

9-25-2023

Aircraft Assignment Method for Optimal Utilization of Maintenance Intervals

Runxia Guo

College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China, rxguoblpl@163.com

Yifu Wang

College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact xtfzxb@126.com.

Aircraft Assignment Method for Optimal Utilization of Maintenance Intervals

Abstract

Abstract: The aircraft assignment problem is studied from a maintenance assurance perspective. In order to ensure its continuous airworthiness, civil aircraft are required to perform maintenance tasks, i. e., scheduled inspections, at specified intervals. The scheduled inspection interval is usually controlled by the number of flight cycles (FC), flight hours (FH), or flight days (FD), whichever comes first. In order to make balanced use of the inspection interval, an aircraft assignment model for a given fleet size is developed to optimize the maintenance interval utilization, and it is solved by a reinforcement learning algorithm to minimize the variance of the FC and FH uniformly discounted at the time of maintenance interval arrival. The proposed method is computationally efficient and can be used to maximize the effectiveness of a single scheduled inspection and maintenance, saving maintenance costs and increasing aircraft utilization. Experiments are conducted by using authentic Chinese airline flight data. The experimental results show that the algorithm can find stable optimal solutions in the data of 761 flight legs in 129.448 seconds, and the Gap value is only 0.122 4%.

Keywords

aircraft assignment, maintenance intervals, reinforcement learning, maintenance interval utilization, flight legs

Recommended Citation

Guo Runxia, Wang Yifu. Aircraft Assignment Method for Optimal Utilization of Maintenance Intervals[J]. Journal of System Simulation, 2023, 35(9): 1985-1999.

以维修间隔利用率最优为目标的飞机派遣方法

郭润夏, 王一府

(中国民航大学 电子信息与自动化学院, 天津 300300)

摘要: 从维修保障的视角研究飞机派遣问题。民用飞机为确保其持续适航性, 均要求在规定的时间内执行检修任务即定检。定检时间间隔通常按照飞行循环(flight cycles, FC)、飞行小时(flight hours, FH)或日历日(flight days, FD)进行控制, 先到为准。为了均衡的对定检时间间隔进行利用, 以最优化维修间隔利用率为目标建立了给定机队规模的飞机派遣模型并通过强化学习算法予以求解, 使得在维修间隔到达时刻, FC 和 FH 统一折算后的方差最小。该派遣方法计算效率高, 能够充分发挥单次定检维修的最大效能, 节约维修成本并增加飞机的利用率。利用某航空公司的真实航班数据进行了实验, 结果表明在最多 761 个航班段数据中, 算法能够在 129.448 s 找到稳定的最优解, 且 Gap 值仅为 0.122 4%。

关键词: 飞机派遣; 维修间隔; 强化学习; 维修间隔利用率; 航班段

中图分类号: TP391.9; TP29; V35 文献标志码: A 文章编号: 1004-731X(2023)09-1985-15

DOI: 10.16182/j.issn1004731x.joss.22-0546

引用格式: 郭润夏, 王一府. 以维修间隔利用率最优为目标的飞机派遣方法[J]. 系统仿真学报, 2023, 35(9): 1985-1999.

Reference format: Guo Runxia, Wang Yifu. Aircraft Assignment Method for Optimal Utilization of Maintenance Intervals[J]. Journal of System Simulation, 2023, 35(9): 1985-1999.

Aircraft Assignment Method for Optimal Utilization of Maintenance Intervals

Guo Runxia, Wang Yifu

(College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract: The aircraft assignment problem is studied from a maintenance assurance perspective. In order to ensure its continuous airworthiness, civil aircraft are required to perform maintenance tasks, i. e., scheduled inspections, at specified intervals. The scheduled inspection interval is usually controlled by the number of flight cycles (FC), flight hours (FH), or flight days (FD), whichever comes first. In order to make balanced use of the inspection interval, an aircraft assignment model for a given fleet size is developed to optimize the maintenance interval utilization, and it is solved by a reinforcement learning algorithm to minimize the variance of the FC and FH uniformly discounted at the time of maintenance interval arrival. The proposed method is computationally efficient and can be used to maximize the effectiveness of a single scheduled inspection and maintenance, saving maintenance costs and increasing aircraft utilization. Experiments are conducted by using authentic Chinese airline flight data. The experimental results show that the algorithm can find stable optimal solutions in the data of 761 flight legs in 129.448 seconds, and the Gap value is only 0.122 4%.

Keywords: aircraft assignment; maintenance intervals; reinforcement learning; maintenance interval utilization; flight legs

收稿日期: 2022-05-24 修回日期: 2022-08-22

基金项目: 国家自然科学基金(62173331); 航空科学基金(2019ZD067007); 天津市特支计划青年拔尖人才基金(TJZJH-QNBJRC-2-19); 天津市教委科研计划(2018KJ238)

第一作者: 郭润夏(1981-), 男, 教授, 博士, 研究方向为民用飞机故障诊断、剩余寿命预测和健康管理。E-mail: rxguoblp@163.com

0 引言

飞机派遣问题 (aircraft assignment problem, AAP) 是航空运输中最常见的问题之一。目前飞机派遣研究的主要目的是在航线结构确定的前提下安排机队中的飞机满足维修要求进而执行相应的航班段 (flight leg)^[1], 相关的维修需求如不超过规定的飞机累计飞行小时 (flight hours, FH) 和飞行循环 (flight cycles, FC), 即维修间隔约束。

近年来, 由于民航业的快速发展, 关于飞机派遣的研究越来越受到重视^[2]。在安排航线运行的基础上, 对各种调度资源进行优化配置是一个复杂的决策问题。对于民用航空来说, 不仅要保持飞机定期的良好维护, 还要对其运行路线进行适当管理^[3]。考虑到 COVID-19 对民航业的影响^[4], 通过合理对飞机进行派遣减少维修资源的浪费以节省成本具有非常重要的现实意义。

在过去 20 多年里, 一些研究人员分析并研究了 AAP, 试图在考虑足够的约束条件下获得相对良好的经济效益^[5-6]。关于 AAP 的研究主要集中在 TAAP (tactical AAP)^[7] 和 OAAP (operational AAP) 两个领域^[8-9], 相比之下前者忽略了如维修间隔的维修操作约束, 然后为机队中的每架飞机生成通用飞行任务。

许多学者对 TAAP 进行了研究^[10-13]。例如文献 [10] 提出了一种启发式算法, 用于解决实际航线规划问题中整合的机队分配和飞机航线问题。文献 [11] 根据现实条件整合飞机派遣问题和机组人员匹配, 以满足多个目标, 提出了一个多目标进化算法来求解这些目标。文献 [12] 提出了一个新的回旋时空网络和一个新的混合整数线性规划方法。这些 TAAP 研究有 2 个特点: ① 允许飞机重复飞行; ② 忽略维修间隔等维护规定。这导致了航空公司可能较难将其应用到实际中, 也使得 OAAP 得以出现并被研究。

同样许多学者也对 OAAP 进行了研究。为了优化机队的总剩余飞行时间, 文献 [5] 更新了连接

网络的表示方法, 并提供了一种新的快速反应算法, 为机队中的每一架飞机给出匹配的飞行任务。文献 [14] 提出了一个精确的混合整数规划模型, 用于最小化飞机的剩余飞行时间, 并基于大规模邻域搜索算法提出的启发式算法进行求解。文献 [15] 建立了以最小化使用飞机数量和剩余飞行时间为目标的飞机派遣模型, 并使用改进的启发式可变邻域搜索算法来求解该模型。还有很多研究人员的目标是最小化通过价值^[4, 16-17]。如文献 [4] 提出了一种混合求解算法, 将 Gasimov 的改进次梯度算法和蚁群优化元启发法结合起来达到最小化通过价值的目标。

然而, 这些 AAP 的研究并没有考虑飞机的维修间隔利用率。因为在飞机执行航班段的任务时维修间隔中的 FH 或者 FC 有任一者不再满足维修间隔约束时就需要对飞机进行维护以满足飞机持续适航性保障飞行安全, 但这可能带来的直接后果便是其累积 FH 和 FC 不匹配而引起过度维护造成维修间隔利用率低下。而飞机维修间隔利用率最优可以通过最小化飞机的累积 FH 和 FC 在同一换算标准下的差来实现, 从而最大化利用飞机的维修间隔避免维修的提前或滞后。需要注意的是累积 FH 和 FC 均有相应的下界, 因为一味追求相对差最小可能导致飞机累积 FH 和 FC 均远远达不到维修间隔限制同样引起过度维护造成维修资源浪费。故旨在研究以维修间隔利用率最优为目标的飞机派遣具有一定的实际意义。

AAP 是一个复杂的序贯决策问题, 它的求解是基于各种调度资源的最佳分配。该特点与强化学习的理念以及其中 Agent 的行为规则极为相似。这是本文提出基于强化学习算法的主要原因之一。在强化学习中, Agent 总试图学习前一次迭代中的经验以应用于下一次迭代, 从而提高解决方案的质量。这是本文用它来求解 AAP 的另一个原因。此外, 随着强化学习的不断发展, 它正被应用于许多决策系统中^[18-21]。其中 Q-learning 是基于价值函数的经典强化学习方法之一^[22-25], 其使用具体的

数值来衡量 Agent 行动的效用, 因为 Agent 的行为有可能不会立即得到奖励, 而是在未来得到更大的奖励, 这与飞机派遣需要考虑全局最优的前提始终保持一致。综上, 本文提出了一种基于 Q-learning 的算法来求解飞机派遣问题。

(1) 建立了一个以维修间隔利用率最优为目标的飞机派遣线性整数规划模型。该模型可以在确保飞机累积 FH 和 FC 上下界的前提下, 最小化每架飞机的累积 FH 和 FC 在同一标准下的差距。

(2) 提出了基于强化学习中的 Q-learning 算法来求解所提出的飞机派遣线性整数规划模型。该算法中 Agent 总是试图从之前的迭代中学习, 并且可以通过采取具有高 Q 值的动作来生成选择飞机的策略。

1 飞机派遣模型构建

研究目标是在航班时刻和机队给定的前提下最小化 FH 和 FC 在同一标准下的差, 然后将机队中的飞机分配至相应的航班段中, 即以维修间隔利用率最优为目标建立飞机派遣模型。相关术语

和名词的含义如图 1 所示。飞机从起飞至降落的过程为一个飞行循环。

连接网络是研究航班调度常用的网络之一^[9]。基于修改的连接网络建立以维修间隔利用率最优为目标的飞机派遣模型, 然后介绍所考虑的约束条件。修改的连接网络如图 2 所示。在连接网络中航班段由节点表示(如图 2 中 f, i, j 和 g), 节点之间的弧代表航班段之间可能的连接。其中定义了两类弧: 对于图 2 中航班段 i 和 j 之间的 M 弧而言, 当飞机在航班段 j 后不进行维护时 M 为普通弧 (ORD arc); 同样, 如果飞机在航班段 j 后进行维护则 M 为维护弧 (MAT arc)。此外, 与 MAT arc 连接的后续节点是为了让飞机在完成维修作业后重新投入使用。故针对 ORD arc 和 MAT arc 提出了 2 个决策变量: X_{ijk}^v 和 Y_{ijk}^v 。在连接网络中, 源节点 (source node) 和汇节点 (sink node) 是虚拟节点, 没有实际意义, 仅是为了帮助数学模型的完整建立^[5]。

数学建模过程中的相关符号定义如表 1 所示。

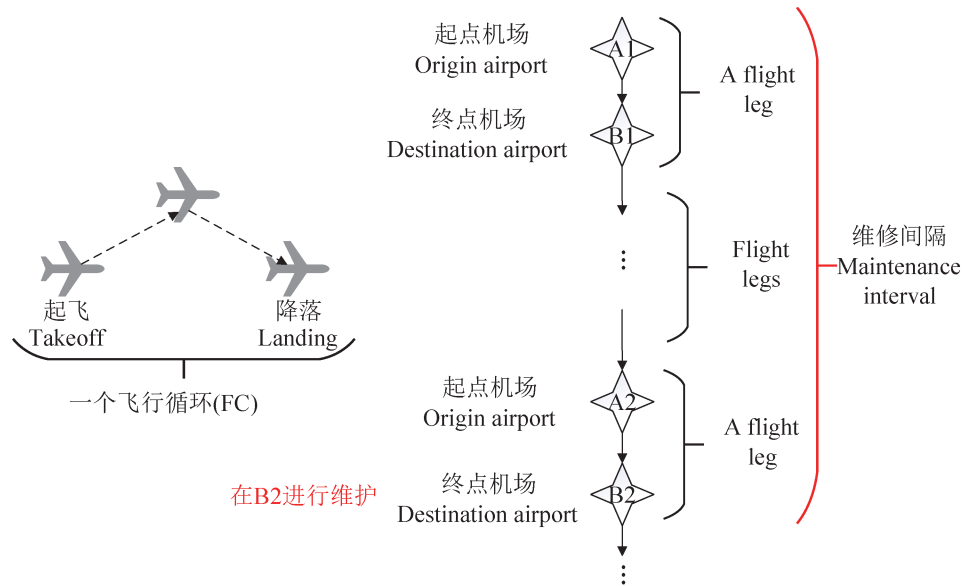


图 1 相关术语介绍
Fig. 1 Introduction to related terms

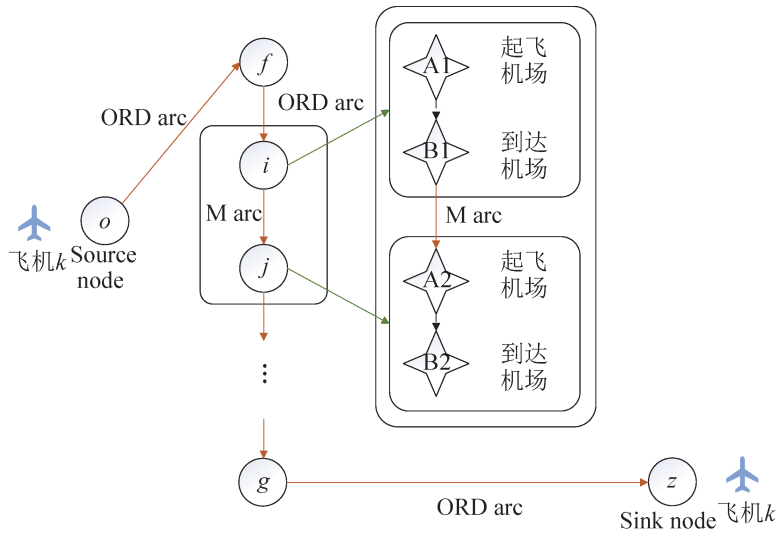
图2 飞机 k 修改的连接网络和航班段连接细节展示Fig. 2 Connection details of flight legs and connection network of aircraft k

表1 相关符号和定义

Table 1 Relevant symbols and definitions

类型	名称	含义
集合	OFL	航班段的集合, $i, j, l \in OFL$
集合	MFL	在到达机场进行维护的航班段的集合, $i, j, l \in MFL$
集合	K	飞机的集合, $k \in K = \{1, 2, \dots, k, \dots\}$
集合	AP	飞机场的集合, $a \in AP$
参数	DT_i	航班段 i 的起飞时间
参数	AT_i	航班段 i 的到达时间
参数	FD_i	航班段 i 的累计飞行时间
参数	O_{ia}	$O_{ia} = \begin{cases} 1, & \text{航班段 } i \text{ 的起飞机场是 } a \\ 0, & \text{其他} \end{cases}$
参数	D_{ia}	$D_{ia} = \begin{cases} 1, & \text{航班段 } i \text{ 的到达机场是 } a \\ 0, & \text{其他} \end{cases}$
参数	v	一个整数, 代表飞机被维护的次数
参数	T_{\max}	每架飞机在两次连续维护之间所允许的最大累积飞行时间
参数	TF_{\max}	每架飞机在两次连续维护之间所允许的最大飞行循环。
参数	TAT	周转时间(Turnaround time)。用于下机, 行李装卸, 更换登机口, 飞机加油等
决策变量	X_{ijk}^v	一个0-1变量。 $X_{ijk}^v = 1$ 如果飞机 k 执飞完连续航班段 i 和 j 后不进行第 v 次维护操作; $X_{ijk}^v = 0$ 其他
决策变量	Y_{ijk}^v	一个0-1变量。 $Y_{ijk}^v = 1$ 如果飞机 k 执飞完连续航班段 i 和 j 后进行第 v 次维护操作; $Y_{ijk}^v = 0$ 其他

为了更好的表述目标函数做如下定义: 将飞机已执飞的FC与所允许最大FC(两次维护之间所允许最大FC, 下同)的相对差定义为 α ; 将飞机已执飞FH和所允许最大FH的相对差定义为 β 。

$$\text{故目标函数为 } \min \sum_{k \in K} \sum_{v \in V} |\alpha_k^v - \beta_k^v|$$

在满足飞机累积FH和FC上下界的前提下, 分别计算飞机FC和所允许最大FC的相对差 α 与飞机FH和所允许最大FH的相对差 β , 来对飞机面向维修间隔的状态进行量化, 通过最小化这两个相对差来达到维修间隔利用率最优的目标进而减少维修资源的浪费。相关处理细节如下:

计算飞机 k 的 FC 和所允许最大 FC 在第 v 次维护和上一次维护之间的相对差:

$$\alpha_k^v = \frac{TF_{\max} - \sum_{i \in OFL} (\sum_{j \in OFL} X_{ijk}^v + \sum_{j \in MFL} Y_{ijk}^v)}{TF_{\max}} \quad (1)$$

计算第 v 次维护与上一次维护之间飞机 k 累积 FH 和所允许最大 FH 之间的相对差:

$$\beta_k^v = \frac{T_{\max} - \sum_{i \in OFL} FD_i (\sum_{j \in OFL} X_{ijk}^v + \sum_{j \in MFL} Y_{ijk}^v)}{T_{\max}} \quad (2)$$

根据图 2 建立模型相关的约束。研究目标是在航班时刻制定完成的前提下满足相关的约束限制, 为每个航班段分配一架飞机以执行相应的飞行任务。本文基于确定的航班时刻所考虑的约束如下:

(1) 覆盖约束(Coverage constraints)

当给飞机分配航班段飞行任务时, 每架飞机都不能同时执行多个飞行任务。

$$\sum_{k \in K} \sum_{j \in OFL} X_{ijk}^v + \sum_{k \in K} \sum_{j \in MFL} Y_{ijk}^v = 1, \forall i \in OFL, v \in V \quad (3)$$

$$\sum_{v \in V} \sum_{j \in OFL} X_{ojk}^v + \sum_{v \in V} \sum_{j \in MFL} Y_{ojk}^v = 1, \forall k \in K \quad (4)$$

$$\sum_{v \in V} \sum_{i \in OFL} X_{izk}^v = 1, \forall k \in K \quad (5)$$

$$X_{ilk}^v + Y_{ijk}^v \leq 1 \forall i, l \in OFL, j \in MFL, k \in K, v \in V \quad (6)$$

覆盖约束确保了某一特定航班段由某一特定飞机执飞。式(3)的构建是确保每个航班段 i 只由一架飞机执飞。式(4)确保每架飞机飞行任务起点都是源节点, 之后的弧线可以是 ORD 弧, 也可以是 MAT 弧。约束(5)保障了每架飞机各自飞行任务的结束。约束(6)确保飞机 k 所执飞的航班段后续所连接的只有 ORD 弧和 MAT 弧两种可能。

(2) 平衡约束(Balance constraints)

当同一架飞机连续执行两个任务时, 前一航班段的终点机场应是下一航班段的起点机场。

$$\sum_{j \in OFL \cup \{z\}} X_{ijk}^v + \sum_{j \in MFL} Y_{ijk}^v = \sum_{j \in OFL \cup \{o\}} X_{jik}^v, \quad \forall i \in OFL, k \in K, v \in V \quad (7)$$

$$\sum_{v \in V} (\sum_{j \in OFL} X_{ojk}^v + \sum_{j \in MFL} Y_{ojk}^v) = \sum_{v \in V} \sum_{i \in OFL} X_{izk}^v, \forall k \in K \quad (8)$$

$$\sum_{k \in K} X_{ijk}^v \leq \sum_{a \in AP} D_{ia} O_{ja}, \forall i, j \in OFL, v \in V \quad (9)$$

$$\sum_{k \in K} Y_{ijk}^v \leq \sum_{a \in AP} D_{ia} O_{ja}, \forall i \in OFL, j \in MFL, v \in V \quad (10)$$

约束(7)意味着如果飞机 k 执飞航班段 i , 则它与下一个执飞航班段之间的连接是 ORD 弧或者 MAT 弧。模型中所有从源节点开始的飞机 k 均必须在汇节点结束其飞行任务, 即式(8)的意义。约束(9)确保在 ORD 弧时飞机需满足起飞降落位置的约束, 即如果飞机 k 连续执飞航班段 i 和 j , 则航班段 i 的到达机场与航班段 j 的起点机场为同一机场。同样, 约束(10)是确保 MAT 弧的位置约束。

(3) 时间衔接约束(Time articulation constraints)

当飞机连续执飞两个航班段时, 其在前一个航班段的到达时间和周转时间 TAT 之和应小于或等于下一航班段的起飞时间。

$$AT_i + TAT - DT_j \leq L \cdot (1 - X_{ijk}^v), \quad \forall i, j \in OFL, k \in K, v \in V \quad (11)$$

$$AT_i + TAT - DT_j \leq L \cdot (1 - X_{ijk}^v), \quad \forall i, j \in OFL, k \in K, v \in V \quad (12)$$

$$Y_{ijk}^v \cdot (AT_i + MT - DT_j) \leq L \cdot (1 - X_{jik}^v) \quad \forall i, l \in OFL, j \in MFL, k \in K, v \in V \quad (13)$$

对于 ORD 弧而言, 约束(11)表示其时间衔接约束, 即当飞机 k 连续执飞航班段 i 和 j 时, 航班段 i 的到达时间与周转时间之和小于等于航班段 j 的起飞时间(其中 L 是一个很大的数)。与约束(11)不同的是约束(12)针对的是 MAT 弧。为了保证飞机进行维护后再次顺利投入使用, 构建相应时间约束(13), 即飞机 k 在执飞完航班段 i 和 j 后进行维护(时间 MT)再执飞航班段 l 的时间约束。

(4) 维护约束(Maintenance constraints)

目标是在给定航班时刻计划的前提下追求维修间隔的最大利用, 其中考虑飞机飞行时间和飞行循环的约束。为了不影响飞机在白天的正常飞行状态, 维护一般在夜间进行。

$$TF_{\min} \leq \sum_{i \in OFL} \left(\sum_{j \in OFL} X_{ijk}^v + \sum_{j \in MFL} Y_{ijk}^v \right) \leq TF_{\max},$$

$$k \in K, v \in V \quad (14)$$

$$T_{\min} \leq \sum_{i \in OFL} FD_i \left(\sum_{j \in OFL} X_{ijk}^v + \sum_{j \in MFL} Y_{ijk}^v \right) \leq T_{\max},$$

$$k \in K, v \in V \quad (15)$$

维护约束(14)和(15)分别规定在连续两次维护之间累积FH和FC的下界和上界。除维护约束外，

其他约束定义为普通约束。

2 所提方法

在本节将描述如何建立基于强化学习的算法来求解以维修间隔利用率最优为目标的飞机派遣模型。整体框架如图3所示。

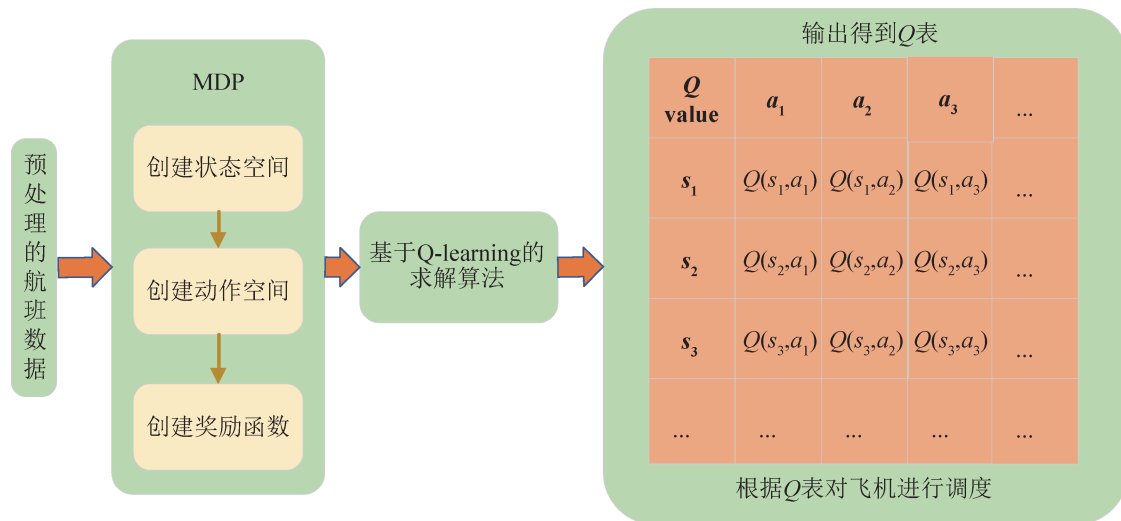


图3 整体框架流程

Fig. 3 Overall framework flow

根据飞机派遣序贯决策的特性将其构造为马尔可夫决策过程(Markov decision process, MDP), 然后基于Q-learning算法来求解该问题模型。

2.1 MDP的建立

2.1.1 创建状态空间

为每个航班段创建相应的五维元组 $FL_i = (c, O_i, D_i, DT_i, AT_i)$ 。每个元组都对应于某个状态 s_i , 每个状态包含着该航班段的具体信息。其中 c 为飞行日, 剩下的4个元素 O_i, D_i, DT_i, AT_i 分别代表航班段 i 的起点机场、到达机场、起飞时间和到达时间。本文所有与时间相关的数据设置为同一时区, 单位 min , 故一天的时间范围为 $0 \sim 1440 \text{ min}$ 。

按照如下规则对MDP状态空间的状态进行排序:

(1) 按照 c 的大小进行排序, 如果多个航班段

处于同一天则按照航班段的起飞时间进行排序。例如以下3种状态:

$$s_1: FL_1 = (2, O_1, D_1, 530, 600),$$

$$s_2: FL_2 = (5, O_2, D_2, 540, 620),$$

$$s_3: FL_3 = (5, O_3, D_3, 720, 810).$$

则3种状态的顺序应为: s_1, s_2, s_3 。

(2) 创建一个列表 $whole_state$ (状态空间) 用来存放所有的航班信息, 内含状态顺序如前所述, 即 $whole_state = [s_1, s_2, s_3, \dots]$ 。

2.1.2 创建动作空间

Agent选择的动作就是为每个航班段分配一架飞机来执行飞行任务。对飞机编号并建立一个列表 $whole_action$ 来记录。当Agent为状态选择动作时, 可选的动作空间并不是完整的列表 $whole_action$, 而是它的一个子集 CU 。该子集 CU

中的动作是满足相关约束条件的。即当 Agent 处于状态 s_t 要选择动作时, 要对动作空间中的动作进行筛选, 需要保证飞机在不同弧时起飞降落机场之间的连接约束(9)和(10); 也需要保证飞机在不同弧时到达时间和起飞时间之间的衔接约束(11)和(12); 同时也需要确保飞机的累积 FH 和 FC 不能超过上下界的约束(14)和(15)。例如一个由 N 架飞机组成的机队, 则 $whole_action=[1, 2, \dots, N]$ 。若此时满足上述相关约束的飞机仅有: 1、3、5、6, 则可选动作空间 CU 为 [1, 3, 5, 6]。MDP 能确保 Agent 在每个状态结束时只被一个动作所覆盖, 映射至飞机派遣模型中确保每个航班段仅由一架飞机执飞, 即约束(3)。

2.1.3 创建奖励函数

在以飞机维修间隔利用率最优为目标的飞机派遣模型中, 目标函数为 α 和 β 之间的最小化差距, 充分利用维修间隔, 减少维修资源的浪费以达到飞机均衡使用的目的。根据所提目标函数创建奖励函数。对于所提目标函数, 希望 α_k^v 和 β_k^v 之

间差的绝对值之和尽可能小。而对于奖励函数, 需要使其与目标函数呈负相关。Q 值一般初始化为 0, 若奖励函数为负值, Agent 在选择动作时会违反累积奖励最大化的原则选择 Q 值为 0 所对应的动作, 因此奖励函数不能为负值。设置奖励函数为

$$r_t = \frac{1}{\sum_{k \in K} |\alpha_k^v - \beta_k^v| + \delta}$$

式中: δ 为一个相当小的值, 此处设置 δ 的目的是为了确保当目标函数为 0 时, Agent 能得到一个巨大且有意义的奖励。当 Agent 处于状态 s_t 时, 飞机对应的维护操作次数 v 已经确定。故奖励函数的设定符合要求。

2.2 基于 Q-learning 的求解算法

利用前述内容, 在 Q-learning 的基础上建立相应的算法细节。图 4 显示了 Q-learning 的算法流程、状态转移的概述以及整个循环方法。首先创建所需的列表和字典, 然后给出确切的算法流程。在算法中所需要用到的工具如表 2 所示。

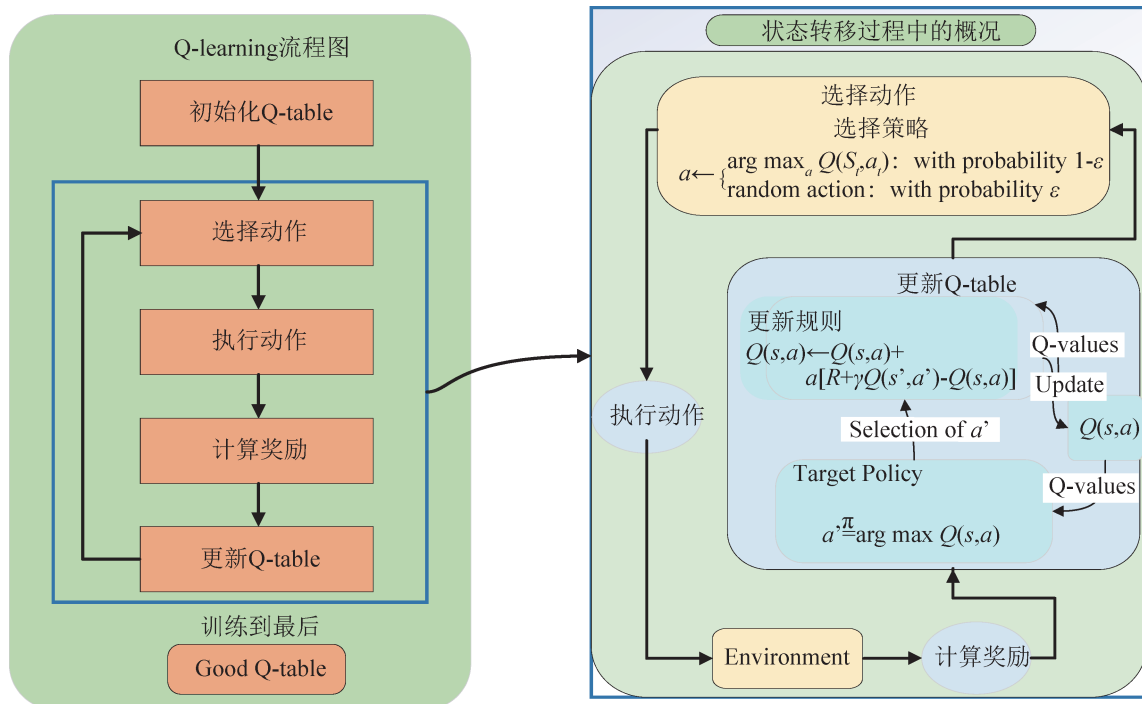


图 4 Q-learning 的算法流程和状态转移示意图
Fig. 4 Algorithmic flow and state transfer of Q-learning

表 2 准备工作
Table 2 Prepared tools

名称	类型	名称
Q-table	DataFrame	状态价值表,行索引是列表 <i>whole_action</i> 中的元素,列索引是列表 <i>whole_state</i> 中的元素。表内元素初始化为 0
Cu-table	DataFrame	用来存储可选动作空间 CU 的状态值
IU	Dictionary	形如 $\{k:[k,v,flight_time,FL_1,FL_2,\dots]\}$,其中 <i>flight_time</i> 是进行维护操作 <i>v</i> 和下一次维护操作之间的累积飞行时间。初始状态为空
NIU	Dictionary	形如 $\{k:v,\dots\}$,这用来存储维修后可投入使用但尚未允许的飞机信息。初始状态为 $\{1:0,2:0,\dots,M:0\}$,其中机队规模为 <i>N</i>
MTN	Dictionary	形如 $\{k:[c,v]\}$,用于存储要维护飞机的相关信息。初始状态为空

图 4 中第二部分对 Q-learning 中的状态转换策略做了说明,包括 Agent 的行为选择策略和 Q 值的更新方法。如图 5 所示,Agent 在每个状态下都要与环境交互一次,在每轮学习中进行交互直至 Agent 达到最终状态时结束。下面结合图 4 详细说明图 5 中 Agent 在状态 s_t 时与环境交互细节。

(1) 根据 IU 中的信息更新 IU、CU、NIU 和 MTN。首先,应确定 IU 中飞机的累积 FH 和 FC 是否处于规定的上下界之间以满足约束条件(14)和(15)。然后判断 IU 中的飞机是否符合在不同弧时相邻两个航班段之间的位置衔接约束(9)和(10);同时判断在不同弧时相邻两个航班段之间是否符合时间衔接约束(11)和(12)。将满足这些约束条件的飞机和 NIU 中飞机均添加至 CU 中。

(2) 为 Agent 的动作选择创建 Cu-table。前一个动作空间 CU 中的元素被用作 Cu-table 的列索引,Cu-table 的行索引是此时的状态。Cu-table 中的“状态动作”的值就是此时 Q-table 中“状态动作”的对应值。之后以 $1-\epsilon$ 的概率从 Cu-table 中选择对应于“状态-动作”最大值的动作;以 ϵ 概率从 CU 中随机选择一个动作,即 ϵ -greedy 策略^[26]。

(3) 根据选定的动作 k 对 IU、NIU、MTN 进行更新。首先,确定飞机 k 在 IU 中是否存在。如果存在,则更新飞机 k 的信息,并将相应的飞行信息添加到飞机的值中。如果 IU 中没有这架飞机,则这架飞机就在 NIU 中,之后需要将 NIU 中的飞机信息添加到 IU 中,同时删除 NIU 中的飞机信

息。根据此时状态,判断 MTN 中的飞机是否满足再次投入飞行的条件满足约束(13)。若满足,则将该飞机移至 NIU;若不满足,则不需要处理。此时可能有多架飞机,其处理方式均相同。

(4) 根据选定的动作 k 计算奖励和目标值。首先,计算每架飞机的 α 和 β ,然后将 IU 中所有飞机的相应 α 和 β 分别求和,得到 α_{total} 和 β_{total} 。

最后计算出奖励 $\frac{1}{|\alpha_{total}-\beta_{total}|+\delta}$ 和目标值。

当达到最终状态时,计算其目标值即可。

(5) 更新 Q-table。更新的原则:当下一个状态不是最终状态时, Q-learning 更新公式为

$$Q_{k+1}(s_t, a_t) = Q_k(s_t, a_t) + \lambda [R_{t+1} + \gamma \max_a Q_k(s_{t+1}, a_{t+1}) - Q_k(s_t, a_t)]$$

当下一个状态是最终状态时,取 $Q_{k+1}(s_t, a_t) = R_{t+1}$ 。

(6) 确定 Agent 是否遍历到最终状态。如果下一个状态不是最终状态,则将下一个状态作为当前状态,并返回到第一步,让 Agent 继续遍历剩余状态。若下一个状态是最终状态,可以按照(5)中的更新模式得到一个更新的、完整的 Q-table。

3 实验和分析

本节介绍了使用基于强化学习算法来求解以维修间隔利用率最优为目标的飞机派遣而进行的实验。测试案例是基于某中国航空公司的数据。

在所有的测试案例中, 首先分别给出数据量最小和最多的案例。然后给出所有测试案例的实验结果并进行分析。最后使用商业求解器 Gurobi 求解飞机派遣模型, 并与基于 Q-learning 的求解算法进行对比以验证实验结果的准确性。所有实验均在 Intel(R) Xeon(R) W-2133 CPU @ 3.60GHz, 操作系统为 Windows 10-64 位的计算机上运行, 其中所提算法采用 Python 编程语言在 Pycharm 软件中编写并调试。

3.1 实验相关数据设置

所有测试案例中的时间数据均按照 2.1.1 中规定设置。其中飞机的周转时间 TAT 设置为 45 min。除最后一组测试案例外, 将飞机维护间隔中的最大允许 FC 取 30, 将最大允许 FH 取 65×60 min, 即 $TF_{max} = 30$ 和 $T_{max} = 65 \times 60$ min。而飞机 FC 和 FH 所取下界分别为 28 和 60.7×60 min。即 $TF_{min} = 28$, $T_{min} = 60.7 \times 60$ min。其中:

$$T_{min} = T_{max} - (TF_{max} - TF_{min}) \cdot (T_{max} / TF_{max})。$$

在算法中所采用的参数学习率和折扣系数分别取 0.2 和 1。关于 ϵ -greedy, 设计等式为

$$\epsilon = \frac{1}{1 + \exp(\frac{10(episode - 0.4 \times episode_max)}{episode_max})}$$

式中: $episode_max$ 是设定的最大学习回合数, 目标是在 Agent 刚开始学习的时候有足够的概率去探索新的动作。

以 $episode_max = 50$ 为例展示一下 ϵ 与学习回合数(episode)的关系曲线。关于 ϵ -greedy 策略, 设定 Agent 探索的概率随着训练次数的增加从接近 1 下降至接近 0, 即如图 6 所示。随着学习回合数的增加, Agent 慢慢获得了经验, 然后以 $1-\epsilon$ 的概率来根据经验选择最佳行为。

在实验中所用实例信息见表 3。在进行实验的测试案例中, 都是 6 天飞行计划。即目标是最大化在 6 天的航班计划中的维修间隔利用率。

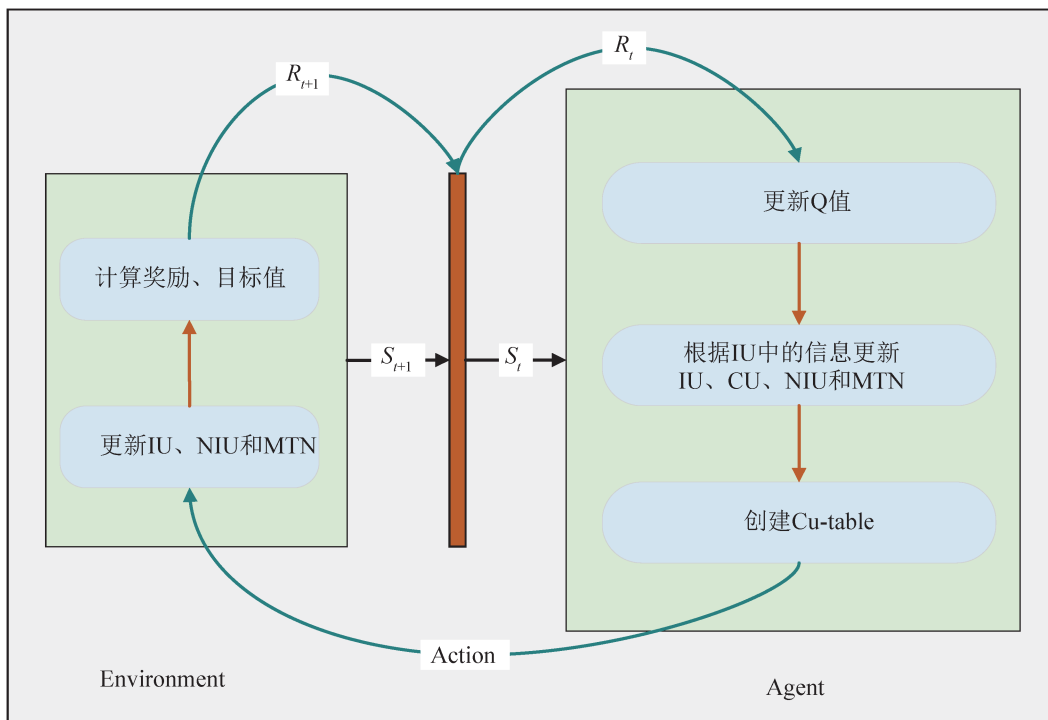


图 5 Agent 与环境交互示意图
Fig. 5 Agent's interaction with environment

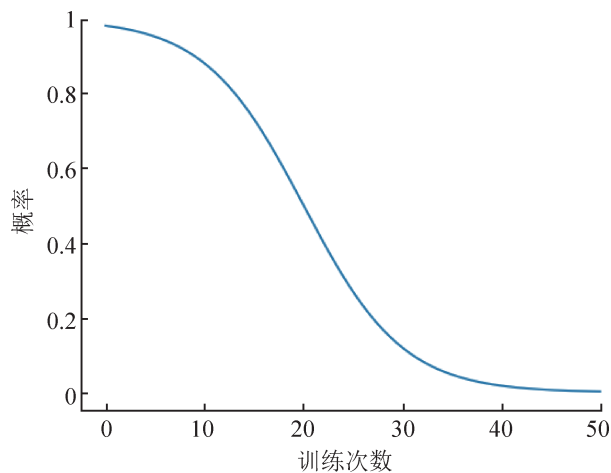


图6 ϵ -greedy策略
Fig. 6 ϵ -greedy strategy

表3 测试案例的信息

Table 3 Information about test cases

案例	航空公司	机型	航班段	机队规模	机场数
1	A	A332	46	20	5
2	A	AT7	104	17	8
3	A	ERG	204	26	20
4	A	M90	297	36	18
5	A	A333	318	35	8
6	B	A320	761	60	26

3.2 具体测试案例

本节以其中2个案例Case 1和Case 6来进行实验说明。在本节中，将最大的学习回合数 $episode_max$ 设定为50。对Case 1进行了20次实验，实验结果如图7所示。这些目标值的平均值

为0.409 744，最优值为0.408 974。需要注意的是，在实验过程中保存了14个有效数字，而在最终结果呈现时，只保留6位有效数字。

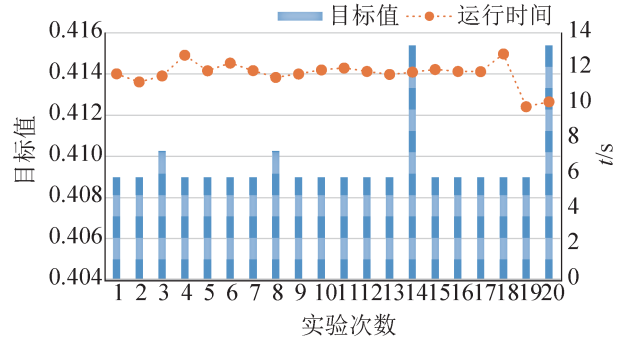
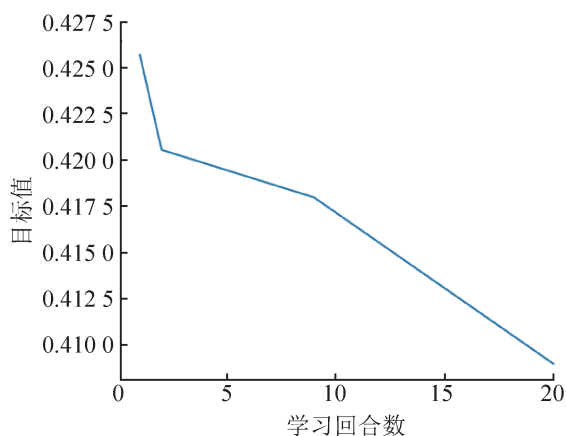
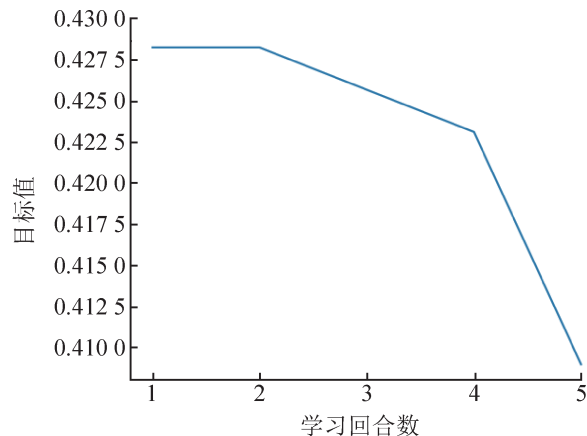


图7 Case 1通过多组实验获得的目标值
Fig. 7 Case 1 Objective values obtained by multiple experiments

在此，做如下处理 $Gap\% = (\bar{m} - m_{best}) / m_{best}$ ，其中 \bar{m} 是多次实验的平均值， m_{best} 是多次实验的最优值。故 $Gap\%$ 可以作为算法的性能指标。对于测试案例Case 1对应的是： $Gap\% = 0.188\%$ 。从Case 1测试案例结果来看，该实验最佳结果是0.408 974。在找到最佳值的几个实验中，目标值随学习回合数的变化如图8所示(图中4个实验的结果的图像分别为第1、2、4、5次实验)。以第4次实验为例，分析实验输出结果：从该实验中得到的Q-table是本文为每架飞机安排飞行任务的主要依据。该次实验输出部分Q-table如表4所示。



(a) 第1次实验



(b) 第2次实验

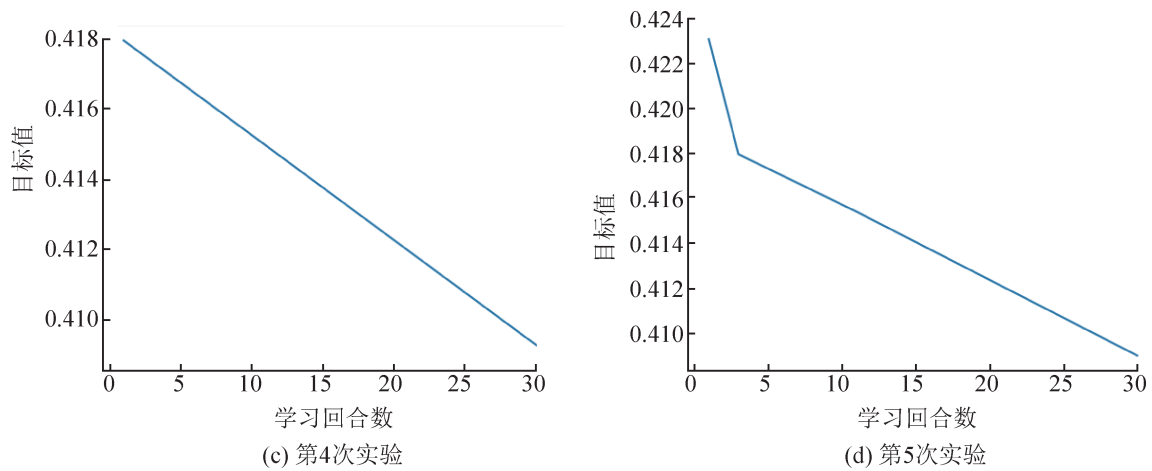


图 8 四次求得最优值时的变化曲线

Fig. 8 Variation curves when optimal value is obtained four times

表 4 Q-table 部分显示
Table 4 Partial display of Q-table

航班段	飞机编号						
	1	2	3	4	5	6	...
1	35.615 770	74.899 057	22.182 434	56.255 023	30.230 163	20.566 729	...
2	23.066 095	40.096 695	14.022 655	24.676 447	14.022 655	40.003 080	...
3	25.138 153	0	35.558 799	23.353 293	19.880 945	22.608 453	...
4	0	25.842 968	18.171 796	36.765 886	18.595 400	33.518 226	...
5	0	26.769 401	7.962 807	13.026 680	6.016 170	6.892 842	...
6	19.287 721	17.136 571	4.150 836	6.983 535	0	4.057 789	...
7	13.914 893	8.576 375	3.251 386	8.346 326	11.842 005	18.997 614	...
8	4.131 718	15.623 605	0	7.608 969	9.590 552	9.538 389	...
...

由于飞机的初始状态没有差别, 每次产生的 Q-table 并不一定保持一致, 但是不影响实验结果。该实验的最优目标值所对应的 IU 为

$IU = \{ '10': ['10', 0, 310, (1.0, 'BJS', 'CAN', 555.0, 740.0), (3.0, 'SHA', 'CAN', 1005.0, 1130.0)], '18': ['18', 0, 190, (1.0, 'BJS', 'CAN', 675.0, 865.0)] \dots \}$

而 NIU 和 MTN 在此时都是空的。这是因为此时所有的飞机都有相应的飞行任务, 但都没有超过设定的飞机维护间隔时间。在该轮实验中, 该机队中所有飞机都不需要进行维护。

之后对 Case 6 做同样实验, 实验结果如图 9 所示。在该实验中, 对 T_{max} 和 TF_{max} 做如下改动:

$T_{max} = 50 \times 60 \text{ min},$
 $TF_{max} = 20$

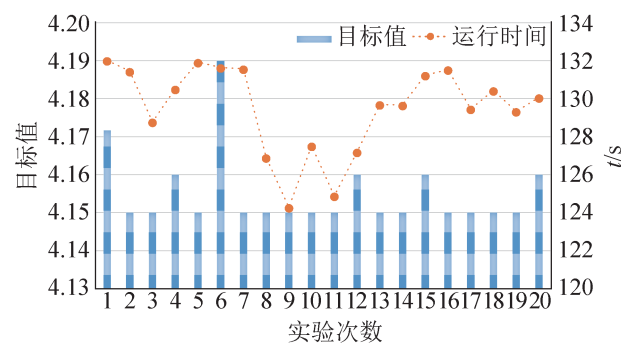


图 9 Case 6 通过多组实验获得的目标值

Fig. 9 Case 6 Objective values obtained by multiple experiments

FH 和 FC 的下界分别为 $45 \times 60 \text{ min}$ 和 18。该测试样例的最优值为 4.150 00, 平均值为 4.155 08, 对应的 Gap% 为 0.122 4。该测试样例实验的结果

与之前的实验不同，在找到最优值时，有飞机不再满足维修约束。当找到最优值时：

$$IU = \{ '15': ['15', 0, 176.5, (1.0, 'CKG', 'BJS', 440.0, 580.0) \dots] \}$$

$$MTN = \{ '35': [6.0, 1], '5': [6.0, 1] \}$$

此时 NIU 仍然为空。由于一些飞机在飞行的最后一天因不再符合维修间隔而需要维修但未完成维护，故维修次数为 1。这些都是与 Case 1 实验结果的不同之处。故能保证当维护间隔相对较小而

飞机需要维护时，所提出的算法仍能找到最优值。

3.3 所有案例实验数据整理

对其他 4 个测试案例也做同样的处理，实验结果如图 10 所示。

图 10 显示，在每个测试案例的实验中，所提出的算法都能找到最优值。然而，运行时间并不稳定，最主要的原因是，当 Agent 选择不同的动作进而计算奖励和更新 Q-table 时，计算量有差异。

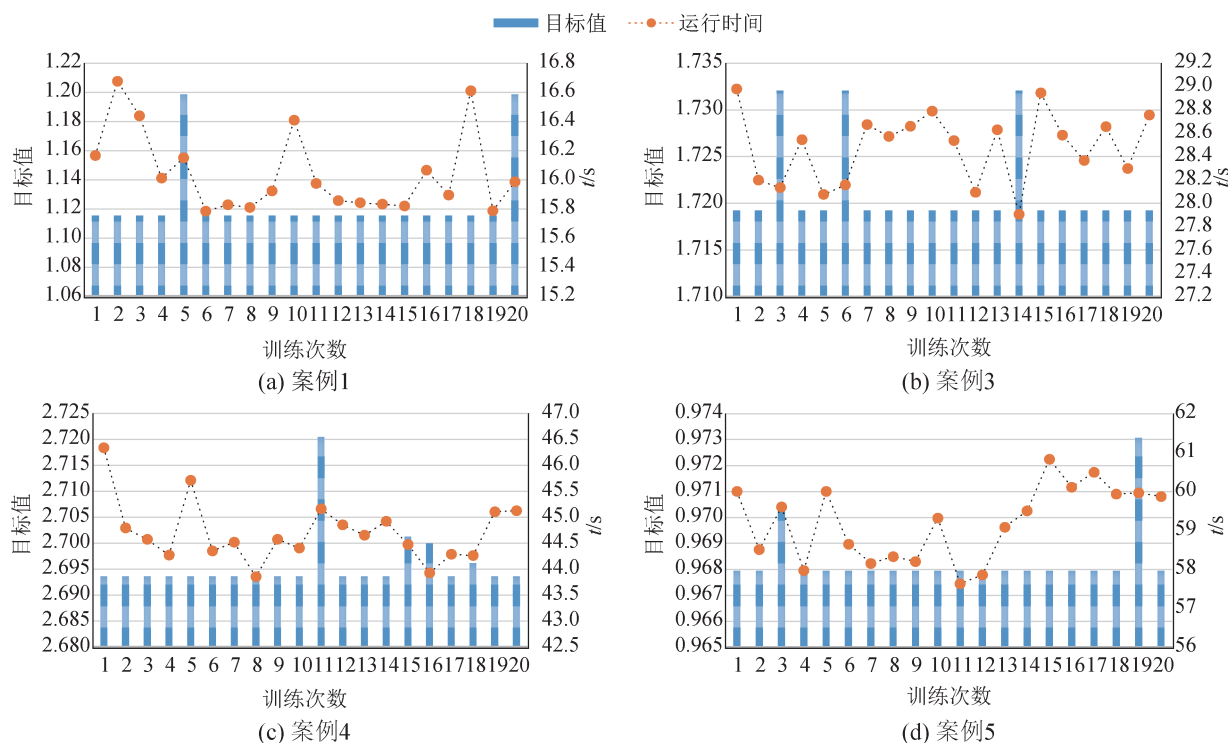


图 10 其他测试案例的实验结果
Fig. 10 Experimental results of other test cases

结合前面 2 个测试案例，整理实验结果，得到表 5。表 5 中： \bar{t} 为 20 次实验的平均运行时间。

表 5 不同测试案例的实验结果
Table 5 Experimental results of different test cases

案例	m_{best}	\bar{m}	\bar{t}/s	Gap%	有无维修
1	0.408 974	0.409 744	11.688 5	0.188 1	No
2	1.115 38	1.123 72	16.043 5	0.747 1	No
3	1.719 23	1.721 15	28.477 3	0.111 9	No
4	2.693 59	2.695 77	44.708 2	0.080 9	No
5	0.967 949	0.968 333	59.201 5	0.039 7	No
6	4.150 00	4.155 08	129.448	0.122 4	Yes

通过表 5 可以看到，对于每个测试案例来说，Gap% 都保持着很低的水平，最低达到了 0.039 7%，这就证明了所提出的算法在求解飞机派遣时性能是比较稳定的。最优值应随着测试用例规模的增加而变大，但测试案例 5 产生的目标值相对较小。这是因为算法所需的动作空间是由机队指派决定的，而算法所需的状态空间是由航班时刻制定决定的，故分析得出，实验结果受前期航班时刻制定和机队指派影响。所以在未来的工作中，可以把航班时刻制定、机队指派和飞机派遣两两结合

或三者结合设计算法。从3.2节可以得知, 在Case 6的实验中, 会有飞机进行维修。

为避免由于学习迭代次数过少而导致最优解不够好, 将Case 1的学习迭代次数由50次改为500次。其实验最优值与学习次数为50时得到的同为0.408 974。然而, 当学习回合数设置为500时, 求解时间为144.163 s。当得到的最优解没有变化时, 没有必要将求解时间扩大124倍以上。

3.4 算法收敛性与准确性

当应用一个新方法来求解给定的优化问题时, 首先想到的是所求结果是否收敛于最优解。分别对六组测试案例进行收敛性的实验, 并对结果收敛性进行说明, 目标值随学习回合数变化如图11所示。该图表明6组测试案例均收敛于最优解。这是因为Agent在学习的过程中探索新行为的概率越来越低逐渐接近0(如图6所示), 随之Agent根据所获得的经验选择最佳行为, 使得目标值收敛于最优解。

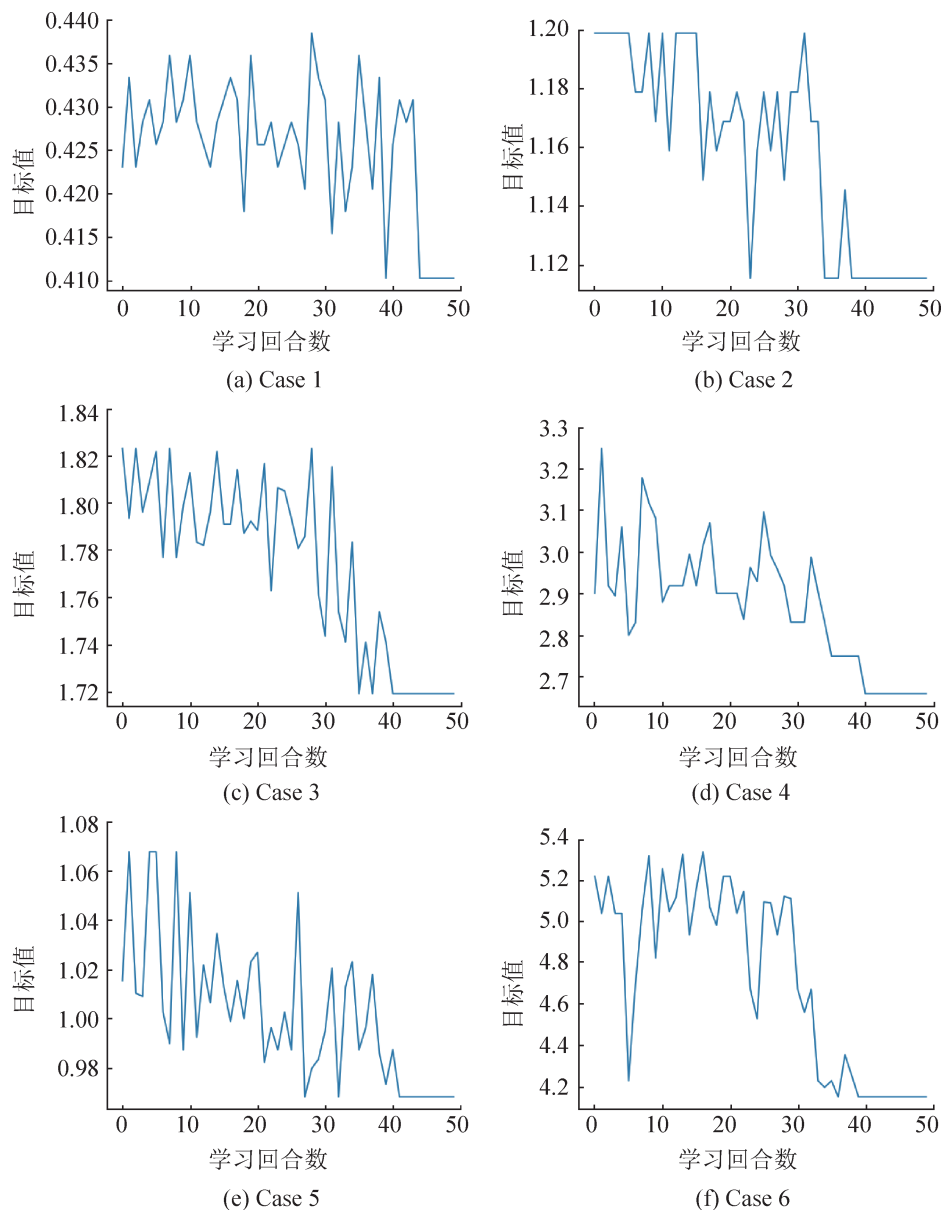


图11 6组测试案例实验结果收敛性

Fig. 11 Convergence of experimental results for six test case groups

其次为了验证实验结果的准确性, 本文使用商业求解器 Gurobi 对飞机派遣线性整数规划模型进行求解。但是商业求解器 Gurobi 对于大中型数据集获得最优解是非常具有挑战性的, 因此案例 4、5、6 的求解时间分别设置为 2 400 s、5 000 s 和 8 000 s。求解实验结果与基于 Q-learning 算法实验结果对比如表 6 所示。

表 6 两种算法实验结果
Table 6 Experimental results of two algorithms

案例	Q-learning-based		Gurobi	
	m_{best}	t/s	M_{g-best}	t/s
1	0.408 974	10.282 5	0.408 974	9.753 39
2	1.115 38	15.121 4	1.115 38	65.755 9
3	1.719 23	27.912 1	1.719 23	652.556
4	2.693 59	42.611 5	2.693 59	2 400.00
5	0.967 949	58.856 4	0.967 949	5 000.00
6	4.150 00	128.125	4.186 45	8 000.00

该结果能够证明基于 Q-learning 算法实验结果的准确性, 同时能够展现其在大中型规模问题上优于商业求解器 Gurobi。案例 6 的实验结果出现了差异是因为 Gurobi 设置的求解时间限制了其找到更好的求解结果。

4 结论

本文从维修保障的角度建立以维修间隔利用率最优为目标的飞机派遣模型。在满足飞机累积 FH 和 FC 上下界的前提下, 最小化飞机的累积 FH 和 FC 在同一换算关系下的相对差距, 为航班段分配飞机执行飞行任务从而充分提高飞机利用率节约维修成本。提出了基于强化学习中的 Q-learning 求解算法。使用该算法在 6 个测试案例上进行了实验, 每个案例都能找到相应的最优值。并验证了该算法的性能, 该算法 Gap% 最低达到了 0.039 7%, 最高只有 0.747 1%。同时该算法的收敛性与准确性也得到了验证。这为研究飞机派遣问题提供了一定的参考。

然而, 本文并没有证明这种最优解的经济效益如何。因此, 考虑飞机派遣模型的经济化是未

来的任务之一。在实际的飞机派遣中, 还有更多的细节约束需要考虑, 如维修人员; 一些不确定因素, 如飞机延误等。

参考文献:

- [1] Temucin T, Tuzkaya G, Vayvay O. Aircraft Maintenance Routing Problem-a Literature Survey[J]. Promet-Traffic & Transportation, 2021, 33(4): 491-503.
- [2] Jamili A. A Robust Mathematical Model and Heuristic Algorithms for Integrated Aircraft Routing and Scheduling, with Consideration of Fleet Assignment Problem[J]. Journal of Air Transport Management, 2017, 58: 21-30.
- [3] Eltoukhy A E E, Wang Z X, Chan F T S, et al. Robust Aircraft Maintenance Routing Problem Using a Turn-around Time Reduction Approach[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 50 (12): 4919-4932.
- [4] Bulbul K G, Kasimbeyli R. Augmented Lagrangian Based Hybrid Subgradient Method for Solving Aircraft Maintenance Routing Problem[J]. Computers & Operations Research, 2021, 132: 105294.
- [5] Mehmet Başdere, Ümit Bilge. Operational Aircraft Maintenance Routing Problem with Remaining Time Consideration[J]. European Journal of Operational Research, 2014, 235(1): 315-328.
- [6] Yan Chiwei, Kung J. Robust Aircraft Routing[J]. Transportation Science, 2018, 52(1): 118-133.
- [7] Maher S J, Desaulniers G, François Soumis. The Daily Tail Assignment Problem Under Operational Uncertainty Using Look-ahead Maintenance Constraints[J]. European Journal of Operational Research, 2018, 264(2): 534-547.
- [8] Eltoukhy A E E, Wang Z X, Chan F T S, et al. Data Analytics in Managing Aircraft Routing and Maintenance Staffing with Price Competition by a Stackelberg-Nash Game Model[J]. Transportation Research Part E: Logistics and Transportation Review, 2019, 122: 143-168.
- [9] Eltoukhy A E E, Chan F T S, Chung S H, et al. A Model with a Solution Algorithm for the Operational Aircraft Maintenance Routing Problem[J]. Computers & Industrial Engineering, 2018, 120: 346-359.
- [10] Vildan Özkır, Mahmud Sami Özgür. Two-phase Heuristic Algorithm for Integrated Airline Fleet Assignment and Routing Problem[J]. Energies, 2021, 14(11): 3327.
- [11] Chen C H, Chou F I, Chou J H. Multiobjective Evolutionary Scheduling and Rescheduling of Integrated Aircraft Routing and Crew Pairing Problems[J]. IEEE

- Access, 2020, 8: 35018-35030.
- [12] Liang Zhe, Chaovalitwongse W A, Huang H C, et al. On a New Rotation Tour Network Model for Aircraft Maintenance Routing Problem[J]. *Transportation Science*, 2011, 45(1): 109-120.
- [13] 李耀华, 王磊. 基于改进遗传算法的飞机排班优化方法研究[J]. *系统仿真学报*, 2016, 28(3): 620-626.
Li Yaohua, Wang Lei. Study on Aircraft Scheduling Optimization Based on Improved Genetic Algorithm[J]. *Journal of System Simulation*, 2016, 28(3): 620-626.
- [14] Al-Thani N A, Ben Ahmed M, Haouari M. A Model and Optimization-based Heuristic for the Operational Aircraft Maintenance Routing Problem[J]. *Transportation Research Part C-Emerging Technologies*, 2016, 72: 29-44.
- [15] Cui Ruyu, Dong Xingye, Lin Youfang. Models for Aircraft Maintenance Routing Problem with Consideration of Remaining Time and Robustness[J]. *Computers & Industrial Engineering*, 2019, 137: 106045.
- [16] Eltoukhy A E E, Chan F T S, Chung S H, et al. Heuristic Approaches for Operational Aircraft Maintenance Routing Problem with Maximum Flying Hours and Manpower Availability Considerations[J]. *Industrial Management & Data Systems*, 2017, 117(10): 2142-2170.
- [17] Ruan J H, Wang Z X, Chan F T S, et al. A Reinforcement Learning-based Algorithm for the Aircraft Maintenance Routing Problem[J]. *Expert Systems with Applications*, 2021, 169: 114399.
- [18] Drakaki M, Tzionas P. Manufacturing Scheduling Using Colored Petri Nets and Reinforcement Learning[J]. *Applied Sciences*, 2017, 7(2): 136.
- [19] Wang Haoxiang, Sarker B R, Li Jing, et al. Adaptive Scheduling for Assembly Job Shop with Uncertain Assembly Times Based on Dual Q-learning[J]. *International Journal of Production Research*, 2021, 59 (19): 5867-5883.
- [20] Chen Ronghua, Yang Bo, Li Shi, et al. A Self-learning Genetic Algorithm Based on Reinforcement Learning for Flexible Job-shop Scheduling Problem[J]. *Computers & Industrial Engineering*, 2020, 149: 106778.
- [21] Zhou Tong, Tang Dunbing, Zhu Haihua, et al. Reinforcement Learning with Composite Rewards for Production Scheduling in a Smart Factory[J]. *IEEE Access*, 2021, 9: 752-766.
- [22] Guo W, Atasoy B, Negenborn R R. Global Synchronodal Shipment Matching Problem with Dynamic and Stochastic Travel Times: a Reinforcement Learning Approach[J/OL]. *Annals of Operations Research*, 2022. (2022-01-21) [2022-03-15]. <https://doi.org/10.1007/s10479-021-04489-z>.
- [23] Zhang Ke, He Fang, Zhang Zhengchao, et al. Multi-vehicle Routing Problems with Soft Time Windows: A Multi-agent Reinforcement Learning Approach[J]. *Transportation Research Part C: Emerging Technologies*, 2020, 121: 102861.
- [24] Eda Köksal Ahmed, Li Zengxiang, Veeravalli B, et al. Reinforcement Learning-enabled Genetic Algorithm for School Bus Scheduling[J]. *Journal of Intelligent Transportation Systems*, 2022, 26(3): 269-283.
- [25] He Shengxue, He Jianjia, Liang Shidong, et al. A Dynamic Holding Approach to Stabilizing a Bus Line Based on the Q-learning Algorithm with Multistage Look-ahead[J]. *Transportation Science*, 2022, 56(1): 31-51.
- [26] Šemrov D, Marsetič R, Žura M, et al. Reinforcement Learning Approach for Train Rescheduling on a Single-track Railway[J]. *Transportation Research Part B: Methodological*, 2016, 86: 250-267.