УДК 004.896

ИССЛЕДОВАНИЕ ОФЛАЙН МЕТОДОВ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В РОБОТОТЕХНИЧЕСКИХ ЗАДАЧАХ

Симонов Н.А. (Университет ИТМО)
Научный руководитель – к.т.н., доцент Ведяков А.А. (Университет ИТМО)

Введение. Значительный интерес сообщества специалистов в области машинного обучения к офлайн методам обучения с подкреплением привел к разработке множества новых алгоритмов, разработанных специально для обучения политик, которые обучаются без возможности взаимодействия с окружающей средой. Эта проблематика особенно актуальна в контексте робототехнических задач, где взаимодействие робота с окружающей средой может требовать значительных затрат ресурсов. Офлайн методы обучения с подкреплением позволяют предварительно обучить робота в симуляции или на предварительно собранных данных, что дает возможность значительно сэкономить ресурсы и не повредить оборудование. Целью этой работы является исследование указанного офлайн метода обучения с подкреплением ReBRAC и выработка методических указаний, для последующего применения в новых задачах.

Основная часть. Офлайн методы обучения, представленные, в частности, в работе [1], позволяют осуществлять обучение политики на основе статического набора данных, полученного от одной или нескольких других политик, вместо обучения на основе активного взаимодействия с окружающей средой. Указанные ограничения порождают ряд сложностей, такие как оценка функции ценности действий, которые не включены в статический набор.
В основе метода ReBRAC, который представляет собой расширение метода TD3+BC [2], лежит метод Behavior Regularized Actor-Critic (BRAC) [3], основная идея которого заключается в применении одинакового штрафа как для актора, так и для критика при выполнении действий, отсутствующих в статическом наборе данных. Авторы метода ReBRAC, в свою очередь, предлагают использовать различные коэффициенты для наложения штрафа на актора и критика. Кроме предложенной алгоритмической модификации, авторы также рекомендуют настраивать параметры, которые играют ключевую роль в реализации большинства офлайн методов обучения. Среди таких параметров можно выделить глубину нейронных сетей актора и критика [4], размер батча (количество сэмплов, используемых в одной итерации обновления модели) [5], коэффициент штрафа (величина штрафа, накладываемого на действия, не включенные в статический набор данных) [6], а также коэффициент дисконтирования γ (важность будущих вознаграждений по сравнению с текущими) [7]. Настройка этих параметров позволяет адаптировать метод к конкретной задаче и достичь более эффективных результатов.

Симуляционные среды
Для исследования метода ReBRAC используются симуляционные среды на основе библиотек gym, MuJoCo и PyBullet. Обучения модели проводилось для нескольких задач с различными наборами гиперпараметров для подбора наилучших конфигураций по качеству обучения.

Выводы. В работе рассматривался актуальный метод офлайн обучения с подкреплением, который является перспективным на сегодня. Ставилась цель получить метод решения робототехнических задач, не используя взаимодействие агента со средой с помощью метода ReBRAC. После проведения экспериментов в симуляционных средах, разработаны методические указания по использованию и настройке исследованного метода для применения в робототехнических системах.

Список использованных источников:

1. Tarasov D. et al. Revisiting the Minimalist Approach to Offline Reinforcement Learning //arXiv preprint arXiv:2305.09836. – 2023.
2. Fujimoto S., Gu S. S. A minimalist approach to offline reinforcement learning //Advances in neural information processing systems. – 2021. – Т. 34. – С. 20132-20145.C. et al. Diffusion policy: Visuomotor policy learning via action diffusion //arXiv preprintarXiv:2303.04137. – 2023.
3. Wu Y., Tucker G., Nachum O. Behavior regularized offline reinforcement learning //arXiv preprint arXiv:1911.11361. – 2019.
4. Kumar A. et al. Conservative q-learning for offline reinforcement learning //Advances in Neural Information Processing Systems. – 2020. – Т. 33. – С. 1179-1191.
5. Nikulin A. et al. Q-Ensemble for Offline RL: Don't Scale the Ensemble, Scale the Batch Size //arXiv preprint arXiv:2211.11092. – 2022.
6. Rezaeifar S. et al. Offline reinforcement learning as anti-exploration //Proceedings of the AAAI Conference on Artificial Intelligence. – 2022. – Т. 36. – №. 7. – С. 8106-8114.
7. Wu J. et al. Supported policy optimization for offline reinforcement learning //Advances in Neural Information Processing Systems. – 2022. – Т. 35. – С. 31278-31291.