Incentivizing Exploration in Curiosity-driven Deep Reinforcement Learning

Author: Reizinger Patrik

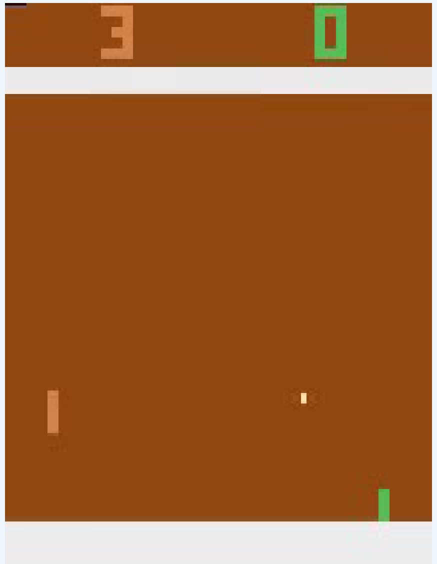
Supervisor: Márton Szemenyei

BME Faculty of Electrical Engineering and Informatics

12.07.2019

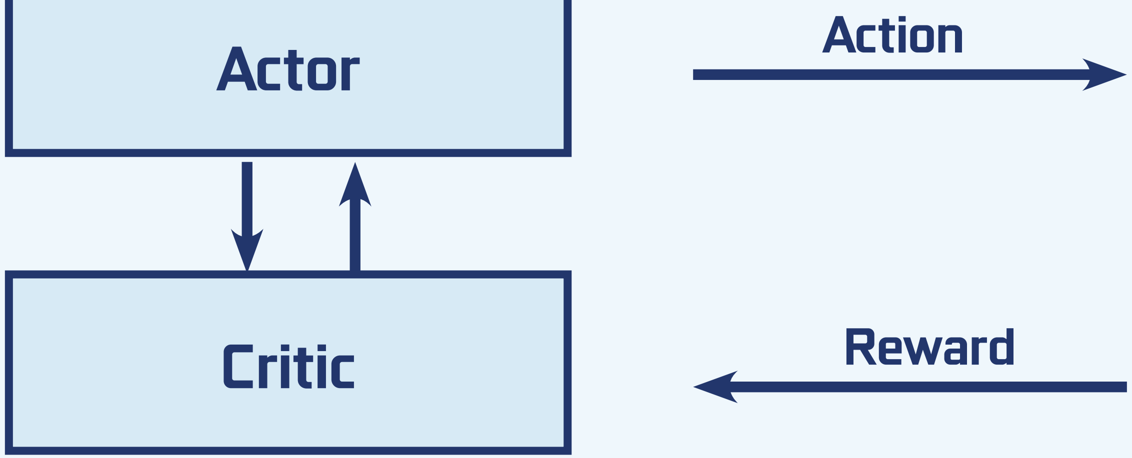
Basic model

Benchmarks & challenges



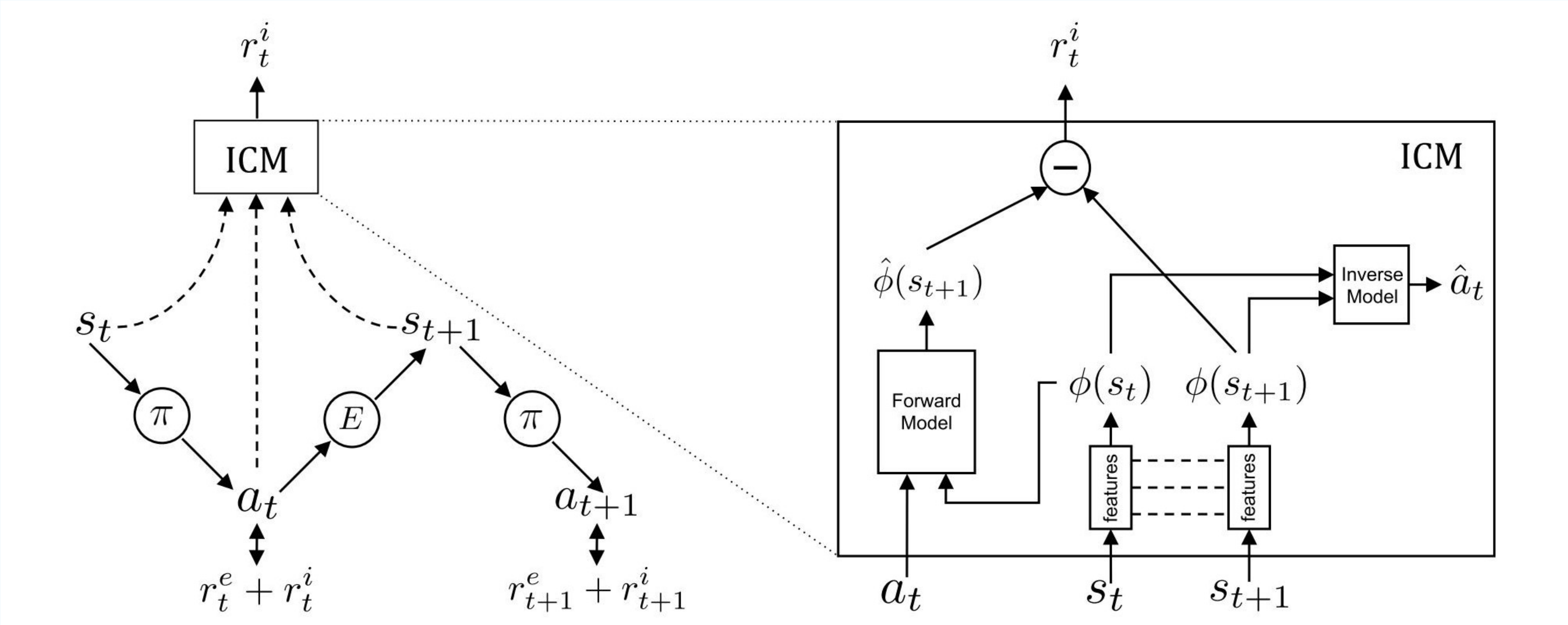
- Exploitation: use the current best
- Exploration: try to discover better options
- Sparse rewards

Advantage Actor Critic (A2C)

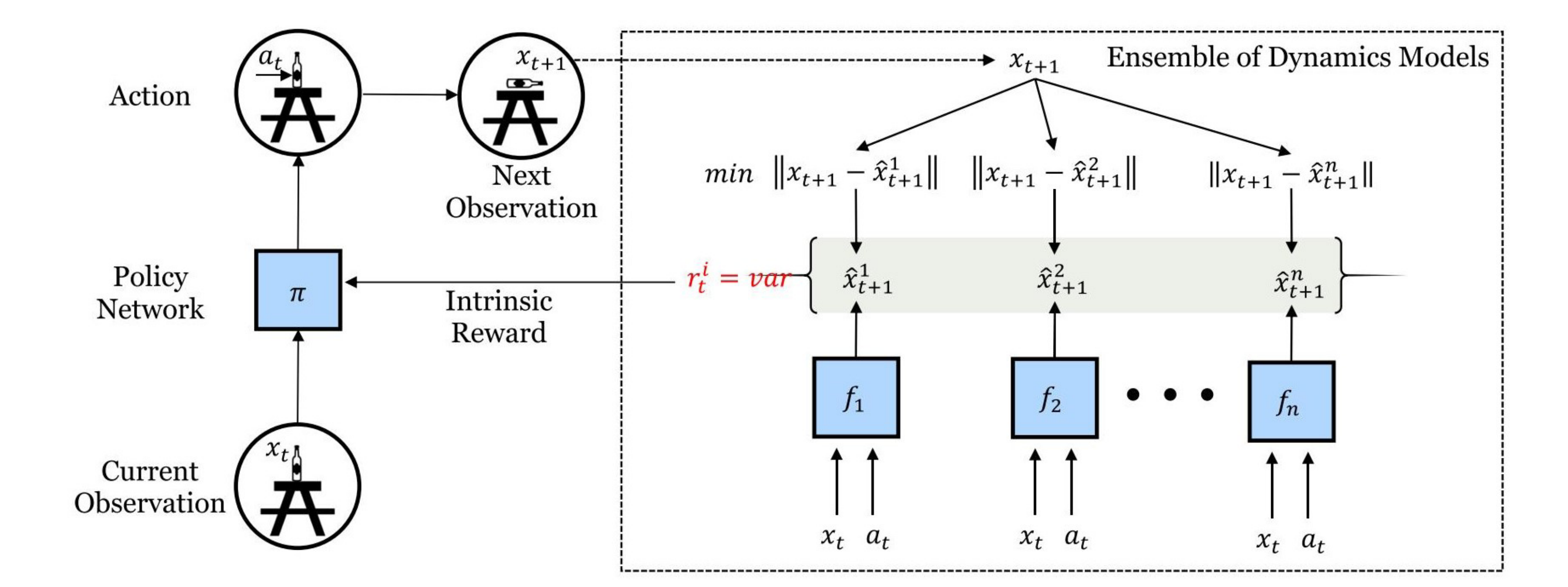


- Actor: learns the optimal policy
- Critic: learns the state value function
- Advantage: improves result with learning relative change

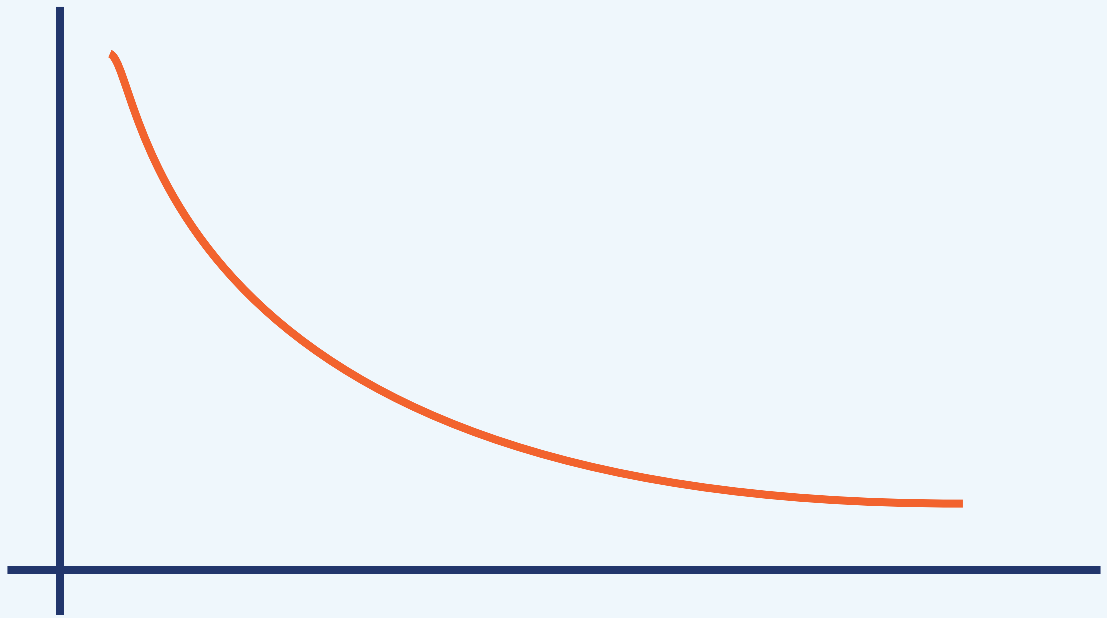
Introducing curiosity (ICM)



Disagreement-based curiosity

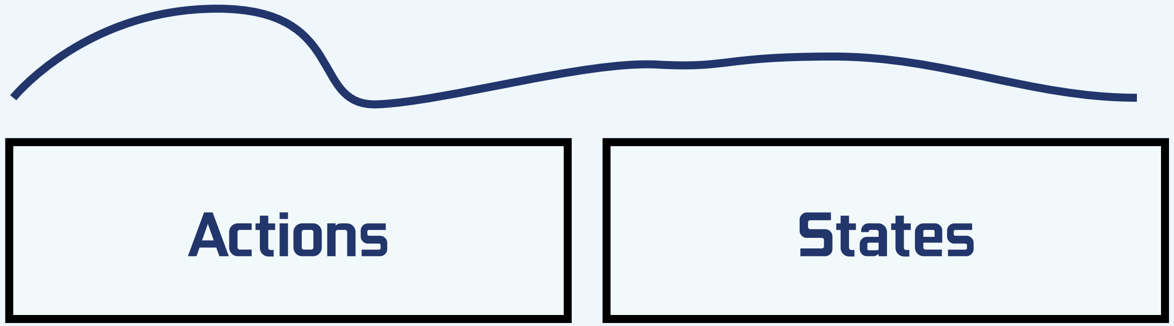


Enforcing exploration in feature space



- Idea: similar as learning rate scheduling
- Decaying weight of incentivizing bad predictions (like epsilon-greedy policy)

Focusing with Attention



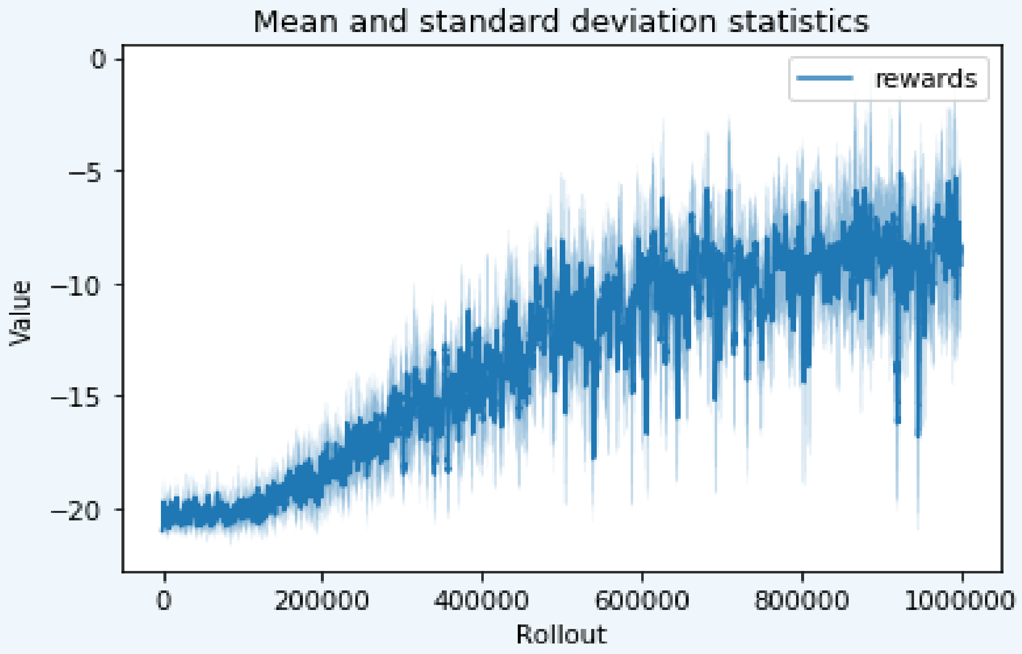
- Self-induced deadlock: TV with re-more control
- Expectation: Attention helps leaving the deadlock

Experiments

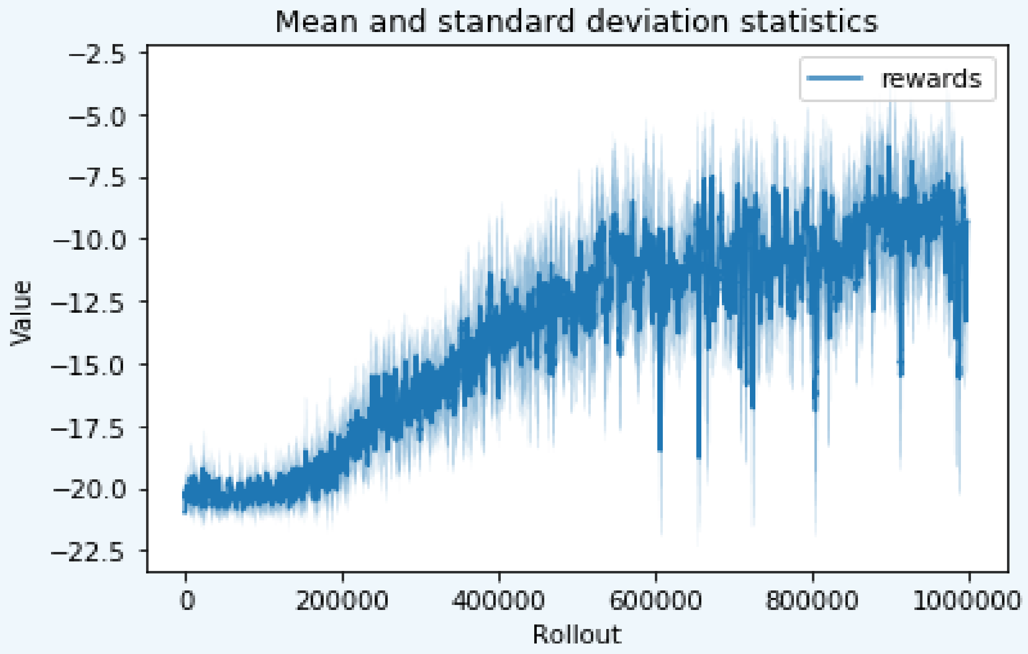
- 1 Titan X GPU
- More than 1 day is needed for each training process
- Currently extrinsic reward is also used
- Deterministic&stochastic (i.e. action repeat) were considered

Results - rewards

ICM

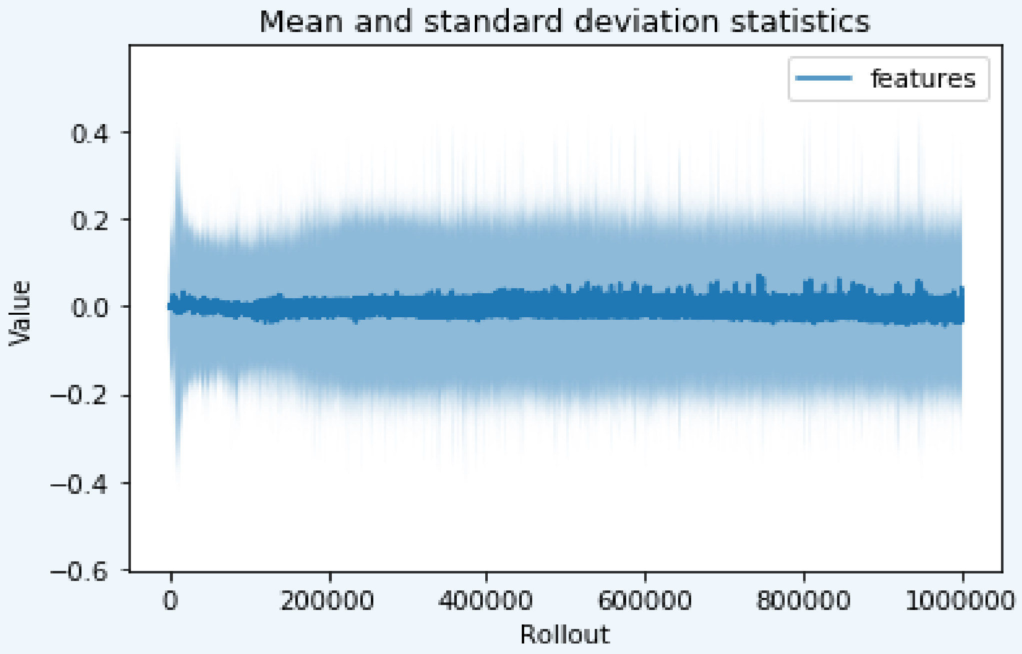


ICM+exploration enforcement

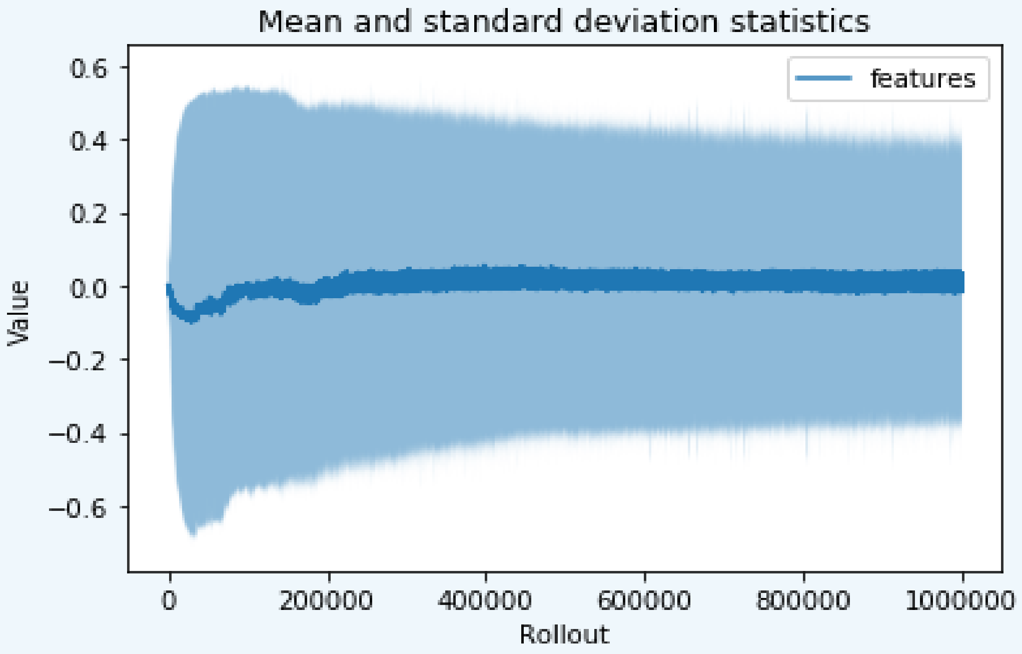


Results – features

ICM

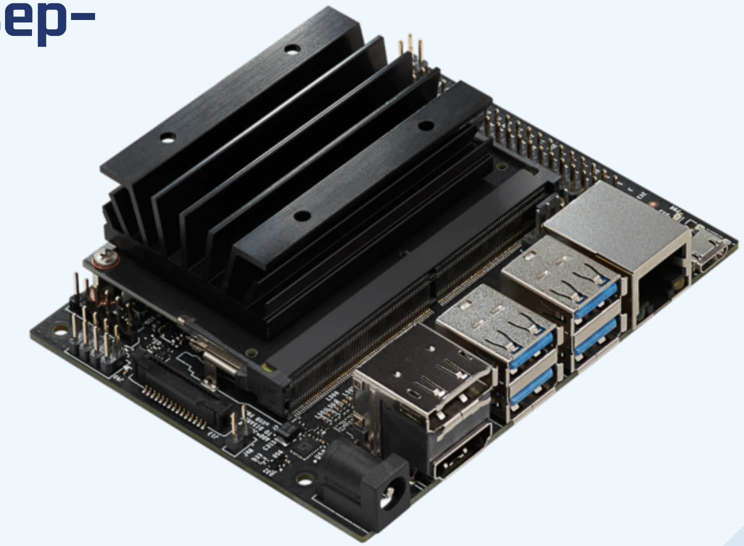


ICM+exploration enforcement



Conclusion

- Bigger spread of the features
- Evaluation without extrinsic reward needed
- Attention: action and feature space can be treated separately



https://github.com/rpatrik96

Incentivizing Exploration in Curiosity-driven Deep Reinforcement Learning

Author: Reizinger Patrik

Supervisor: Márton Szemenyei

12.07.2019