

0.1. Бобровская О.П. Агент транспортного потока, обученный с подкреплением

Активно разрабатываемые различными производителями системы беспилотного транспорта будут сильно отличаться друг от друга. При их внедрении в существующий транспортный поток возможны проблемы, которые можно заранее смоделировать в рамках вычислительных экспериментов.

Как известно, сейчас разработки в основном сфокусированы на обучении автоматических систем подражанию реальным водителям. Поэтому имеет смысл проверить последствия ситуации, при которой множество отдельно обученных агентов-беспилотников будут двигаться по одной дороге.

Для этого было решено применить методы обучения с подкреплением к агентам транспортного потока. В качестве среды для обучения использовалась разработанная ранее модель транспортного потока, основанная на потенциале действия [1]. Обучение осуществлялось с помощью модели DDPO, впервые предложенной в работе [2] и реализованной в библиотеке keras.rl2 для языка Python3.

У агента (нейронной сети) обучалось 5 параметров:

- $v \in \mathbb{R}; \quad -1 \leq v \leq 1$ — скорость;
- $d_i \in \mathbb{R}; \quad i = 1..4; \quad -1 \leq d_i \leq 1$ — направления перестроения: при всех $d_i < 0$ полоса не меняется, в противном случае выбирается d_{max} .

В обучении с подкреплением важно задать функцию вознаграждения, чтобы агент поощрялся за совершение полезных действий и наказывался в противном случае. Цель, которой мы хотели достичь — агент, движущийся вперед на максимальной скорости, при необходимости перестраивающийся и избегающий столкновений. Поэтому для оценки совершенных действий каждый такт времени использовалась функция награждения, которая состоит из следующих составляющих (с указанием баллов в скобках): сохранение безопасной дистанции с впереди (0.4) и позади (0.05) идущими агентами на полосе, перестроение на другую полосу (0.4) только для увеличения дистанции до впереди идущего агента, сохранение положительной скорости (0.2).

Результатом является программное обеспечение, предоставляющее среду для обучения нейросетевого агента, имитирующего движение транспортного средства. Обучение и тестирование может осуществляться в гибридной среде, состоящей как из исходных автоматов, так и из агентов, прежде обученных в этой среде.

Работа выполнена в рамках государственного задания ФГУ ФНЦ НИИСИ РАН по теме № FNEF-2024-0001 «Создание и реализация доверенных систем искусственного интеллекта, основанных на новых математических и алгоритмических методах, моделях быстрых вычислений, реализуемых на отечественных вычислительных систе-

мах» (1023032100070-3-1.2.1).

Научный руководитель — к.т.н. Гавриленко Т. В.

Список литературы

- [1] БОБРОВСКАЯ О.П., ГАВРИЛЕНКО Т.В., ГАЛКИН В.А. Модель транспортного потока, основанная на взаимодействии частиц с потенциалом действия // Вестник КРАУНЦ. Физ.-мат. науки. 2022. Т. 40. № 3. С. 72–87.
- [2] VAN HASSELT H., GUEZ A., SILVER D. Deep reinforcement learning with double Q-learning // Proceedings of the AAAI Conference on Artificial Intelligence. 2016. Vol. 30. N. 1. P. 2094–2100. DOI:10.1609/aaai.v30i1.10295.