NATIONAL TAIWAN UNIVERSITY
COMPUTER SCIENCE & INFORMATION ENGINEERING
Computational Learning Lab

Active Reinforcement Learning from Demonstration in Continuous Action Spaces

Ming-Hsin Chen, Si-An Chen, Hsuan-Tien Lin

Active Reinforcement Learning from Demonstration (ARLD)

- Learning from Demonstration (LfD) is a popular approach to improve the efficiency and safety of RL by collecting expert's demonstration.
- To reduce the cost of collecting demonstration, ARLD [1] propose a human-in-the-loop paradigm where agents can actively query for critical demonstration, as shown in Figure 1.
- Existing ARLD algorithms focus on discrete action space.
- **Goal: Explore the solutions of ARLD in continuous space.**

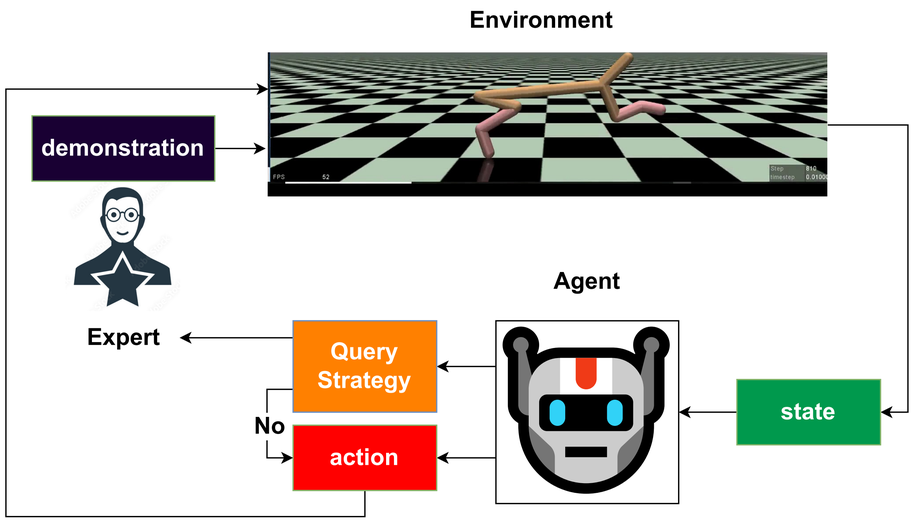

Figure 1: The ARLD process, including three key component, the Agent, the Expert and the Environment.

1 Uncertainty Sampling in ARLD

- Uncertainty Sampling
  - A simple and effective query framework for Active Learning
  - Decide whether to query by estimating the agent's uncertainty based on current states
- Challenge: It is not easy to define the uncertainty estimation for RL in continuous action space, figure 2 shows the difference.

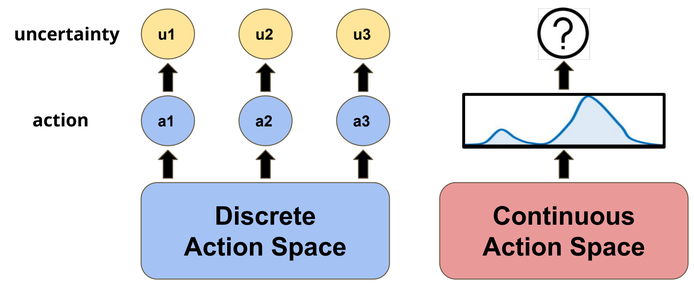

Figure 2: The uncertainty in discrete and continuous action space

2 Uncertainty Estimation for Continuous Actions

- We take Soft Actor-Critic (SAC [2]) as RL agent for ARLD, which is a SOTA algorithm for complex, high-dimensional tasks.
- ARLD in discrete action spaces builds upon Noisy DQN, a noisy variant of DQN that injects parameter noise to alter action decisions, we follow the similar way to derive the uncertainty from NoisySAC [3]

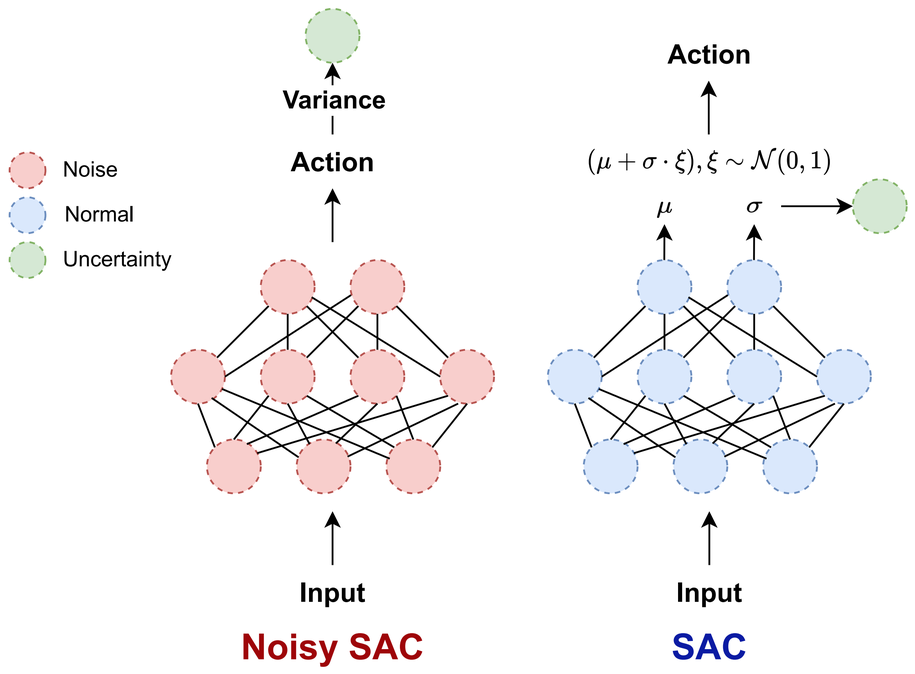

Figure 3: The uncertainty of SAC and Noisy SAC

3 Experimental Results

3.1 Historical Model Uncertainty

- Model Uncertainty = Epistemic Uncertainty
- As training on more data, the uncertainty should decrease

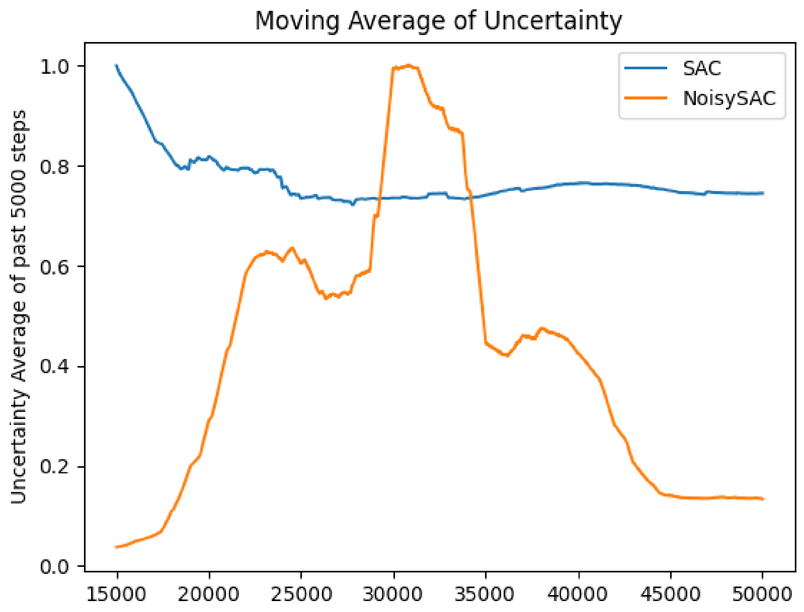

Figure 4: Moving average of the normalized uncertainty of SAC and NoisySAC.

- Intrinsic SAC uncertainty fit the trend of decreasing while Extrinsic Noisy SAC uncertainty not.

3.2 Empirical Results

- We evaluate the two models with different query approaches on simulated environment, HalfCheetah.

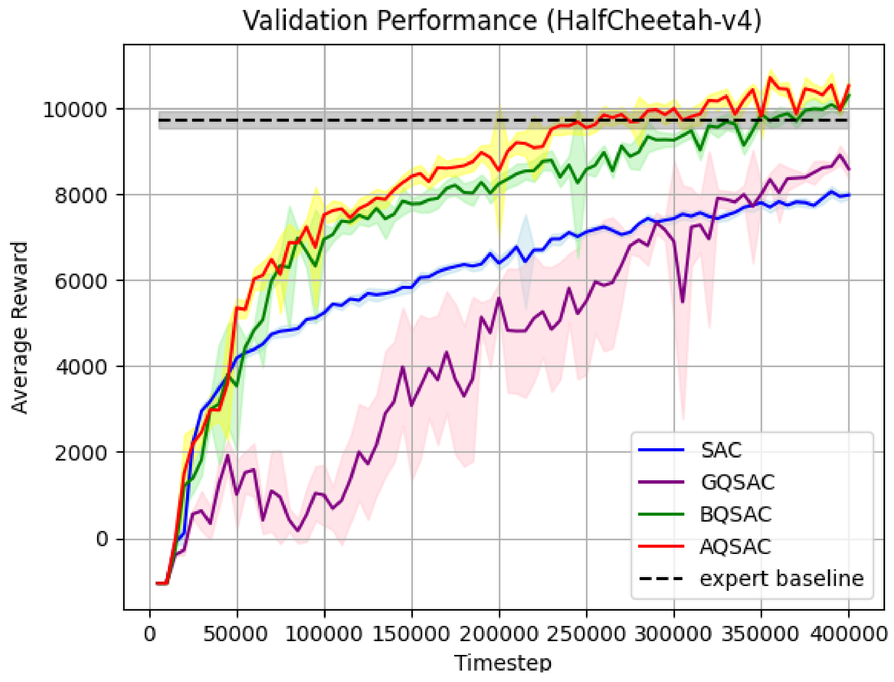

Figure 5: Validation results for SAC.

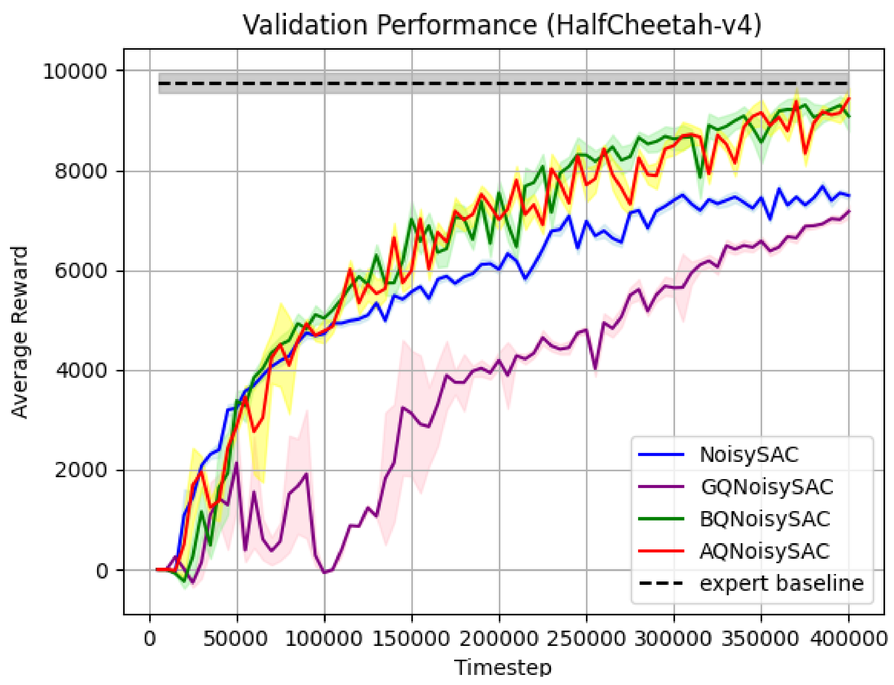

Figure 6: Validation results for NoisySAC.

- Active learning query outperform other heuristic approaches in SAC, while not in Noisy SAC

4 Conclusions || Discussion

- We propose a first attempt to extending ARLD to continuous action space.
- The work could be extended to more environments and more realistic tasks.
- How to better explain that the query demonstration can really help improving the training is one of the future directions.

References

[1] Si-An Chen, Voot Tangkaratt, Hsuan-Tien Lin, and Masashi Sugiyama. Active deep q-learning with demonstration. *Machine Learning*, 109:1699–1725, 2020.

[2] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR, 2018.

[3] Matthias Plappert, Rein Houthooft, Prafulla Dhariwal, Szymon Sidor, Richard Y Chen, Xi Chen, Tamim Asfour, Pieter Abbeel, and Marcin Andrychowicz. Parameter space noise for exploration. *arXiv preprint arXiv:1706.01905*, 2017.