（1）引言和相关工作

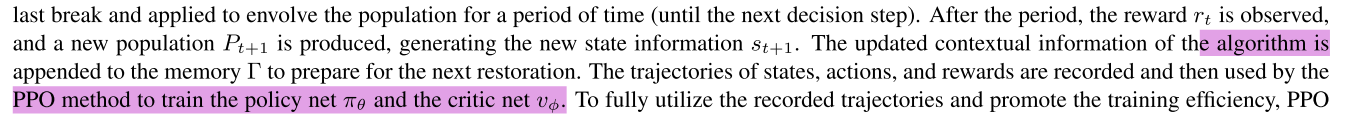
怎么用DRL的方法求解MDP过程的？公式没太懂。

（2）论文方法

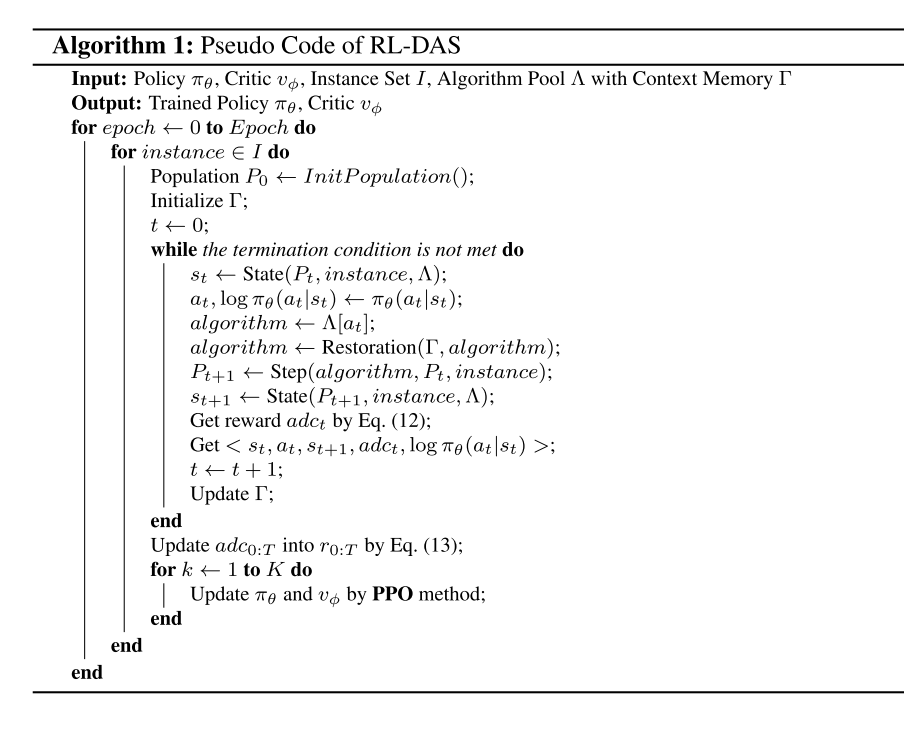
DAS的模型公式不太清楚，这个公式是建立了一个有M个问题实例，每个问题实例有T个时间状态的MDP？那论文解决的不是针对一个问题的优化过程？（我一直想的是针对一个问题，把这一个问题分成T各时间段代表T个状态，由每个状态产生动作（算法）这样动态解决一个问题）。



（3）3.2 RS-DAS Overview 里“种群”的概念不太清楚。看描述是，在t0开始阶段初始化种群信息，pt时间段的种群信息产生st的状态。状态（种群+优化问题+算法池），t时刻状态通过Π策略产生动作at（从算法池选择算法），reward Rt被计算，Pt+1被生产，以此循环。Pt怎么被更新的？PPO方法怎么用的？

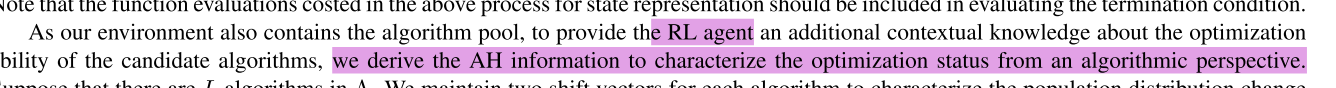


每个状态的算法池的算法信息一样吗？会变化吗？ 这俩是什么关系怎么用？



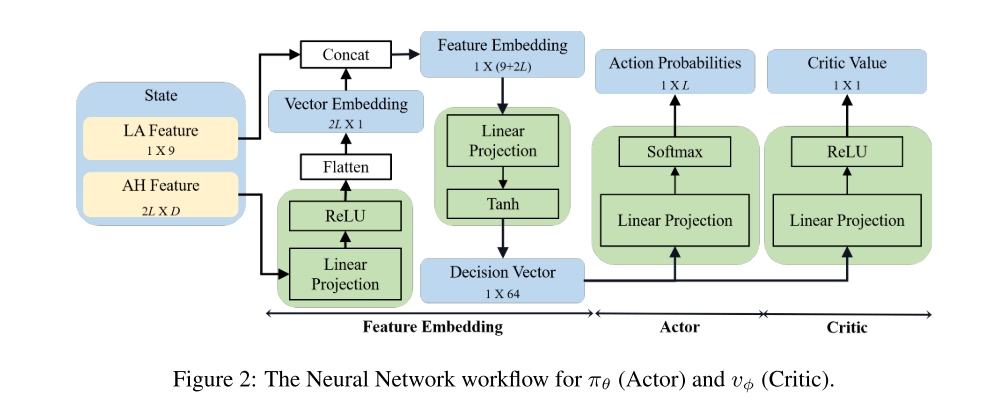
这一个表格的算法过程

AH信息（表征优化状态）？ Shift？





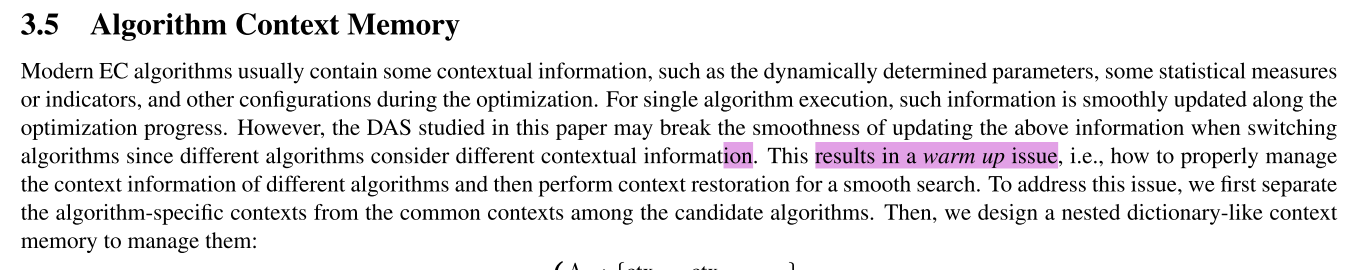
中的D是什么？





Critic是什么？

算法上下文处理



RL-agent强化学习代理？