

ИССЛЕДОВАНИЕ АЛГОРИТМОВ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В ЗАДАЧЕ УПРАВЛЕНИЯ ШАГАЮЩИМ РОБОТОМ ПРИ ПРЕОДОЛЕНИИ ПРЕПЯТСТВИЙ

Ворков Н.Р. (Университет ИТМО)

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

Введение. Работа посвящена сравнительному анализу алгоритмов машинного обучения с подкреплением для решения задачи преодоления препятствий четвероногим шагающим роботом. Шагающие роботы все чаще начинают применяться в решении задач связанных с картографированием, поисковыми работами, а также доставкой грузов внутри предприятий. Все эти задачи требуют от робота уверенного преодоления различного рода препятствий на пути следования. Это порождает необходимость искать и внедрять все более надежные в работе алгоритмы обучения роботов.

Среди современных методов для решения задачи передвижения шагающих роботов можно выделить два основных:

- метод прогнозирующего управления (MPC) – метод, основанный на использовании теории оптимального и адаптивного управления для построенной модели объекта,
- метод обучения с подкреплением (RL) – метод, основанный на принципах машинного обучения.

Для крупных робототехнических компаний свойственно непрерывно модернизировать технические характеристики своих продуктов. С коммерческой точки зрения требуется использовать такое решение, которое позволило бы настроить систему правления модернизированных роботов без существенных затрат человеческих ресурсов. Для данной задачи может оказаться проще переобучить модель с помощью машинного обучения, чем править математическую модель для метода прогнозирующего управления [4]. В случае использования машинного обучения при изменении технических характеристик робота будет достаточно просто запустить обучение в симуляторе по уже используемым алгоритмам, что довольно сильно экономит человеческие ресурсы.

Основная часть. На первом этапе был проведен сравнительный анализ наиболее популярных в последние три с половиной года алгоритмов обучения с подкреплением [1, 2] и актуальность совместного с ними использования методов прогнозирующего управления [3]. Из проанализированных алгоритмов было выбрано несколько потенциально наиболее успешных для задачи преодоления препятствий четвероногим роботом.

Чтобы оценить эти алгоритмы на практике, был проведен анализ пригодности и популярности существующих симуляторов [1]. Был изучен исходный код выбранного симулятора. На основе исходного кода была создана необходимая для тестирования среда с автоматически генерируемым препятствием случайного размера. Для сгенерированной среды были созданы функции вознаграждения, а также добавлена система логирования текущих параметров среды в tensorboard. Данная система логирования упростила слежение за процессом обучения модели на удаленном сервере, потому что позволяла транслировать данные на персональный компьютер.

После в разработанной среде были проведены испытания алгоритма ближайшей оптимизации политик (PPO) для двух моделей четвероногих роботов с применением низкоуровневого метода прогнозирующего управления и без него.

Выводы. По итогам проведения практических испытаний алгоритма РРО для задачи преодоления препятствий четвероногим роботом был сделан следующий вывод: многообразие способов преодоления препятствий вносит в систему множество локальных минимумов, это не позволяет системе достичь приемлемого способа выполнения задачи, потому необходимо применять такие алгоритмы, которые позволяют повысить стохастические свойства агента. Также стоит тщательно подходить к выбору функции награды, чтобы минимизировать возможность получения наград за потенциально неудачные действия. Обновленный код [5] симулятора выложен на платформе GitHub в виде форка репозитория raisimLab.

Список использованных источников:

1. Zhang H., He L., Wang D. Deep reinforcement learning for real-world quadrupedal locomotion: a comprehensive review // *Intelligence & Robotics*. – 2022. – 2(3). – pp. 275-97.
2. Mastrogeorgiou A.S., Elbahrawy Y.S., Machairas K., Kecskemethy A., Papadopoulos E.G. Evaluating deep reinforcement learning algorithms for quadrupedal slope handling // In 23rd International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR 2020), Moscow, Russian Federation. – 2020. – pp. 24-26.
3. Yang Y., Zhang T., Coumans E., Tan J., Boots B. Fast and efficient locomotion via learned gait transitions // In Conference on Robot Learning, PMLR. – 2022. – pp. 773-783.
4. Suryana L., Which one is better: Reinforcement Learning or Model Predictive Control? Inverted Pendulum — Case* [Электронный ресурс] // Medium. – 2020. Режим доступа : <https://medium.com/analytics-vidhya/which-one-is-better-reinforcement-learning-or-model-predictive-control-inverted-pendulum-case-7fc29e52bbfb>, свободный. – Загл. с экрана.
5. Ворков Н. Форк RAISIM [Электронный ресурс] // GitHub, Inc. – 2023. Режим доступа : <https://github.com/VorkovN/raisimLib/tree/myBranch>, свободный. – Загл. с экрана.