离线到在线的上下文对手建模

我们的方法就是提出缓冲区的概念，先利用离线的PEL和ICD然后应对未知的对手策略，并利用得到的轨迹来更新PEL，再利用自适应的g online 来探索得到新的轨迹，并更新我们的ICD网络，在测试的时候，用每个ICD 先跑几轮，然后利用得到的return值自适应的选择合适的ICD网络。

离线到在线的对手建模

由于直接对离线数据集中训练得到的PEL,ICD微调会导致对已知的对手策略识别变得不好，或者不能更好的应对对手策略，本文提出一个应对对手扩展策略集的概念，当在线微调时，遇到未见过的策略，单独微调一个PEL,ICD 来构成策略集中的其中一个策略。当进行对未知策略进行微调时，以adaptive RTG来进行和环境交互，并将得到的优质轨迹加入到所构造的优质轨迹缓冲区中，以优质缓冲区的数据对其进行微调。当在测试时候加入了evaluate阶段，将策略集中的所有策略都与对手策略一起加载到环境中，快速的选择策略集中的针对这个对手策略的更好策略。