**基于多智能体协作的停机位分配算法**

卢敏1,2▲, 张振华1,2，李维康1,2

(1.中国民航大学 计算机科学与技术学院,天津300300；

2.中国民航大学民航智慧机场理论与系统重点实验室,天津 300300)

**摘 要：**停机位分配旨在统筹机型匹配、到达时间和飞机冲突等约束，将机位合理分配给航班，提高机场运营效率。现有停机位分配主要采用运筹优化方法解决单个机场的机位分配问题，存在学习效率低、缺乏泛化性和稳定性等问题。为此，将停机位分配建模为部分可观测马尔可夫决策过程，以最大化停机位利用率为最终目标，提出基于多智能体协作的停机位分配算法，其核心是：为解决停机位部分可观测情况，采用DRQN(Deep Recurrent Q-Network)模型记录各个智能体状态变化从而更好的了解环境,提高智能体的决策性能；针对机场环境中停机位之间复杂的协作关系，设计一个混合网络，将各个智能体的状态动作值（Q值函数）通过一个超网络线性组合得到全局Q值函数，帮助智能体在全局层面上协同合作，达到全局最优分配方案的目的；为解决训练中智能体可能存在的具有不同动作空间的问题，通过添加动作掩码层提高学习效率。基于厦门机场运行数据的实验数据集，实验结果表明：在不同规模情境下，所提出的停机位分配算法在近机位时间利用率方面表现显著：相对于线性规划算法，该算法最多提升了13.8%的时间利用率；相对于贪婪算法和遗传算法，提升幅度最高达11.9%；相对于单智能体强化学习算法，提升最多达到9%。此外，在当前的环境下近机位使用率达到100%。

**关键词**：智能交通；停机位分配；多智能体系统；多智能体强化学习；协同决策

**中图分类号**: TP391 **文献标志码:** A

**Multi-Agent Collaborative Airport Gate Assignment Algorithm**

LU Min1,2▲ ,ZHANG Zhenhua1,2 ,LI Weikang1,2

（1.College of Computer Science and Technology, Civil Aviation University of China ,Tianjin 300300,China;

(2.Key Laboratory of Smart Airport Theory and System, Civil Aviation University of China, Tianjin 300300,China)

**Abstract：**Airport gate assignment aims at efficiently assign gates to flights by considering constraints such as aircraft type compatibility, arrival times, and potential conflicts. Current methods for airport gate assignment primarily rely on optimization algorithms from operations research, focusing on solving the allocation problem for individual airports. However, encountering issues such as low learning efficiency, lack of generalization, and instability. To overcome these challenges, proposes modeling airport gate assignment as a partially observable Markov decision process, with the ultimate goal of maximizing gate utilization. a multi-agent collaborative airport gate assignment algorithm for is proposed: Addressing the partially observable nature of airport gate availability, employs a Deep Recurrent Q-Network model to record state changes for each agent, enhancing its understanding of the environment and improving their decision-making performance. To handle the complex collaborative relationships among gate, a mixing network is designed. This network combines the local State-Action-Value Function (Q-value functions) of individual agents through a hypernetwork, resulting in a linear combination that produces a global Q-value function, this facilitates cooperative coordination among agents at a global level, aiming to achieve a globally optimal allocation solution. To tackle potential discrepancies in the action spaces of intelligent agents during training, employs an action mask layer to enhance learning efficiency. Based on the operational data from Xiamen Airport, an experimental dataset was obtained. The research results demonstrate that the proposed airport gate assignment algorithm exhibits significant performance in near-gate time utilization across various operational scales: in comparison to linear programming, the algorithm achieves a maximum improvement of 13.8% in time utilization. When contrasted with greedy algorithms and genetic algorithms, the algorithm attains a maximum enhancement of 11.9%. Moreover, relative to single-agent reinforcement learning algorithms, the algorithm achieves a maximum improvement of 9%. Furthermore, in the given environment, the near-gate spatial utilization rate reaches 100%.

**Key words**：intelligent transportation; airport gate assignment problem; multi agent systems; multi agent reinforcement learning; collaborative decision making

0 引 言

机场停机位是飞机地面保障的关键资源，机场运行控制中心根据航班的到达时间、机型以及机场停机位属性和占用情况，为每个进港航班分配停机位[1]。有效的停机位分配算法可以将停机位公平分配，最大化利用率，从而实现机场运营的优化[2]。

传统停机位算法主要包括数学优化算法[3]和启发式算法[4]。数学优化算法将停机位分配建模为0-1整数优化问题，在飞机机型匹配、到达时间和飞机冲突等约束限制下,优化停机位利用率。启发式算法通过探索搜索空间来确定的解决方案。上述两类优化算法不能适用于复杂变化的停机位分配问题。主要表现在：（1）针对同一机场动态变化的航班，需要反复实验选择合适的求解策略，且求解复杂度随着停机位规模数而指数级增长。（2）单一机场的停机位分配优化方案无法快速迁移到其它机场，尽管逻辑上诸多机场的停机位分配问题类似，各自却有独特规则。

针对上述问题，使用深度强化学习解决机场停机位问题，一方面利用强化学习建模停机位分配这一运筹优化问题，将停机位分配问题抽象为序列决策过程，即当前的停机位占用情况将直接影响后续停机位分配[5]。通过与环境（或所有航班计划）动态交互，学习最大化累计奖励的停机位分配策略。另外一方面，为了停机位分配策略能够适用不同机场的各类场景，增强策略泛化性，在策略构建中引入深度学习以表征各类场景，也突破运筹优化算法无法迁移的问题。

为了解决停机位之间相互协作问题，本文进一步将停机位分配建模为多智能体协作问题。其中每一个停机位视为一个智能体，通过构建多智能体强化学习算法，以学习具有最大累计奖励的分配策略。其主要面临的挑战是：多智能体强化学习算法中智能体存在部分可观测问题，如何将其直接应用到算法中训练；由于各智能体之间复杂的关系，如何设计信用分配机制，以有效分配各智能体与环境交互的贡献，从而促进各智能体协作。为此，使用DRQN模型[6]来编码每个智能体的局部观察信息，记录智能体交互轨迹，以克服部分可观测性问题。针对多智能体信用分配，使用一个混合网络来估计全局 Q 值函数，更好的学习智能体之间的合作策略。在基于厦门机场运行情况的数据集上，实验结果表明我们的方法在停机位利用率、训练稳定性和模型泛化方面都优于人工设计的规则以及单智能体强化学习算法。

本文的主要贡献如下：

（1）将停机位分配问题建模为部分可观测马尔可夫决策过程，使用多智能体强化学习算法得出分配方案并训练可以快速响应和高泛化能力的模型。（2）使用DRQN网络模型建模智能体交互轨迹，编码每个智能体的局部观察信息，解决部分可观测性问题。（3）将各个智能体的局部Q值函数通过一个混合网络得到全局Q值函数，从而学习最优的协作决策。

1 相关工作

* 1. **组合优化的停机位分配研究**

停机位分配问题在航空运输领域属于组合优化问题。组合优化问题是在离散状态下，求极值的最优化问题，可以通过穷举法得到问题的最 优解，但是由于可行解数量随问题规模呈指数型增长，无法在多项式时间内穷举得到问题的最优解[8]。数十年来学者对组合优化问题的求解算法进行了大量的研究,当前求解组合优化问题分为数学优化算法和启发式算法。数学优化算法包括线性规划和整数规划[8]。文献[8]中，作者综合考虑航空器机型匹配、缓冲时间和航空器冲突等约束限制,科学合理地权衡机场各种利益需求,并不断实时更改及更新停机位数据,采用分支界定法求解模型优化函数值,将停机位调度问题简化为线性化模型求解。启发式算法包括贪婪算法[9,10]、模拟退火算法[11]、禁忌搜索[12,13]、遗传算法[14]等。文献[10]中，作者先根据贪婪算法先来先服务准则使“丢失”的航班数最少得到模拟退火算法的初始解，然后运用经典模拟退火算法求解出最优指派结果。文献[14]中，作者提出了一种基于遗传算法的停机位分配算法，采用进化和遗传的思想，设计相应的遗传编码方式，通过种群的进化过程来探索最优的停机位分配方案。虽然这些算法可以求解到停机位分配问题的全局最优解，但是停机位环境是不断变化的，在面对不同规模情况时，需要对参数进行不断调整以获得更好的效果，计算成本过高。

* 1. **基于强化学习的停机位分配研究**

强化学习用于描述和解决智能体与环境交互过程中学习策略的问题，达到最大化奖励或实现特定目标的效果[15]。文献[16]中，作者提出了基于 K 最近邻（K nearest neighbor，KNN）和深度确定性策略梯度（deep deterministic policy gradient，DDPG）的机位分配算法，即DDPG\_KNN，实验验证了DDPG\_KNN算法的有效性，其求解的停机位利用率结果高于传统的停机位算法求解的结果。但智能体做出决策会面临动作空间维度过大的问题，这样不利于智能体与环境进行交互学习，影响智能体对环境的了解程度,并且停机位环境是一个多智能体场景，单智能体强化学习算法体现不出停机位之间协作关系，影响最终的结果。

多智能体强化学习是强化学习的一种，用于解决多个智能体在相互作用环境中的协同决策问题。所有的智能体的联合动作决定环境的变化和奖励[17]。与传统的强化学习不同，多智能体强化学习需要考虑多个智能体之间的相互影响与协作。智能体的目标不仅仅是最大化自身的奖励，还需要平衡与其他智能体的协同行为，最终达到纳什均衡全局最优。本文提出多智能体协作的停机位分配算法，将停机位分配建模为多智能体环境，各个智能体之间相互协作，以最大化近机位利用率为最终目标。在这种环境中，将每个停机位当作一个智能体，通过强化学习算法控制所有的智能体选择最适合的航班进行机位分配。

2 主要算法

**2.1 问题形式化描述**

问题表述的参数列出如下：

：航班集合F，航班编号m，航班数量M

：停机位集合G，停机位编号n,停机位数量N

：航班m开始占用近机位n的时间

：航班m驶离近机位n的时间

：航班m的飞机型号, 例如B737、A320、A300等机型

：停机位n可停放机型的集合

：机位分配状态，若航班m被分配近机位n上时，为1，否则为0（）

:所有航班占用机位的数量()

：所有航班占用近机位的数量()

：首个航班开始占用停机位时间

：最后一架航班驶离停机位时间

2.1.1 目标函数

（1）最大化近机位时间利用率：



（2）最大化近机位空间利用率：



2.1.2 约束条件

（1）机型约束

各个停机位只能停放特定的机型，不允许停放非可容纳机型。



（2）占用约束

当前时刻只可以允许空闲的停机位进行停放，

当停机位被占用，则不可被别的航班进行停放，直到停机位变为空闲状态。



（3）唯一性约束

每个航班只能被分配到一个停机位，不可出现重复分配停机位情况。



（4）机位进出间隔约束

同一机位上后一架航班的进港时间与前一架航班离港时间的时间间隔应大于最小时间隔。

**2.2 基于多智能体协作的停机位分配建模**

停机位分配被建模为一个部分可观测马尔可夫决策过程，由组成，表示全局状态。在每一步中，每个智能体a选择一个动作，组成一个联合动作。之后进行环境状态更新，所有的智能体共享相同的奖励函数。智能体的观察值，表示智能体a的策略函数。表示折扣因子。2.2.1到2.2.3将介绍状态空间、动作空间以及奖励函数。

2.2.1 状态空间

（1）全局状态空间

全局状态，表示当前停机位环境的全局信息，包括所有航班的信息和所有停机位的信息。全局状态表示为：



如图1所示，表示当前所有停机位的状态：各个停机位当前是否空闲、停机位属性、可容纳飞行型号、当前停放的飞机编号。表示当前所有航班的信息：航班的飞机型号、到达时间、离开时间、选择的停机位编号。全局状态为算法中混合网络提供当前环境的全部信息。



图1 停机位资源图

Fig.1 Gate resource chart

（2）局部状态空间

观测值是智能体策略网络的直接输入，它提供了智能体周围环境的局部观测信息，策略网络可以利用观测值在当前状态下选择适当的动作，智能体的观测值表示为：



其中表示停机位可容纳的飞机型号；表示停机位的属性，具体来说，表示其为近机位还是远机位；表示停机位当前是否空闲，若当前停机位已停放飞机，设置为0，否则为1；用来表示停放的飞机的编号。

2.2.2 动作空间

动作表示智能体a在当前环境下将采取的动作。在停机位分配问题中，动作集表示为,M为总航班数。我们将动作定义为两类：当智能体a选择被航班id为m的飞机停放，=1，否则为0。

2.2.3 奖励函数

我们的最终目标是最大化近机位利用率，因此定义奖励函数使算法在满足约束条件下获得最大的近机位利用率。

停机位利用率奖励：



其中表示停机位占用时间，表示机位被航班占用时间。

机型匹配奖励：

根据停机位可容纳飞机型号，若停机位选择合适的航班,若选择的飞机不可被停放，。



所有智能体总奖励为：



**2.3 算法详细介绍**

这一节将详细介绍多智能体协作的停机位分配算法。算法基于集中式训练、分散式执行（Centralized Training with Decentralized Execution, CTDE）框架，训练两个网络：智能体策略网络和混合网络，根据当前的观察，智能体策略网络选择合适的动作，混合网络为每个智能体提供全局控制。在CTDE框架下，混合网络仅存在于训练过程中，用来学习智能体之间的合作、建模和状态表示。2.3.1与2.3.2将对这两个网络的结构进行详细介绍。

图2多智能体协作的停机位分配算法整体框架

Fig.2 The Framework of Multi-Agent collaborative airport gate assignment algorithm

2.3.1 智能体策略网络

如图2右侧所示，智能体策略网络是一个具有三层神经网络的模型,用来表示智能体策略网络中需要训练的权重参数。策略模型的输入由观测值和动作组成,并对其进行one-hot编码，经过一个线性输入层之后，采用DRQN模型[6]的思想，用一个循环神经网络（Recurrent Neural Network, RNN）来记录历史数据的观测序列信息。最后经过线性输出层得到。这三层网络的计算过程表示为：

其中作为激活函数，*V,U,W*分别是输出线性层、输入线性层和RNN层的权重参数，表示经过t-1次循环后的隐藏信息，表示向量串联运算符。被用于通过混合网络模型计算并选择合适的动作。但是在停机位分配问题中，并非所有从策略模型中选择的动作都是合理的，例如当停机位被飞机停放时，或当前机位进行飞机进出时在这段时间内其不可分配给别的航班，策略模型训练这种情况是非常低效的，因此我们添加一个动作掩码层（action mask）来限制策略网络选择可行的动作。之后通过贪婪策略（-greedy）选择动作：



其中表示当前时刻t的随机变量（0，1），（0，1）表示探索率。表示当前可选择动作。然而为多个智能体训练不同的策略模型是很低效的，因此我们利用参数共享技巧来提高训练效率，所有的智能体将共享相同的策略网络。

2.3.2 混合网络

智能体策略网络对决策过程至关重要，如果没有混合网络的帮助就会导致智能体只考虑自身的局部奖励，难以得到全局最优解，无法对其进行训练，因此混合网络是算法的核心部分。我们用来表示混合网络，表示权重参数。混合网络仅在训练过程中使用，用于将策略网络和混合网络中神经网络反向传播的损失梯度反向传播到每个智能体上，从而训练每个智能体的策略网络。如图2左侧所示，混合网络就是根据每个智能体的q值和全局状态生成,其过程可以表示为：

使用DQN算法[18]中的方法，采用TD error来计算损失。通过训练智能体策略网络与混合网络的参数最小化损失：





其中b表示从重放缓冲区采样的批量大小，表示折扣因子。通过梯度下降的反向传播来训练策略网络和混合网络。整个训练过程如算法1所示。

|  |
| --- |
| 算法1：基于多智能体协作的停机位分配算法 |
| **循环** 轮次 从 1 到 M  重置停机位分配环境  初始化：初始化随机变量  **循环** t= 1 到 最大次数  输入：全局状态，当前观测值 =  **循环** 智能体= 1 到 N  接收当前可执行动作集  选择动作  **结束循环**  执行动作并获得奖励值  将储存到这一轮缓冲区  **结束循环**  将存储到经验池D中  从经验池中随机采样组  计算  通过最小化损失函数更新模型参数：    更新网络参数  **结束循环** |

3 实验部分

这一节对停机位分配算法进行实验部分介绍。首先我们介绍实验设置，包括实验环境、实验所用的数据、实验参数，简单介绍了对比算法。然后对实验得到的结果进行分析，测试其收敛性。其次，改变航班和停机位数量，进行拓展性实验，最后与传统算法做比较。

**3.1 实验设置**

所有的实验均在配备Intel Xeon Silver 4215R @3.20GHz八核CPU，64GB RAM和Nvidia RTX 3090 24GB GPU的计算机上以Python 3.7.1的环境上运行。深度学习平台是PyTorch[19]1.12.1和CUDA 11.4。OpenAI Gym[20]是一个用于开发和比较强化学习算法的工作包，它为强化学习环境定义提供了接口和规范使得开发、测试强化学习算法变得更加简单和高效，因此我们基于OpenAI gym设计停机位分配环境。

**3.2 实验数据集**

本研究所用的数据是基于厦门机场运行数据构建的停机位分配数据集。如表1，2所示，数据集中包括航班的进、离港时刻、航班类型、机型以及停机位相关信息：停机位属性、机位可容纳机型。训练参数如表3所示。

表1部分航班进、离港时刻表

Table 1 Partial flight departure and arrival schedule

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 航班号 | 进港时间 | | 离港时间 | 机型 | 航司 |
| MF8261 | | 10:30 | 12:30 | B737 | 四川航空 |
| 3U2255 | | 13:30 | 15:30 | A320 | 中国南方航空 |
| CZ4403 | | 13:30 | 15:30 | A300 | 长隆航空 |
| GJ3403 | | 13:30 | 15:30 | B737 | 厦门航空 |
| MF8491 | | 13:30 | 15:30 | A300 | 中国东方航空 |
| MU3335 | | 13:30 | 15:30 | B767 | 中国东北航空 |
| NS8491 | | 13:30 | 15:30 | A321 | 中国国际航空 |
| CA4566 | | 15:00 | 17:00 | B767 | 山东航空 |
| SC5150 | | 15:00 | 17:00 | A319 | 深圳航空 |
| ZH4566 | | 15:00 | 17:00 | A321 | 中国南方航空 |
| CZ3724 | | 16:00 | 18:00 | B747 | 重庆航空 |
| MF8261 | | 10:30 | 12:30 | B737 | 四川航空 |
| 3U2255 | | 13:30 | 15:30 | A320 | 中国南方航空 |
| CZ4403 | | 13:30 | 15:30 | A300 | 长隆航空 |
| GJ3403 | | 13:30 | 15:30 | B737 | 厦门航空 |
| MF8491 | | 13:30 | 15:30 | A300 | 中国东方航空 |
| MU3335 | | 13:30 | 15:30 | B767 | 中国东北航空 |
| NS8491 | | 13:30 | 15:30 | A321 | 中国国际航空 |
| CA4566 | | 15:00 | 17:00 | B767 | 山东航空 |
| SC5150 | | 15:00 | 17:00 | A319 | 深圳航空 |
| ZH4566 | | 15:00 | 17:00 | A321 | 中国南方航空 |
| CZ3724 | | 16:00 | 18:00 | B747 | 重庆航空 |
| OQ2024 | | 16:00 | 18:00 | B757 | 海南航空 |

表2 部分停机位信息表

Table 2 Partial gate information Table

|  |  |  |
| --- | --- | --- |
| 停机位编号 | 机位属性 | 可容纳机型 |
| 1 | 近机位 | A321 A320 A319 B737 A300 |
| 2 | 近机位 | B757 B767 A321 A320 A319 A300 A310 B737 |
| 3 | 近机位 | B787-8 B757 B767 A321 A320 A319 A300 A310 B737 |
| 4 | 近机位 | A321 A320 A319 A300 A310 B737 B757 B747 |
| 5 | 近机位 | B757 B767 A321 A320 A319 A300 A310 B737 B747 |
| 6 | 近机位 | B757 B767 A321 A320 A319 A300 A310 B737 |
| 7 | 近机位 | B737 A320 A321 EMB145 EM190 CRJ700 ARJ21-700 |
| 8 | 近机位 | B737 A320 A321 EMB145 EM190 CRJ700 ARJ21-700 |
| 9 | 近机位 | B737 A320 A321 EMB145 EM190 CRJ700 ARJ21-700 |
| 10 | 近机位 | B757 B767 A321 A320 A319 A300 A310 B737 |
| 11 | 近机位 | A321 A319 A320 B737 |
| 12 | 近机位 | B757 B767 A300 A319 A320 |
| 13 | 远机位 | A320 A319 A321 B737 ARJ21-700 |

表3 训练参数说明

Table 3 Experiment parameter configuration

|  |  |
| --- | --- |
| 参数名称 | 参数 |
| 优化器 | RMS优化器 |
| 学习率 | 5e-4 |
| 经验缓存区大小 | 5e3 |
| 批量大小 | 32 |
| 折扣因子 | 0.99 |
| 每回合最大迭代步长 | 20 |

**3.3 对比算法**

我们将人工分配方法、线性规划算法、贪婪算法、遗传算法，以及DQN算法作为对比算法。

1）**人工分配方法** 参照机场运营实际情况综合考虑飞机型号、停机位状态、航空公司要求等多种因素来对停机位进行人工分配。

2）**线性规划算法** 构建线性目标函数以最大化停机位利用率，并设置约束条件确保分配合理，通过线性规划求解器得到最优解。

3）**贪婪算法** 优先分配给最早到达的航班，最大化近机位利用率为最终目标。

4）**遗传算法** 在实验中设置50个种群，使用整数编码方法，为每个停机位分配一个数字，根据停机位利用率、停机位可停放机型设计一个适应度函数来评估每个停机位分配方案的合理程度和效率，不断迭代优化找到最优解。

4）**DQN算法** 在相同的停机位分配环境下，动作空间、观测空间、奖励函数保持一致，使用神经网络作为值函数近似状态动作值，与环境不断地进行交互，学习最优的停机位分配方案。

**3.4 算法收敛性**

本文算法通过最大化奖励函数值来学习停机位分配策略，如图3所示，奖励值随着迭代次数的增加逐渐收敛，并且奖励值随着训练的继续而提高，表明本文算法的收敛有效性。

图4为我们的算法与DQN算法在训练过程中的收敛性对比。结果显示，我们的算法在最终效果和稳定性都具有一定的优势。

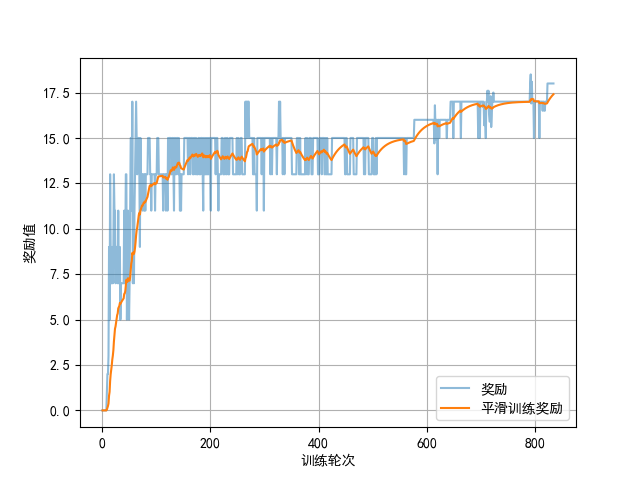


图3 算法收敛性

Fig.3 Algorithm Convergence

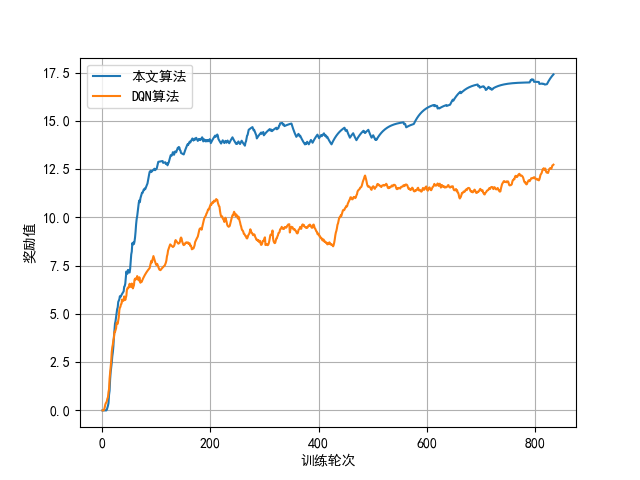


图4本文算法与DQN算法收敛性对比

Fig.4 Convergence Comparison with DQN

**3.5 拓展性实验与分析**

由于停机位环境的多变性，我们设置三种场景来比较我们的算法与传统算法的泛化能力。场景1机场环境为20个航班，16个停机位进行分配，其中12个近机位，4个远机位；场景2机场环境为28个航班，16个停机位，其中12个近机位，4个远机位，场景3有40个航班，30个停机位，其中25个近机位，5个远机位。如表4所示，我们的算法与传统的算法之间的差距随着航班数和机位数的增加效果更加明显：三个场景中我们的算法近机位时间利用率分别达到44.6%、43.4%、42.5%，相较于人工分配的方案提升9.4%、10%、11%，相较于线性规划算法提升12.1%、12.6%、13.8%，相较于贪婪算法提升11.3%、13.7%、11.9%，相较于遗传算法提升11.8%、10.8%、11.9%，相较于DQN算法提升7.1%、7.5%、9%，并且我们的算法在近机位空间利用率高达100%，结果表明在迭代次数相近的情况下，智能体之间的合作关系变得更加重要，在面对不同场景的情况下依然稳定保持较高的性能。

我们绘制甘特图将场景2中停机位分配方案表示出来，以展示详细的结果，如图5所示。结果表示，本文算法学会充分利用近机位，保证近机位空间利用率最大化。



图5停机位分配甘特图

Fig.5 Gate assignment gantt chart

表4 实验结果对比

Table 4 Comparison of Experimental Results

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 算法 | 近机位空间利用率（%） | | | | 近机位时间利用率（%） | | | | 达到收敛迭代次数 | | | |
| 场景1 | 场景2 | 场景3 | 场景1 | | 场景2 | 场景3 | 场景1 | | 场景2 | 场景3 |
| 人工分配 | 100 | 88.7 | 89.6 | 35.2 | | 33.4 | 31.5 | - | | - | - |
| 线性规划算法 | 86.3 | 85.4 | 88.8 | 32.5 | | 30.8 | 28.7 | 20000 | | 40000 | 40000 |
| 贪婪算法 | 92.5 | 88.4 | 86.6 | 33.3 | | 29.7 | 30.6 | - | | - | - |
| 遗传算法 | 84.3 | 84.3 | 72.5 | 32.8 | | 32.6 | 30.6 | 12000 | | 20000 | 30000 |
| DQN算法 | 100 | 92.8 | 92.5 | 37.5 | | 35.9 | 33.5 | 30000 | | 45000 | 60000 |
| **多智能体协作算法** | **100** | **100** | **100** | **44.6** | | **43.4** | **42.5** | **30000** | | **45000** | **60000** |

**4 结束语**

停机位分配是影响机场运行效率的重要因素之一，本文提出了一种基于多智能体协作的停机位分配算法，将多智能体强化学习算法应用于解决停机位分配问题中。使用DRQN模型编码每个智能体的局部观察信息，以解决部分可观测性问题，将各个智能体的局部Q值函数通过一个混合网络得到全局Q值函数，并且考虑到有些情况下策略网络产生的动作是无效的，训练这种情况是很低效的，因此我们加入动作掩码层，对智能体产生的动作加以约束。通过实验表明，我们的方法在训练稳定性、模型泛化能力都具有不错的效果。

我们未来的工作主要集中在三个方面：（1）针对不同机场的规模、运营条件，设计不同的场景并将算法部署到实际场景中。（2）在未来工作中，考虑在算法中加入不确定性因素，以应对机场实际运营中面临的天气、航班延误等动态因素，使得算法更具鲁棒性和适应性。（3）将应用更多的多智能体强化学习算法来解决停机位分配问题，并总结相应的训练经验。

参考文献

References

1. Cong D,Jun B,Yongxing W. A Hybrid Genetic Algorithm Based on Imitation Learning for the Airport Gate Assignment Problem.[J]. Entropy (Basel, Switzerland),2023,25(4).
2. Bouras A, Ghaleb M A, Suryahatmaja U S, et al. The airport gate assignment problem: a survey[J]. The scientific world journal, 2014, 2014.
3. 李凯文,张涛,王锐等.基于深度强化学习的组合优化研究进展[J].自动化学报,2021,47(11):2521-2537.DOI:10.16383/j.aas.c200551.

Li Kai-Wen, Zhang Tao, Wang Rui, Qin Wei-Jian, He Hui-Hui, Huang Hong. Research reviews of combinatorial optimization methods based on deep reinforcement learning. Acta Automatica Sinica, 2021, 47(11): 2521−2537 doi: [10.16383/j.aas.c200551](http://dx.doi.org/10.16383/j.aas.c200551)

1. Khalid O W, Isa N A M, Sakim H A M. Emperor penguin optimizer: A comprehensive review based on state-of-the-art meta-heuristic algorithms[J]. Alexandria Engineering Journal, 2023, 63: 487-526.
2. Muhafız Yıldız M, Avcı U, Örnek M A, et al. Flight Gate Assignment Problem with Reinforcement Learning[C]//International Conference on Intelligent and Fuzzy Systems. Cham: Springer Nature Switzerland, 2023: 189-196.
3. Hausknecht M, Stone P. Deep recurrent q-learning for partially observable mdps[C]//2015 aaai fall symposium series. 2015.
4. Festa P. A brief introduction to exact, approximation, and heuristic algorithms for solving hard combinatorial optimization problems[C]//2014 16th International Conference on Transparent Optical Networks (ICTON). IEEE, 2014: 1-20.
5. 张红颖,申荣苗,罗谦.基于混合整数规划的停机位优化调度研究[J].计算机科学,2020,47(08):278-283.

ZHANG Hong-ying, SHEN Rong-miao, LUO Qian. [Study on Optimal Scheduling of Gate Based on Mixed Integer Programming](https://www.jsjkx.com/EN/Y2020/V47/I8/278)[J]. Computer Science, 2020, 47(8): 278-283. https://doi.org/10.11896/jsjkx.190400154.

1. Paldrak M, Örnek M A. A GRASP algorithm for multi-objective airport gate assignment problem[C]//International Conference on Intelligent and Fuzzy Systems. Cham: Springer International Publishing, 2022: 548-557.
2. 鞠姝妹,许俐.基于GSAA的停机位指派优化问题的研究[J].交通运输系统工程与信息,2008(01):138-143.DOI:10.16097/j.cnki.1009-6744.2008.01.001.

JU Shu-mei, XU Li. Airport Gate Assignment Problems Based on GSAA[J]. Journal of Transportation Systems Engineering and Information Technology, 2008, 8(1): 138-143 .

1. Teoh E J, Tang H, Tan K C. A columnar competitive model with simulated annealing for solving combinatorial optimization problems[C]//The 2006 IEEE International Joint Conference on Neural Network Proceedings. IEEE, 2006: 3254-3259.
2. Wesley Barnes J, LAGUNA M. Solving the multiple-machine weighted flow time problem using tabu search[J]. IIE transactions, 1993, 25(2): 121-128.
3. Basu S. Tabu search implementation on traveling salesman problem and its variations: a literature survey[J]. 2012.
4. 曾琛,王润东.基于遗传算法的动态飞机停机位分配模型研究[J].计算机测量与控制,2023,31(04):199-204.DOI:10.16526/j.cnki.11-4762/tp.2023.04.031.

Zeng Chen, Wang Run-dong. Research on Dynamic Aircraft Stand Allocation Model Based on Genetic Algorithm,2023,31(04):199-204.DOI:10.16526/j.cnki.11-4762/tp.2023.04.031.

1. Li S E. Deep reinforcement learning[M]//Reinforcement Learning for Sequential Decision and Optimal Control. Singapore: Springer Nature Singapore, 2023: 365-402.
2. 顾存昕.基于马尔科夫的机场机位分配建模与强化学习算法[D].华中科技大学,2021.DOI:10.27157/d.cnki.ghzku.2021.000929

Gu Cun-xin. Research on Modeling of Airport Gate Assignment based on Markov Process and Deep Reinforcement Learning[D]. Huazhong University of Science & Technology,2021.DOI:10.27157/d.cnki.ghzku.2021.000929

1. Zhang K, Yang Z, Başar T. Multi-agent reinforcement learning: A selective overview of theories and algorithms[J]. Handbook of reinforcement learning and control, 2021: 321-384.
2. Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning[J]. arXiv preprint arXiv:1312.5602, 2013.
3. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in neural information processing systems, 2019, 32.
4. Brockman G, Cheung V, Pettersson L, et al. Openai gym[J]. arXiv preprint arXiv:1606.01540, 2016.

**附录：变量说明表**

|  |  |  |
| --- | --- | --- |
| 符号 | 符号意义及格式标注 | 上/下标意义及格式标注 |
|  | 航班集合，斜体，英文大写 |  |
|  | 停机位集合，斜体，英文大写 |  |
|  | 航班占用近机位的时间，斜体，英文大写，单位为：h：min | m为航班编号，n为停机位编号，斜体，英文小写 |
|  | 航班驶离停机位的时间，斜体，英文大写，单位为：h：min | m为航班编号，n为停机位编号，斜体，英文小写 |
|  | 航班的飞机型号，斜体，英文小写 | m为航班编号，斜体，英文小写 |
|  | 停机位可停放机型集合，斜体，英文大写 | m为停机位编号，斜体，英文小写 |
|  | 机位分配状态，若航班被分配近机位上时，为1，否则为0，斜体，英文大写 | m为航班编号，n为停机位编号，斜体，英文小写 |
|  | 所有航班占用机位的数量，斜体，英文大写 | F为航班集合，斜体，英文大写 |
|  | 所有航班占用近机位的数量，斜体，英文大写 | F为航班集合，斜体，英文大写 |
|  | 首个航班开始占用停机位时间，斜体，英文大写，单位为h：min | a为首个航班编号，斜体，英文小写 |
|  | 最后一架航班驶离占用停机位时间，斜体，英文大写，单位为h：min | u为首个航班编号，斜体，英文小写 |
|  | 全局状态，包括停机位状态和航班状态，斜体，英文大写 |  |
|  | 所有停机位的全局状态信息，包括各个停机位当前是否空闲、停机位属性、可容纳飞行型号、当前停放的飞机编号，斜体英文小写 | gates表示停机位，斜体，英文小写 |
|  | 所有的航班全局状态信息，包括航班的飞机型号、到达时间、离开时间、选择的停机位编号 | flights表示停机位，斜体，英文小写 |
|  | 局部观测状态，斜体，英文小写 | i表示停机位编号，斜体，英文小写 |
|  | 停机位可容纳飞机型号集合，斜体，英文小写 | i.model表示停机位编号i可容纳机型，斜体，英文小写 |
|  | 停机位状态，空闲为0，否则为1，斜体，英文小写 | i.isIdle表示停机位i的状态，斜体，英文小写 |
|  | 停机位属性，包括近机位、远机位，斜体，英文小写 | i.attr表示停机位i的属性，斜体，英文小写 |
|  | 当前停机位停放航班编号，斜体，英文小写 | i.plane\_id表示停机位i停放的航班编号，斜体，英文小写 |
|  | 所有智能体的动作集合，斜体，英文大写 |  |
|  | 智能体的动作集合，斜体，英文小写 | a表示智能体编号，斜体，英文小写 |
|  | 智能体动作，斜体，英文小写 | m为航班编号，若选择分配，为1，否则为0 |
|  | 智能体策略网络模型，斜体 | 为智能体策略网络中需要训练的权重参数，斜体 |
|  | 混合网络模型，斜体，英文小写 | 表示权重参数，斜体 |