

航空混合动力系统能量管理策略研究综述

张丁予¹, 沈挺^{1,2}

(1. 重庆交通大学 航空学院, 重庆 402200; 2. 绿色航空能源动力重庆市重点实验室, 重庆 401120)

摘要: 能量管理策略作为航空混合动力系统的顶层控制, 用于对混合动力系统不同动力源进行能量分流, 是保证系统高效运行的基础。详细论述了各类航空混合动力系统能量管理策略, 系统总结了基于规则、优化和智能 3 类能量管理策略的特点和研究现状。通过阐述强化学习原理, 分析了深度 Q 网络算法和深度确定性策略梯度算法的奖励原理、神经网络更新原理、以及各自优缺点及适用场景, 并提出基于规则类能量管理策略对于专家经验依赖性较高等缺陷, 可以通过将其与基于智能算法内部创新融合进行缓解的措施。在此基础上, 展望了能量管理策略的未来发展趋势为智能内部算法、智能与其他类型算法的融合创新等, 可以为后续航空混动系统能量管理策略研究提供一定的参考。

关键词: 能量管理策略; 深度 Q 网络算法; 深度确定性策略梯度算法; 强化学习; 航空混合动力系统

中图分类号: V233.7

文献标识码: A

doi: 10.13477/j.cnki.aeroengine.2025.01.002

Review of Research on Energy Management Strategies for Hybrid Aircraft Propulsion System

ZHANG Ding-yu¹, SHEN Ting^{1,2}

(1. School of Aeronautics, Chongqing Jiaotong University, Chongqing 402200, China;

2. Chongqing Key Laboratory of Green Aviation Energy and Power, Chongqing 401120, China)

Abstract: The energy management strategy, as the top-level control of a hybrid aircraft propulsion system, is utilized to distribute energy among different power sources. It is considered the foundation for ensuring the efficient operation of the system. The energy management strategies of various types of hybrid aircraft propulsion systems were elaborately discussed, and the characteristics and research status of three types of energy management strategies based on rules, optimization, and learning were systematically summarized. By describing the principle of reinforcement learning, the reward performance, neural network updating principle, respective advantages and disadvantages, and applicable scenarios of the deep Q-network algorithm and deep deterministic policy gradient algorithm were analyzed. It is pointed out that the deficiency of rule-based energy management strategies, which relied heavily on expert experiences, can be mitigated by integrating with learning-based approaches. On this basis, the future development trends of energy management strategies are envisioned to focus on internal innovation of learning-based algorithms and integration innovation with different types of algorithms. This can provide references for subsequent research on energy management strategies in hybrid aircraft propulsion systems.

Key words: energy management strategy; deep Q-network; deep deterministic policy gradient algorithm; reinforcement learning; hybrid aircraft propulsion system

0 引言

全球航空业在过去几十年里取得了巨大的进步, 成为了全球经济和人类社会的重要支柱之一。然而, 随着航空业的快速发展, 其对环境的负面影响也日益凸显。在此背景下, 混合动力系统凭借碳排放量低、噪声小和效率高等优势, 成为应对航空业碳排放挑战

的方法之一^[1-2]。由于混合动力系统具有多个动力源, 在提高飞行器动力性能的同时, 也增大了对能量合理分配与管理的需求^[3-4]。因此, 良好的能量管理策略 (Energy Management Strategy, EMS) 是保证系统高效运行的重要基础, 对于提高系统能量利用率、延长系统使用寿命、降低碳排放率都有着重要的作用^[5-6]。

混合动力系统的 EMS, 通常根据基于动力装置效

收稿日期: 2024-04-09 基金项目: 绿色航空能源动力重庆市重点实验室开放基金 (GATRI2021F01002B)、重庆市教委科学技术研究项目 (KJQN202100743) 资助

作者简介: 张丁予 (1999), 女, 在读硕士研究生。

引用格式: 张丁予, 沈挺. 航空混合动力系统能量管理策略研究综述[J]. 航空发动机, 2025, 51(1): 12-20. ZHANG Dingyu, SHEN Ting. Review of research on energy management strategies for hybrid aircraft propulsion system[J]. Aeroengine, 2025, 51(1): 12-20.

率或燃料消耗量的经济性优化、动力源退化机理的耐久性优化、成本函数或损失函数的多目标优化,进行功率分配、电池系统管理、可靠性和控制优化的同时,控制供给负载的功率输出^[7-9]。EMS 可以分为基于规则、优化以及智能 3 大类^[10],但目前应用较多的是将不同种类算法相结合的 EMS。张雪等^[11]为解决传统模糊控制策略控制精度较低、动态求解最优解能力较差的问题,提出了一种基于粒子群优化的模糊逻辑控制算法,最大化排除了在策略制定过程中由于专家经验导致的主观局限性,降低了电池损耗以及燃油消耗;Chen 等^[12]提出了一种基于双神经网络的等效消耗最小策略,以及一种等效因子自调节的修正方法,可以在不需要电池荷电状态参考的情况下,实现接近最优燃油经济性的目标;Anselma^[13]提出了一种使用斜率加权的快速动态规划算法,与传统动态规划算法相比,该算法在显著提高计算效率的同时,保持了动态规划算法对于燃油消耗控制的有效性;Hegde 等^[14]提出了一种新的等效消耗最小策略,包含了混合动力传动系统的功率损失图,并设计了基于模型的能量流监控器,用于瞬时最优的功率分配,以减少燃料消耗,并使用遗传算法整定。混动系统能量管理在汽车领域已经得到较为深入的研究。考虑到在实际航空应用中,动力系统应具有更高的安全性以及可靠性,飞机应具有更高的功重比、自适应等要求,有必要根据油电混合动力系统的特殊性,结合不同 EMS 的使用场合以及优缺点,对其进行设计与优化^[15]。但在航空领域,基于混合动力系统的新能源飞机能量管理策略——尤其是具备自适应识别的智能能量管理策略研究较少。

本文对深度 Q 网络算法、深度确定性策略梯度算法这 2 种常用的基于智能 EMS 算法进行底层原理梳理,介绍了国内外学者对于基于规则、优化和智能 3 类 EMS 的研究进展,并分析比较各自特点,以期为航空混合动力系统 EMS 的设计与优化提供参考。

1 基于规则的能量管理策略

基于规则的 EMS 主要来自于实际的应用以及经验总结,由一系列仅依赖于过去及实时系统状态的逻辑状态组成,通过这些逻辑状态设定运行规则,进而决定系统的运行方式,可以实现实时监督控制^[16]。根据规则的性质,可以将其划分为确定规则控制和模糊

规则控制。其中,前者的计算量较小,技术难度较低;后者的鲁棒性及实时性更优,适应性更好。但由于二者对工程师经验的依赖性较高,难以实现全局最优。

1.1 状态机控制策略

状态机也被称为有限状态机 (Finite-State Machine, FSM),当状态机的状态集、输入集以及转移函数都确定时,可以将其归类为基于确定规则的 EMS 中。FSM 模型通过检测电池组的荷电状态 (State of Charge, SOC) 以及负载需求功率^[17],按照提前设定的规则,通过 if-then 条件管理受控对象,使其切换不同的能量分配模式^[18]。

FSM 控制的优点为简单、可靠、可扩展性强。但由于只适用于描述有限个状态的问题,且在运行过程中,系统的行为是通过分散为离散的状态,并根据输入转换状态,所以在面对复杂性较高的问题时,采用 FSM 控制会存在大量状态并产生大量状态转化,此时便会降低程序的理解性以及维护性。同时,FSM 控制存在无法应对连续性问题、对专家经验依赖性较高等缺陷。但其对工程师经验的过度依赖性可以通过与其他 EMS 结合解决。Li 等^[19]提出了一种基于等效消耗最小策略的状态机控制法,通过将 2 种控制策略有效结合,提高了等效消耗最小策略的性能,同时也弱化了状态机策略的此类缺点,最终通过搭建 RT-LAB 半实物硬件平台,验证所提策略的氢气消耗量比纯状态机控制策略的降低了 4.18%。

1.2 模糊逻辑控制策略

在油电混合动力系统中,功率源较传统动力系统有所增大,因而增加了状态机控制策略的设计难度,系统的鲁棒性难以满足,需要进一步发展状态机控制策略,因此,产生了模糊逻辑控制策略 (Fuzzy Logic Control, FLC)。FLC 不需要制定确定规则,只需要根据实际情况制定模糊规则,因此,FLC 可以解决精确度不高的控制问题,增强系统的鲁棒性,通过计算模糊函数实现能量分配^[20],FLC 流程如图 1 所示^[11]。

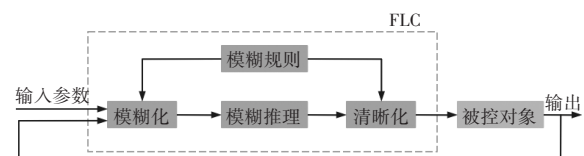


图 1 FLC 流程

李元旭等^[21]提出了一种可以加强超级电容器供电作用的模糊逻辑控制策略,并基于该控制策略设计

了一种削减波动平抑现象的功率修正方案,通过仿真证明超级电容器在无人机重载状态下作为辅助电源的有效性,以及无人机处于功率突变状态时,吸收波动功率,维护系统平稳供电的可靠性。

由于 FLC 控制仍归类于基于规则类 EMS 策略,因此在制定模糊规则时仍对专家经验存在较高的依赖性,仍然无法实现全局优化。基于规则的 EMS 存在的这一问题都可以通过将其与基于智能的 EMS 相结合得以改善。此外,针对 FLC 存在的这一问题还可以通过扩展经典模糊逻辑控制解决:(1)将 FLC 与粒子群算法^[22]、遗传算法^[23]、神经网络^[24]等结合,调整定义隶属函数以及模糊规则,提高动力系统的性能;(2)将神经网络模型与 FLC 控制模型相结合,重新定义最佳隶属函数以及重新制定系统的模糊规则,进而更改输入和输出之间的关系,可以达到改善 FLC 策略性能的目的^[25]。

2 基于优化的能量管理策略

基于优化的 EMS 需要采集大量的数据,建立目标函数,通过对其进行全局或局部的最优解求解得到系统的最佳功率,进而实现提高混合动力系统性能的目的。

基于优化的控制策略可以分为全局优化型和实时优化型 2 种。动态规划(Dynamic Programming, DP)^[26]是采用较多的全局优化型算法,计算量较大,但应用条件较为苛刻,多应用于离线控制;等效消耗最小(Equivalent Consumption Minimization Strategy, ECMS)算法^[27]是应用较广泛的实时优化型算法,可以实现系统的实时控制,在保证能耗或功率损耗最小的前提下,控制系统可以保持在瞬时最佳工作点有效运转。

2.1 动态规划算法

DP 采用自底向上的方法,将系统中复杂的非线性问题转化为数个子问题,进而解决能量管理中的非因果问题,是应用最为广泛的全局优化型 EMS,DP 流程如图 2 所示^[11]。DP 主要可以实现 2 种功能:(1)分配系统功率,提高系统燃油经济性;(2)对其他 EMS 性能进行评估,优化控制规则^[27-28]。

胡春明等^[29]通过商业软件对无人机油电混合动力系统进行模拟仿真,对比了 DP、确定规则控制策略和模糊逻辑控制策略的性能,仿真结果表明,DP 的累

积燃油消耗量相比于二者分别下降了 4.6% 和 6.5%,在随机紊流扰动情况下,DP 可以更好的发挥蓄电池“削峰填谷”的作用,减小发动机的最大转速波动。

DP 属于非因果控制方法,即在非因果系统中,输出与未来的输入有关,因此,将其应用在航空系统 EMS 中时,需要提前预知整个飞行工况^[16],仅通过 DP 难以实现。若要解决此问题,可以将 DP 与机器学习相结合,根据历史飞行数据判断当前飞行情况。

2.2 基于等效消耗最小的 EMS

基于 ECMS 的 EMS 最早被应用在传统油电混合动力汽车上。在油电混动系统中,整机的能量消耗来自发动机,电池发挥其“削峰填谷”作用被当作一种能量缓冲装置,而电机的电量消耗被折算为油耗,通过不断调节油电等效因子的取值保证电池电量的基本平衡,进而优化燃油消耗量以及碳排放。需要建立最小化实时成本函数,由此得到等效成本函数,将全局优化问题转换为实时优化问题,达到发动机油耗与电机等效油耗之和实时处于最小值的目的。

Boukoberine 等^[30]出于增加飞行器续航时间以及节约氢能的目的,采用 ECMS 算法对飞行器进行能量管理,并通过实际飞行情况所得数据进行仿真模拟,验证该算法在实现上述目的的同时降低了系统的运营成本。

ECMS 由庞特里亚金最小值原理发展而来,因此其部分缺陷也会体现在 ECMS 中。在复杂的非线性系统中,存在随机或不确定因素的影响时,需要进行多目标优化或者实时控制,ECMS 都不再适用。例如,对需要在强烈湍流中飞行的飞机进行 EMS 设计时,飞机会受到强烈且随机的外部扰动,会导致其运动方程具有较高的非线性,难以进行 ECMS 设计。

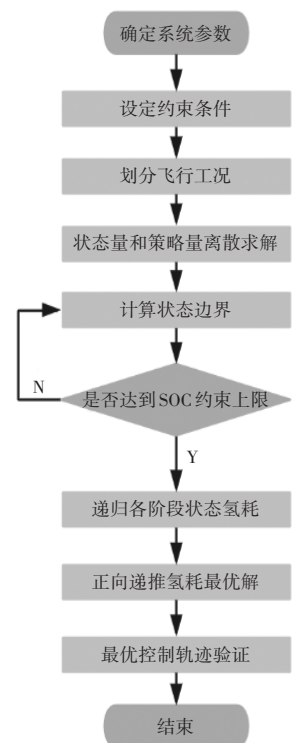


图 2 DP 流程

3 基于智能的能量管理策略

能量管理需要在大量的数据信息中抓取并整合数据集,机器学习(Machine Learning, ML)在数据挖掘及处理方面优势明显,可以将其应用在航空混合动力系统 EMS 中。

ML是实现人工智能的方法之一,是一种从经验中自动改变系统性能的程序,根据反馈信息的不同对其分类:当反馈信息中包含有标签的数据时,被称为监督学习(Supervised Learning, SL);当反馈信息中不包含任何关于模型输出值的信息,即训练集中的数据均为无标签数据时,需要人为设定目标展开学习,被称为无监督学习(Unsupervised Learning, UL);当反馈信息仅包含对于模型的打分值时,被称为强化学习(Reinforcement Learning, RL)。而深度学习(Deep Learning, DL)是一种基于人工神经网络(Artificial Neural Network, ANN)的机器学习方法,其优势在于可以通过多层次的神经网络结构学习并提取数据的特征。ML与DL关系如图3所示。

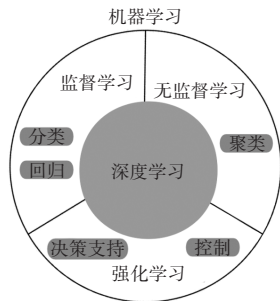


图3 ML与DL关系

3.1 强化学习

RL是一种通过智能体与环境进行交互,根据所获得的反馈不断调整既定策略,将智能体从环境中得到的回报最大化,最终使其获得最优策略的算法,RL原理框架如图4所示。RL问题都可以通过马尔可夫决策过程(Markov Decision Process, MDP)转化为数学概率问题。RL算法包括以下7部分:智能体(Agent)、环境(Environment)、状态(State)、动作(Action)、策略(Policy)、奖励与回报(Reward & Return)、价值函数(Value Function)^[31]。

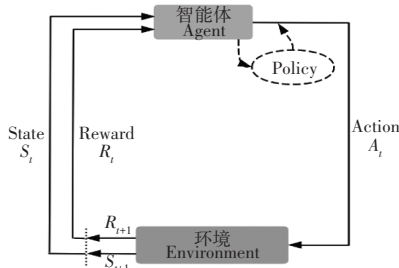


图4 RL原理框架

其中,策略 $\pi(a'|s_t)$ 用于指导智能体在当前状态应采取的动作,本质上是一种概率密度函数,表示智能体在状态 s_t 时,采取任一动作 a' 的概率

$$\pi(a'|s_t) = P(A_t = a'|S_t = s_t) \quad (1)$$

智能体在采取动作 a_t 之后可能进入多种状态,进入任一状态 s' 的概率用状态转移函数表示为

$$p(s'|s_t, a_t) = P(S_{t+1} = s'|S_t = s_t, A_t = a_t) \quad (2)$$

奖励 R_t 表示智能体在采取了动作 a_t 之后来自环境的正负反馈,用于评估该动作的好坏,回报 U_t 表示从 t 时刻的状态 s_t 开始,到终止状态时的奖励之和为

$$U_t = R_t + R_{t+1} + R_{t+2} + \dots \quad (3)$$

考虑到未来状态的奖励对于当前时刻的重要性不一定相同,因此引入折扣率 $\gamma \in [0,1]$ 后,将回报 U_t 重新定义为

$$U_t = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \gamma^3 R_{t+3} + \dots = \sum_{k=0}^n \gamma^k R_{t+k} \quad (4)$$

由式(1)~(4)可得, U_t 取决于 $A_t, A_{t+1}, A_{t+2}, \dots$ 以及 $S_t, S_{t+1}, S_{t+2}, \dots$,且具有随机性,为消除这种随机性,可以采用数学期望,引入动作价值函数 $Q_\pi(s_t, a_t)$ 以及状态价值函数 $V_\pi(s_t)$

$$Q_\pi(s_t, a_t) = E(U_t | S_t = s_t, A_t = a_t) \quad (5)$$

$$V_\pi(s_t) = E_A(Q_\pi(s_t, A)) \quad (6)$$

式(5)表示在状态 s_t 且采取策略 π 的前提下,智能体采取动作 a_t 的好坏;式(6)表示在采取策略 π 的前提下,智能体当前状态 s_t 对于达到设定目标的好坏程度。对于所有的动作空间以及状态空间,一定会存在一种策略 π ,使得智能体在执行任务时可以得到最高的回报,这种策略被称为最优策略 $Q^*(s_t, a_t)$ 。采取最优策略时,动作价值函数即为最优动作价值函数,其数学表达式为

$$Q^*(s_t, a_t) = \max Q_\pi(s_t, a_t) \quad (7)$$

$$V_\pi(s_t) = R_{t+1} + \gamma \sum_{s' \in S} p(s'|s_t, a_t) V_\pi(s_{t+1}) \quad (8)$$

根据贝尔曼方程(式(8))可以进一步求解出最优动作价值函数 Q^* 。

RL是一种决策系统,其底层核心思想在于通过智能体不断与环境进行实时交互,在数次失败与成功的过程中学习,获得最大回报值后学到最优策略,因此,RL常被用来解决状态空间和动作空间都相对较小的任务,而无法解决高维数据输入的问题。

3.2 深度强化学习

2013年 Mnih等^[32]首次提出将深度学习中的卷积神经网络(Convolutional Neural Networks, CNN)应用

在强化学习中,提出深度Q网络(Deep Q Learning Network, DQN)算法,拉开了深度强化学习(Deep Reinforcement Learning, DRL)帷幕。DRL是在RL算法框架中,将DL的特征提取能力与RL的决策能力相结合^[33],便可直接根据高维输入数据学习最优决策,属于端对端决策控制系统,DRL原理框架如图5所示。

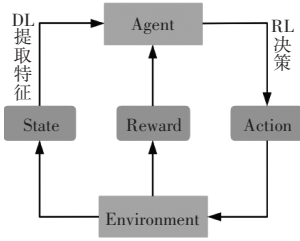


图5 DRL原理框架

与RL算法划分相同,DRL算法也可以分为3类:基于值函数(Value Based)、基于策略梯度(Policy Gradient-Based)和演员-评论家(Actor-Critic, AC)算法。其中,较多应用于能量管理领域的算法为基于值函数的DQN、基于AC的深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)算法。

3.2.1 深度Q网络算法

Q-Learning是RL中一种基于价值函数的算法,其基本思想在于使智能体通过学习1个函数近似得到 $Q^*(s,a)$,进而指导智能体在状态 s_t 时做出回报最高的动作 a_t^* 。由于Q-Learning需要通过Q-Table表格存储每个状态下不同动作的奖励,该算法只能用于处理低维空间任务。

DQN算法则是通过将Q-Learning算法与神经网络相结合,通过神经网络 $Q(s,a;w)$ 近似学习 $Q^*(s,a)$,其中, w 为神经网络的参数,DQN的输入为智能体所处状态 s_t ,输出 $Q(s,a;w)$ 表示该状态下智能体所有动作的价值。该神经网络的更新原理为在智能体执行任务之前,将 w 随机初始化,之后通过学习不断以梯度下降的方式更新 w 实现算法迭代(式(12)),便可以改善Q-Learning算法这一缺陷,处理较高维数据信息。

DQN算法结构主要由经验回放池、评估网络、目标网络这几方面构成,DQN原理框架如图6所示。经验池用于存放智能体每次执行任务时获得的4元组 (s_t, a_t, r_t, s_{t+1}) ,直到经验池饱和,智能体开始从经验池中随机采样4元组,将 s_t 和 a_t 送入评估网络,得到DQN的预测值为

$$q_t = Q(s_t, a_t; w_t) \quad (9)$$

将 s_{t+1} 和 r_t 送入目标网络,得到目标值为

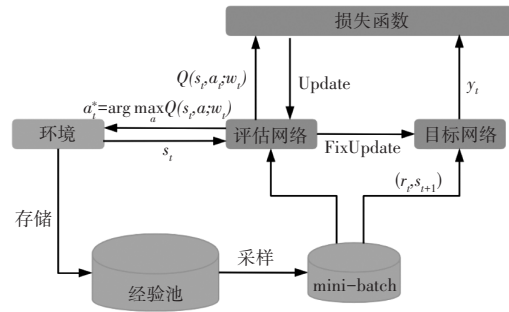


图6 DQN原理框架

$$y_t = r_t + \gamma \max_a Q(s_{t+1}, a; w_t) \quad (10)$$

进而计算损失函数为

$$L(w_t) = \frac{1}{2} (q_t - y_t)^2 \quad (11)$$

通过梯度下降得到 w_{t+1} ,完成1次算法迭代

$$w_{t+1} \leftarrow w_t - \alpha \cdot (q_t - y_t) \cdot \left. \frac{\partial Q(s_t, a_t; w_t)}{\partial w} \right|_{w=w_t} \quad (12)$$

接着智能体继续从经验池中随机抽取4元组,并存放入新的4元组,同时继续上述过程直到DQN完成收敛。

从DQN算法结构中可见,在状态 s_t 时的最优动作 a_t^* 为

$$a_t^* = \arg \max_a Q(s_t, a; w_t) \quad (13)$$

DQN优势在于利用延后更新的目标网络计算Q值,极大的提高了神经网络训练的收敛性和稳定性;利用神经网络拟合Q-Learning中的Q-Table,避免由维数过高导致的维数灾难;通过经验池重复利用4元组避免浪费的同时,打破了数据之间的相关性,使网络更新时的数据独立分布。由于式(10)、(13)采用取最大值的计算方式,若将其应用于连续动作中且不通过求导获取变化梯度,连续动作空间将会提供无限种计算结果,极大地增大损失函数的计算难度。因此,DQN不适用于处理连续动作问题,同时存在稳定性差、高估问题等不足。

针对上述缺陷,学者们提出了一些改进DQN的新算法。为解决DQN算法Q值高估问题,Granacher等^[34]提出了一种基于DDQN的算法,使用在线目标Q-Network估计Q值;为进一步提高DDQN的算法性能,李茹杨等^[35]提出了一种优先经验回放DQN(PER-DQN)算法,充分考虑到各环节不同数据的重要性不同,针对重要性高的数据进行多次回放;Yi等^[36]提出了一种竞争式的DQN(DU-DQN)算法,通过2个独立

的估算器表示状态价值函数 $V_{\pi}(s_t)$ 以及与之相关的动作函数,提高 DQN 运行过程的稳定性;为解决 DQN 无法处理连续动作问题,Lillicrap 等^[37]提出一种 DDPG 算法,这也是目前航空混动系统能量管理领域使用较多的 DRL 算法之一。

3.2.2 深度确定性策略梯度

DDPG 是一种将 Actor-Critic 框架与确定性策略梯度算法相结合,并同时继续采用经验回放机制和目标神经网络的算法。不同于仅采用单一类型的神经网络 $Q(s,a;w)$ 近似 $Q^*(s,a)$ 的 DQN 算法,DDPG 算法采用 Actor 网络 $\pi(s; \theta)$ 近似确定性策略函数 $\pi(s)$,Critic 网络 $q(s,a;w)$ 近似动作价值函数 $Q_{\pi}(s,a)$,前者直接与环境交互,并通过当前学习得到的 $\pi(s; \theta)$ 直接输出控制智能体行动的确定动作;后者则是通过对智能体执行动作的打分,训练及更新 Actor 网络。DDPG 算法的确定性体现在智能体通过策略函数 $\pi(s; \theta)$ 计算得到的是 1 个确定的动作,而非某状态下所有动作的概率密度。因此,可以简化在动作空间进行采样积分,得到具体输出动作这一环节,提高了系统的运行效率。DQN 与 DDPG 对比见表 1。

表 1 DQN 与 DDPG 对比

	DQN	DDPG
优点	可应对高维状态及动作空间	可应对高维且连续的状态及动作空间
缺点	对超参数敏感性高,存在高估、过拟合问题,训练速度较慢	对超参数敏感性高,训练复杂,可能陷入局部最优
应用场景	离散控制问题,高维空间	连续控制问题,高维空间,延迟奖励问题

从式(12)中可见,需要原始神经网络的估计值计算出更新后神经网络的参数,存在高估或低估问题,因而,DDPG 分别将 Actor 网络和 Critic 网络继续各自细分为 online 网络和 target 网络

$$\text{Actor:} \begin{cases} \text{online:} q(s, a; w) \\ \text{target:} q^-(s, a; w^-) \end{cases} \quad (14)$$

$$\text{Critic:} \begin{cases} \text{online:} \pi(s; \theta) \\ \text{target:} \pi^-(s; \theta^-) \end{cases} \quad (15)$$

式中: w 和 w^- 分别为 online-Actor 和 target-Actor 网络的参数; θ 和 θ^- 分别为 online-Critic 和 target-Critic 网络的参数。

DDPG 算法的更新是通过更新这 4 个神经网络参数实现的。

DDPG 原理框架如图 7 所示。智能体从经验池中随机抽取一部分样本(mini-batch),通过 online-Critic 网络计算在状态 s_t 且执行动作 a_t 的预测值为

$$q_t = q(s_t, a_t; w_t) \quad (16)$$

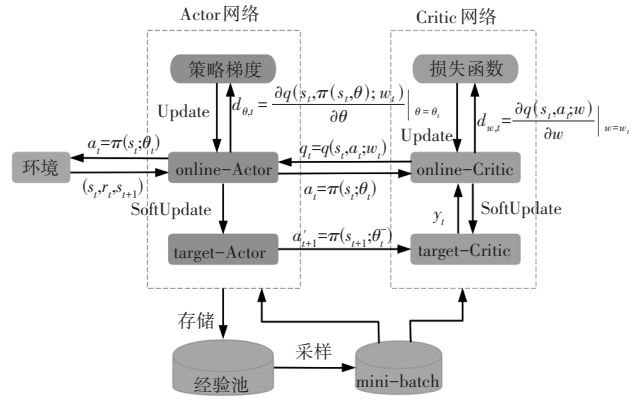


图 7 DDPG 原理框架

通过 target-Actor 计算出 $t+1$ 时刻的动作 a_{t+1}' ,智能体并不会执行该动作,只用于计算状态 s_{t+1} 时对应的动作价值 q_{t+1}

$$a_{t+1}' = \pi(s_{t+1}; \theta_t^-) \quad (17)$$

将上述结果送入 target-Critic 得到目标值

$$y_t = r_t + \gamma \cdot \underbrace{q(s_{t+1}, a_{t+1}'; w_t^-)}_{q_{t+1}} \quad (18)$$

通过梯度上升更新 online-Actor 网络参数 θ

$$\theta_{t+1} \leftarrow \theta_t + \beta \cdot \left. \frac{\partial q(s_t, \pi(s_t, \theta); w_t)}{\partial \theta} \right|_{\theta = \theta_t} \quad (19)$$

式中: β 为 online-Actor 网络的学习率。

通过梯度下降更新 online-Critic 网络参数 w

$$w_{t+1} \leftarrow w_t - \alpha \cdot (q_t - y_t) \cdot \left. \frac{\partial q(s_t, a_t; w)}{\partial w} \right|_{w = w_t} \quad (20)$$

式中: α 为 online-Critic 的学习率。

设定超参数 $\tau \in (0,1)$,通过滑动平均(soft update)的方法更新 2 个 target 网络的参数^[38]

$$\text{softupdate}_{\tau \in (0,1)} = \begin{cases} \theta_{t+1}^- \leftarrow \tau \cdot \theta_t + (1 - \tau) \cdot \theta_t^- \\ w_{t+1}^- \leftarrow \tau \cdot w_t + (1 - \tau) \cdot w_t^- \end{cases} \quad (21)$$

3.3 基于智能的 EMS 研究现状

近年来,在航空工程及车辆工程能量管理领域也出现大量与 DRL 相关的研究,总体上可以分为 2 类:基于智能内部算法创新与融合;基于智能外部算法创新与融合。

3.3.1 基于智能内部算法创新与融合

考虑到飞机在执行飞行任务过程中,动作空间是连续和离散并存的,而DDPG更擅长处理连续动作空间问题,因此,Liu等^[39]提出了一种将DDPG与双深度Q网络算法相结合的EMS,将能量管理问题转化为马尔可夫决策过程求解,将目标函数的整数变量和连续变量分别视为离散动作空间和连续动作空间,并分别由双深度Q网络算法以及DDPG算法求解,最终通过硬件在环实验证明其有效性。为解决强化学习过程中因为状态选择过多导致的“维数灾难”以及过估计问题,Zhang等^[40]选择在DQN的基础上新增1个神经网络,即双深度Q网络算法,2个神经网络分别用于动作决策和目标值计算,并将该算法与基于规则的EMS相对比,证明其优势。为了对多模式车辆进行能量管理,Hua等^[41]提出了一种多智能体深度强化学习策略,即使2个智能体在DDPG中协同工作,通过引入相关度并对影响2个智能体学习性能的因素进行分析,得到了2个智能体的统一设置,之后通过相关度的参数化研究得到了在该策略下的最优工作模式。

此外,Maroto等^[42]在商业软件的仿真下开发了一种双深度Q网络算法对系统进行能量管理,同时采用卷积神经网络表征混动系统的污染排放,进而实现降低油耗,提高燃油经济性的目的。汪钰皓^[43]针对燃料电池无人机提出了一种基于DDPG的EMS,并通过与功率跟随策略和模糊逻辑控制策略对比,证明该策略可以兼顾燃料经济性和锂电池性能保护。

3.3.2 基于智能外部算法创新与融合

为解决电推进无人机在不同飞行状态下推进功率的不确定性,Min等^[44]针对分布式燃料电池无人机提出了一种将模型预测控制与神经网络结合的EMS,采用神经网络对分布式电推进系统建模,并将多次试飞的数据用于训练神经网络,同时推导优化模型,最终将神经网络对推进功率的预测值和优化模型计算结果组成的2维向量,作为模型预测控制的轨迹,便可兼顾推进功率需求以及最小氢耗要求。

由于大多数DRL控制算法属于无模型控制,更依赖于环境中大量的真实样本,需要不断试错以获得能量管理最优策略,因而存在采样效率较低的问题^[45]。为解决该问题,Lian等^[46]提出了一种基于智能和基于规则相融合的EMS,在进行DDPG神经网络训练时,通过嵌入专家知识,为DDPG的神经网络训练

提供训练偏好,并证明该策略确实提高了神经网络的学习效率。

在进行EMS设计时,还有学者将几种不同类型的算法相结合,对不同的算法设定不同的控制目标,最大化各算法的优势。Chen等^[47]提出了一种基于DRL的融合控制性EMS,采用DDPG连续控制发动机节气门开度,DQN离散控制AMT换挡,并设计1套基于规则的发动机启停策略,并通过对比试验证明其优势。

4 结束语

目前,能量管理综述性研究主要集中在混合动力汽车领域和电力储能领域,对航空混合动力系统领域则少有涉及。

针对航空混动系统EMS存在计算量大、对复杂工况适应性差或最优性差等问题,基于智能的EMS凭借其较强的学习能力、自适应能力等优势成为解决上述问题的一种有前途的方案。

因此本文以RL底层原理入手,详细介绍了DQN和DDPG这2种深度强化学习算法的运行以及神经网络更新原理,详细分析这2种算法的优缺点。并提出未来需根据每种DRL算法的优化侧重点,将其与基于智能内部算法、基于规则算法、基于优化外部算法优势互补,或形成新算法。为DRL算法深度应用于航空混合动力系统的EMS中提供有力支持。

参考文献:

- [1] 刘振敏.航空业的绿色之路[J].大飞机,2020(7):50-55.
LIU Zhenmin.The green path of the aviation industry[J].Big Aircraft, 2020,(7):50-55.(in Chinese)
- [2] 王哲,郭迎清.混合动力系统建模及能量管理策略应用[J].航空发动机,2019,45(5):7-11.
WANG Zhe, GUO Yingqing. Hybrid power system modeling and energy management strategy application[J]. Aeroengine, 2019, 45(5):7-11.(in Chinese)
- [3] 马睿,宋剑,王宇昂,等.计及飞行任务与能耗分析的航空燃料电池推进系统能量管理策略[J].中国电机工程学报,2023,43(1):221-236.
MA Rui, SONG Jian, WANG Yuang, et al.Fuel cell electric propulsion system energy management strategy based on flight mode and energy consumption analysis[J].Proceedings of the CSEE, 2023, 43(1):221-236.(in Chinese)
- [4] LV S Q, Wu Y B, Lian J, et al.Energy management of hybrid electric vehicles: a review of energy optimization of fuel cell hybrid power sys-

- tem based on genetic algorithm[J]. *Energy Conversion and Management*, 2020, 205: 112474.
- [5] 赵洋, 潘天宇, 郑孟宗, 等. 考虑变负载效应的油电混合系统能量管理研究[J]. *航空动力学报*, 2024, 39(11): 1-10.
ZHAO Yang, PAN Tianyu, ZHENG Mengzong, et al. Research on energy management of electric hybrid system considering variable load effect[J]. *Journal of Aerospace Power*, 2024, 39(11): 1-10. (in Chinese)
- [6] Ting C C, Tsai D Y, Hsiao C C. Developing a mechanical roadway system for waste energy capture of vehicles and electric generation[J]. *Applied Energy*, 2012, 92: 1-8.
- [7] Zeng Y, Xu J, Zhang R. Energy minimization for wireless communication with rotary-wing UAV[J]. *IEEE transactions on wireless communications*, 2019, 18(4): 2329-2345.
- [8] Karunarathne L, Economou J T, Knowles K. Power and energy management system for fuel cell unmanned aerial vehicle[J]. *Proceedings of the Institution of Mechanical Engineers Part G Journal of Aerospace Engineering*, 2012, 226(4): 437-454.
- [9] Wei Z, Ma Y, Xiang C, et al. Power prediction-based model predictive control for energy management in land and air vehicle with turboshaft engine[J]. *Complexity*, 2021(1): 24.
- [10] 程智博. 四旋翼无人机燃料电池混合动力系统研究[D]. 长春: 吉林大学, 2022.
CHENG Zhibo. Research on fuel cell hybrid power system of quadrotor UAV[D]. Changchun: Jilin University, 2022. (in Chinese)
- [11] 张雪, 覃永新, 蓝红莉, 等. 基于优化模糊控制的混联式混合动力汽车的能量管理策略[J/OL]. *广西科技大学学报*, 2024: 1-14[2024-12-06]. <http://kns.cnki.net/kcms/detail/45.1395.T.20240411.0851.013.html>.
ZHANG Xue, QIN Yongxin, LAN Hongli, et al. Energy management strategy for series-parallel hybrid electric vehicles based on optimized fuzzy control[J/OL]. *Journal of Guangxi University of Science and Technology*, 2024: 1-14[2024-12-06]. <http://kns.cnki.net/kcms/detail/45.1395.T.20240411.0851.013.html>. (in Chinese)
- [12] Chen Z H, Liu Y G, Zhang Y J, et al. A neural network-based ECMS for optimized energy management of plug-in hybrid electric vehicles[J]. *Energy*, 2022, 243: 122727.
- [13] Anselma P G. Dynamic programming based rapid energy management of hybrid electric vehicles with constraints on smooth driving, battery state-of-charge and battery state-of-health[J]. *Energies*, 2022, 15(5): 1665
- [14] Hegde S, Bonfitto A, Galluzzi R, et al. Equivalent consumption minimization strategy based on belt drive system characteristic maps for P0 hybrid electric vehicles[J]. *Energies*, 2023, 16(1): 487.
- [15] 张钰凡, 马睿, 张羽翔, 等. 航空燃料电池推进系统能量管理策略研究综述[J]. *中国电机工程学报*, 2023, (13): 5012-5024.
ZHANG Yufan, MA Rui, ZHANG Yuxiang, et al. A review on the energy management strategy for aviation fuel cell propulsion system[J]. *Proceedings of the CSEE*, 2023, (13): 5012-5024. (in Chinese)
- [16] Xie Y, Antonios T. Review of hybrid electric powered aircraft, its conceptual design and energy management methodologies[J]. *Chinese Journal of Aeronautics*, 2021, 34(4): 432-450.
- [17] 陈宗科, 杨向宇. 并联式油电混合无人机的能量管理策略研究[J]. *电源技术*, 2022, 46(9): 1066-1070.
CHEN Zongke, YANG Xiangyu. Research on energy management of parallel hybrid UAVs[J]. *Chinese Journal of Power Sources*, 2022, 46(9): 1066-1070. (in Chinese)
- [18] 陈洪伟. 混合动力无人飞机的能源管理系统设计与研究[D]. 镇江江苏: 江苏大学, 2020.
CHEN Hongwei. Design and research of energy management system for hybrid unmanned aerial vehicle[D]. Jiangsu Zhenjiang: Jiangsu University, 2020. (in Chinese)
- [19] Li Q, Su B, Pu Y, et al. A state machine control based on equivalent consumption minimization for fuel cell/ supercapacitor hybrid tramway[J]. *IEEE Transactions on Transportation Electrification*, 2019, 5(2): 552-564.
- [20] Buticchi G, Costa L F, Barater D, et al. A quadruple active bridge converter for the storage integration on the more electric aircraft[J]. *IEEE Transactions on Power Electronics*, 2017, 33(9): 8174-8186.
- [21] 李元旭, 乔培玉, 彭发醇. 复合电源型无人机能量管理策略研究[J]. *电源技术*, 2022, 46(6): 672-675.
LI Yuanxu, QIAO Peiyu, PENG Fachun. Study on energy management strategy of hybrid power UAV[J]. *Chinese Journal of Power Sources*, 2022, 46(6): 672-675. (in Chinese)
- [22] Wu J, Zhang C H, Cui N X. Fuzzy energy management strategy for a hybrid electric vehicle based on driving cycle recognition[J]. *International Journal of Automotive Technology*, 2012, 13(7): 1159-1167.
- [23] Wang A H, Yang W Z. Design of energy management strategy in hybrid vehicles by evolutionary fuzzy system Part I: fuzzy logic controller development[C]//2006 6th World Congress on Intelligent Control and Automation. Dalian: IEEE Press, 2006: 8324-8328.
- [24] Chrenko D, Gan S, Gutenkunst C, et al. Novel classification of control strategies for hybrid electric vehicles[C]//2015 IEEE Vehicle Power and Propulsion Conference. Montreal: IEEE Press, 2015: 1-6.
- [25] 朱炳杰, 朱莹涛, 李建奇. 混合电推进能源管理系统模糊逻辑控制[J]. *国防科技大学学报*, 2023, 45(6): 32-39.
ZHU Bingjie, ZHU Yingtao, LI Jianqi. Fuzzy logic control for hybrid electric propulsion energy management system[J]. *Journal of National University of Defense Technology*, 2023, 45(6): 32-39. (in Chinese)
- [26] Leite J P S P, Voskuijl M. Optimal energy management for hybrid-electric aircraft[J]. *Aircraft Engineering and Aerospace Technology*, 2020, 92(6): 851-861.
- [27] Sun H, Tao F, Fu Z, et al. Driving-behavior-aware optimal energy management strategy for multi-source fuel cell hybrid electric vehicles based on adaptive soft deep-reinforcement learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(4): 4127-4146.

- [28] Majed C, Karaki S, Jabr R. Neural network technique for hybrid electric vehicle optimization[C]//2016 18th Mediterranean Electrotechnical Conference. Lemosos: IEEE Press, 2016: 1-6.
- [29] 胡春明, 闫丁洋, 刘娜, 等. 油电混合动力无人机能量管理策略的对比仿真研究[J]. 内燃机工程, 2022, 43(4): 74-83.
HU Chunming, YAN Dingyang, LIU Na, et al. Comparison and simulation research on energy management strategies of oil-electric hybrid unmanned aerial vehicle[J]. Chinese Internal Combustion Engine Engineering, 2022, 43(4): 74-83. (in Chinese)
- [30] Boukoberine M, Zia M F, Benbouzid M, et al. Hybrid fuel cell powered drones energy management strategy improvement and hydrogen saving using real flight test data[J]. Energy Conversion and Management, 2021, 236: 113987.
- [31] Tan H, Zhang H, Peng J, et al. Energy management of hybrid electric bus based on deep reinforcement learning in continuous state and action space[J]. Energy Conversion and Management, 2019, 195: 548-560.
- [32] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518: 529 - 533.
- [33] 赵冬斌, 邵坤, 朱圆恒, 等. 深度强化学习综述: 兼论计算机围棋的发展[J]. 控制理论与应用, 2016, 33(6): 701-717.
ZHAO Dongbin, SHAO Kun, ZHU Yuanheng, et al. Review of deep reinforcement learning and discussions on the development of computer go[J]. Control Theory and Applications, 2016, 33(6): 701-717. (in Chinese)
- [34] Granacher J, Nguyen T, Castro-Amoedo R, et al. Overcoming decision paralysis—a digital twin for decision making in energy system design [J]. Applied Energy, 2022, 306: 117-128.
- [35] 李茹杨, 彭慧民, 李仁刚, 等. 强化学习算法与应用综述[J]. 计算机系统应用, 2020, 29(12): 13-25.
LI Ruyang, PENG Huimin, LI Rengang, et al. A review of reinforcement learning algorithms and applications[J]. Computer Systems Applications, 2020, 29(12): 13-25. (in Chinese)
- [36] Yi C, Qi M. Research on virtual path planning based on improved DQN [C]//2020 IEEE International Conference on Real-Time Computing and Robotics. Asahikawa: IEEE Press, 2020: 387-392.
- [37] Lillicrap T, Hunt J, Pritzel A, et al. Continuous control with deep reinforcement learning[J]. IEEE Transactions on Software Engineering, 2020, 15(11): 217-224.
- [38] 王树森, 黎彧君, 张志华. 深度强化学习[M]. 北京: 人民邮电出版社, 2023: 150-215.
WANG Shusen, LI Yujun, ZHANG Zhihua. Deep reinforcement learning[M]. Beijing: People's Posts and Telecommunications Press, 2023: 150-215. (in Chinese)
- [39] Liu B, Xu B, He T, et al. Hybrid deep reinforcement learning considering discrete-continuous action spaces for real-time energy management in more electric aircraft[J]. Energies, 2022, 15(17): 6323.
- [40] Zhang J, Jiao X, Yang C. A double-deep Q-network-based energy management strategy for hybrid electric vehicles under variable driving cycles[J]. Energy Technology, 2021, 9(2): 2000770.
- [41] Hua M, Zhang C, Zhang F, et al. Energy management of multi-mode plug-in hybrid electric vehicle using multi-agent deep reinforcement learning[J]. Applied Energy, 2023, 348: 121526.
- [42] Maroto E P, Delima D, Bauer P H, et al. Deep learning in the development of energy management strategies of hybrid electric vehicles: a hybrid modeling approach[J]. Applied Energy, 2023: 329: 120231.
- [43] 汪钰皓. 燃料电池无人直升机混合动力系统设计[D]. 杭州: 浙江大学, 2021.
WANG Yuhao. Design of hybrid power system for fuel cell unmanned helicopter[D]. Hangzhou: Zhejiang University, 2021. (in Chinese)
- [44] Min Z, Lei T, Zhang X, et al. A data-driven MPC energy optimization management strategy for fuel cell distributed electric propulsion UAV [C]//2022 4th Asia Energy and Electrical Engineering Symposium. Chengdu: IEEE Press, 2022: 571-578.
- [45] Schulman J, Levine S, Abbeel P, et al. Trust region policy optimization [C]//ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille: JMLR Press, 2015: 346-367.
- [46] Lian R, Peng J, Wu Y, et al. Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle[J]. Energy, 2020, 197: 117297.
- [47] Chen J, Shu H, Tang X, et al. Deep reinforcement learning-based multi-objective control of hybrid power system combined with road recognition under time-varying environment[J]. Energy, 2022, 239: 122123.

(编辑: 兰海青, 孙明霞)