

# Сравнение алгоритмов обучения с подкреплением для управления движением автономного робота в симуляторе Gazebo

Д.А. Козлов<sup>1</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34а, Самара, Россия, 443086

## Аннотация

Развитие методов машинного обучения привело к тому, что способности компьютерных систем классифицировать изображения или распознавать объекты приближаются к способностям человека, а иногда превосходят их. Сегодня методы машинного обучения предоставляют множество инструментов для решения различных задач робототехники, таких как автономное управление движением, навигация. Так, обучение с подкреплением является основным методом машинного обучения, используемым в робототехнике.

В данной работе производится сравнение различных алгоритмов обучения с подкреплением. Также рассматриваются влияние различных параметров этих алгоритмов на результат обучения. Так как создание и настройка реальной робототехнической системы является длительным процессом, то первоначальная отладка робота может производиться при помощи соответствующего программного обеспечения, симулирующего реальную среду. В нашем случае был использован симулятор Gazebo. Используя его, можно проводить исследования, не имея реальной системы.

## Ключевые слова

Обучение с подкреплением, робототехника, моделирование, симуляция, ROS, Gazebo, Gym, Q-Learning, SARSA, DQN

## 1. Введение

Обучение с подкреплением – область машинного обучения, в которой испытываемая система, также называемая агентом, учится, взаимодействуя с окружающей средой, наблюдая за результатом взаимодействия для достижения максимально возможного награждения. Обучение с подкреплением предлагает основу и набор инструментов для проектирования сложных для разработки моделей поведения.

В обучении с подкреплением ключевую роль играет функция  $Q(s, a)$ , которая является оценкой действия  $a$ , в состоянии  $s$ . Эта функция необходима агенту для принятия решения. С другой стороны, агент постоянно изменяет эту функцию, в чём и заключается процесс обучения.

Целью данного исследования является сравнение эффективности различных алгоритмов обучения с подкреплением с использованием в качестве окружающей среды симулятора Gazebo.

## 2. Исследования

Для исследований в данной работе был использован набор инструментов [1], сочетающий в себе следующие компоненты:

- OpenAI Gym – популярный набор инструментов для исследования обучения с подкреплением [2].

- Gazebo – трёхмерный симулятор робототехники.
- ROS (Robot Operation System) – набор библиотек и инструментов, которые помогают создавать приложения для роботов, обеспечивая структурированный уровень взаимосвязи отдельных систем робота [3].

Исследование содержало следующие этапы:

- Разработка модели робототехнической системы. Наша система представляет собой достаточно простого робота, имеющего тело и две подвижных ноги.
- Формулировка задачи для робота, которая будет использоваться для обучения с подкреплением. Поощрение робота было равно количеству секунд, которое тело робота находится выше некоторой заданной высоты. В нашем случае эта высота была равна половине высоты робота в положении, когда он стоит на выпрямленных ногах. Цель робота состоит в управлении углом поворота каждой ноги.
- Проведение экспериментов с различными алгоритмами и параметрами.
- Анализ результатов.

### 3. Заключение

В результате данного исследования были сравнены эффективности наиболее распространенных методов обучения с подкреплением, а именно таких как Q-Learning, SARSA, Deep Q-Learning. Данное сравнение было выполнено в среде симулятора Gazebo. Все алгоритмы испытаны на симулируемом роботе под управлением промежуточного программного обеспечения ROS. Полученные результаты возможно перенести на реальную робототехническую систему, обучив её автономному движению в физической среде. При этом обучение возможно начать как с нуля, так и использовать предобученные модели. Использование предобученных в условиях симулятора моделей в реальных системах является важным преимуществом систем, подобных Gazebo.

### 4. Литература

- [1] Zamora, I. Extending the openai gym for robotics: a toolkit for reinforcement learning using ros and gazebo // arXiv preprint: 1608.05742. – 2016.
- [2] Brockman, G. Openai gym // arXiv preprint: 1606.01540. – 2016.
- [3] Quigley, M. ROS: an open-source Robot Operating System // ICRA workshop on open source software. – 2009. – Vol. 3(3.2). – P. 5.
- [4] Koenig, N. Design and use paradigms for gazebo, an open-source multi-robot simulator / N. Koenig, A. Howard // IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566). – 2004. – Vol. 3. – P. 2149-2154.
- [5] Sutton, R.S. Generalization in reinforcement learning: Successful examples using sparse coarse coding // Advances in neural information processing systems. – 1996. – P. 1038-1044.
- [6] Barto, A.G. Neuronlike adaptive elements that can solve difficult learning control problems / A.G. Barto, R.S. Sutton, C.W. Anderson // IEEE transactions on systems, man, and cybernetics. – 1983. – Vol. 5. – P. 834-846.
- [7] Kober, J. Reinforcement learning in robotics: A survey / J. Kober, J.A. Bagnell, J. Peters // The International Journal of Robotics Research. – 2013. – Vol. 32(11). – P. 1238-1274.
- [8] Geramifard, A. Rlpy: a value-function-based reinforcement learning framework for education and research // Journal of Machine Learning Research. – 2015. – Vol. 16. – P. 1573-1578.