

基于专家示教聚类经验池的高效深度强化学习

王士珉¹, 赵彬琦¹, 张政锋¹, 张军平¹, 浦剑²

¹复旦大学计算机科学技术学院上海市智能信息处理重点实验室, 中国上海市, 200433

²复旦大学类脑智能科学与技术研究院, 中国上海市, 200433

摘要: 作为强化学习领域最基本的主题之一, 样本效率对于深度强化学习算法的部署至关重要。与现有大多数从不同类型的后验分布中对动作进行采样的探索方法不同, 我们专注于策略的采样过程, 提出一种有效的选择性采样方法, 通过对环境的内部层次结构建模来提高样本效率。具体来说, 首先在策略采样过程中使用聚类方法生成动作候选集, 随后引入一个用于对内部层次结构建模的聚类缓冲区, 它由同轨数据、异轨数据以及专家数据组成, 用于评估探索阶段动作候选集中不同类别动作的价值。通过这种方式, 我们的方法能够更多地利用专家示教数据中的监督信息。在6种不同的连续运动环境中进行了实验, 结果表明选择性采样方法具有卓越的强化学习性能和更快的收敛速度。特别地, 在LGSVL任务中, 该方法可以减少46.7%的收敛步数和28.5%的收敛时间。代码已开源, 见<https://github.com/Shihwin/SelectiveSampling>。

关键词: 强化学习; 采样效率; 采样过程; 聚类方法; 自动驾驶

<https://doi.org/10.1631/FITEE.2300084>