УДК 004.89:004.4

НАХОЖДЕНИЕ ОПТИМАЛЬНОЙ СТРАТЕГИИ В ПРОЦЕССЕ ВЫПОЛНЕНИЯ ПОСЛЕДОВАТЕЛЬНОСТИ КОМАНД МОБИЛЬНЫМ РОБОТОМ В УСЛОВИЯХ ЧАСТИЧНО ОБОЗРИМОЙ СРЕДЫ С ИСПОЛЬЗОВАНИЕМ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Розанов М.С., Прокопович Г.А.
Объединенный институт проблем информатики Национальной академии наук Беларуси, лаборатория «Робототехнические системы», Минск, Беларусь

Введение. С развитием идеи интеллектуализации робототехнических платформ растет количество прикладных применений методов машинного обучения и искусственного интеллекта (ИИ) в области робототехники к таким типовым задачам, как поиск пути на местности, долгосрочное или краткосрочное планирование, выполнение технологических операций и т.д. Среди всех областей ИИ в робототехнике особо выделяется обучение с подкреплением (reinforcement learning) вследствие наличия удобного фреймворка для описания множества "состояние-действия", которым возможно представить практически любую задачу робототехники из-за высокого уровня дискретизации состояний робота и, соответственно, возможного множества действий, доступных роботу из данного состояния. Обучение с подкреплением позволяет реализовывать сценарии обучения мобильного робота в качестве агента, при этом фактические знания агента о внешней среде отсутствуют и должны приобретаться в процессе обучения посредством метода проб и ошибок (trial-and-error). При этом внешняя среда является частично обозримой, т.е. в каждый момент времени агенту доступно ограниченное количество информации об окружающем мире. Однако при этом агенту доступно некоторое ограниченное множество действий, которые ему позволено использовать в любой последовательности и в любой момент времени для достижения цели. С помощью специального набора правил обучения, агент в итоге строит оптимальную стратегию действий, именуемую также в теории обучения с подкреплением «политикой».

Следует отметить, что важной особенностью обучения агента является способ моделирования среды, с которой взаимодействует агент. Наиболее рапространенной практикой является использование т.н. марковского процесса принятия решений (МППР), который, в данном случае, является частично обозримым МППР.

Цель работы: создание программной модели, способной продемонстрировать процесс обучения выполнению простейшей последовательности действий при отсутствии начальных знаний о внешней среде.

Описание программной части. Для моделирования агента используется виртуальная среда OpenAI Gym, предоставляющая достаточный набор инструментов для тестирования алгоритмов обучения с подкреплением, а также программный продукт под названием Gazebo, использующийся в робототехнике для моделирования физических процессов робота в ходе выполнения действий (симуляция работы сенсоров, симуляция движения, моделирование и расчет коллизий и т.д.).

Следует отметить, что, хотя наличие физического робота при тестировании на данном этапе не предусматривается, в Gazebo будет смоделирован мобильный робот, имеющий на борту несколько инфракрасных дальномеров SHARP, одноплатный компьютер вместе с камерой для распознавания объектов, также в некоторых тестах используется манипулятор в качестве надстройки к основной базе.

Команды роботу подаются в консольном режиме, после чего идет их формализация и интерпретация относительно имеющегося множества действий робота. Вследствие того, что множество действий робота зачастую достаточно велико, обучение в реальном времени неприемлемо долгое. В связи с этим, определенные периоды обучения моделируются программно и среди всех листовых вершин дерева вероятных исходов для данной политики выбирается одно-два с наибольшими показателями награды.

1. Kober, Jens Reinforcement Learning in Robotics: A Survey / Jens Kober, J. Andrew Bagnell, Jan Peters // The International Journal of Robotics Research. — 2013. — №32. — P. 1-38.

2. Sutton, Richard S. Reinforcement Learning: An Introduction / Richard S. Sutton, Andrew G. Barto; . — London : MIT Press, 2018. — 444 p.

3. Leslie Pack Kaelbling, Michael L. Littman, Andrew W. Moore, Reinforcement Learning: A Survey, JAIR, 1996

4. Прокопович, Г.А. Адаптивная нейросетевая система управления автономным мобильным роботом на основе метода обучения с учителем в online режиме / Г.А. Прокопович // Весці нацыянальнай акадэміі навук беларусі. Сер. фіз.-мат. навук. – 2015. -№1. – С. 117-122.

5. Sutton, Richard Between MDPs and Semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning / Richard Sutton, Satinder Singh // Artificial Intelligence. — 1999.

6. Peters, Jan Natural Actor-Critic / Jan Peters, Sethu Vijayakumar, Stefan Schaal // ECML. — 2005.

7. Szepesvari, Csaba Algorithms for Reinforcement Learning of Synthesis Lectures on Artificial Intelligence and Machine Learning series / Csaba Szepesvari: Morgan & Claypool Publishers, 2009. — 78 с.