马里奥DQN智能体的特征工程探究

第14组 李瑞堃[1], 黎睿曦[1]

[1]清华大学深圳国际研究生院

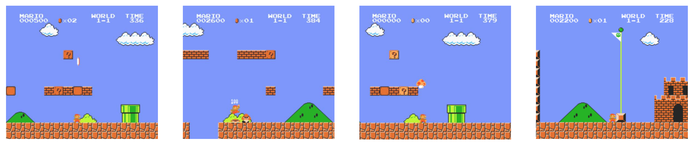
Environment

超级马里奥

➢游戏环境API

- ✓ 状态：画面帧（3×240×256）
- ✓ 动作：2-7-12，三种集合
- ✓ 奖励：前进×1 - 时耗×1
- ✓ 信息：金币、Score、位置、状态、剩余时间 等

玩法机制



收集金币　碾压敌人　采集道具　抵达终点

Baseline

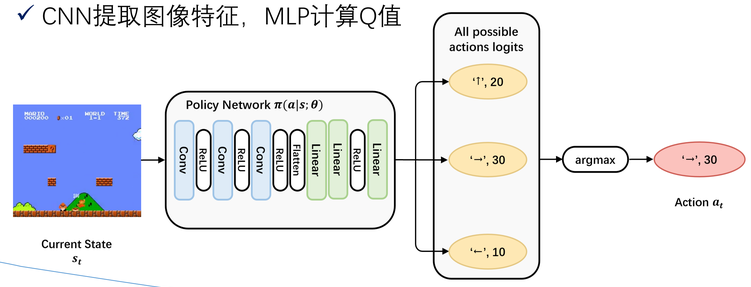
算法

➢ 默认DQN

- ✓ DQN[1]：使用神经网络来估计Q值（动作值函数），解决了深度强化学习中的Q值精确估计问题
- ✓ Double[2]：用两个独立的DQN网络来解决由于估计Q值而产生的估计误差问题，提高稳定性和性能
- ✓ Prioritized Replay[3]：根据样本的重要程度赋予不同的权重，优先采样重要的样本，提高数据利用效率和智能体的学习效率

网络架构

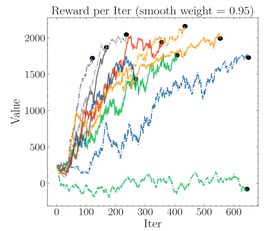
- ✓ CNN提取图像特征，MLP计算Q值



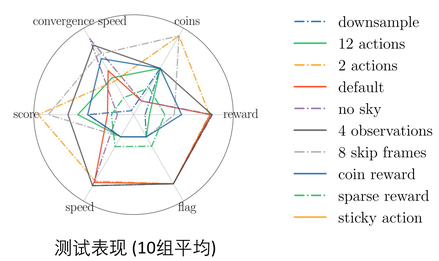
Feature Engineering

基础特征工程

- ✓ 动作, 叠帧, 跳帧, 降采样, 去除细节, 金币, 稀疏奖励, 粘性动作
- ✓ 默认超参数, 随机种子612, 训练对比
- ✓ 测试指标：平均奖励, 通关时间, 金币数, 得分, 通过率, 收敛速度



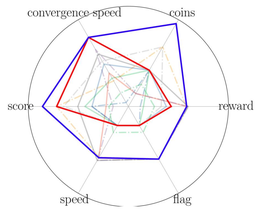
训练曲线



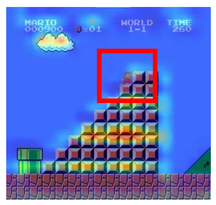
测试表现 (10组平均)

最优组合

- ✓ Action=2版本收敛慢，性能六边形战士
- ✓ 叠帧=4版本收敛快，表现也不错



无法通关



CAM激活图，注意力分散

去除无关细节

Algorithm Advances

改进DQN

- ✓ Dueling DQN[4]：将Q值函数分解为状态值函数和优势函数，提高了Q值的估计精度，从而提高了强化学习算法的性能
- ✓ Noisy DQN[5]：向网络参数添加随机噪声来增加网络探索性
- ✓ Rainbow DQN[6]：聚合了Double, Dueling, Prioritized Replay 和 Noisy 的 Rainbow DQN 青春版

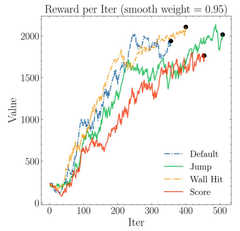
算法对比	1-1 关					1-2 关				
	收敛轮次	奖励	通关时间	金币	得分	收敛轮次	奖励	通关时间	金币	得分
Baseline DQN	355	3030	91	0	200	567	2837	57	2	1100
+Dueling DQN	276	3059	62	0	100	539	2836	58	3	1100
+Noisy DQN	1037	1883	–	1	500	1034	2450	–	1	2400
Rainbow DQN -	1017	3065	56	0	400	997	2831	55	2	2500

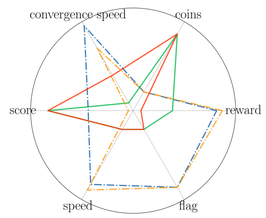
Reward Shaping

鼓励跳跃 + 得分

- ✓ 撞墙惩罚
- ✓ 跳跃奖励
- ✓ 得分奖励

行为	撞墙	跳跃	踩敌人	吃金币	吃蘑菇	踩龟壳
Reward	-1	+1	+3	+5	+10	+5





- ✓ 得分增加收敛难度
- ✓ 罚撞墙加速了收敛

Conclusion

特征-算法-奖励

➢ 特征工程

- ✓ 减少动机集中注意、叠帧引入速度信息能够提升智能体性能
- ✓ 叠帧、去除天空信息有利于提升收敛速度

➢ 算法改进

- ✓ Dueling DQN 收敛速度更快，且提升了智能体的通关速度
- ✓ Noisy DQN 效果差，智能体无法在有限时间内通关
- ✓ Rainbow DQN(青春版) 收敛慢，测试表现最好

➢ 奖励设计

- ✓ 撞墙惩罚有利于提高收敛速度，且不影响测试表现
- ✓ 得分奖励大幅提升了测试表现，但是通关能力被削弱

最终效果

1-1 关					1-2 关				
收敛轮次	奖励	通关时间	金币	得分	收敛轮次	奖励	通关时间	金币	得分
924	3071	58	2	600	867	2911	62	3	2700

Reference

[1] Mnih, Volodymyr, et al. "Human-level control through deep reinforcement learning." nature 518.7540 (2015): 529-533.
[2] Van Hasselt, Hado, Arthur Guez, and David Silver. "Deep reinforcement learning with double q-learning." Proceedings of the AAAI conference on artificial intelligence. Vol. 30. No. 1. 2016.
[3] Schaul, Tom, et al. "Prioritized experience replay." arXiv preprint arXiv:1511.05952 (2015).
[4] Wang, Ziyu, et al. "Dueling network architectures for deep reinforcement learning." International conference on machine learning. PMLR, 2016.
[5] Fortunato, Meire, et al. "Noisy networks for exploration." arXiv preprint arXiv:1706.10295 (2017).
[6] Hessel, Matteo, et al. "Rainbow: Combining improvements in deep reinforcement learning." Proceedings of the AAAI conference on artificial intelligence. Vol. 32. No. 1. 2018.

作者信息

李瑞堃
专业：电子与通信工程，深无硕232班
地址：信息楼 21-216，清华大学，SIGS，518000，中国

黎睿曦
专业：互联网+创新设计，深中法硕232班
地址：能源楼 8层，清华大学，SIGS，518000，中国