doi: 10.7690/bgzdh.2023.06.008

基于 Gym 与 Flight Gear 的 AI 模拟飞行训练平台搭建

刘剑超[1]，董　斐[1]，林亚军[1]，俞艺涵[1]，姚　杰[2]
（1. 中国人民解放军 91475 部队，辽宁 葫芦岛 125000；2. 中国人民解放军 92543 部队，山西 长治 046000）

摘要：针对 AI 模拟飞行研究，提出应用平台 Gym 与 Flightgear 模拟飞行软件相结合，构建 AI 模拟飞行训练平台。通过平台优化，可不断增加飞行动作的难度系数，重新设计奖励机制与神经网络，实现由 AI 操控模拟飞行软件向 AI 反馈训练数据的交互闭环。训练结果验证了该训练平台的有效性。

关键词：模拟飞行；强化学习；Flight Gear；Gym 平台
中图分类号：TJ85　　**文献标志码：**A

Building of AI Simulation Flight Training Platform Based on Gym and Flight Gear

Liu Jianchao[1], Dong Fei[1], Lin Yajun[1], Yu Yihan[1], Yao Jie[2]
(1. *No*. 91475 *Unit of PLA*, *Huludao* 125000, *China*; 2. *No*. 92543 *Unit of PLA*, *Changzhi* 046000, *China*)

Abstract: Aiming at the high research significance of AI simulation flight research, it is pointed out that the application platform Gym is combined with Flightgear flight simulation software to construct an AI simulation flight training platform. Through platform optimization, the difficulty coefficient of flight action can be continuously increased, and the reward mechanism and neural network can be designed to realize the interactive closed-loop feedback of training data from AI control simulation flight software to AI. The training results verify the effectiveness of the training platform.

Keywords: flight simulation; reinforcement learning; Flight Gear; Gym platform

0　引言

模拟飞行训练具有数据素材量大、评分标准明确的特点，与强化学习向来具有较高的契合性[1]。目前基于强化学习的 ALPHA 系列模拟飞行智能体不断在新闻中出现，具有较高的研究意义。

Gym 强化学习研究平台由 OpenAI 公司研发，为目前较为通用的 AI 研究平台。平台基于 Python 语言搭建，为降低强化学习的学习门槛，平台中对强化学习中较为繁琐的算法部分进行了高度包装，显著降低了编程难度；但由于系统兼容性问题[2]，Gym 官方只提供部分小游戏的 AI 训练支持，如过山车、平衡杆等，为适应所需的仿真软件开展强化学习研究，必须构建自己的训练平台。

Flightgear(FG)为目前认可度较高的模拟飞行软件，软件开源程度高，便于获取飞行数据并下达飞行指令。如果能够将强化学习应用到该游戏中去，则会对 AI 飞行员研究与辅助决策飞行领域产生较高的指导意义[3]。

笔者将这 2 个平台相结合，构建一个 AI 模拟飞行训练平台，通过 AI 操控仿真飞机飞行，并根据要求完成简单的飞行动作。

1　FG 结构分析

笔者采用 2020.3.11 版本 FG 进行仿真。FG 软件总体结构较为复杂，笔者只针对初始化和主函数中需要改动的部分进行分析，软件运行逻辑如图 1 所示。

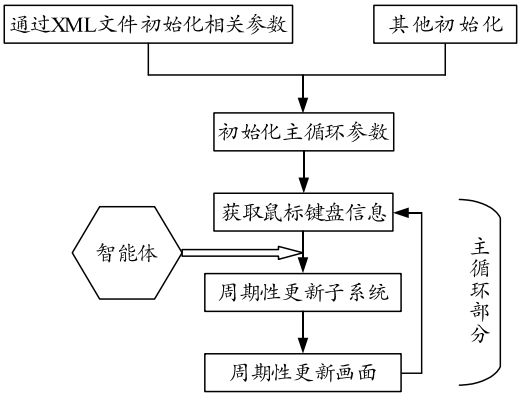

图 1　FG 架构分析

首先飞行初始化参数设置的核心思路为降低训练成本，缩短训练周期长度，如飞行起点与初始姿等初始参数不变、起飞时飞机状况良好、不考虑风

收稿日期：2023-02-13；修回日期：2023-03-05
作者简介：刘剑超(1982—)，男，河北人，硕士，高级工程师，从事模拟飞行与软件工程研究。E-mail: dfgyt@163.com。

向或飞机故障等异常情况。通过合理设置飞行初始化参数可提高 AI 训练的稳定性，避免出现长时间奖励值不能收敛提高的现象。相关初始化参数均保存在 FG 的 data 文件夹的 xml 文件中[4]，可根据需求修改。

AI 的操控指令建议在获取按键信息代码 fireValueChanged()与子系统更新代码 updata(time)之间加入。为避免 AI 操作过于频繁，可以根据计算机运算能力对操作频率加以限制，如最多每 0.2 s 操作一次。软件的操控指令分为 2 类，而由于通过 UDP 传来的信息只能是 char 类型信息，因此需要根据指令类型不同分别做信息处理机制。一类是飞机操作指令，即对飞机操纵杆和脚舵的操控，该系列指令可以直接从 UDP 传输数据的固定位置获取，并通过 FG 自带的赋值函数(global->set_value())实现，这样可以避免采用键盘响应机制，保证了在 2 次操作之间不会产生空档期；另一类是游戏指令，包括游戏开始、重启等指令，该指令需要通过 if 等判断语言响应。

FG 游戏画面更新通过 get_render()系列函数实现，在训练期间可将相关画面更新函数注释掉，以降低显卡负担，提升训练效率。

2　训练平台搭建

强化学习思路的重点在于“交互”，即智能体不断通过动作 action，与环境 environment 交互，以得到环境对于动作的奖赏评价 reward，并通过算法分析奖赏与状态 state 的映射关系优化神经网络。由于 2 个交互平台所采取的语言不同，并且对于显卡资源的占用都很高，因此采用 2 台计算机同时运行的方式进行训练。计算机之间将通过 UDP 协议传输数据。平台工作流程如图 2 所示。

优化神经网络的过程本需要较高的数学基础，但 Gym 平台将强化学习所需的大量张量算法进行了较好的封装，使用户不太需要了解算法背后的复杂数学逻辑，只需要少量的 Python 代码便可完成智能体训练策略的设计工作[5]。其中大部分代码不需要更改，如神经网络类函数 layers、滤波函数、缓冲池函数、梯度函数 gradients 都可直接使用。需改动的部分主要存在于 AI 对于游戏的控制与游戏反馈部分。在 Gym 平台中，打开游戏将作为 env 类的对象进行处理，后续的所有关于游戏的操作利用 env 的成员函数实现，例如游戏重启 reset、画面更新 render、下达游戏指令并获取返回数据 step，由于游戏运行不在该电脑内，所以所有游戏相关代码均需按照代码更换为 UDP 函数(socket、sendto、recvfrom)，根据实际使用情况建立创建自己的套接字，套接字解码函数，最终实现两平台一闭环。

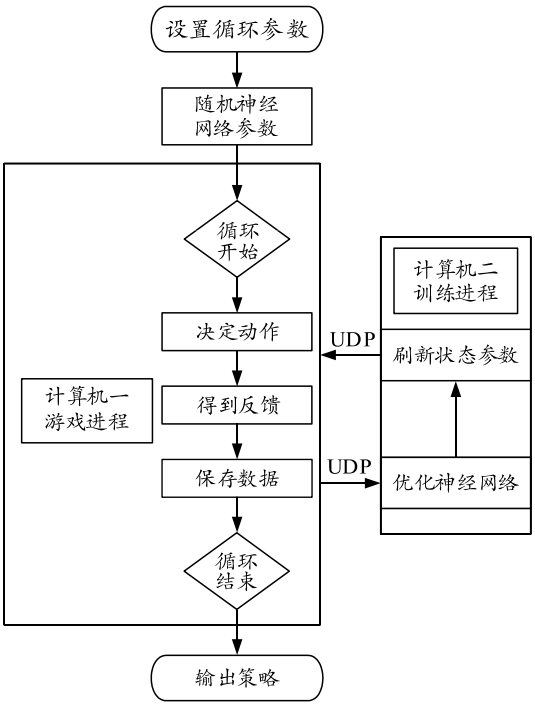

图 2　训练平台运行流程

3　训练仿真实例

本次仿真目标为验证训练平台有效性，因此会尽可能降低本次训练的计算量，快速达到 AI 操控飞行的效果。本次训练目标为将飞行操纵杆的左右运动权限交给 AI 操控，要求 AI 通过控制操纵杆保持飞机无翻滚平飞，共训练 5 000 回合，每回合最长时间为 20 s，每 100 ms 训练平台与 FG 交互一次，也就是每回合交互 200 次，每次交互的奖励值域为[0, 1]，当飞机“翻车”时，也就是翻滚角绝对值接近 180°时游戏提前结束，后续奖励值全部计算为 0。预计当飞行时间可以稳定达到 20 s，总奖励值在 170 分以上且分数不再有较大起伏时，便应认定为训练成功。双平台重要参数设置思路如下：

1) FG 中参数设置思路为尽可能采用默认参数，飞机机型、机场、环境等无关参数皆不做改动。飞行操纵杆的俯仰控制 elevator 锁死为 0(同样利用 set_value 系列函数)，脚舵控制 rudder 锁死为 0，留下操纵杆俯仰控制 aileron 交给 AI 控制。游戏中自带飞行教程功能，该功能可以保证每次启动飞行教程时飞机的初始参数不变。游戏正式启动后，在主循环内计算每次交互后得出的奖励值、飞行状态参

数、游戏状态参数并打包发送训练平台。

2) Gym 中的训练算法设置大致与 DQN 算法训练平衡杆游戏的思路类似，训练神经网络根据自己的状态参数矩阵形状做简单更改即可。若游戏刚开始后，AI 长时间没有动作就会导致最开始的奖励分数过高，为了让 AI 操作更活跃，学习过程与奖励分数提高过程更明显，需要在前期给予一个较大的贪心参数 ε，鼓励 AI 做更多的随机选取动作。值得注意的是FG经过多次重启后可能出现卡死的现象，因此 Gym 需要设置判断函数，如果 10 s 以上没有收到 FG 数据，就停止训练，及时保存数据。

本次训练的奖励值曲线如图 3 所示，训练经过约 2 000 回合后，曲线基本收敛，飞行状态基本稳定。当人类玩家在 FG 中需要完成平飞动作时，会选择先不移动操纵杆，当有明显翻滚幅度时则小幅调整操纵杆，直到恢复平飞。经过训练后的 AI 对于控制翻滚角的思路为一直高频率小幅度反复左右摆动控制杆，使得飞机始终不会出现明显翻滚的现象。虽然这个策略对于飞行的现实意义有待商榷，但是 AI 确实在毫无飞行经验、毫无参数指引的前提下，仅通过游戏本体与奖励参数指导下实现了成长，“学会”了飞机的平飞动作，并在现有游戏规则下，超过了作者飞行时的奖励值，验证了在该强化学习平台中可以进行 AI 的研发工作。

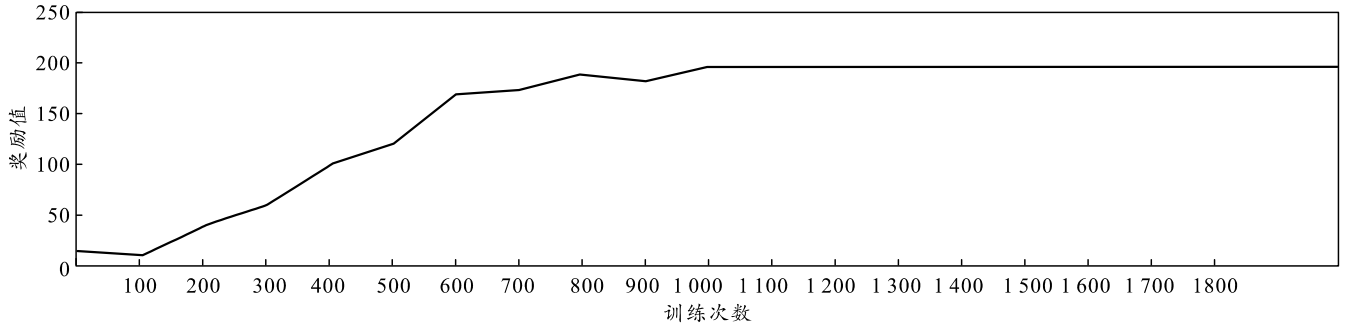

图 3 奖励值曲线分布

4 结论

经验证，该强化学习训练平台可以完成 AI 训练。为进一步提高 AI 训练质量，帮助 AI 完成更具有现实意义的飞行动作需要大幅优化技术细节。如将 FG 代码进行进一步分析与精简，在游戏内存较低的基础上增加多 AI 飞行功能，通过遗传机制保留高分 AI 的决策参数；优化游戏时间系统，增加“变速齿轮”功能，使游戏可以在 5 倍速甚至更快速度下进行等。在平台优化的基础上，可以不断增加飞行动作的难度系数，重新设计奖励机制与神经网络，提高 AI 飞行的现实指导意义。

参考文献：

[1] 王浩楠, 刘苧, 章艺云. 深度强化学习综述(英文)[J]. Frontiers of Information Technology & Electronic Engineering, 2020, 21(12): 63-82.

[2] BROCKMAN G, CHEUNG V, PETTERSSON L, et al. OpenAI Gym[J]. arXiv, 2016: 1606. 01540.

[3] TERRY J K, BLACK B, HARI A, et al. PettingZoo: Gym for Multi-Agent Reinforcement Learning[J]. arXiv, 2020: 2009. 14471.

[4] Perry A R. The FlightGear Flight Simulator[C]//USENIX Annual Techincal Conference. USENIX Association, 2004.

[5] 晋帅, 李煊鹏, 何嘉颖, 等. 基于强化学习的两轮模型车控制仿真分析[J]. 测控技术, 2019, 38(12): 7.

****

（上接第 15 页）

[16] DENG H, SUN X, LIU M, et al. Small infrared target detection based on weighted local difference measure[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(7): 4204-4214.

[17] OTSU N. A Threshold Selection Method from Gray-Level Histograms[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.

[18] 回丙伟, 宋志勇, 王琦, 等. 空中弱小目标检测跟踪测试基准[J]. 航空兵器, 2019, 26(6): 56-59.

[19] BORJI A, CHENG M M, JIANG H, et al. Salient Object Detection: A Benchmark[J]. IEEE Transactions on Image Processing, 2012, 24(12): 5706-5722.