了航电系统内执行的包括航迹控制、交通通信、飞行警告等一系列航空电子功能,在综合化设计中可根据飞行需求选择并集成相应的功能。任务集合 T 表示为:

$$T = \{T_1, T_2, \cdots, T_{N_T}\} \tag{2}$$

式中 N_T 为系统任务集合 T 内所有任务数量之和。

每个任务所具有的属性包括:处理能力需求、内存需求和安全关键等级。因此,使用三元数组 $(\chi_j, \kappa_j, \mathbf{s}_j)$ 表示当前选择的任务序号为 j 的任务 T_j, χ_j 表示第 j 个任务具有的处理能力需求, κ_j 表示第 j 个任务具有的内存需求, $\mathbf{s}_j$ 表示第 j 个任务具有的安全关键等级。

1.3　航电任务安全关键等级量化模型构建

在航电系统中,许多功能被描述为安全关键[12],其关键程度取决于其故障后果以及故障导致乘员或机组人员死亡的风险。基于文献[13-14],笔者拟提出一种安全关键等级的量化方法,通过定量评估任务对功能丧失的影响为每个任务分配一个安全关键等级,以便均衡分区风险,达到提高系统整体安全性的目标。

为了在综合化设计中均衡分区失效风险,首先对航电任务安全关键等级量化建模。基于文献[15],航电功能的失效状态分类是定性的,以此为据,建立航电任务定量安全关键等级模型。航电任务 T_j 的安全关键等级为:

$$\mathbf{s}_j = \sum_{m=1}^{3} n_j^m \times \omega^m \tag{3}$$

式中: n_j^m 为第 j 个航电任务所属功能的 m 类失效状态数量,对于一般不会造成不可接受风险的较小的Ⅳ类失效和无安全影响的Ⅴ类失效不作考虑,因此, $m = \{1,2,3\}$; ω^m 为 m 类失效状态风险权重,权重的赋值取决于采取的风险建模策略,主要有:

1) 灾难级失效不可容忍策略,即对Ⅰ类失效状态赋值量级远超其他失效状态类别,如取 $\omega^1 = 10\,000, \omega^2 = 10, \omega^3 = 1$。

2) 失效风险度差异赋值策略,赋值体现不同类别的风险差异,同时,各类别失效状态风险差异是可

比拟的,如取 $\omega^1 = e^3, \omega^2 = e^2, \omega^3 = e$ 。

2 安全性优化方法设计方案

2.1 安全性设计优化问题建模

基于文献[16],多分区多任务的分配问题又可描述为任务集合 T 在分区集合 P 上的一个映射问题,即:

$$T \to P \tag{4}$$

为确保制定最优综合化设计,必须有效地同步每个分区的资源能力情况,并在分配过程中考虑处理能力、内存资源可用性等约束。即将任务分配到分区时,需要保证分区的处理资源和内存资源的利用率满足标准:

$$\Im(P_i) \leqslant 1, i \in [1,2,\cdots,N_P] \tag{5}$$

$$\aleph(P_i) \leqslant 1, i \in [1,2,\cdots,N_P] \tag{6}$$

$$\Im(P_i) = \sum_{j=1}^{N_T} \chi_j \cdot \zeta_{i,j} / C_i \tag{7}$$

$$\aleph(P_i) = \sum_{j=1}^{N_P} \kappa_j \cdot \zeta_{i,j} / M_i \tag{8}$$

式中:$\Im(P_i)$ 为第 i 个分区处理资源的利用率,$\aleph(P_i)$ 为第 i 个分区内存资源的利用率;$\zeta_{i,j}$ 为二元决策变量,当第 j 个任务 T_j 被分配到第 i 个分区 P_i 时,其值为 1,否则为 0。

为降低系统潜在的安全风险,并提高对系统资源能力的高效利用,实现考虑安全性的综合化设计,定义分区风险均衡和负载均衡 2 个优化目标。

随着综合化航电系统内驻留功能数量的剧增,模块化设计对系统运行安全和性能造成极大影响,传统综合化设计方法难以保证系统配置方案满足高安全性需求,即某一分区相较于系统中其他分区而言,承载着更多高安全关键等级的任务。若由于某些原因导致分区失效,那么此分区失效对航电系统造成的影响更为严重,系统就存在更高的安全风险。因此,基于分区内驻留任务的安全关键等级,使用标准差来衡量分区潜在风险的离散程度,从而提高系统安全:

$$B_{\mathrm{R}} = \frac{1}{N_P} \sum_{i=1}^{N_P} (\rho_i - \bar{\rho})^2 \tag{9}$$

$$\rho_i = \sum_{j=1}^{N_T} \mathbf{s}_j \cdot \zeta_{i,j},\ \bar{\rho} = \frac{1}{N_P} \sum_{i=1}^{N_P} \rho_i \tag{10}$$

式中:B_{R} 为分区风险离散程度的衡量结果;ρ_i 为第 i 个分区内的潜在风险;$\bar{\rho}$ 为系统内所有分区的平均风险。

均衡各个分区负载有助于确保系统对于资源的有效利用,提高系统运行效率,同样,使用标准差的形式来具体化衡量分区负载的离散程度:

$$B_{\mathrm{L}} = \frac{1}{N_P} \sum_{i=1}^{N_P} (L_i - \bar{L})^2 \tag{11}$$

$$L_i = \varphi_1 \times \Im(P_i) + \varphi_2 \times \aleph(P_i),\ \varphi_1 + \varphi_2 = 1 \tag{12}$$

$$\bar{L} = \frac{1}{N_P} \sum_{i=1}^{N_P} L_i \tag{13}$$

式中:B_{L} 为分区负载离散程度的衡量结果;L_i 为第 i 个分区内的负载,$\bar{L}$ 是系统内所有分区的平均负载;φ_1 为处理资源的负载权重因子;φ_2 为内存资源的负载权重因子。

综上,该问题的优化目标函数定义为:

$$\min(\mu_1 \times B_{\mathrm{R}} + \mu_2 \times B_{\mathrm{L}}), \mu_1 + \mu_2 = 1 \tag{14}$$

式中:μ_1 为风险离散程度优化权重系数;μ_2 为负载离散程度优化权重系数,两者之和固定为 1。

2.2 安全性设计优化问题求解

安全性设计优化问题是一个需要考虑任务属性、分区能力等多个维度的复杂问题,且由于在优化系统安全性的同时还要考虑系统资源的高效利用,使得问题耦合程度加深,为求解带来进一步的困难。因此,选择将问题表述为 MDP 问题[17],并使用基于深度强化学习的(Soft Actor-Critic,SAC)算法,求解满足所有需求的最优综合化设计方案。

综合化安全性设计问题可正式描述为由五元组 (S,A,Pr,R,γ) 组成的 MDP 问题,其中,状态为 S、动作为 A、转移概率 Pr、奖励函数 R 和折扣因子 γ 。

2.2.1 状态空间

状态空间的设计应该综合考虑综合化设计问题的具体需求,并且需要避免状态空间过大导致计算复杂度过高的问题。因此,在实际设计中需要进行适当抽象和简化,以确保能够有效地进行强化学习训练和决策。状态空间设计如下:

1)需要分配的第 j 个任务及其具有的属性:

$$T_j = (\chi_j, \kappa_j, \mathbf{s}_j) \tag{15}$$

2)系统内的第 i 个工作分区及其占有的系统资源:

$$P_i = (C_i, M_i) \tag{16}$$

因此,t 时刻下状态空间定义为:$S_t = \{T, P\}$ 。

2.2.2 动作空间

动作属于综合化设计中的分配过程,即选择当前状态下需要处理的任务以及正常工作的分区,根据指定的策略对其进行分配。该动作定义为:

$$A_t = (T_j \to P_i \mid 1 \leqslant j \leqslant N_T, 1 \leqslant i \leqslant N_p) \tag{17}$$

2.2.3　转移概率

在马尔可夫过程问题建模中,评估执行特定动作时从一种状态转移到另一种状态的概率是非常关键的。有时,从一种状态到另一种状态的转变是完全可预测的,表明该动作与后续状态之间存在直接对应关系。然而,许多现实场景表现出更多的随机行为(即存在的动态性),其中,转移概率是非确定性的。在解决具有非确定性转移概率的 MDP 问题时,通常会采用随机性建模来模拟现实世界中的动态过程。这种情况下,传统的确定性转移概率被替换为非确定性转移概率,使得在执行特定动作后,从一个状态转移到另一个状态的概率成为一个随机变量。因此,为最大程度模拟综合化设计中的动态过程,在求解过程中采取非确定性的转移概率 $Pr(s'|s,a)$ 设计。

2.2.4　奖励函数

奖励函数的设计在强化学习中起着至关重要的作用,用于提供智能体执行特定动作后所获得的奖励信号,以评估状态转移的质量。在综合化安全性设计中,奖励函数的设计应该能够体现对系统安全性和资源利用效率的关注,以便在训练过程中引导智能体更好地学习如何均衡各个分区内驻留任务的安全关键等级来最小化系统潜在的安全风险,并确保计算资源和内存资源的高效利用。因此,t 时刻下奖励函数定义如下:

$$R_t=\begin{cases}-\sum_{i=1}^{N_P}\left[\rho_i-\sum_{i=1}^{N_P}\rho_i/N_P\right]^2 & \text{满足约束}\\ -M_i & \exists\aleph(P_i)>1\\ -C_i & \exists\Im(P_i)>1\end{cases} \tag{18}$$

2.3　基于 SAC 算法的优化方法设计生成

当前,由于基于在线策略的主流的近端策略优化(Proximal Policy Optimization, PPO)算法和异步优势演员-评论家算法(Asynchronous Advantage Actor-Critic, A3C)算法在每个梯度步骤都需要大量样本来学习,导致算法取样效率较低。此外,虽然如深度确定性策略梯度(Deep Deterministic Policy Gradient,DDPG)之类的离线策略算法比 PPO 算法的取样效率更高,但它们对其超参数敏感并且收敛效果较差。为获得最优策略,必须选择合适的深度强化学习算法彻底探索环境,而不是只优先考虑具有最高奖励的行动,从而导致算法陷入局部最优。

SAC 算法是 HAARNOJA 等[18]提出的一种基于最大熵的强化学习算法。通过在目标函数中引入最大熵正则化项促使算法在学习过程中保持探索性,来提高算法的鲁棒性和取样效率[19]。工作原理是重复选择一个动作会导致熵降低,使得智能体在学习过程中不仅考虑奖励最大化,还要考虑策略的多样性,从而扩大算法的探索范围。同时,这种探索性的引入有助于避免算法陷入局部最优,进一步提高算法的收敛性和稳定性。

Actor-Critic 框架下,SAC 算法部署了 2 组神经网络来构建价值网络和策略网络。价值网络的训练目标是最小化(Temporal Difference,TD)误差,以使其能够准确地估计状态-动作对的长期回报。策略网络的训练目标是最大化预期回报,以使得在不同状态下选择的动作能够最大化长期回报。图 2 描述了策略网络、价值网络架构。

策略网络的输入为归一化后的航电系统任务和

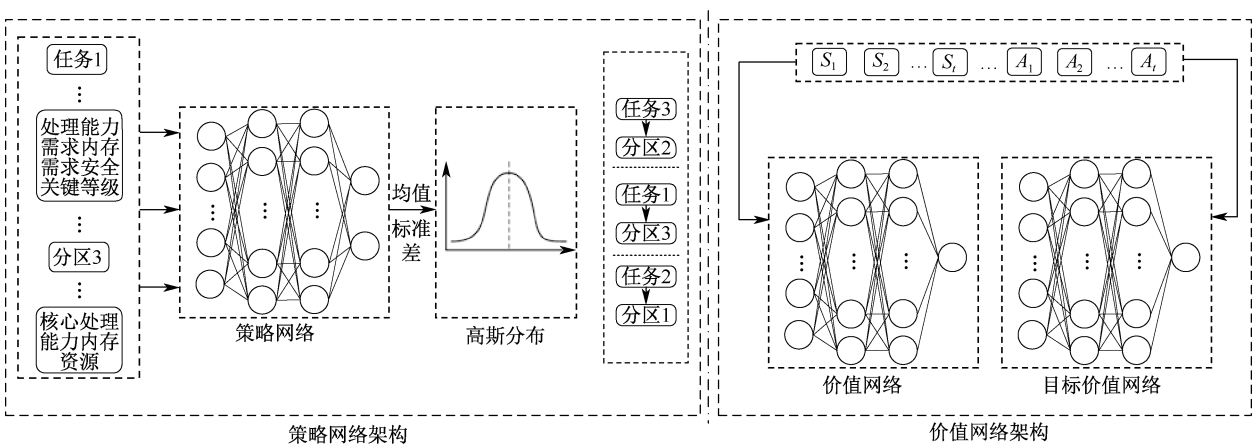

图 2　策略网络和价值网络架构

Fig. 2　Policy network and value network architecture

分区的状态信息,输出为高斯采样得到的[-1,1]区间内的动作值,并实时映射为分配的任务序号及目的分区序号。价值网络和目标价值网络在 SAC 算法中通常采用相同的结构,这 2 个网络都用于评估状态-动作对的价值函数,以帮助目标网络通过软更新的形式更新参数。

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta' \tag{19}$$

式中:θ 为目标网络参数;θ' 为在线网络参数;τ 为软更新参数。

相比于 PPO 和 A3C 算法等基于在线策略的主流强化学习算法,SAC 算法更适用于连续动作空间的环境,在处理这种类型的问题时表现更加优秀。SAC 算法的目标是通过权衡预期奖励和熵,最大化奖励和探索性之间的关系,从而更好地解决深度强化学习中的探索与利用的平衡问题。考虑最大熵机制的航电系统综合化设计策略优化问题表述如下:

$$\pi^* = \arg\max_{\pi} E_{\vartheta\sim\pi}\left[\sum_{t=0}^{T}\gamma^t(R_t + \alpha H(\pi(\cdot\mid S_t)))\right] \tag{20}$$

式中:$E_{\vartheta\sim\pi}[\cdot]$ 为 MDP 中基于策略 π 的状态价值函数;ϑ 为轨迹,$\vartheta = (S_0,A_0,S_1,A_1,\cdots,S_t,A_t)$;$R_t$ 为 t 时刻下预期奖励;γ^t 为时刻 t 下折扣因子;S_t 为 t 时刻的状态;$H(\pi(\cdot\mid S_t))$ 为策略熵的定义;α 为用于制定预期奖励和策略熵之间权衡的温度系数,可通过下述公式自动调整:

$$J(\alpha) = E_{A_t\sim\pi}[-\alpha\ln\pi(A_t\mid S_t) - \alpha H_0] \tag{21}$$

式中:H_0 为策略初始熵,A_t 为 t 时刻采取的动作。

此外,策略熵的定义表述如下:

$$H(\pi(\cdot\mid S_t)) = -\sum_{a}\pi(\cdot\mid S_t)\ln\pi(\cdot\mid S_t) \tag{22}$$

在最大熵机制的框架下,软状态-动作值和软状态值可以重定义为:

$$\begin{aligned} Q(S_t,A_t) &= E_{\vartheta\sim\pi}\left[\sum_{t=0}^{T}\gamma^t(R_t + \alpha H(\pi(\cdot\mid S_t)))\right] \\ V(S_t) &= E_{\vartheta\sim\pi}[Q(S_t,A_t) - \alpha\ln\pi(\cdot\mid S_t)] \end{aligned} \tag{23}$$

式中:$Q(S_t,A_t)$ 为 t 时刻状态和动作下算法的软状态-动作值;$V(S_t)$ 为 t 时刻状态下算法的软状态值。

最后,根据 KL(Kullback-Leibler)散度公式推导出实现最优策略的闭式解:

$$\pi_{\mathrm{B}} = \arg\min_{\pi^*} D_{\mathrm{KL}}\left\{\pi^*(\cdot\mid S_t)\,\|\,\exp\left(\left(\frac{1}{\alpha}(Q(S_t,\cdot) - V(S_t))\right)\right)\right\} \tag{24}$$

式中:D_{KL} 为 KL 散度计算公式;π_{B} 为最优策略的闭式解。

3　基于优化方法的综合化设计案例

3.1　IMA 系统仿真试验设置

用于仿真试验的任务案例是基于 IMA 系统的分区配置生成,总共设置有 15 个任务和 8 个分区。表 1 为与综合航电系统设计相关的分区基本配置参数,包括核心处理能力和内存资源。

表 1　分区配置相关参数

Table 1　Partition configuration related parameters

分区 ID	核心处理能力/GHz	内存资源/kb
P_1	16	128
P_2	8	256
P_3	9.6	128
P_4	9.6	128
P_5	12.8	64
P_6	11.2	512
P_7	16	128
P_8	16	256

为验证综合化设计方法对不同的系统配置方案的支持能力,构造 3 组任务案例,见表 2。分别标记为组 1、组 2 和组 3,其中包括每个任务的处理能力需求和内存需求。这 3 组试验中,每个任务的安全关键等级相同,区别是任务的处理能力需求和内存需求不同。

表 2　仿真试验的航电任务相关参数设置

Table 2　Avionics mission-related parameter settings for simulation experiments

任务 ID	组 1		组 2		组 3	
	处理能力/GHz	内存/kb	处理能力/GHz	内存/kb	处理能力/GHz	内存/kb
T_1	7.2	90	3.7	35	5.1	60
T_2	7.7	50	4.0	60	1.0	50
T_3	3.1	35	3.9	50	5.6	30
T_4	7.3	60	1.7	55	4.2	35
T_5	5.4	50	1.4	45	7.4	65
T_6	8.3	30	1.3	60	7.6	35
T_7	5.9	35	4.3	70	2.7	50
T_8	6.7	65	2.0	40	1.6	35
T_9	2.8	55	7.7	30	4.1	40

续表 2

任务 ID	组 1		组 2		组 3	
	处理能力/GHz	内存/kb	处理能力/GHz	内存/kb	处理能力/GHz	内存/kb
T_{10}	5.4	95	2.8	55	4.9	70
T_{11}	9.2	70	1.6	50	9.5	30
T_{12}	4.1	40	7.6	35	4.8	75
T_{13}	1.5	35	3.0	60	4.8	60
T_{14}	6.9	50	8.9	80	1.7	50
T_{15}	2.6	60	4.2	55	3.8	60

表 3 为任务所属航电功能的失效状态数量。失效状态风险权重的赋值采取失效风险度差异赋值策略,即 $w^1=25, w^2=5, w^3=1$。

表 3　任务所属航电功能的失效状态数量

Table 3　Number of failure states of avionics function to which task belongs

任务 ID	所属功能的失效状态数量			任务 ID	所属功能的失效状态数量		
	Ⅰ类	Ⅱ类	Ⅲ类		Ⅰ类	Ⅱ类	Ⅲ类
T_1	1	5	4	T_9	1	0	3
T_2	0	4	5	T_{10}	3	0	5
T_3	2	1	4	T_{11}	2	3	0
T_4	2	4	2	T_{12}	0	4	5
T_5	0	1	3	T_{13}	1	0	4
T_6	2	0	5	T_{14}	2	1	4
T_7	2	5	6	T_{15}	2	2	3
T_8	1	0	4	—	—	—	—

依据式(3)计算可得各任务的安全关键等级,见表 4。由于文中重点是解决综合化设计中的安全性问题,优化目标更注重于分区潜在风险的离散程度,权重系数 μ_1 和 μ_2 分别设置为 0.7 和 0.3。

表 4　航电任务安全关键等级量化结果

Table 4　Quantified results of avionics task safety criticality levels

任务 ID	安全关键等级	任务 ID	安全关键等级	任务 ID	安全关键等级
T_1	54	T_6	55	T_{11}	65
T_2	25	T_7	81	T_{12}	25
T_3	59	T_8	29	T_{13}	29
T_4	72	T_9	28	T_{14}	59
T_5	8	T_{10}	80	T_{15}	63

3.2　参数灵敏度分析

在强化学习中,灵敏度分析可通过改变算法的参数值,观察对最终结果的影响程度,来确定最优的参数选择。具体来说,选择需要进行灵敏度分析的价值网络 Q 学习率参数 λ_Q,策略网络 π 学习率参数 λ_π,权衡温度系数为 α 的熵相关学习率参数 λ_α 以及奖励折扣因子 γ。通过在不同参数取值下运行算法,观察算法的性能变化(即奖励曲线的变化)。

首先,以任务案例组 1 为仿真对象,在固定奖励折扣因子 $\gamma=0.9$ 的情况下,取不同学习率训练,结果如图 3 所示。

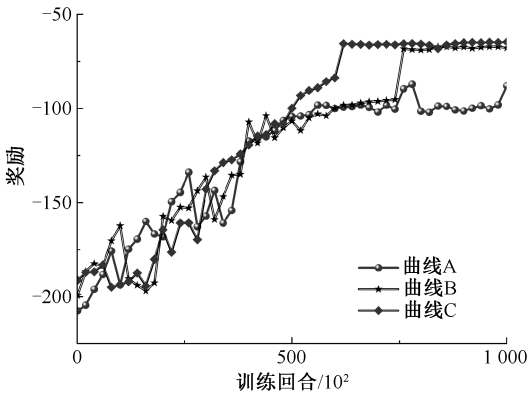

图 3　多学习率参数的奖励曲线对比

Fig. 3　Comparison chart of reward curves for multiple learning rate parameters

其中,曲线 A 的学习率参数设置为 $\lambda_Q=0.001$, $\lambda_\pi=0.002$, $\lambda_\alpha=0.002$,曲线 B 的学习率参数设置为 $\lambda_Q=0.01$, $\lambda_\pi=0.02$, $\lambda_\alpha=0.02$,曲线 C 的学习率参数设置为 $\lambda_Q=0.005$, $\lambda_\pi=0.015$, $\lambda_\alpha=0.015$。由图 3 中可以看出,学习率参数设置过小的情况下,当训练回合达到 100 000 次时,曲线 A 仍然没有达到稳定收敛的状态,还在震荡上升。同时,曲线 C 的收敛时间相较于曲线 B 缩短 1/6,且收敛效果也有略微提升,这表明在学习率参数设置适当的情况下,模型能够更快地收敛并达到更好的效果。因此,学习率的参数确定为相对最佳的曲线 C 参数。

在相同环境和任务案例下,使用确定的最佳学习率参数分别运行不同奖励折扣因子的仿真试验,结果如图 4 所示。

其中,曲线 D、E、F、G 和 H 的奖励折扣因子 γ 分别设置为 0.9, 0.75, 0.5, 0.3, 0.01。深色和浅色曲线分别表示 5 次试验下的累积奖励的平均值和边界。根据折线图分析,奖励曲线大致都在 70 000 训练回合内完成上升,进入稳定收敛状态,但曲线 D 和 F 的收敛速度相较于其他曲线有明显提高。同时,在收敛效果上曲线 F 的奖励稳定最大值相比其他曲线也有 8%左右的提升。因此,在后续仿真试验中设置奖励折扣因子 $\gamma=0.5$ 可加快收敛速度,并且在一定程度上提高最终的稳定收敛效果。

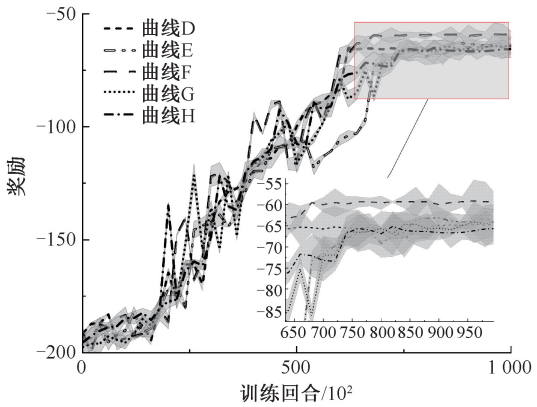

图 4　奖励折扣因子的奖励曲线对比

Fig. 4　Reward curve comparison chart of discount factors

3.3　试验结果分析

基于灵敏度分析的试验参数,使用 SAC 算法综合化设计航电系统。为对比分析突显深度强化学习算法以及 SAC 算法的优势,提出基于 DDPG 的深度强化学习算法以及传统分配算法等多种优化算法。深度强化学习算法方面,在 SAC 和 DDPG 算法稳定收敛的情况下,选择最优奖励的综合化设计方案并分析在此方案下的航电系统配置。根据 3.1 节设置的 3 组任务案例进行仿真分析,3 组试验的平均结果见表 5。其中,传统分配算法的选择依据经典的最佳适应算法和循环首次适应算法,因此,对于两者的最大奖励以及约束累计违背率不作考虑。同时,深度强化学习算法的求解时间,以达到稳定收敛状态的时间为依据。

表 5　算法优化效果

Table 5　Algorithm optimization effect

算法选择		综合化设计优化效果				
		最大奖励	约束累计违背率/%	分区风险标准差	分区资源利用率标准差	求解时间/s
深度强化学习	SAC	−59.73	67.91	32.85	16.83	7.69
	DDPG	−96.52	5 137.35	123.74	90.27	67.33
传统分配算法	最优适应	—	—	2 286.5	374.23	2.3
	循环首次适应	—	—	2 782.3	17.09	1.7

由表 5 可知:在相同的参数设置下,使用基于 SAC 算法的优化方法获得的最大奖励、约束累计违背率、分区均衡风险效果、分区资源利用及求解时间,相较于 DDPG 算法均有显著优势。此外,虽然求解时间的耗费略高,但是,对于分区风险的均衡效果远高于传统的分配算法,而且,也显著提高了航电系统分区资源的利用。基于上述分析可知:基于 SAC 算法的优化方法对于考虑安全性的航电系统综合化设计方法优化效果明显。

4　结　论

1) 基于 SAC 算法的综合航电系统优化方法能够有效优化系统设计,显著改善系统性能,更好地管理系统的约束条件并降低潜在风险。

2) 基于 SAC 算法的优化方法可显著提高分区资源的利用效率,保障系统的整体效能。

3) 基于 SAC 算法的优化方法在求解过程中展现出了更高的效率,具有较短的收敛时间,在实际应用中更具可行性和实用性。在未来的工作中,可以考虑使用更先进的问题建模方式以及优化算法。

参 考 文 献

[1] WANG Hongli, ZHONG Deming, ZHAO Tingdi, et al. Integrating model checking with SysML in complex system safety analysis[J]. IEEE Access, 2019, 7: 16 561-16 571.

[2] 赵长啸,汪克念,张伟,等. 民机航电系统功能-信息安全一体化分析方法[J]. 中国安全科学学报,2022,32(9): 49-56.
ZHAO Changxiao, WANG Kenian, ZHANG Wei, et al. Integrated analysis method of function safety and cyber security of avionics system for civil aircraft[J]. China Safety Science Journal, 2022, 32(9): 49-56.

[3] SAE ARP4761A, Guidelines for conducting the safety assessment process on civil aircraft, systems, and equipment[S]. 2023.

[4] KHAMVILAI T, SUTTER L, BAUFRETON P, et al. Decentralized task reallocation on parallel computing architectures targeting an avionics application[J]. Journal of Optimization Theory and Applications, 2021, 191(2/3): 874-898.

[5] LU Hui, ZHOU Qianlin, FEI Zongming, et al. Scheduling based on interruption analysis and PSO for strictly periodic and preemptive partitions in integrated modular avionics[J]. IEEE Access, 2018, 6: 13 523-13 540.

[6] ZHOU Tianran, XIONG Huagang, ZHANG Zhen. Hierarchical resource allocation for integrated modular avionics systems[J]. Journal of Systems Engineering and Electronics, 2011, 22(5): 780-787.

[7] ZHOU Xuan, XIONG Huagang, HE Feng. Hybrid partition-and network-level scheduling design for distributed integrated modular avionics systems[J]. Chinese Journal of Aeronautics, 2020, 33(1): 308-323.

[8] POLYDOROS A S, NALPANTIDIS L. Survey of model-based reinforcement learning: applications on robotics[J]. Journal of Intelligent & Robotic Systems, 2017, 86(2): 153-173.

[9] LI Dong, ZHAO Dongbin, ZHANG Qichao, et al. Reinforcement learning and deep learning based lateral control for autonomous driving[J]. IEEE Computational Intelligence Magazine, 2019, 14(2): 83-98.

[10] BARRETT E, HOWLEY E, DUGGAN J. Applying reinforcement learning towards automating resource allocation and application scalability in the cloud[J]. Concurrency and Computation: Practice and Experience, 2013, 25(12): 1 656-1 674.

[11] 魏明,孙雅茹,孙博,等. 基于深度强化学习的无人机线路及航迹协同规划[J]. 中国安全科学学报,2023,33(8): 68-76.
WEI Ming, SUN Yaru, SUN Bo, et al. UAV distribution route and flight path collaborative planning based on deep reinforcement learning[J]. China Safety Science Journal, 2023, 33(8): 68-76.

[12] BARON C, LOUIS V. Towards a continuous certification of safety-critical avionics software[J]. Computers in Industry, 2021: DOI: 10.1016/j.compind.2020.103382.

[13] GAO Yuan, LIU Hu, TIAN Yongliang. Inverse design of mission success space for combat aircraft contribution evaluation[J]. Chinese Journal of Aeronautics, 2020, 33(8): 2 189-2 203.

[14] GAO Yuan, TIAN Yongliang, LIU Hu, et al. Entropy based inverse design of aircraft mission success space in system-of-systems confrontation[J]. Chinese Journal of Aeronautics, 2021, 34(12): 99-109.

[15] 赵长啸,何锋,阎芳,等. 面向风险均衡的 AFDX 虚拟链路路径寻优算法[J]. 航空学报,2018,39(1):261-272.
ZHAO Changxiao, HE Feng, YAN Fang, et al. Path optimization algorithm of AFDX virtual link to balance the network risk[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(1): 261-272.

[16] 赵长啸,李道俊,汪鹏辉,等. 基于 DDPG 的综合化航电系统多分区任务分配优化方法[J]. 电讯技术,2024,64(1): 58-66.
ZHAO Changxiao, LI Daojun, WANG Penghui, et al. A DDPG-based optimization method for multi-partition task assignment of IMA[J]. Telecommunication Engineering, 2024, 64(1): 58-66.

[17] PUTERMAN M L. Markov decision processes[J]. Handbooks in Operations Research and Management Science, 1990, 2: 331-434.

[18] HAARNOJA T, ZHOU Aurick, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor[C]. Proceeding in International Conference on Machine Learning, PMLR, 2018: 1 861-1 870.

[19] 付宇鹏,邓向阳,朱子强,等. 基于模仿强化学习的固定翼飞机姿态控制器[J]. 海军航空大学学报,2022,37(5): 393-399.
FU Yupeng, DENG Xiangyang, ZHU Ziqiang, et al. Imitation reinforcement learning based attitude controller for fixed-wing aircraft[J]. Journal of Naval Aviation University, 2022, 37(5): 393-399.

作者简介: 赵长啸 (1989—),男,山东临清人,博士,副教授,主要从事综合化航电系统性能评估与适航设定技术研究。E-mail:cxzhao@cauc.edu.cn。