

МЕТОДЫ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В ЗАДАЧЕ ГРУППОВОГО ПРЕСЛЕДОВАНИЯ

Артамонов Георгий Игоревич

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: s02220323@gse.cs.msu.ru

Научный руководитель — Фомичев Василий Владимирович

Рассматривается задача группового преследования на плоскости в ограниченной прямоугольной области с препятствиями с полной информацией о состоянии агентов и с локальной информацией о препятствиях в окрестностях преследующих агентов. Ставится задача поимки одного или нескольких уклоняющихся агентов группой преследователей с использованием централизованного управления всеми преследователями. Динамика агентов описывается системой разностных уравнений (1). Предполагается наличие преимущества в ограничениях скорости v_{max} и динамики для уклоняющихся агентов.

$$\begin{cases} x_{i+1} = x_i + v_i \cos \varphi_i \\ y_{i+1} = y_i + v_i \sin \varphi_i \\ \varphi_{i+1} = \varphi_i + \Delta\varphi_i, & |\Delta\varphi_i| \leq \delta_1 \\ v_{i+1} = \text{sat}(v_i + \Delta v_i, 0, v_{max}), & |\Delta v_i| \leq \delta_2 \\ (x_0, y_0, \varphi_0, v_0) = (x^0, y^0, 0, 0) \end{cases} \quad (1)$$

Здесь (x_i, y_i) — координаты агента в дискретный момент времени i . (x^0, y^0) — стартовая точка агента. φ_i, v_i — угол поворота и скорость агента в момент времени i соответственно. Управление мобильным агентом осуществляется изменением его скорости и угла поворота.

Существуют классические подходы к решению задачи преследования и уклонения, основанные на теории дифференциальных игр. В статье [1] представлен обзор последних достижений в этой области. Для решения поставленной задачи предлагается использование методов обучения с подкреплением [2–4]. В статье [3] рассматривается применение метода Q-обучения для упрощенной клеточной игры преследования и уклонения. Были отдельно рассмотрены случаи, когда преследующие агенты выбирают действия на каждой итерации коллективно и независимо друг от друга. Обученная модель смогла успешно преследовать и ловить убегающего агента, почти всегда избегая столкновений между преследователями.

Авторы статьи [2] адаптируют идеи метода Q-обучения для случая непрерывного пространства действий и представляют метод DDPG, работающий по схеме Actor-Critic. Алгоритм DDPG, используя один и тот же способ обучения, ту же архитектуру нейронной сети и те же гиперпараметры, смог качественно решить более 20 различных физических задач, включая такие классические задачи, как стабилизация перевернутого маятника на тележке и управление манипулятором с захватом. В статье [4] рассматривается дифференциальная игра «Собака–овца», в которой среда представляет собой круг, первый агент находится внутри круга и пытается его покинуть, а второй агент перемещается по окружности, преследуя первого. Были применены метод глубокого Q-обучения и алгоритм DDPG, позволившие обучить уклоняющегося агента успешно покидать круг.

В рамках данной работы был реализован алгоритм DDPG на языке программирования Python с использованием библиотеки PyTorch и применен для управления группой преследователей. Модели нейросетей актёра и критика содержат два скрытых полносвязных слоя с функцией активации ReLU. Была разработана программная среда моделирования, позволяющая задавать карту с препятствиями и начальные точки для агентов и визуализировать процесс преследования. Для управления уклоняющимся агентом использовалось случайное блуждание с постоянной скоростью и меняющимся углом поворота, а также методы потенциальных полей для отталкивания от стен и преследователей. Были проведены эксперименты для трех преследующих и одного уклоняющегося агента и разных начальных конфигураций агентов. В большинстве запусков была осуществлена поимка убегающего агента.

Литература

1. Weintraub I., Pachter M., Garcia E. An Introduction to Pursuit-evasion Differential Games // American Control Conference 2020.
2. Lillicrap T., Wierstra D., Hunt J., Pritzel A., Heess N., Erez T., Tassa Y., Silver D. Continuous Control with Deep Reinforcement Learning // Published as a conference paper at ICLR 2016.
3. Paczolay G., Harmati I. A Simplified Pursuit-evasion Game with Reinforcement Learning // Periodica Polytechnica Electrical Engineering and Computer Science, 65(2), pp. 160–166, 2021.
4. Xu C., Zhang Y., Wang W., Dong L. (2022) Pursuit and Evasion Strategy of a Differential Game Based on Deep Reinforcement Learning // Front. Bioeng. Biotechnol. 10:827408.