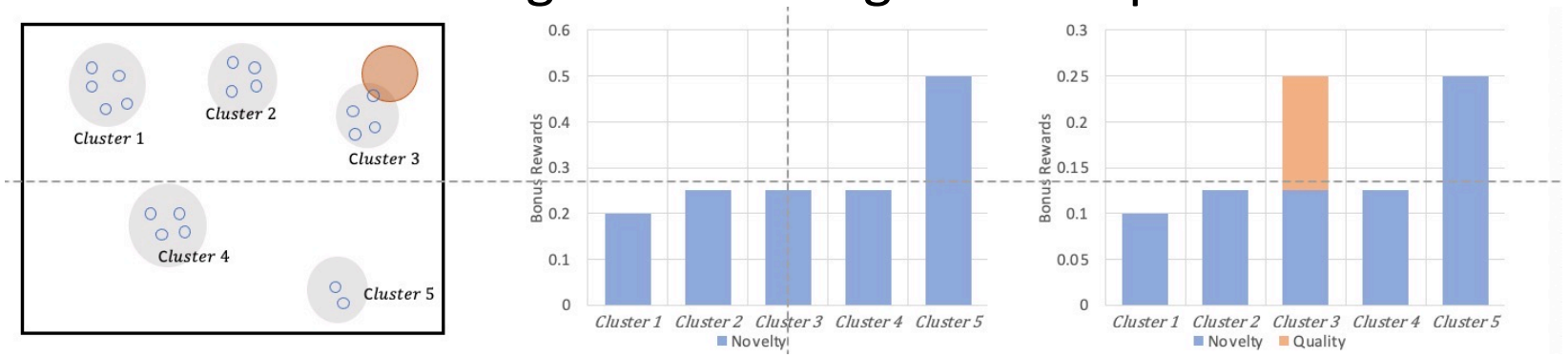
Clustered Reinforcement Learning

Xiao MA, Shen-Yi ZHAO, Zhao-Heng YIN, Wu-Jun LI

Frontiers of Computer Science, DOI: 10.1007/s11704-024-3194-1

Problems & Ideas

- Problems of exploration in reinforcement learning (RL):
  - During exploration, the agent tries to discover unexplored (novel) areas or high reward (quality) areas.
  - However, the novelty and quality in the neighboring area of the current state have not been well utilized to guide the agent's exploration simultaneously.
- Ideas: Using clustering to divide the collected states into several clusters, based on which a bonus reward reflecting both novelty and quality in the neighboring area (cluster) of the current state is given to the agent for exploration.



Left: Using clustering to divide the collected states (blue dots) into 5 clusters. The agent is rewarded with 1 in the orange area and receives no reward in other areas. Middle: The clustering-based bonus rewards with novelty alone ($\eta = 1.0$). Right: The clustering-based bonus rewards ($\eta = 0.5$). The blue bar represents the portion of bonus rewards reflecting the novelty of states, and the orange bar represents the portion reflecting the quality of states.

Main Contributions

- Contributions:
  - We propose a novel RL framework, called clustered reinforcement learning (CRL), for efficient exploration in RL.
  - CRL adopts clustering to divide the collected states into several clusters and uses a novel bonus reward that reflects both novelty and quality in the neighboring area of the current state to guide the agent to perform efficient exploration.
  - CRL can be combined with existing exploration strategies that use the novelty of states as bonus rewards to guide the agent's exploration.
  - Experimental results on multiple environments demonstrate that our method can outperform other state-of-the-art methods to achieve the best performance in most cases.

Method	Freeway	Frostbite	Gravitar	Montezuma	Solaris	Venture
TRPO [58]	17.55	1229.66	500.33	0	2110.22	283.48
CRL	**30.80** (0.75)	**4337.98** (0.1)	552.46 (0.1)	0 (0.75)	**3672.55** (0.5)	312.40 (0.1)
Hash [24]	22.29	2954.10	577.47	0	2619.32	299.61
CRL-Hash	28.38 (0.75)	4148.90 (0.1)	585.79 (0.1)	0.0 (0.75)	2741.48 (0.5)	328.50 (0.1)
RND [41]	21.52	2837.70	867.30	2188.8	765.47	966.00
CRL-RND	20.85 (0.9)	4076.60 (0.9)	**1002.40** (0.75)	**2453.30** (0.5)	1021.60 (0.5)	**981.20** (0.9)
NovelD [42]	21.39	3476.46	677.9	1744.8	975.52	283.60
CRL-NovelD	19.97 (0.9)	3520.06 (0.9)	971.5 (0.5)	2323.4 (0.5)	980.16 (0.5)	498.6 (0.9)

The mean average return of CRL and baselines on six Atari-2600 games over 5 random seeds. For CRL and its variants, the numbers in parentheses indicate the values of η . The Boldface numbers are the best results among all methods.