ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ АЛГОРИТМОВ ОБУЧЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ АГЕНТОВ В УСЛОВИЯХ МИНИМИЗАЦИИ ОБРАБАТЫВАЕМОЙ ИНФОРМАЦИИ

Е. И. Пантюхин
Томский политехнический университет
E-mail: eip2@tpu.ru

Введение

За последние несколько лет было проведено множество исследований методов машинного обучения с целью улучшения функциональных возможностей мобильных роботов. Одним из важнейших вопросов при проектировании и разработке интеллектуальной мобильной системы является баланс между быстродействием и полнотой обрабатываемой информации. Чем сложнее задачи, поставленные перед интеллектуальными агентами, тем больший объём информации необходимо обрабатывать, что в свою очередь увеличивает затрачиваемые ресурсы, время обучения и как правило снижает быстродействие системы [1-3].

Целью данной работы является проверка возможности различных алгоритмов обучения интеллектуальных агентов выполнять поставленные задачи в условиях сильного ограничения в предоставляемой для обработки информации.

Среда для обучения

С целью проверки различных алгоритмов обучения интеллектуальных агентов была смоделирована задача перемещения объекта агентом в указанную зону. На рисунке 1 продемонстрирована сцена, в которой агентом является синий квадрат. Белым квадратом является объект, который необходимо переместить в зелёную зону. Перемещение объекта возможно только методом его толкания агентом, никаких функций сцепки элементов не предусмотрено. Лучи, исходящие от агента, являются сенсорами, способными определять тип объекта, а также расстояние до него.

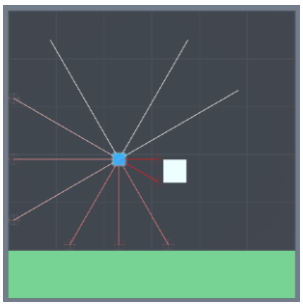

Рис. 1. Смоделированная среда для обучения интеллектуальных агентов

Каждая симуляция ограничена по числу пройденных эпох. Это сделано с целью выхода симуляции из тупиковой ситуации, когда объект расположен в углу сцены, и агент не имеет возможности сдвинуть его. В начале каждой новой симуляции агент и объект располагаются в случайных местах.

Результаты обучения

Для достижения разумной степени автономии интеллектуального агента два основных требования — это восприятие и рассуждение. Первый обеспечивается сенсорной системой, которая собирает информацию об агенте относительно окружающей среды. Количество и качество сенсоров может сильно отличаться, и подбирается индивидуально для каждой задачи. На рисунке 2 продемонстрированы различные конфигурации сенсоров, которые отличаются углом обзора и максимальным измеряемым расстоянием. В качестве алгоритмов обучения интеллектуальных агентов применялись Soft Actor-Critic (SAC) и Proximal Policy Optimization (PPO).

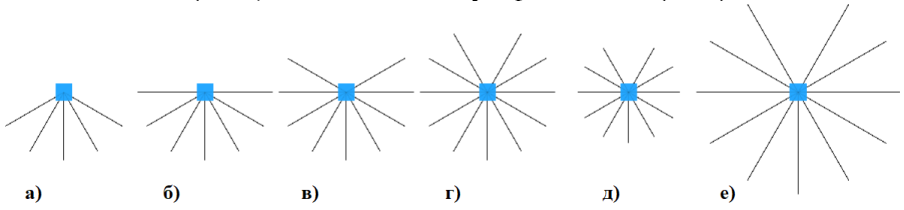

Рис. 2. Различные конфигурации сенсоров интеллектуальных агентов (а — 120 градусов, б — 180 градусов, в — 240 градусов, г — 300 градусов, д — 300 градусов, максимальная длина луча L уменьшена на 25%, е — 300 градусов, максимальная длина луча L увеличена на 50%)

Сконфигурированные искусственные нейронные сети для алгоритмов SAC и PPO имеют по 2 скрытых полносвязных слоя прямого распространения, имеющих по 256 нейронов. Период обучения агентов составляет 2.5 миллиона эпох. Результаты обучения интеллектуальных агентов с различными конфигурациями сенсоров представлены в таблице 1.

Таблица 1. Результаты обучения интеллектуальных агентов

Конфигурации сенсоров	SAC		PPO	
	Среднее вознаграждение	Функция потерь	Среднее вознаграждение	Функция потерь
а — 120 градусов	2.63	9.95e-4	4.908	0.389
б — 180 градусов	3.012	2.81e-3	4.918	0.307
в — 240 градусов	3.917	3.39e-3	4.981	0.019
г — 300 градусов	4.742	3.97e-3	4.986	0.015
д — 300 градусов, L -25%	1.278	1.29e-3	4.872	0.345
е — 300 градусов, L+50%	4.979	3.39e-4	4.985	0.014

Среднее вознаграждение и функция потерь высчитываются из последних 50000 эпох симуляции. Среднее вознаграждение характеризует то, насколько хорошо выполняется поставленная задача, и может варьироваться от -1 до 5. Из таблицы 1 видно, что с увеличением угла обзора сенсоров агенты выполняются задачу быстрее, и допускают меньше ошибок, однако алгоритм обучения PPO способен найти сложную последовательность решений, которая приводит к хорошим результатам даже в условиях неполноты предоставляемой информации.

Функция потерь характеризует то, как хорошо агент способен предсказывать последствия своих действий. В алгоритме PPO заметна зависимость между предоставляемой информацией и функцией потерь, на основе которой можно подбирать типы сенсоров для различных задач. В алгоритме SAC на основе функции потерь возможен анализ результатов обучения только при полноте предоставляемой информации. Из всех проведённых опытов, только в последнем функция потерь алгоритма SAC в процессе обучения имела характерную тенденцию роста, а затем уменьшения. Все остальные конфигурации приводили к хаотичным скачкам показателя на каждом этапе обучения.

Стоит отметить, что оба алгоритма значительно теряют эффективность при уменьшении максимального измеряемого лучом расстояния, а при увеличении данного показателя алгоритм SAC получает достаточно информации для эффективного пространственного восприятия сцены и поиска эффективного решения задачи.

Заключение

Проанализировав полученные результаты были сделаны следующие выводы:

1. В условиях неполноты или нарочной минимизации входных данных, алгоритм PPO показывает лучшие результаты за более короткое время обучения, нежели SAC;
2. Алгоритм SAC в условиях неполноты предоставляемой информации не способен обнаружить сложные, многоуровневые решения задач, найденных алгоритмом PPO;
3. На основе функции потерь алгоритма PPO можно судить о полноте предоставляемой информации. Алгоритм SAC такой возможности не предоставляет.

Список использованных источников

1. Researchers ran a simulator to teach this robot dog to roll over. [Электронный ресурс] / Интернет-издание о стартапах, интернет-бизнесе, инновациях и веб-сайтах – URL: https://techcrunch.com/2019/01/17/researchers-ran-a-simulator-to-teach-this-robot-dog-to-roll-over/ (дата обращения 01.03.2021)
2. Применение нейронных сетей в робототехнике: перспективы и преимущества [Электронный ресурс] / IT форум Хабр – URL: https://habr.com/ru/post/239543 (дата обращения 01.03.2021)
3. Janglova D. Neural Networks in Mobile Robot Motion. Int. journal Adv. Robot. Syst. – 2004. – vol. 1. – P. 15–22.