

0.1. Плаксин А.Р., Миргородский Н.В. Адаптивный шаг дискретизации в алгоритмах обучения с подкреплением при решении задач оптимального управления

В последние годы теория обучения с подкреплением [1] стала основой для развития большого количества алгоритмов, способных решать сложные задачи принятия решений (Шахматы, Го, игры Atari, StarCraft II и т.д.). При этом, один из основных подходов в обучении с подкреплением является Q-learning, в котором оптимальное действие выбирается по Q-функции являющейся решением уравнения Беллмана. В работе [2], был предложен алгоритм Deep Q-Networks (DQN), в котором Q-функция аппроксимируется нейронной сетью, что позволяет использовать подход Q-learning для задач с бесконечным (многомерным) пространством состояний. Отметим, что алгоритм DQN, также как и другие алгоритмы обучения с подкреплением, может быть использован для решения дискретных по времени задач оптимального управления. Если же изначально задача является непрерывной по времени, то ее необходимо дискретизировать. При этом, как было исследовано в [3], работа алгоритма существенно зависит от шага дискретизации и, в основном, становится хуже при уменьшении этого шага.

В данной работе предлагается модификация алгоритма DQN способная преодолевать эти сложности путем нахождения оптимального шага дискретизации. Для этого в алгоритм добавляется еще одна нейронная сеть, которая по текущему состоянию назначает длину следующего шага для управления. Предложенный алгоритм сравнивается с DQN и, с одной стороны, показывает сравнимые с DQN результаты при большом шаге дискретизации, а с другой стороны, при малом шаге является намного более эффективным, чем DQN алгоритм.

Работа выполнена в рамках исследований, проводимых в Уральском математическом центре при финансовой поддержке Министерства науки и высшего образования Российской Федерации (номер соглашения 075-02-2022-874).

Список литературы

- [1] SUTTON R. S., BARTO A. G., Reinforcement Learning An Introduction (Second Edition) / Cambridge, Massachusetts: The MIT Press, 2018. 526 p.
- [2] MNIN V., KAVUKCUOGLU K., SILVER D., AT EL. Human-level control through deep reinforcement learning // Nature. 2015. Vol. 518. N. 7540. P. 529–533.
- [3] TALLEC C., BLIER L., OLLIVIER Y. Making deep q-learning methods robust to time discretization // Proceedings of the 36th International Conference on Machine Learning. 2019. Vol. 97. P. 6096–6104