1. 引入

机器人学和强化学习是水乳交融的关系——强化学习可以帮助机器人做出复杂的决策，而机器人能验证强化学习在物理世界中的表现。强化学习在机器人学中的应用包括自动飞行器、机械臂、自动驾驶汽车和人形机器人等。但是，和在游戏世界中不同，机器人学的问题往往是高维、连续的，而且不能假设世界是完全可以被观察的、观察是没有噪声的，同时现实世界中的实验往往很难复现，例如对于打网球机器人，想要让机器人在完全相同的位置，以完全相同的力度从完全相同的角度击出，是做不到的，并且我们不希望机器人采取某些行为，比如会导致其自身毁坏的行为。

1. 案例分析

自动飞行直升机是一个具有挑战性的控制问题，同时由于直升机在低速时不稳定，为直升机在低速状态下设计控制引擎更是困难，而在这个应用实例中，作者成功地设计出可以使得直升机自动倒飞的控制引擎。

作者采用Bergen双旋翼式直升机，这种直升机长约150cm，高约55cm。由于智能体操纵直升机，很容易将直升机损坏，而直升机价格昂贵，所以让智能体直接操纵直升机的方案不可行。作者的方案是：先用有监督学习学出一个可以模拟直升机飞行的模型，进而建立模拟器，从而在上面使用强化学习训练智能体。

直升机的状态可以用一个八维向量进行表示：其中两维表示直升机中线的方向，三维表示直升机的速度，最后三维表示直升机中线的角速度。而直升机飞行的标注数据由人类操纵直升机获得，这样就可以通过有监督学习得到，即在真实环境中的状态转移，从而建立一个模拟器。

在强化学习问题中，奖励函数设置为一个二次惩罚项，反映直升机与目标位置的距离直升机中线角度与目标中线角度的距离，使用神经网络来表示策略，用蒙特卡洛方法进行探索。

这样，作者就成功地设计出可以使得直升机自动倒飞的控制引擎。

1. 将强化学习应用到机器人领域存在的困难

首先是维数灾难。根据Bellman提出的“维数的诅咒”，随着维数的增加，需要收集的数据和所需的计算时间呈指数增长。而由于机器人动作的灵活性，往往有很高的维数。

再者，收集真实世界的数据要耗费更多的劳动力、时间、财力等。如果要像案例一样建立一个模拟环境，而建立模拟环境本身又是一件耗时耗力的事情。

并且，建立合适的目标也存在困难。由于强化学习很容易利用规则的漏洞，如何建立一个好的奖励函数是一个问题。

1. 参考文献

[1] Kober, J., Bagnell, J.A. and Peters, J. (2013). Reinforcement learning in robotics: A survey. The International Journal of Robotics Research, 32(11), pp.1238–1274. doi:10.1177/0278364913495721.

[2] Ng, A.Y., Coates, A., Diel, M., Ganapathi, V., Schulte, J., Tse, B., Berger, E. and Liang, E. (2006). Autonomous Inverted Helicopter Flight via Reinforcement Learning. Springer Tracts in Advanced Robotics, pp.363–372. doi:10.1007/11552246\_35.

互评人：1951106 孙艺珈

评分：8/8