

Как казаки retro контекст решали

OpenAI Retro Contest 4th place

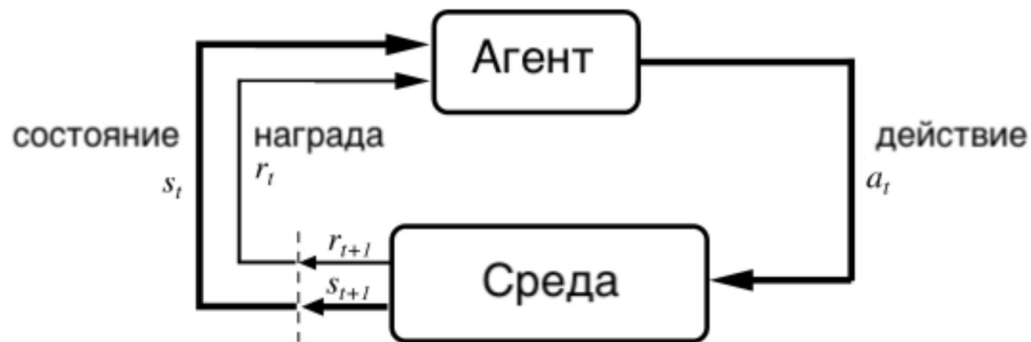
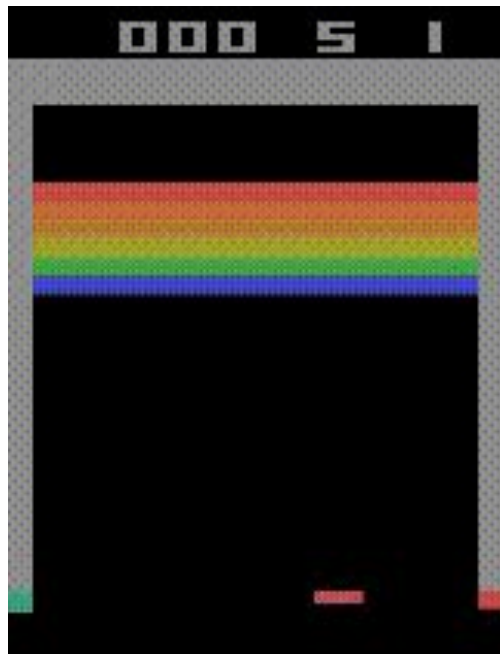
Колесников Сергей
Senior Data Scientist @ Dbrain

План доклада



- Intro в обучение с подкреплением
- Про задачу
- Baselines
- Особенности соревнования
- Подходы и решение
- Bloopers
- Итоги
- Послесловие

Intro в обучение с подкреплением

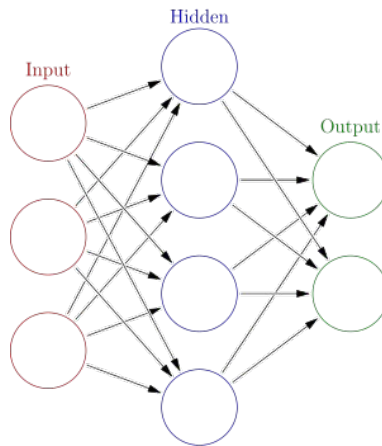


Про задачу

RGB картинка
(s_t)



CNN



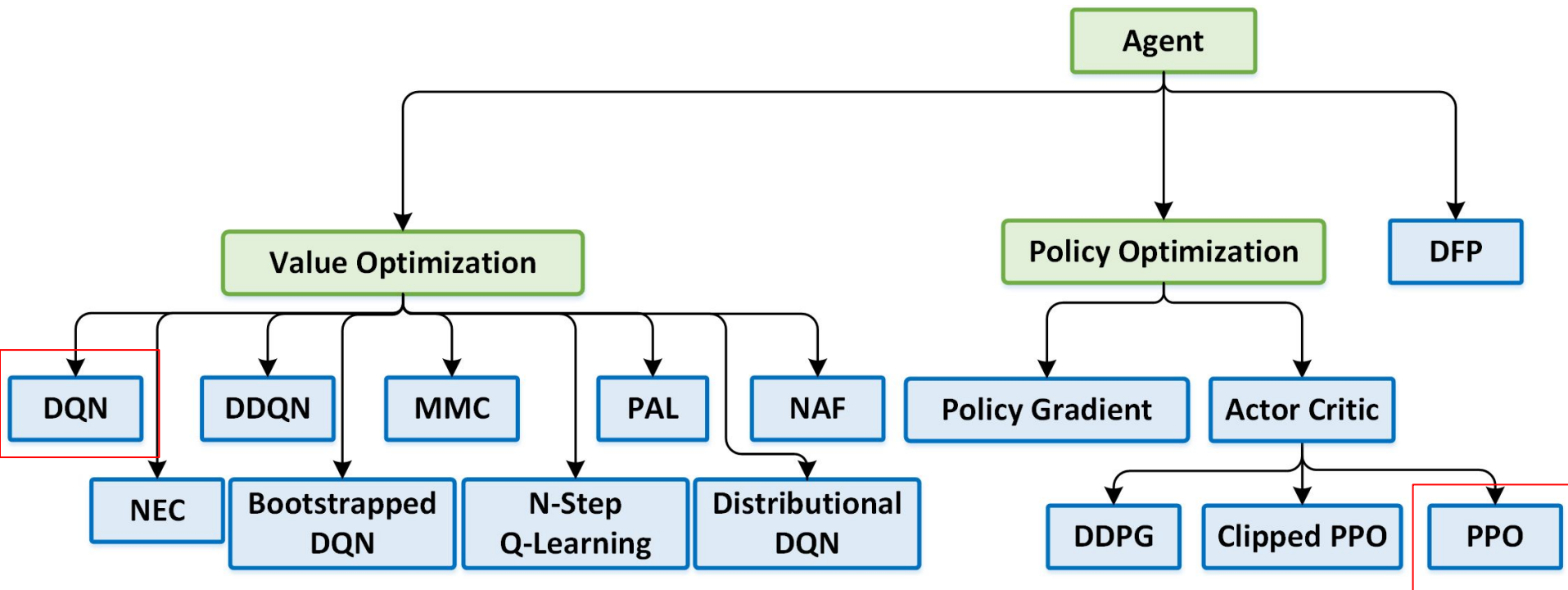
Вероятности действий
(a_t)



Награда
(r_t)

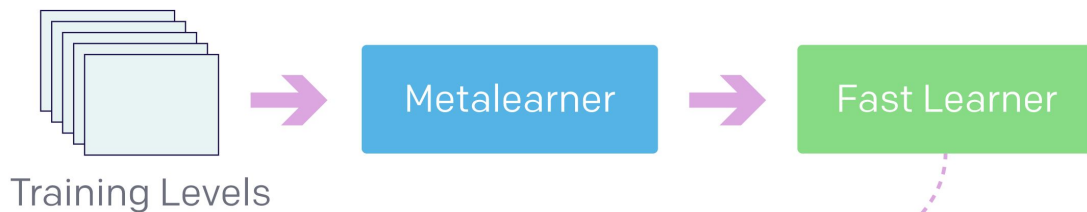
Новая
картинка
(s_tp1)

Baselines

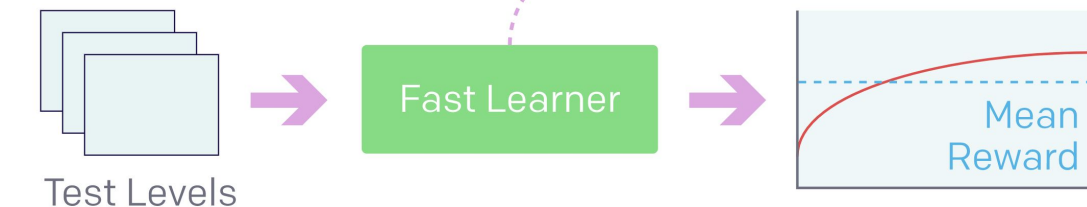


Особенности соревнования

Train on your hardware



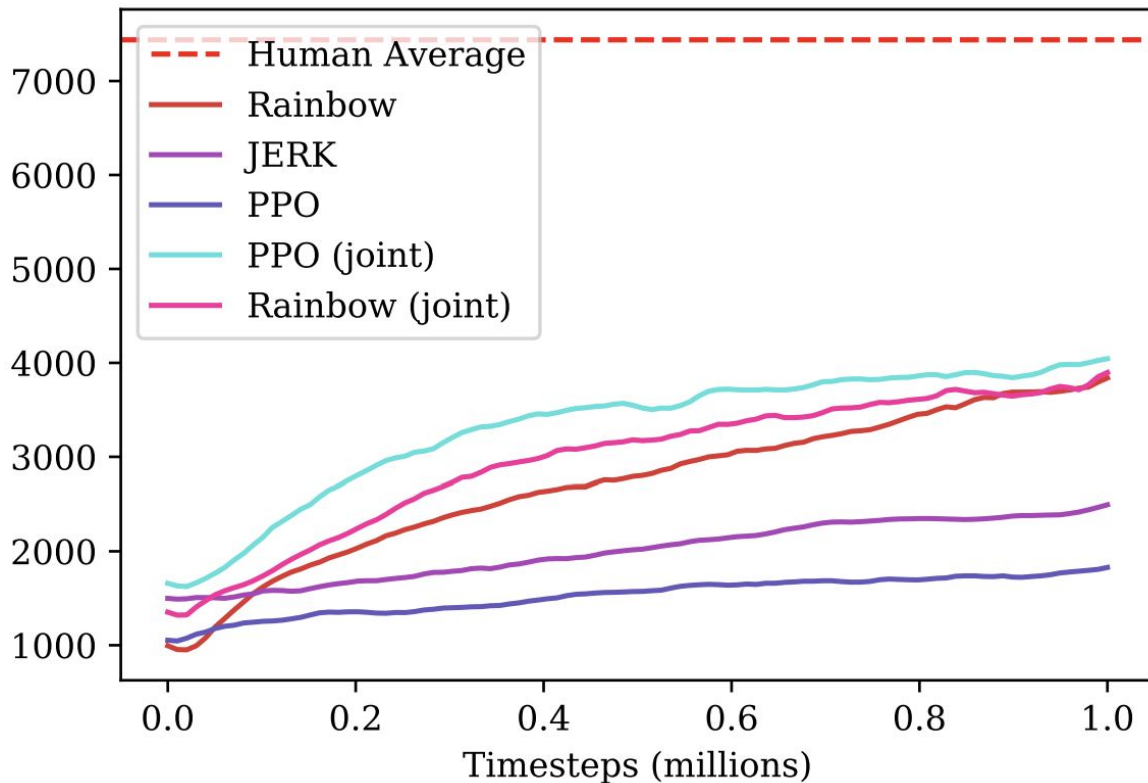
Test on our hardware



1. Обучаем агента локально на доступных уровнях.
 2. Запускаем в Docker контейнер с учетом API и отправляем.
 3. Тестирование алгоритма проводилось на **новых** уровнях Sonic'a.
 4. Финальная метрика - средний reward.
-
1. На тестовых уровнях можно было обучаться "вслепую".
 2. Были ограничения по ресурсам и времени.

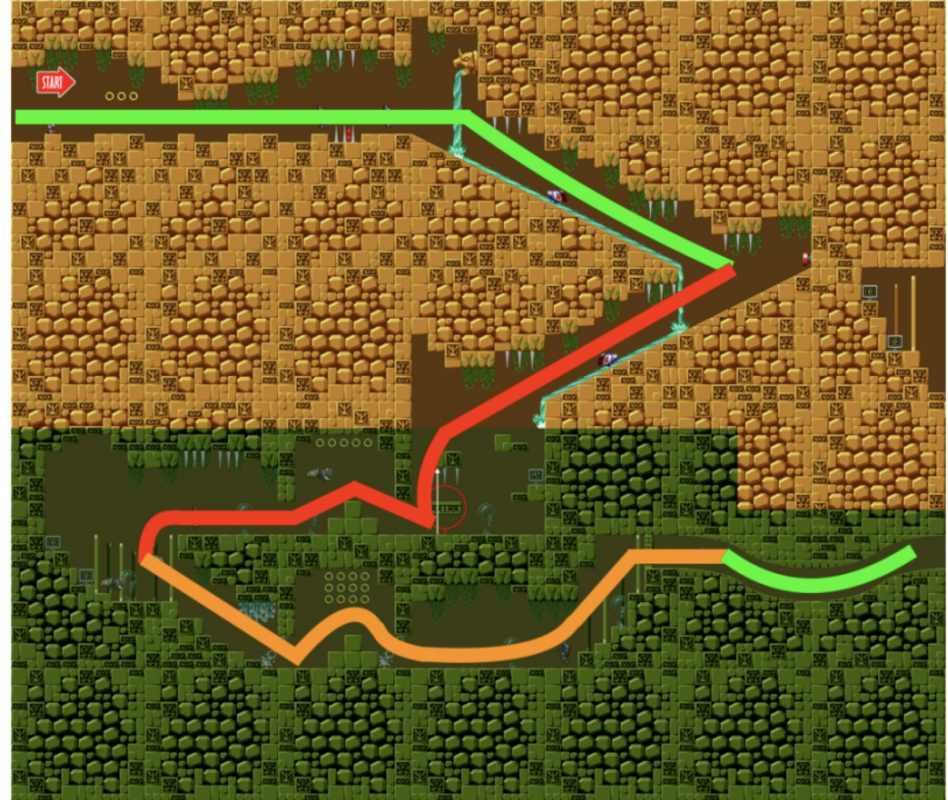
Подходы и решение

1. Совместное обучение PPO на всех доступных уровнях
2. Дообучение в процессе тестирования



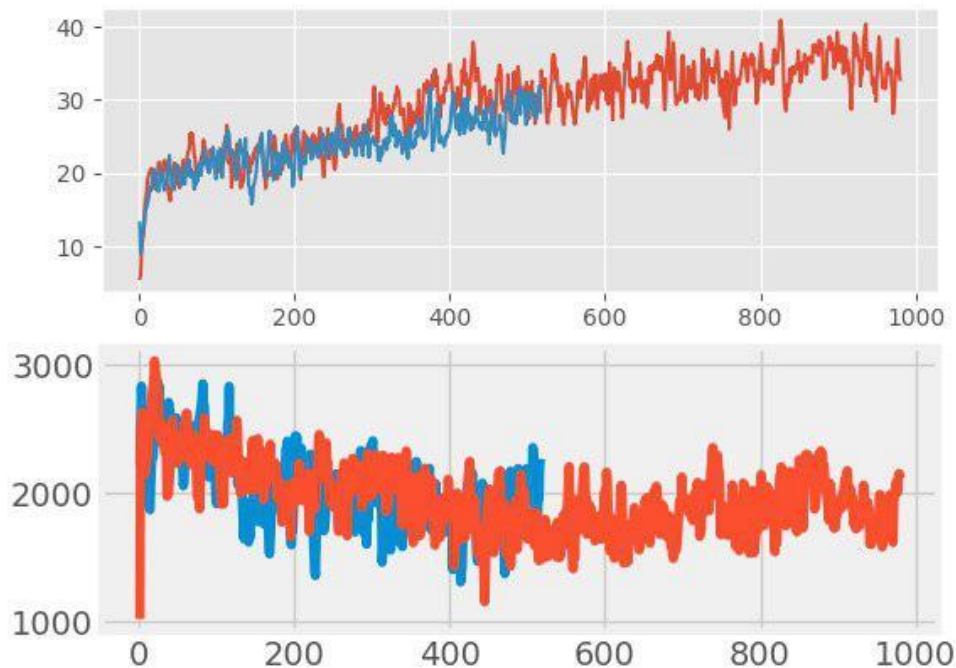
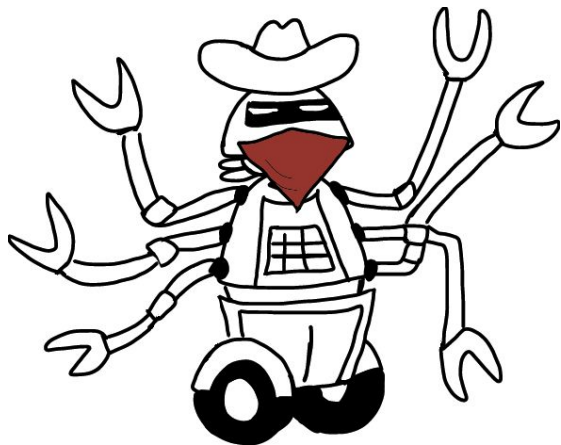
Подходы и решение

3. Exploration bonus to reward based on observation and x coordinate



Подходы и решение

- 4. Эксперименты с микс
- 5. Подбор лучшей начальной политики



Bloopers

1. Изменение архитектуры сети:
 - a. [SELU activation](#)
 - b. Self-attention
 - c. [SE blocks](#)
2. [Neuroevolution](#)
3. Создание своих уровней Sonic
4. Мета-обучение:
 - a. [MAML](#)
 - b. [REPTILE](#)
5. Ансамблирование нескольких моделей и дообучение с importance sampling.



ИТОГИ

RANK	TEAM	SCORE
#1	Dharmaraja	4692
#2	mistake	4446
#3	aborg	4430
#4	whatever	4274
#5	Students of Plato	4269
	Joint PPO baseline	4070
	Joint Rainbow baseline	3843
	Rainbow baseline	3498

Основные отличительные моменты, которые “полетели” у первой 3ки:

1. Доработанная система действий: придумали свои кнопки, убрали лишние
2. Исследование состояний через hash от входной картинки
3. Больше количество обучающих уровней

Послесловие

Code & contacts – https://github.com/fgvbrt/retro_contest

Papers – <https://github.com/Scitator/papers>

Text version – <https://habr.com/company/ods/blog/421585/>

AnyRL – <https://github.com/unixpickle/anyrl-py>

OpenAI baselines – <https://github.com/openai/baselines>



Спасибо за внимание!