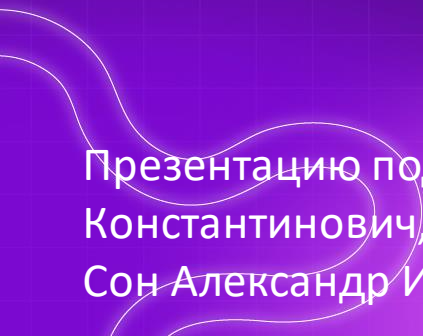




ІІТМО

Q-learning и обучение с подкреплением



Презентацию подготовили: Нечаева Анна Анатольевна, Иванов Александр Константинович, Велюго Кирилл Олегович, Воротников Андрей Андреевич, Сон Александр Игоревич из R3138

Обучение с подкреплением (RL)

Обучение с подкреплением (англ. *reinforcement learning*) — один из способов машинного обучения, в ходе которого испытываемая система (агент) обучается, взаимодействуя с некоторой средой

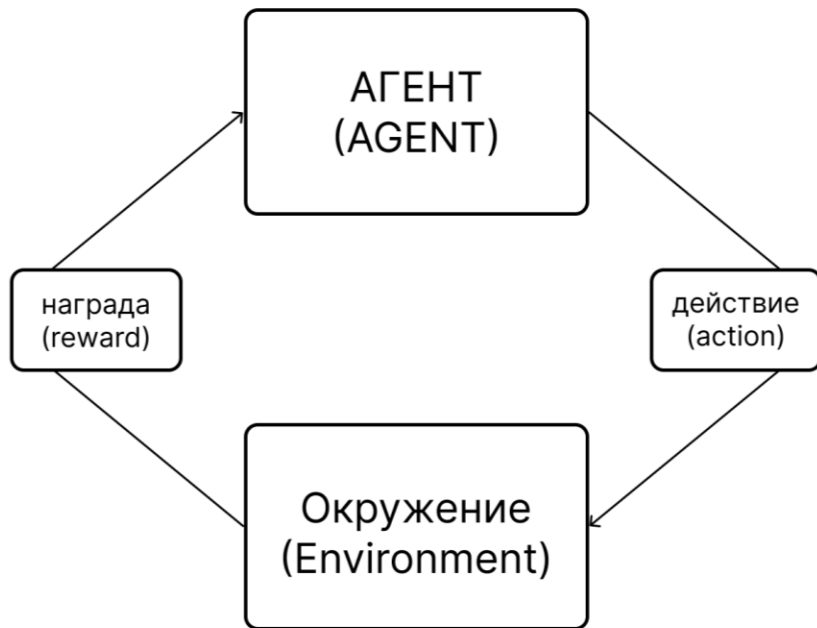


Обучение с подкреплением (RL)



- ✓ Агент знает конечную цель, но ему неизвестен алгоритм ее достижения
- ✓ Обучение происходит путем проб и ошибок
- ✓ Системе заранее неизвестны правильные действия
- ✓ Цель RL - обучить агента определенному поведению

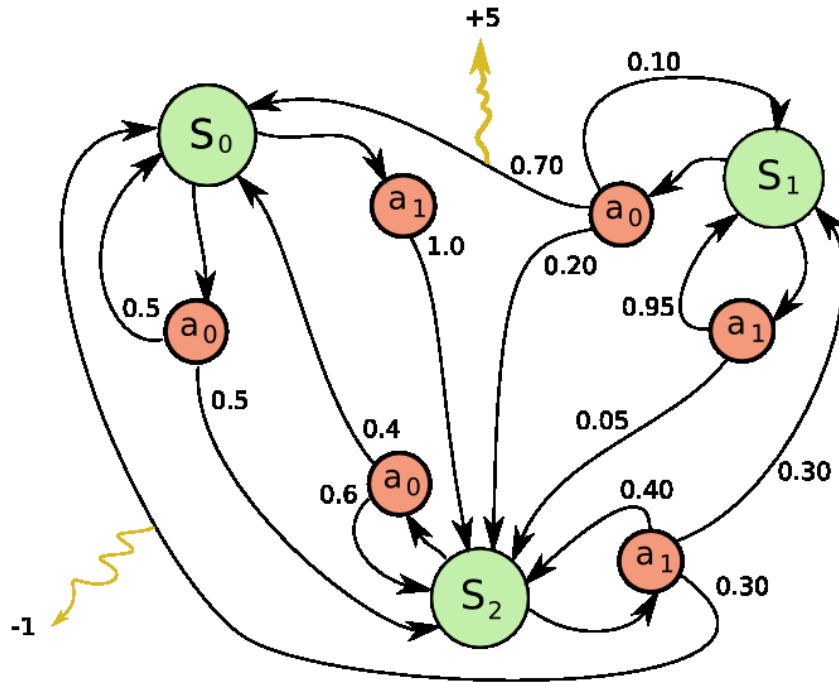
Как работает обучение с подкреплением ИТМО



Основные элементы RL системы:

- Агент (обучающийся)
- Окружающая среда, с которой взаимодействует Агент
- Действие Агента, которое следует заданной стратегии
- Награда, которую Агент получает в зависимости от успешности выполнения

Марковский процесс принятия решений ИТМО



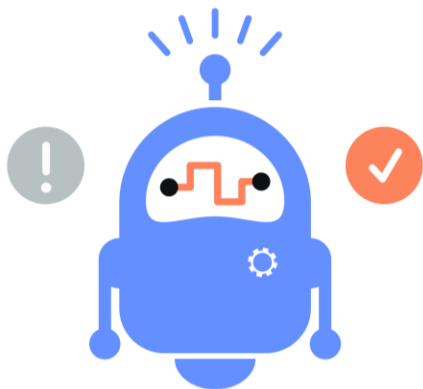
From: <https://en.wikipedia.org/>

Для определения МППР используется:

- S конечное множество состояний
- A конечное множество действий
- Вероятность $P_a(s, s') = \Pr(s_{t+1} = s' | s_t = s, a_t = a)$
- $R_a(s, s')$ вознаграждение

Для определения МППР задают 4-кортеж вида $(S, A, P(*, *), R(*, *))$

Как работает обучение с подкреплением ИТМО



Простейшая модель состоит из:

- Множества состояний окружения S
- Множества действий A
- Множество «выигрышей» r

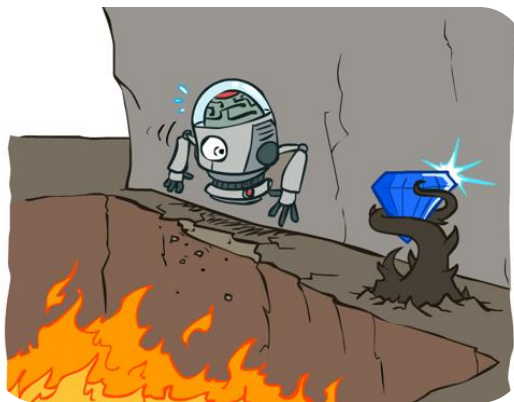
В произвольный момент времени t агент характеризуется состоянием $s_t \in S$ и множеством возможных действий $A(s_t)$

Выбирая действие $a \in A(s_t)$, он переходит в состояние s_{t+1} и получает выигрыш r_t

Основываясь на таком взаимодействии с окружающей средой, агент, обучающийся с подкреплением, должен выработать стратегию $\pi: S \rightarrow A$, которая максимизирует величину:

$$R = \sum \gamma^t r^t,$$

где $\gamma \in [0; 1]$. Величины могут быть различными.



Оценка эффективности стратегии



$$\sum_{i=1}^{\infty} r_i$$

$$\sum_{i=1}^t r_i$$

$$\sum_{i=1}^{\infty} \gamma^{1-i} r_i$$

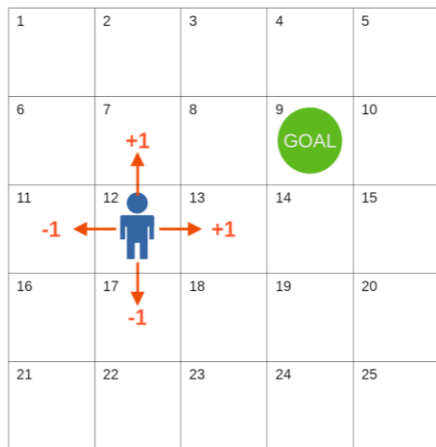
**Простое суммирование всех
наград**

+ за конечное время

+ с учетом дисконтирования

Задача Агента является максимизировать вознаграждение за минимальное время.
Вознаграждением считается сумма получаемых наград (чаще всего на практике
используются суммы наград с дисконтированием)

Q-Learning. Q-Table. Что это?



	↑	↓	←	→
1	-	+1	-	+1
2	-	+1	-1	+1
3	-	+1	-1	+1
4	-	+1	-1	-1
5	-	+1	+1	-
...				
23	+1	-	-1	+1
24	+1	-	-1	-1
25	+1	-	+1	-

From: <https://blog.spiceai.org/>

На основе получаемого от среды вознаграждения Агент формирует функцию полезности Q , что впоследствии дает ему возможность уже не случайно выбирать стратегию поведения, а учитывать опыт предыдущего взаимодействия со средой.

Можно построить таблицу со всевозможными парами состояний/действий для более эффективного выбора.

Преимущества и недостатки RL

Преимущества

- ✓ Фокусируется на проблеме в целом
- ✓ Не требует специального этапа сбора данных
- ✓ Способен подстраиваться под динамичные, новые среды

Недостатки

- ✓ Большое количество времени на обучение
- ✓ Отсутствие возможности повлиять на принимаемые агентом решения

Кто занимается этим в ИТМО



Ведяков Алексей Алексеевич

Кандидат технических наук,
ассистент кафедры систем
управления и информатики



Евстафьев Олег Александрович

Ассистент, факультет систем
управления и робототехники

Преподает дисциплины: «Глубокое
обучение», «Прикладной
искусственный интеллект»



Асадулаев Арип Амирханович

Сотрудник Международной
лаборатории «Компьютерные
технологии» ИТМО

Ведет курс «Обучение
с подкреплением»

Применение RL в задачах робототехники ИТМО

Self-Inspection Method of Unmanned Aerial Vehicles in Power Plants Using Deep Q-Network Reinforcement Learning



Figure 7. A path navigated using the converged model in simulated space

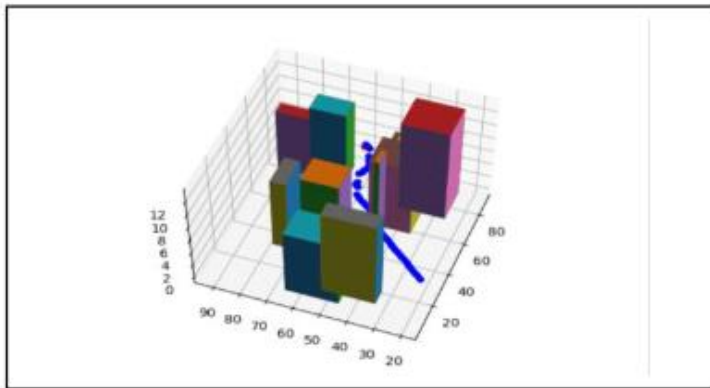
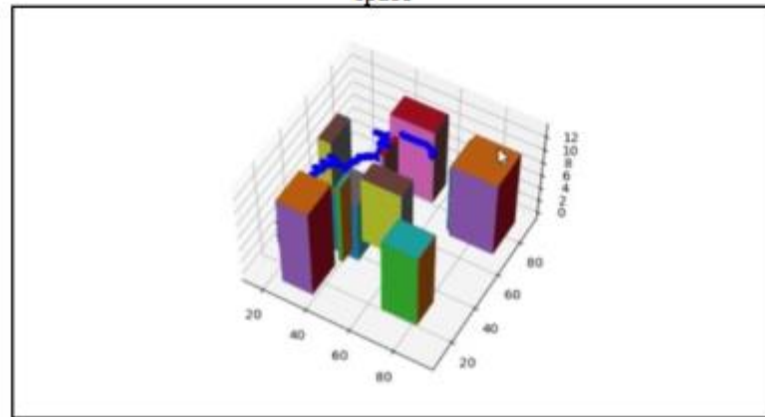


Figure 6. A path navigated using the pre-trained model in simulated space



Применение RL в задачах робототехники ИТМО

Исследование алгоритмов повышающих качество обучения с подкреплением в симуляции для использования на реальных робототехнических системах



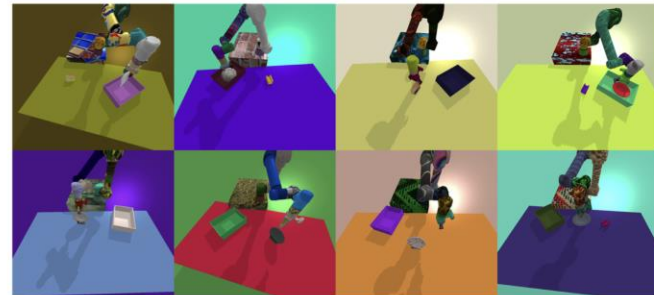
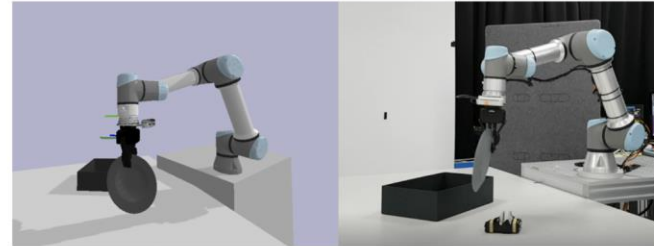
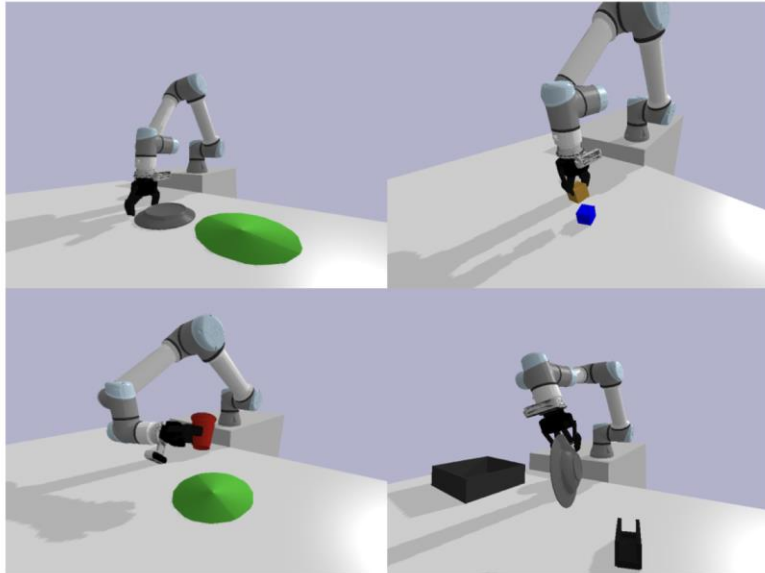
Применение RL в задачах робототехники ИТМО

Reinforcement Learning based Autonomous Multi-Rotor Landing on Moving Platforms



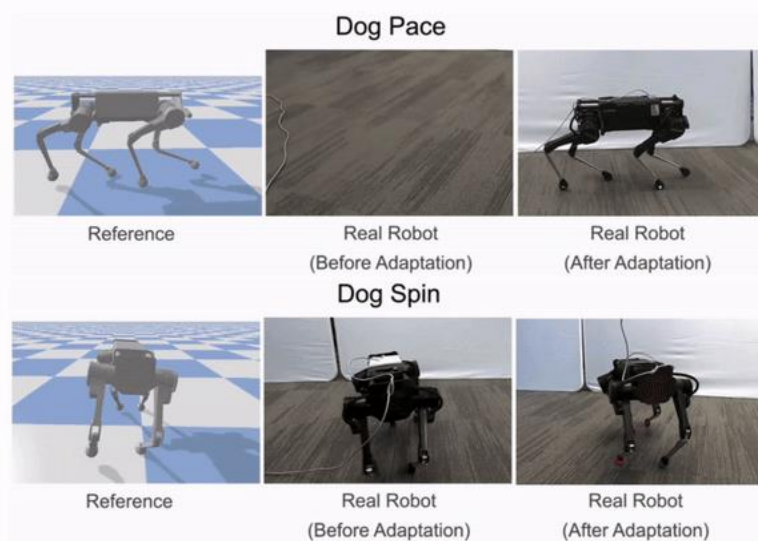
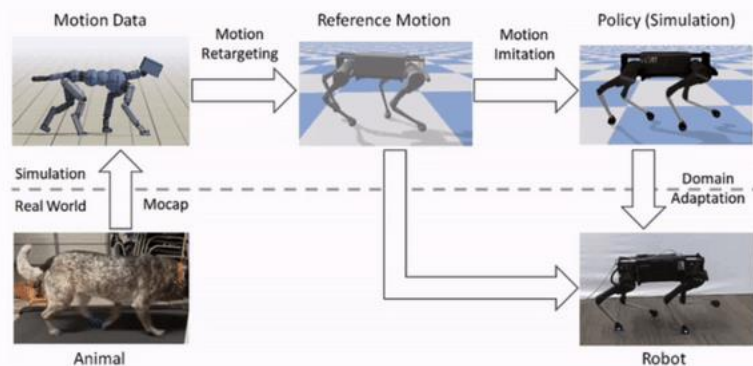
Применение RL в задачах робототехники ИТМО

Reinforcement Learning for Vision-based Object Manipulation with Non-parametric Policy and Action Primitives



Применение RL в задачах робототехники ИТМО

Robots Learning to Move like Animals



1. Т. Ю. Ким. Применение алгоритма DDPG обучения с подкреплением для мобильного робота
2. Haoran Guan. Self-Inspection Method of Unmanned Aerial Vehicles in Power Plants Using Deep Q-Network Reinforcement Learning
3. Dongwon Son*, Myungsin Kim, Jaecheol Sim, and Wonsik Shin. Reinforcement Learning for Vision-based Object Manipulation with Non-parametric Policy and Action Primitives
4. Pascal Goldschmid and Aamir Ahmad. Reinforcement Learning based Autonomous Multi-Rotor Landing on Moving Platforms
5. Труфанова А.А., Симонов Р.А., Симонов Н.А. (науч. рук. Ведяков А.А.). Исследование алгоритмов повышающих качество обучения с подкреплением в симуляции для использования на реальных робототехнических системах
6. Xue Bin (Jason) Peng. Robots Learning to Move like Animals
7. Corentin Risselin. Understanding Q-learning: How a Reward Is All You Need

Спасибо
за внимание!

it's **MO** *re than a*
UNIVERSITY

бу люди, честно
разбирающиеся
в этой теме