Intrinsically Motivated Exploration for Reinforcement Learning in Robotics

University of Hamburg
Faculty of Mathematics, Informatics and Natural Science
Department of Informatics
Technical Aspects of Multimodal Systems

Outline

- Intro to RL: successes and problems
- Directed exploration and why RL in Robotics needs it
- Three recent approaches:
    1. Intrinsic Curiosity Module (ICM)
    2. Random Network Distillation (RND)
    3. Episodic Curiosity Through Reachability (EC)
- Discussion

Reinforcement Learning - Introduction

- Algorithms maximize discounted cumulative reward
- Exploration essential, usually used: epsilon-greedy

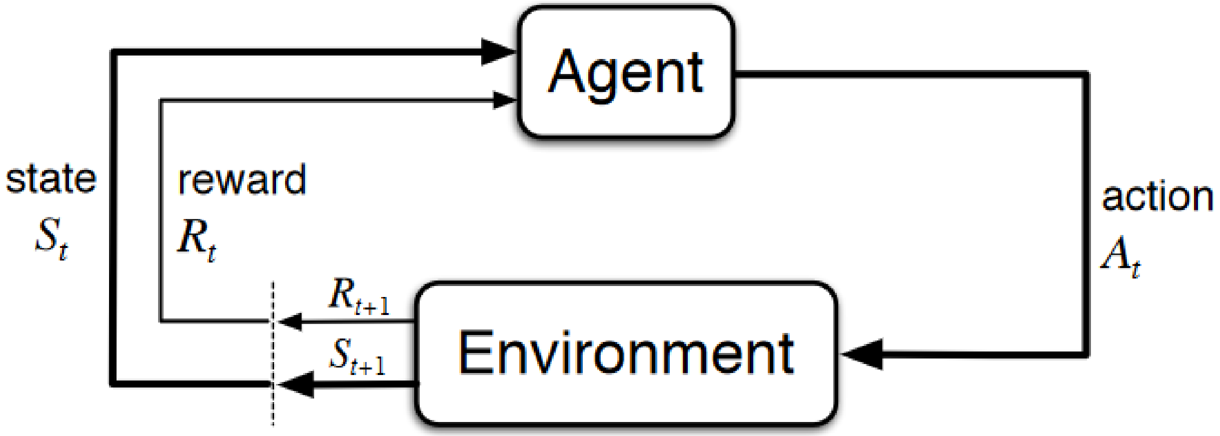


Figure 3.1: The agent–environment interaction in a Markov decision process.

From: "Reinforcement Learning: An Introduction" by Sutton and Barto [1]

RL – Successes

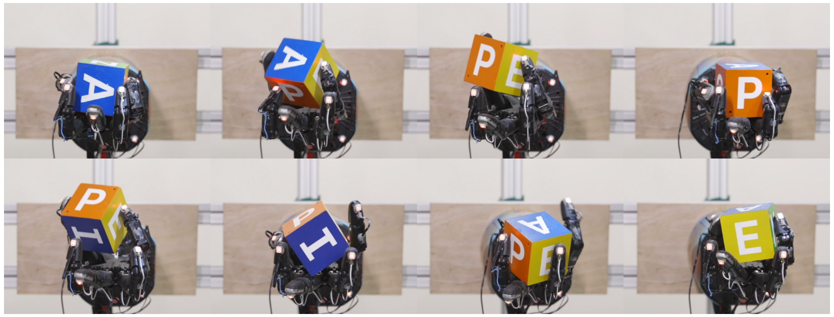
AlphaGo

https://foreignpolicy.com/2016/03/18/china-go-chess-west-east-technology-artificial-intelligence-google/

OpenAI Five

https://blog.openai.com/openai-five/

Learning Dexterous In-Hand Manipulation



From the whitepaper [4]

RL – Limitations

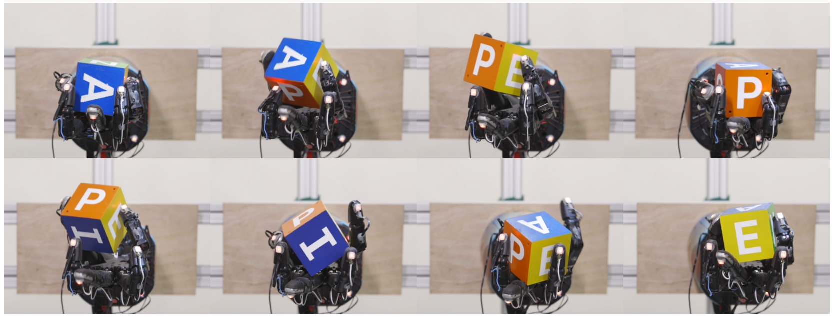
World known + Self-play

https://foreignpolicy.com/2016/03/18/china-go-chess-west-east-technology-artificial-intelligence-google/

Self-play + Simulation

https://blog.openai.com/openai-five/

Domain Randomization + Simulation



From the whitepaper [4]

RL – Problems in Robotics

Ideally learn without simulation but:

https://blog.openai.com/faulty-reward-functions/

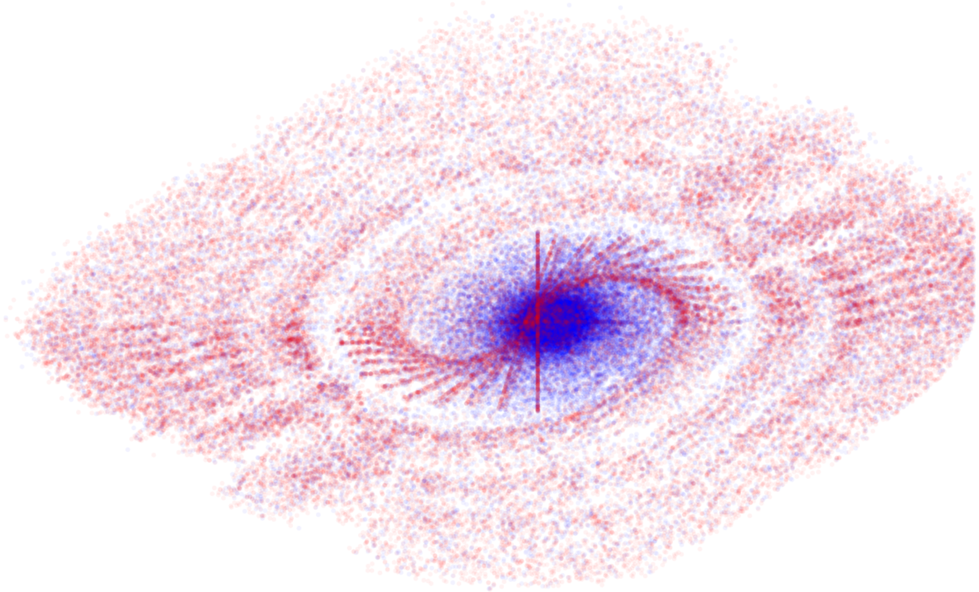
- Sparse Rewards:
  - necessary, but difficult to reach

- Sample Efficiency:
  - hardware limits

Directed Exploration - Introduction

- Helps with sparse rewards
- Makes exploration efficient

In general:
Total reward =
intrinsic + extrinsic reward



Comparison of TRPO+VIME (red) and TRPO (blue) on MountainCar: visited states until convergence. Source: “VIME: Variational Information Maximizing Exploration” [2]

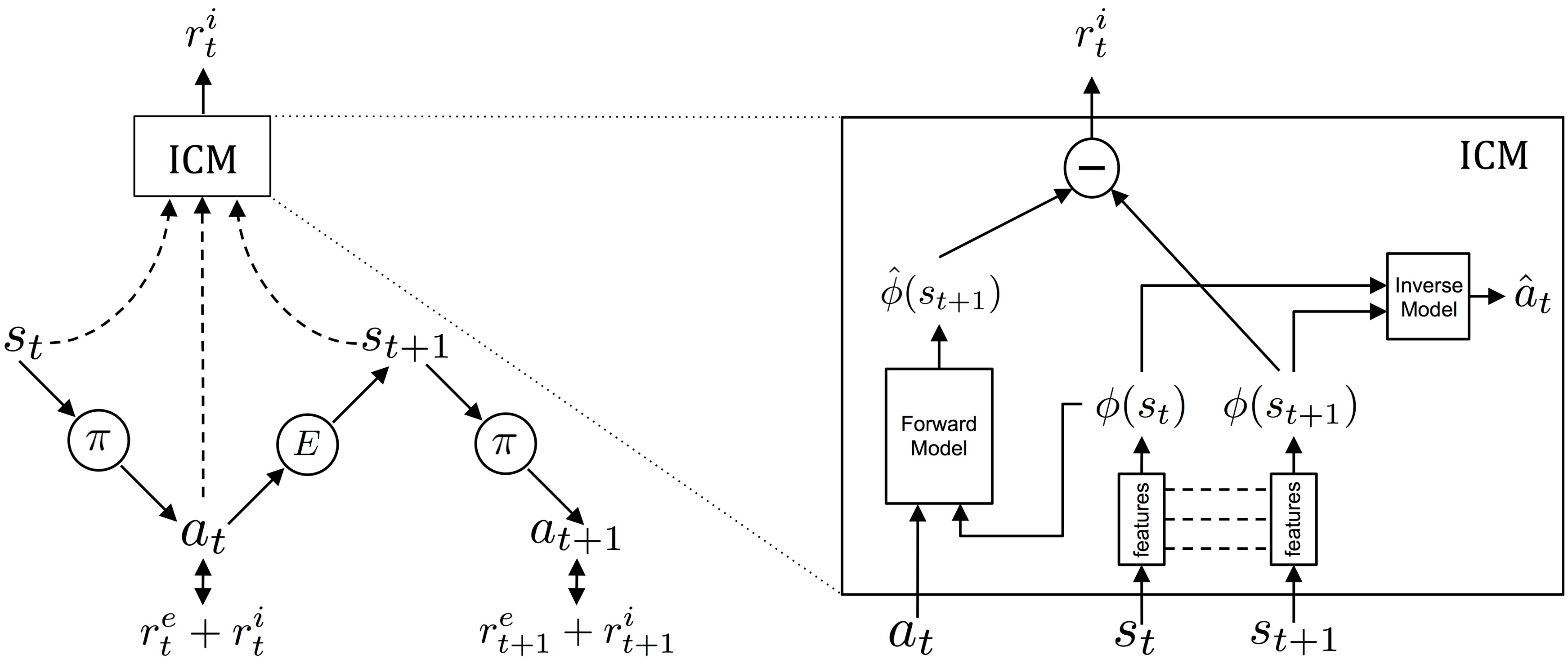
Environment → extrinsic reward

Exploration Algorithm → intrinsic reward

Intrinsic Curiosity Module (ICM) - Overview

- Train world model:
  - Predicts next state from current state
- Magnitude of prediction error of this model = intrinsic reward
- World model predicts relevant features
  - Use features that are necessary for inverse dynamics

Intrinsic Curiosity Module (ICM) - Details



Intrinsic Curiosity Module (ICM) - Demo

See: https://pathak22.github.io/noreward-rl/

Intrinsic Curiosity Module (ICM) - Problems

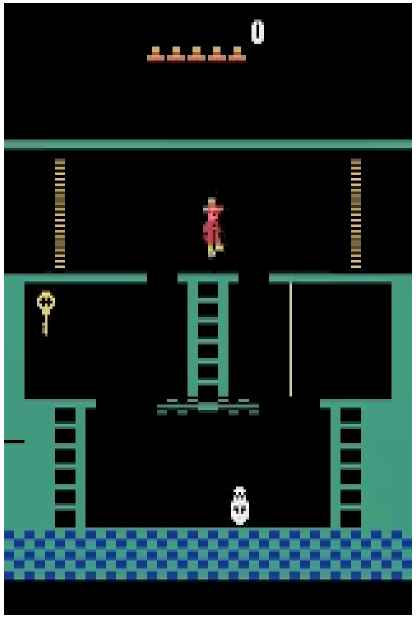
- Four factors that influence predictability of next states:

1) States similar to next state not yet encountered often
2) Stochastic environment
3) World model is too weak
4) Partial observability

- Only first one is a desired source of unpredictability

Intrinsic Curiosity Module (ICM) – Problems

“Montezuma’s Revenge” is a difficult atari game:

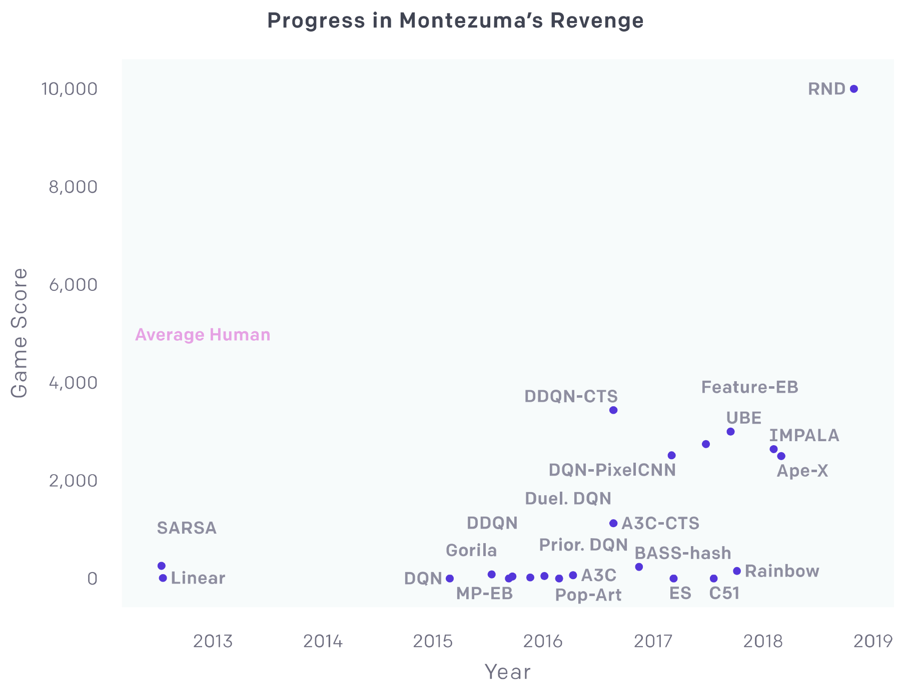


https://blog.openai.com/reinforcement-learning-with-prediction-based-rewards/

Problems can be mitigated: large models, Bayesian networks, LSTM

Random Network Distillation (RND) - Motivation

Deals with three previous problems by only using current state



https://blog.openai.com/reinforcement-learning-with-prediction-based-rewards/

Random Network Distillation (RND) - Overview

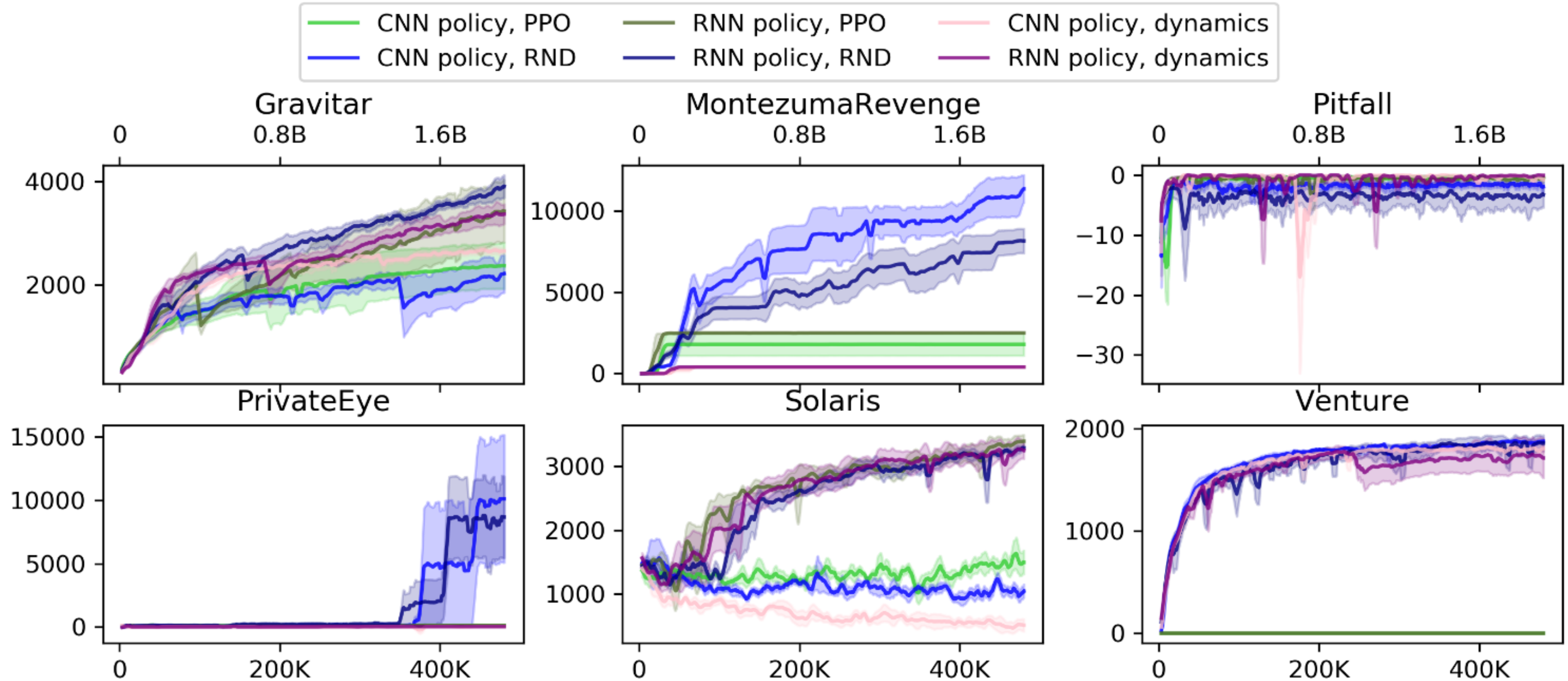
- Initialize Random Network (RN) and Predictor Network (PN) with random weights

- PN and RN have the same architecture and map the state representation to a vector

- PN is trained to predict output of RN for current state:
  - The prediction error is the intrinsic reward

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Random Network Distillation (RND) - Results



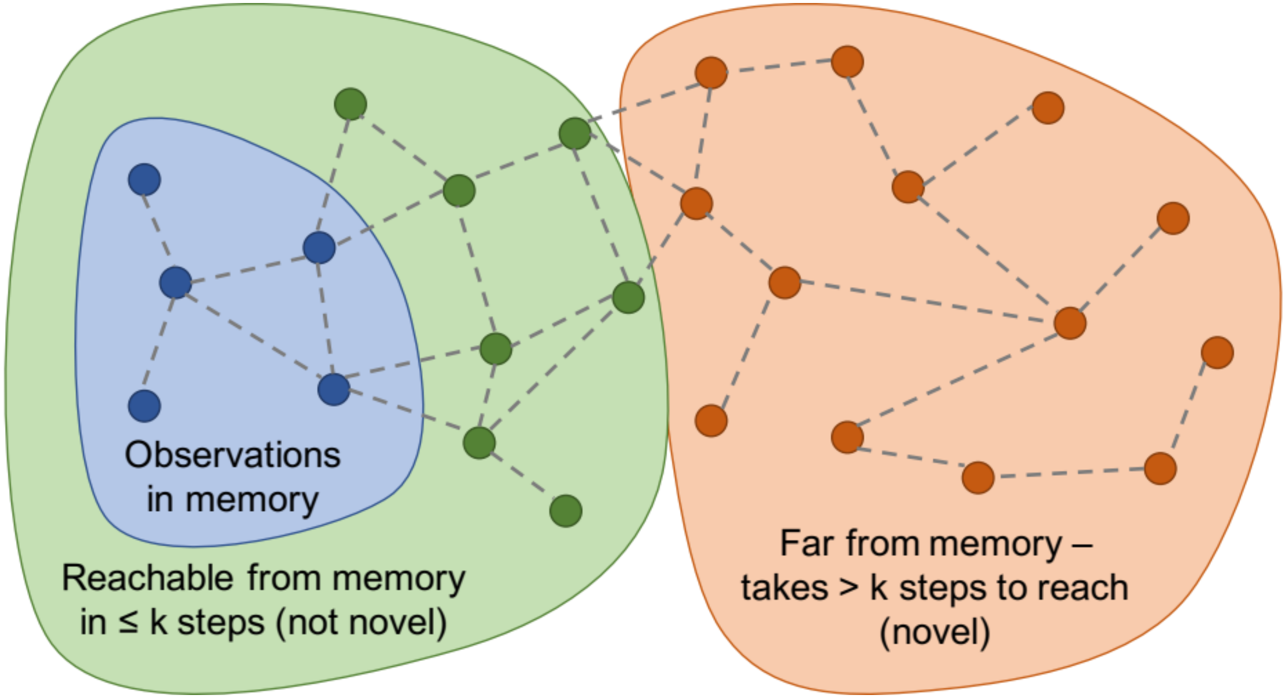
From the whitepaper [6]

Random Network Distillation (RND) - Drawbacks

- Simple, but not flexible
- No evidence for sample efficiency (trained for 1.6 Billion frames)
- No filtering of irrelevant state features
- Does not return to states it has seen before within episode

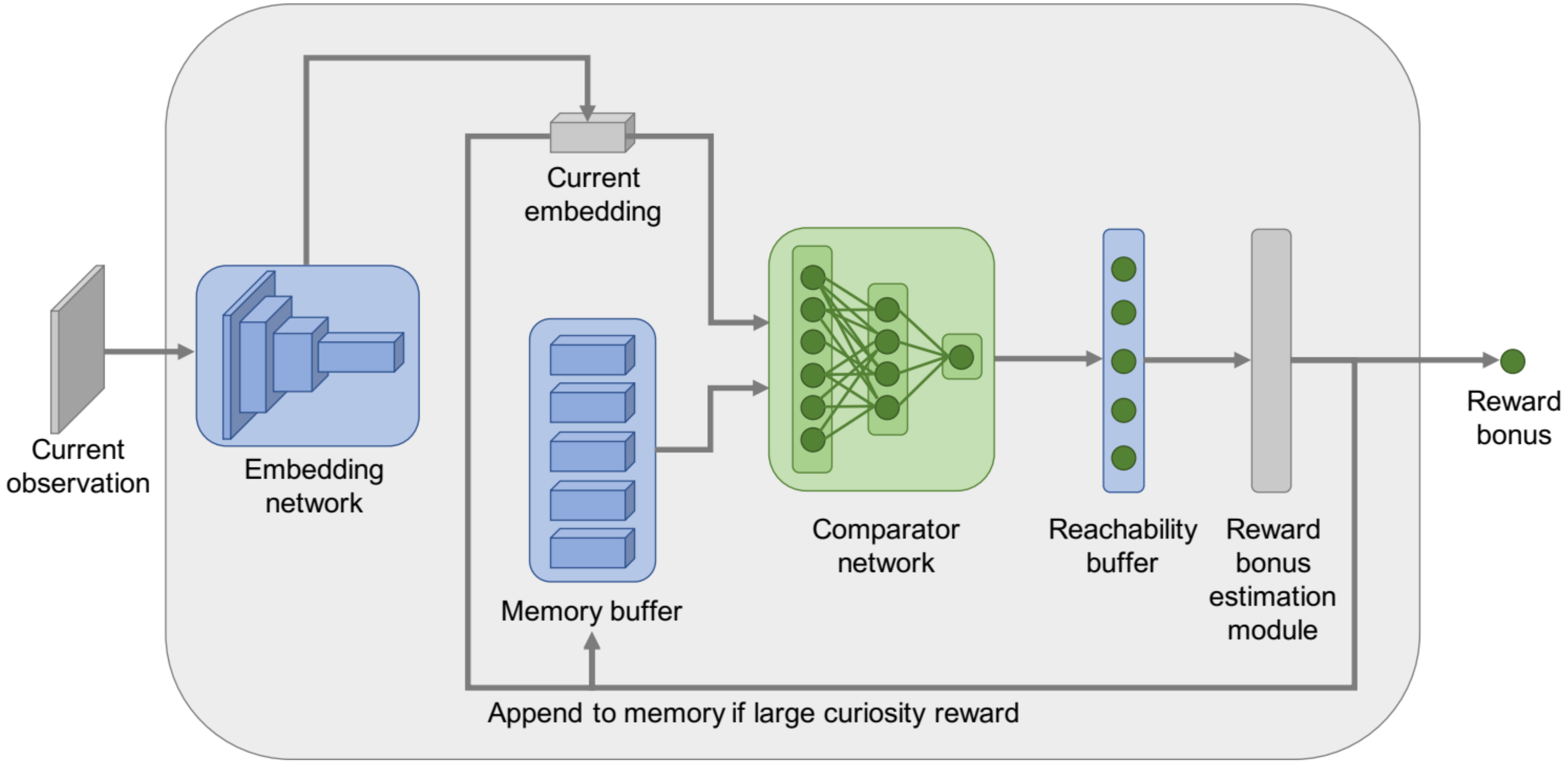
Episodic Curiosity Through Reachability (EC) - Idea

Incorporates acting into curiosity



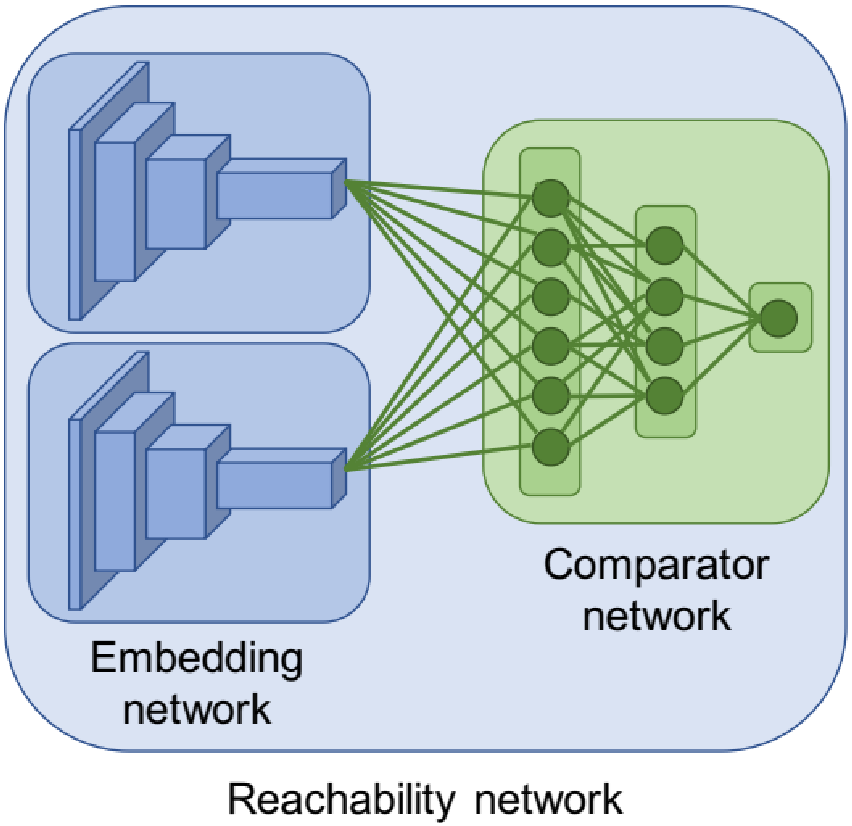
All figures from the whitepaper [7]

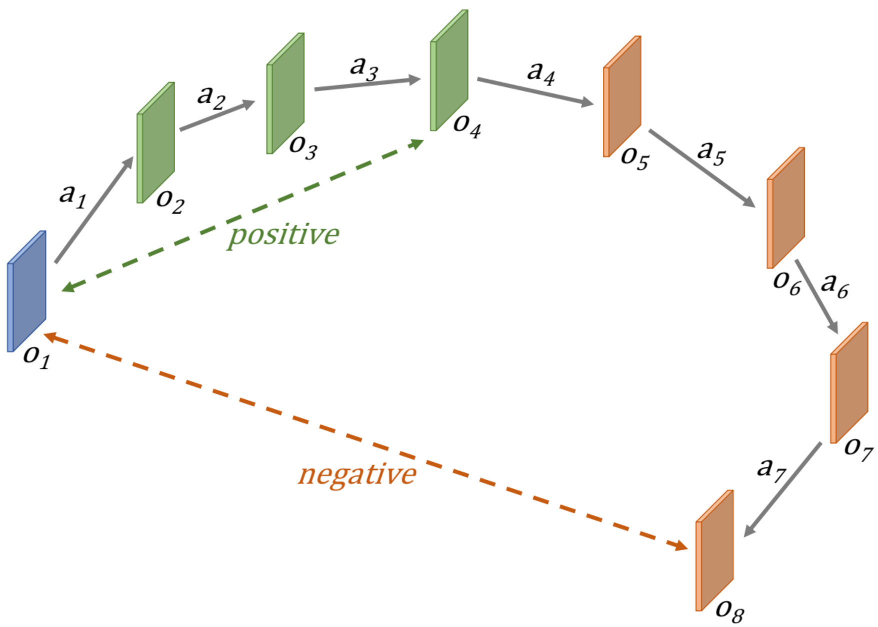
Episodic Curiosity Through Reachability (EC) - Overview



All figures from the whitepaper [7]

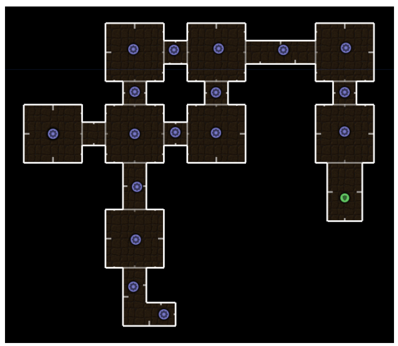
Episodic Curiosity Through Reachability (EC) – How to Embed and Compare

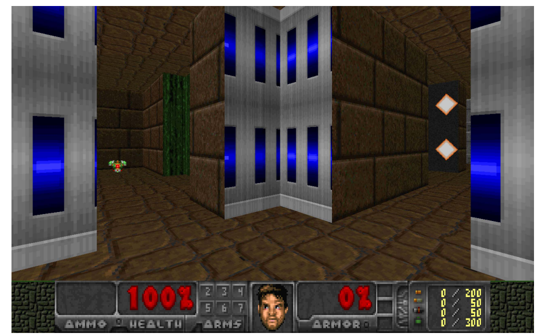




All figures from the whitepaper [7]

Episodic Curiosity Through Reachability (EC) – Results on VizDoom

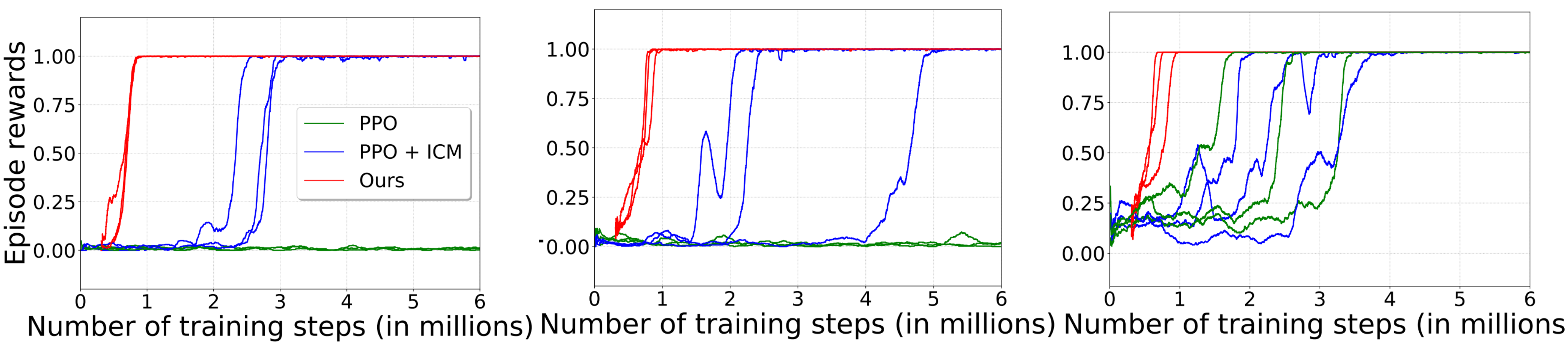




Very sparse rewards

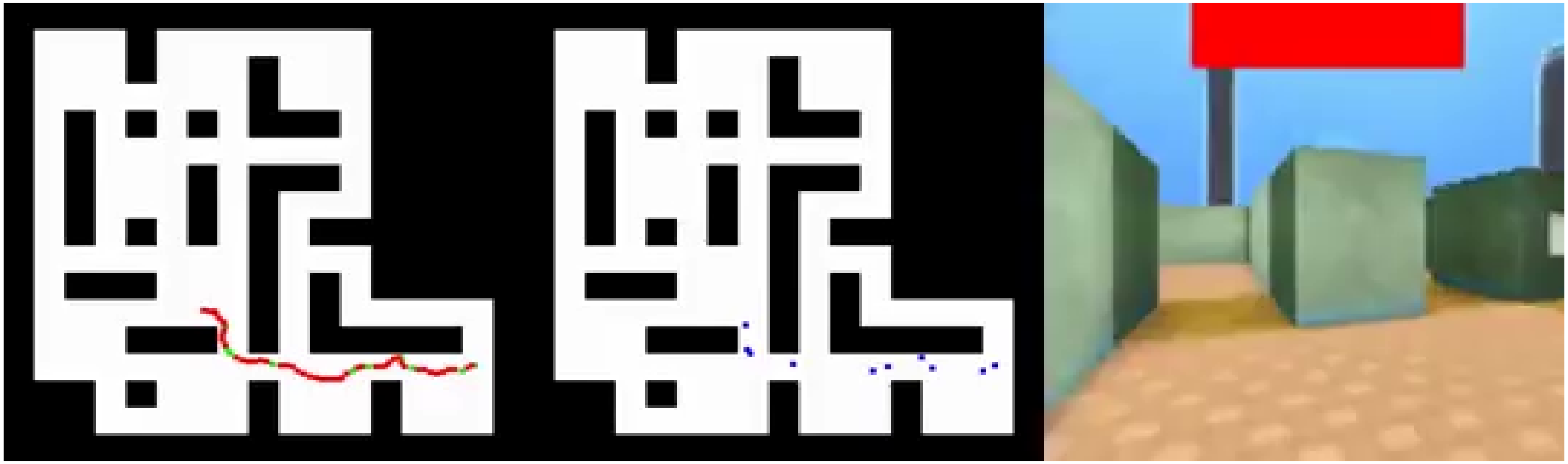
Sparse rewards

Dense rewards



All figures from the whitepaper [7]

Episodic Curiosity Through Reachability (EC) – Reward visualization



https://www.youtube.com/watch?v=mphIRR6VsbM&feature=youtu.be

Conclusion of these approaches

- ICM:
  - Works on state predictability

  \- Requires powerful world model
- RND:
  - Uses form of pseudo-state-count
  - Simple

  \- Not flexible
- EC:
  - Uses episodic memory to determine reachability
  - Incorporates acting in curiosity

  \- Has many moving parts

Drawbacks of Intrinsic Motivation in Robotics in General

Safety: It might be interesting for the robot to destroy parts of the environment, itself, or possibly humans.

Maybe fixable by:

- letting robots experience pain on extremities [3]
- training supervisor agent that identifies unsafe behavior

Complex intrinsic motivation might lead to unexpectable behavior:

http://terminator.wikia.com/wiki/Skynet

Outlook and Final Conclusion

Intrinsic motivation important for real intelligence, as obtaining extrinsic reward is “only” optimization problem.

Unclear which motivation is best!

Combine motivation approaches?

What are your intrinsic motivations?

Is there high and low-level curiosity?

Thank you for listening!
Any Questions?

References

- [1] Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. MIT Press, 2017
- [2] Rein Houthooft, Xi Chen, Yan Duan, John Schulman, Filip De Turck, and Pieter Abbeel. VIME: Variational information maximizing exploration. In NIPS, 2016.
- [3] J. Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. In J. A. Meyer and S. W. Wilson, editors, Proc. of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats, pages 222-227. MIT Press/Bradford Books, 1991.
- [4] A. Gupta, C. Eppner, S. Levine, and P. Abbeel, "Learning dexterous manipulation for a soft robotic hand from human demonstrations," in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2016, Daejeon, South Korea, October 9-14, 2016, pp. 3786–3793, 2016
- [5] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In International Conference on Machine Learning (ICML), volume 2017, 2017.
- [6] Burda, Y., Edwards, H., Storkey, A., and Klimov, O. (2018). Exploration by random network distillation. ArXiv preprint arXiv:1810.12894
- [7] Savinov, N., Raichuk, A., Marinier R., Vincent, D., Pollefeys, M., Lillicrap, T. Episodic Curiosity through Reachability. ArXiv preprint arXiv:1810.02274