Playing Overcooked-AI with deep reinforcement learning

Final project presentation

Leonardo Monti

Implementation

- Training algorithm: multi agent ppo, with clip
- Architecture: centralized critic and decentralized policies
- CTDE framework
- Policies share parameters: agents share the same policy
- Shallow dense neural networks
- 1000 episodes per training, for the baseline
- 10 episodes in the buffer for training

Results

Single layout per agent

LAYOUTS	SCORE	STD
cramped_room	127	17.06
asymmetric_advantages	165	21.79
coordination_ring	62	32.80
forced_coordination	72	18.33
counter_circuit	0	0.00

Multiple layout: curriculum learning

LAYOUT	EPISODES
cramped_room	333
asymmetric_advantages	250
cramped_room	125
asymmetric_advantages	250
cramped_room & asymmetric_advantages	250

LAYOUT	SCORE	STD
cramped_room	127	14.52
asymmetric_advantages	119	31.28

Conclusions and suggestions

- The implementation met the requirements
- The policies are not general
- Cooperation is hard to achieve
- Maintaining exploration is needed to find better solutions
- A CNN would have capacity to encode for more complex policies