Алгоритм управления мобильным роботом в среде с динамическими препятствиями на основе метода обучения с подкреплением

Валенков Алексей Дмитриевич

Студент

Кафедра проблем управления РТУ МИРЭА, Москва, Россия

E-mail: valenkov.a.d@edu.mirea.ru

Научный руководитель *— Диане Секу Абдель Кадер*

Современное развитие робототехники связано с необходимостью создания автономных мобильных систем, способных эффективно функционировать в сложных средах и в условиях потенциального контакта с людьми. Одним из ключевых факторов успешной работы таких систем является способность избегать столкновений как со ститическими обектами, так и с динамическими.

Большинство подходов к планированию целенаправленных движений мобильных роботов требуют настройки параметров экспертами для получения хороших навигационных характеристик [1]. В этом исследовании рассматривается применение глубокого обучения с подкреплением для обеспечения автономного перемещения мобильного робота в динамической среде.

Использование нейросетевых технологий для оптимизации расчета управляющих воздейсвий на робота позволит уменьшить потребность в высоких вычислительных мощностях на борту устройства. Кроме того, подобный подход повысит адаптивность робота к изменчивым факторам внешней среды (вариациям скоростей движения людей и априорно неизвестной геометрии статичных препятствий на пути движения робота).

Пусть робот использует данные лидарных датчиков, имеет три степени подвижности (Рис. 1) и глубокую нейронную сеть для генерации управляющих сигналов a_t, которые направляют его к заданной цели, избегая столкновения с препятствиями. Тогда задачу можно описать следующей формулой:

$$a_t = f(l_t, p_t, v_{t-1}, \Delta\theta_t), \tag{1}$$

где a_t — вектор управляющего воздействия на робота, сстоящий из линейных скоростей v_{xt}, v_{yt} и угловой скорости ω_t , l_t — вектор, содержащий информацию от дальнометрических сенсоров, p_t — координаты точки цели в системе координат робота, v_{t-1} — вектор

линейных скоростей агента в предыдущий момент времени, $\Delta\theta_t$ — разница между поворотом агента и углом на цель.

В работе проводится обзор существующих методов навигации мобильных роботов на основе нейросетевых технологий, включающих в себя глубокое Q-обучение (DQN), алгоритм Proximal Policy Optimization (PPO) и Deep Deterministic Policy Gradients (DDPG). В результате сравнительного анализа сделан вывод о приемуществах алгоритма PPO в контексте данной задачи.

В работе рассмотрены функции оценки мгновенной награды, представленные в публикациях [1, 2]. Учтены факторы приближения агента к цели и поворот агента в заданную сторону, столкновение с препятствием, и ограничение времени на выполнение задачи.

Разработана функция награды с дополнительным параметром, влияющим на итоговую награду при приближении агента к препятствию на заданное расстояние (4).

$$d_{reward} = (d_{t-1} - d_t)2^{(d_{t-1}/d_t)}, \tag{2}$$

$$\theta_{reward} = \theta_{target} - \theta_{robot}, \tag{3}$$

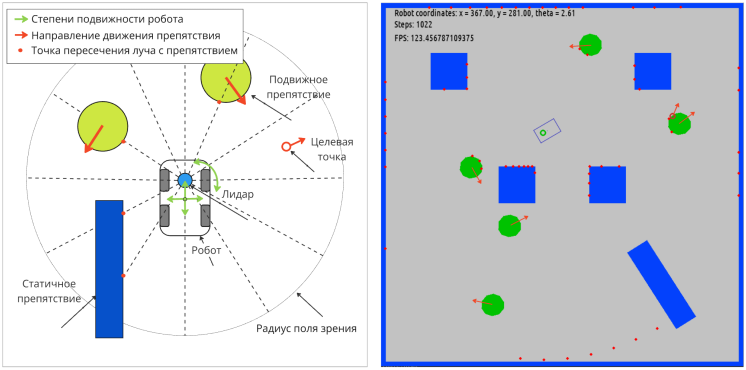
$$c_{reward} = (x_{t-1} - x_t)2^{(x_{t-1}/x_t)}, \tag{4}$$

$$r_1(s_t, a_t) = \begin{cases} r_{\text{arrive}}, & \text{if } d_t < c_d \\ r_{\text{collision}}, & \text{if } x_t < c_o \\ r_{\text{timeout}}, & \text{if } timeout \\ c_r d_{reward} - c_p \theta_{reward}, & \text{otherwise.} \end{cases} \tag{5}$$

$$r_2(s_t, a_t) = \begin{cases} r_{\text{arrive}}, & \text{if } d_t < c_d \\ r_{\text{collision}}, & \text{if } x_t < c_o \\ r_{\text{timeout}}, & \text{if } timeout \\ c_r d_{reward} - c_p \theta_{reward} - c_k c_{reward}, & \text{if } x_t < c_h \\ c_r d_{reward} - c_p \theta_{reward}, & \text{otherwise.} \end{cases} \tag{6}$$

где r_{arrive} — награда за достижение цели, $r_{collicion}$ — отрицательная награда за столкновение с препятствием, d_t — реальное расстояние от робота до цели, x_t — $\min(l_t)$, c_d — Порог достижения цели, c_o — порог столкновения, c_h — Порог получения награды за приближение к препятствию, c_r — коэффициент награды за приближение к цели, c_p — коэффициент награды за удержание оринтации цели, c_k — коэффициент награды за приближение к препятствию.

Иллюстрации



а) б)

Рис. 1. Описание окружения робота: а) схема движения робота и обработка сенсорной информации; б) рабочее простанство робота в симуляторе.

В ходе проведения исследования была разработана преобразованая функция награды (6), с учетом чувствительности приближения к препятствиям, что повысило адаптивность модели к динамически изменяющимся средам.

Проведено экспериментальное исследование по оценке эффективности моделей поведения, обченных алгоритмом PPO с использованием функций награды r_1 (5) и r_2 (6).

Литература

1. Taheri H. Hosseini S. R. Nekoui M. A. Deep Reinforcement Learning with Enhanced PPO for Safe Mobile Robot Navigation // ArXiv abs/2405.16266, 2024, P. 1–5.
2. Flögel D. Fischer L. Rudolf T. Schürmann T. Socially Integrated Navigation: A Social Acting Robot with Deep Reinforcement Learning // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2024, P. 4–5.