Исследование и переработка системы навигации виртуального ровера в среде Unity на основе обучения с подкреплением

В. В. Маносин*, М. И. Кумсков‡

Проверен и переработан прототип навигации виртуального ровера на основе PPO. Вместо исходного наблюдения без дальномерных данных реализован вектор из 52 компонент, включая 36 признаков LiDAR. Три независимо обученные политики оценены на 500 эпизодах каждая: средняя доля достижения цели — 46,8%, стандартное отклонение между запусками — 6,8 процентного пункта. Результаты относятся к одной симуляционной сцене.

Ключевые слова: обучение с подкреплением, навигация, виртуальный ровер, PPO, LiDAR.

1. Постановка задачи и аудит

Рассмотрена навигация ровера к цели без карты в Unity [1, 2]; выполнены аудит наблюдений и оценка PPO.

В `CollectObservations()` обнаружены 23 входа: направление на цель в координатах ровера (3), вертикаль ровера в мировой системе (3), скорость (1), направление и модуль угловой скорости (3+1), уклоны (4), нули в позициях 16–23 (8). Последние восемь — сектора *камерной маски* (0 — свободно, 1 — препятствие); её обработка отключена. Дальности использовались в награде, но не поступали политике.

2. Переработанная постановка

LiDAR посылает 360 горизонтальных лучей длиной 5 м с высоты 0,5 м. Локальные углы $\alpha_j = -180^\circ + j^\circ$ ($j = 0, \ldots, 359$). Сектор $k = 0, \ldots, 35$ объединяет лучи $10k, \ldots, 10k+9$; признак равен $\min d_j/5$, где d_j — дальность

* *Маносин Владислав Валерьевич* — студент мех.-мат. ф-та МГУ имени М. В. Ломоносова, каф. вычислительной математики, e-mail: `vladislav.manosin@math.msu.ru`.

Manosin Vladislav Valerievich — Student, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University.

‡ *Кумсков Михаил Иванович* — д-р физ.-мат. наук, профессор каф. вычислительной математики мех.-мат. ф-та МГУ имени М. В. Ломоносова, зав. лабораторией вычислительных методов, e-mail: `mikhail.kumskov@math.msu.ru`.

Kumskov Mikhail Ivanovich — Doctor of Physical and Mathematical Sciences, Professor, Department of Computational Mathematics, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University; Head of the Laboratory of Computational Methods.

Опубликовано на условиях лицензии CC-BY 4.0
Published under CC-BY 4.0

до препятствия, либо 5 м без пересечения. Учитывается слой препятствий Unity.

Вектор из 52 признаков содержит: локальное направление на цель (x, z) (2), расстояние до цели (1), локальные линейную и угловую скорости (3+3), мировую вертикаль в системе ровера (3), уклоны (4) и LiDAR (36). Дистанция делится на 50 м и ограничивается единицей; скорости — на 4 м/с и 1 рад/с с ограничением $[-1, 1]$. Уклоны измеряются в четырёх точках на расстоянии 1,5 м от ровера; углы нормалей поверхности к вертикали делятся на 45°, обрезаются до $[0, 1]$ (без пересечения — 0). Направления единичные.

Тяга, поворот и тормоз подаются в `RoverController.SetInputs` в $[-1, 1]$ (торможение в $[0, 1]$). Пределы [1]: 500 Н·м, 45°, 1000 Н·м. Преобразование внутри контроллера не проверено.

Плотная награда v5 за прогресс к цели и время:

$$r_t^{\text{prog}} = \text{clip}\big(0{,}1(d_{t-1} - d_t), -0{,}05, 0{,}05\big), \quad r_t^{\text{time}} = -0{,}0002.$$

Столкновение даёт $-0{,}25$, а минимальная дальность q_t в переднем секторе $\pm 45°$ — штраф

$$r_t^{\text{near}} = -0{,}0005 \left[\text{clip}\left(\frac{2 - q_t}{1{,}5}, 0, 1\right)\right]^2;$$

При $q_t \geq 2$ м штраф нулевой. Цель в радиусе 1,5 м завершает эпизод с $+1$ без плотных наград этого шага; лимит — 8 столкновений. Curiosity отключён.

3. Эксперименты и результаты

Для предварительного подбора конфигурации выполнено по одному диагностическому запуску v2–v5. Показатели по последним 25 тыс. шагов: v2 — штраф за столкновение $-0{,}25$, лимит 8, успех 62,1%, 2,05 столкновения на эпизод; v3 — $-0{,}50$, 3, 42,9%, 0,87; v4 — $-0{,}50$, 8, 54,0%, 1,22; v5 — $-0{,}25$, 8, 62,9%, 1,45. В v5 дополнительно введён штраф за близость; успех без столкновений составил 58,1%.

По сочетанию долей успешных и безопасных эпизодов v5 принята как рабочая конфигурация для дальнейшей оценки.

Для выбранной v5 выполнены три независимых обучения PPO [3] (seed 1, 2, 3; ML-Agents 1.0.0 [4]): батч 1024, буфер 10240, шаг 0,0002, энтропия 0,002, отсечение 0,15, GAE $\lambda = 0{,}95$, дисконтирование 0,995; четыре эпохи, три слоя по 256 нейронов, горизонт 64, по 500000 шагов. Расписания линейные, нормирование выключено; прочие настройки, включая активацию, — по умолчанию.

Оценка: по 500 эпизодов без изменения весов. Успех — расстояние до цели $\leq 1{,}5$ м, отказ — восемь столкновений, тайм-аут — достижение лимита шагов (для v5 значение не уточнено; в сохранённой сцене v2 — 4000). Стартовая позиция фиксирована, угол ровера равномерно распределён на $[0°, 360°)$; координаты цели $x \in [-15, 24]$, $z \in [-10, 11]$ м равномерны, высота — рельеф +0,3 м. Требования: удаление от старта больше 3 м, свободный радиус 1,75 м; до 100 попыток. Карта неизменна. При seed 20260927 проверено совпадение стартовых расстояний; полные начальные состояния не сверялись.

Таблица 1: Оценка v5: по 500 эпизодов на политику

Метрика	Seed 1	Seed 2	Seed 3	Среднее ± SD
Успех, %	54,6	43,2	42,6	46,8 ± 6,8
Успех без столкн., %	41,6	31,8	30,0	34,5 ± 6,2
Лимит столкн., %	7,4	17,2	17,4	14,0 ± 5,7
Тайм-аут, %	38,0	39,6	40,0	39,2 ± 1,1
Столкновений/эп.	1,476	2,224	2,126	1,942 ± 0,407

Из 1500 эпизодов: 702 успеха (517 без столкновений), 210 отказов, 588 тайм-аутов. SD долей — в процентных пунктах [5].

Без столкновений завершились 517 из 702 успешных эпизодов (73,6%); доля 34,5% в табл. 1 рассчитана относительно всех эпизодов. Среди неуспешных завершений преобладают тайм-ауты: 39,2% всех эпизодов при 14,0% завершений по лимиту столкновений. Возможные причины тайм-аутов — излишне осторожное поведение под действием штрафа за близость, зацикливание и недостаточная длительность эпизода.

4. Заключение

LiDAR включён во вход политики, средняя доля успеха — 46,8% в одной сцене. Сравнения со старой политикой нет; камера отключена, цель определяется координатами Unity. Используются земная гравитация и `WheelCollider`, поэтому перенос на реальный ровер не установлен.

Список литературы

[1] А. А. Жабицкая, *Реализация и анализ модели автономной навигации лунного ровера с использованием методов Reinforcement Learning*, ВКР специалиста, МГУ имени М. В. Ломоносова, М., 2025.

[2] L. Tai, G. Paolo, M. Liu, “Virtual-to-Real Deep Reinforcement Learning: Continuous Control of Mobile Robots for Mapless Navigation”, IEEE/RSJ Int. Conf.

on Intelligent Robots and Systems (IROS), 2017, 31–36. DOI: 10.1109/IROS.2017.8202134.

[3] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, "Proximal Policy Optimization Algorithms", arXiv: arXiv:1707.06347.

[4] A. Juliani, V.-P. Berges, E. Teng et al., "Unity: A General Platform for Intelligent Agents", arXiv: arXiv:1809.02627.

[5] P. Henderson, R. Islam, P. Bachman et al., "Deep Reinforcement Learning that Matters", AAAI Conference on Artificial Intelligence, **32**:1. DOI: 10.1609/aaai.v32i1.11694 (2018).

Статья поступила 5 октября 2026 г.

Investigation and Refinement of a Virtual Rover Navigation System in Unity Based on Reinforcement Learning

V. V. Manosin, M. I. Kumskov

We audit and refine a PPO-based virtual rover navigation prototype. The original observation lacked obstacle distances; the revised 52-dimensional vector contains 36 LiDAR features. Three independently trained policies were evaluated over 500 episodes each. The mean success rate was 46.8%, with a between-run standard deviation of 6.8 percentage points. The results apply to one simulated scene only.

Keywords: reinforcement learning, navigation, virtual rover, PPO, LiDAR.

References

[1] A. A. Zhabitskaya, *Implementation and Analysis of a Lunar Rover Autonomous Navigation Model Using Reinforcement Learning Methods*, Specialist thesis, Lomonosov Moscow State University (In Russian), Moscow, 2025.

[2] L. Tai, G. Paolo, M. Liu, "Virtual-to-Real Deep Reinforcement Learning: Continuous Control of Mobile Robots for Mapless Navigation", IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2017, 31–36. DOI: 10.1109/IROS.2017.8202134.

[3] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, "Proximal Policy Optimization Algorithms", arXiv: arXiv:1707.06347.

[4] A. Juliani, V.-P. Berges, E. Teng et al., "Unity: A General Platform for Intelligent Agents", arXiv: arXiv:1809.02627.

[5] P. Henderson, R. Islam, P. Bachman et al., "Deep Reinforcement Learning that Matters", AAAI Conference on Artificial Intelligence, **32**:1. DOI: 10.1609/aaai.v32i1.11694 (2018).

Received on October 5, 2026