**УДК 004.8**

# Solving Continuous Control with Episodic Memory

**Кузнецов И.С.** (Университет ИТМО)
**Научный руководитель – к.ф.-м.н. Фильченков А.А.**
(Университет ИТМО)

Previous works on memory mechanisms show benefits of using episodic-based data structures for discrete action problems in terms of sample-efficiency. The application of episodic memory for continuous control with a large action space is not trivial. Our study aims to answer the question: can episodic memory be used to improve agent's performance in continuous control?

The general idea of episodic memory in reinforcement learning setting is to leverage long-term memory reflecting data structure containing information of the past episodes to improve agent performance. The existing works [Blundell et al.,2016; Pritzel et al., 2017], show that episodic control (EC) can be beneficial for decision making process. In the context of discrete action problems, episodic memory stores information about states and the corresponding returns in a table-like data structure. With only several actions the proposed methods store past experiences in multiple memory buffers per each action. During the action selection process the estimate of taking each action is reconsidered with respect to the stored memories and may be corrected taking into the account past experience. The motivation of using episodic-like structures is to latch quickly to the rare but promising experiences that slow gradient-based models cannot reflect from a small number of samples. The notion of using episodic memory in continuous control is not trivial. Since the action space may be high dimensional, the methods that operate discrete action space become not suitable. Another challenge is the complexity of the state space. The study of [Blundell et al.,2016] shows that some discrete action environments (e.g. Atari) may have high ratio of repeating states, which is not the case for complex continuous control environments.

We present Episodic Memory Actor-Critic (EMAC), a deep reinforcement learning algorithm that exploits episodic memory in continuous control problems. EMAC uses non-parametric data structure to store and retrieve experiences from the past. The episodic memory module is involved in the training process via the additional term in the critic's loss. The motivation behind such an approach is that the actor is directly dependent on the critic, therefore improving critic's quality ensures the stronger and more efficient policy. We do not exploit episodic memory during the policy evaluation, which means that memory module is used only within the network update step. The loss modification demonstrates sample-efficiency improvement over the DDPG baseline. We further show that introducing prioritization based on the episodic memories improves our results. Experimental study of Q-value overestimation shows that proposed approach has less tendency in critic overestimation thus providing faster and more stable training

Our experiments show that leveraging episodic memory gave superior results in comparison to the baseline algorithm DDPG and TD3 on all tested environments and also outper- formed SAC on 3 out of 5 environments. We hypothesize that the applicability of the proposed method is dependent from environment complexity. As a result, we struggle to outperform SAC for such a complicated environment as Humanoid which has bigger action space than the other OpenAI gym domains.

Кузнецов И.С. (автор) Подпись

Фильченков А.А. (научный руководитель) Подпись