**问题0**：直接依据reward学习效果十分不明显，reward震荡不收敛

解决方案：采用已有的fix\_rule agent 在对战中的回放（experience learning）作为前期学习依据，具体而言需要搜集对战经验，要求每一个状态下fix rule agent所采取的动作以及采取动作的概率分布（难点在于该动作的概率计算，拟从大量采样中计算该分布（已完成））

———————————————————————————————————————

**学习中的问题：**

**问题1**：1V1和2V2的fix\_rule agent对战中，二者直接对冲而死，大概率平手（其余情况也是1s定胜负直接冲撞），不能用作学习数据

解决方案1：无论1v1,2v2，10v10，目前所做的都是单智能体强化学习。采集经验数据，越多的智能体越能探索空间，采样效率越高。采用10v10的对战数据进行训练，相当于10个agent在空间中同时采样，采样效率更高（已完成）

#解决方案2：agent采用随机的初始位置，探索的状态动作更多（出生位置random参数指左边还是右边）。

**问题2**：fix\_rule agent 采用的观测是raw\_obs，而目前没有针对raw\_obs进行建模(our\_obs可以根据raw\_obs设计，策略网络需要针对our\_obs来设计)

解决方案1：采用别人设计好的simple\_obs转换fix的raw\_obs,智能agent在simple\_obs基础上构建策略网络学习网络参数，使得our\_agent和fix\_rule agent使用同一套state，action（已完成）

#解决方案2：对raw\_obs分析，设计our\_obs，构建基于raw\_obs或our\_obs的决策网络

**问题3**：raw\_obs中course为360°即360个角度动作，加上长短导弹共计目标，动作空间过大。

解决方案1：每15°为一个动作，划分方向为24方位，fix\_rule agent中的动作转换成新设计动作（已完成）

#解决方案2：动作分层选择

问题4：fix rule 的经验需要大量样本，转换为simple\_obs（图像矩阵），数据进一步扩大10倍以上，内存或显存均难以负载

解决方案1：分文件存储fix rule的经验数据，agent分文件读入（内存足够），读入的每一个文件分batch读入显存（已完成）。

---------------------------------------------------------------------------

问题5：依据fix\_rule agent学习的小飞机our agent在四周空白的状态时乱飞

解决方案1：乱飞的根本原因在于fix—rule agent本身的策略在四周空白状态就是不定的大量采样下趋于平均即在一个空白状态下向任意方向飞行是均等概率的，因而我们学习的小飞机最终学到的方向动作概率是均等的。简单的解决方案，应该在状态信息中加入上一动作的信息，使得agengt决策当前方向时考虑上一方向，这样可以学习平滑的飞行轨迹，然而本质上依旧是无目的飞行(待实现)。

解决方案2：引入特定的搜索策略，例如索敌加成reward等。或者引入侦察机视野，使得战斗机的移动方位与目标有关，而不是单纯的空白视野。

问题6：多GPU和多CPU使用，实验中仅使用一个cpu和一个gpu

解决方案:（待考虑）

-------------------------------------------------------------------------------