# Reinforcement Learning

DataFest 2020

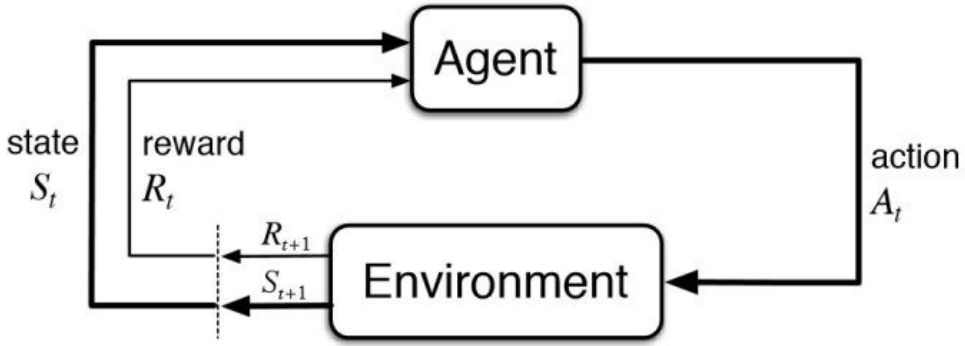
# Tl;dr - track organizer

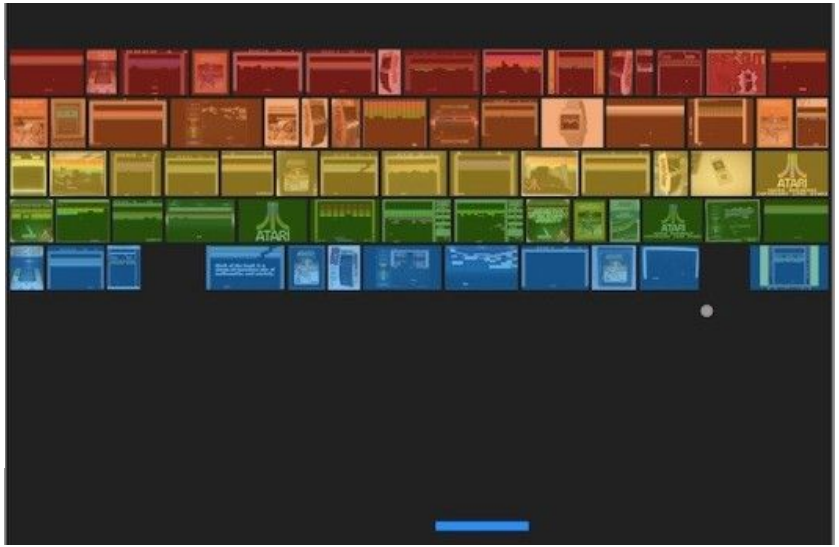
**Scitator**

scitator.com

- Sergey Kolesnikov
- MSc in Math and Computer Science, 2018
- Interested in sequential decision making and lifelong learning
- RnD Lead @Tinkoff
- Researcher @MIPT
- Lead @Catalyst-Team
- NeurIPS RL competition winner
  - 3rd place in 2017 and 2018
  - 2nd place in 2019
  - Working on 2020 one :)

# Tl;dr - Reinforcement Learning

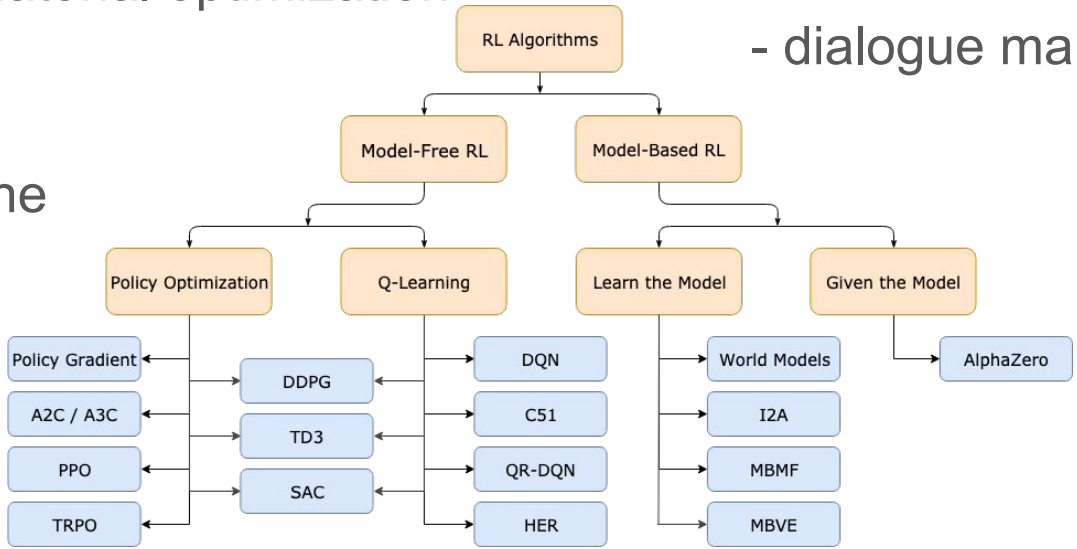


# Tl;dr - Reinforcement Learning

- games
- robotics
- combinatorial optimization
- trading
- medicine

- recommendations, personalization
- text summarization, image captioning
- dialogue management



# Real Pendulum: обучи своего DIY робота с помощью RL

Антон Печенко (parilo)

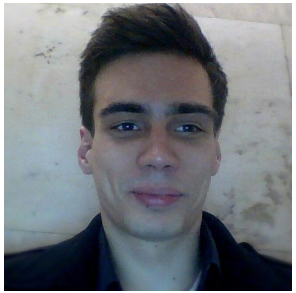
Робототехника - это область, которая становится все более распространенной и интересной. Применение искусственного интеллекта делает это направление еще более перспективным, поскольку позволяет серьезно снизить количество труда на программирование робота, одновременно давая возможность решать гораздо более сложные задачи.

Вместо того, чтобы тратить время и силы на программирование и отладку деталей отдельно взятой задачи не будет ли проще и удобнее предоставить роботу возможность самому научиться решать эту задачу методом проб и ошибок? Предлагаю вам свой вариант ответа на этот вопрос на примере игрушечной задачки: обратный маятник.

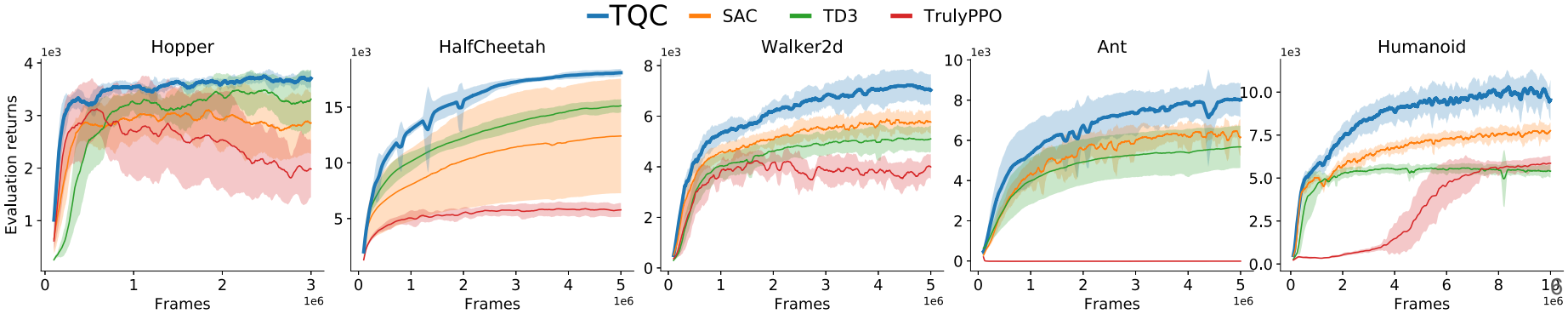
# Controlling Overestimation Bias

## with **Truncated Mixture** of **Continuous Distributional Quantile Critics (ICML 2020)**

Arsenii Kuznetsov  Pavel Shvechikov  **Alexander Grishin (alexgri)**  Dmitry Vetrov



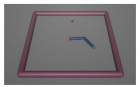
- Novel way to alleviate the overestimation bias in a continuous control
- Distributional RL + truncation of critics prediction + ensembling of critics
- Outperforms SOTA on MuJoCo, 2x speed boost to Humanoid robot

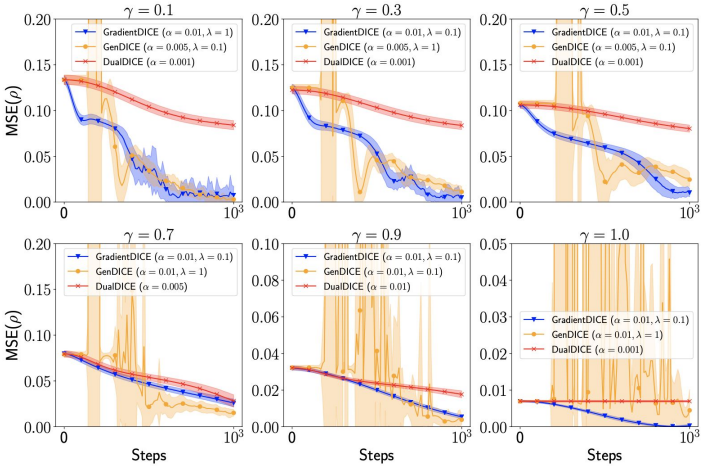


# Off-policy Evaluation with GradientDICE

Shangtong Zhang (alpha.rl)

In this talk, I will present my recent work on off-policy evaluation, where we want to estimate the performance of a policy with only a given dataset without executing the policy. Off-policy evaluation has broad real world applications such as recommendation systems. I will start with a brief introduction to reinforcement learning and discuss main challenges in Off-policy evaluation. Then I will present our work GradientDICE at ICML 2020 and discuss how and why it is better than previous methods like DualDICE and GenDICE, both theoretically and empirically.



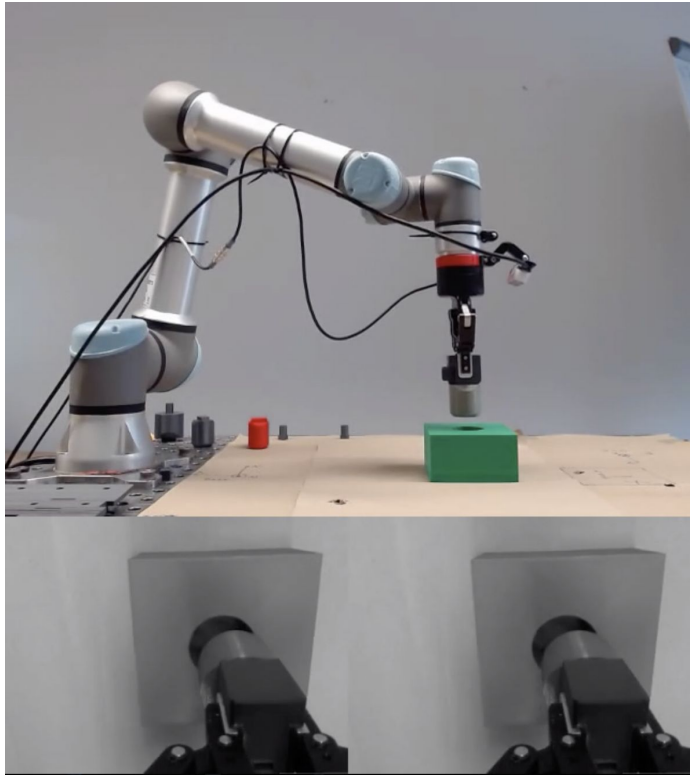


# RL in Industrial Robotics

No, in Arrival Robotics we are not using RL in production yet, but we could be pretty close. In this talk I will make an overview of the state of RL in industrial robotics, where and why could it be useful, including unexpected cases like solving NP-hard optimization problems.

I will show a real case of training a real robot to perform a smooth assembly operation, and discuss most promising areas of research (in my subjective opinion) to bring us closer to basic income economy!

Fedor Chervinskii (fedor.chervinskii)



# RL for the adaptive speed regulation on a metallurgical pickling line

Boris Voskresenskii
Kseniia Kingsep
Anna Bogomolova

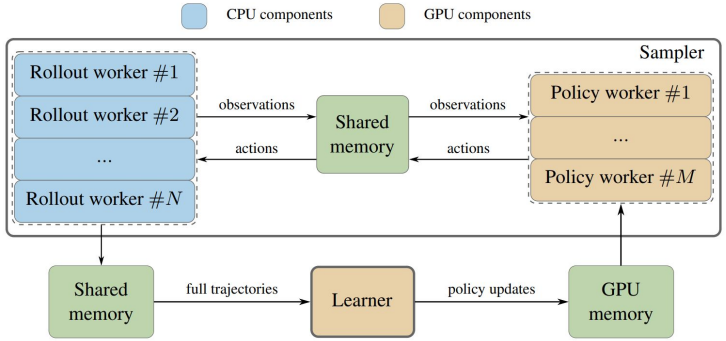
RL controls the speed of the continuous pickling line (NTA-3) at Cherepovets Steel Mill. Previously, the control speed was set once for each roll manually.

A mathematical model controls the speed of the unit, taking into account in real time about a hundred parameters including the length, width and thickness of the roll, the steel grade, the temperature of the metal and many others. In the end of March Severstal improved this model with a RL module. Technological processes of the pickling line heavily depend on the parameters of steel strips going through the line. Our RL based agent uses steel strip parameters synthesized by generative adversarial network (GAN). Today, a mathematical model and the RL based agent work cooperatively in the real-time producing more than 5% additional steel on this unit.

# SampleFactory and high-throughput reinforcement learning

Aleksei Petrenko (a.petrenko)


The quest for sample efficiency in general-purpose RL algorithms has proven to be rather challenging. The level results in RL has been growing largely due to the increased amount of compute research labs are willing to use in their projects. As a result, SOTA-level results have become increasingly unreachable for regular researchers.




Our goal is to bring the deep RL back to the community by improving the efficiency of training and reducing the cost of data collection. We present the SampleFactory - an on-policy RL training system optimized for speed. By maximizing the hardware utilization of our algorithm we approach 150000 FPS of training on a single machine, 10x faster than many popular frameworks. Our agents trained with SampleFactory APPO approach human level of performance in challenging and immersive 3D games.

# Reinforcement Learning

DataFest 2020

#reinforcement_learning

@Scitator