Breadcrumbs: Unaliasing in Reinforcement Learning

Philip Raeisghasem

Reinforcement Learning

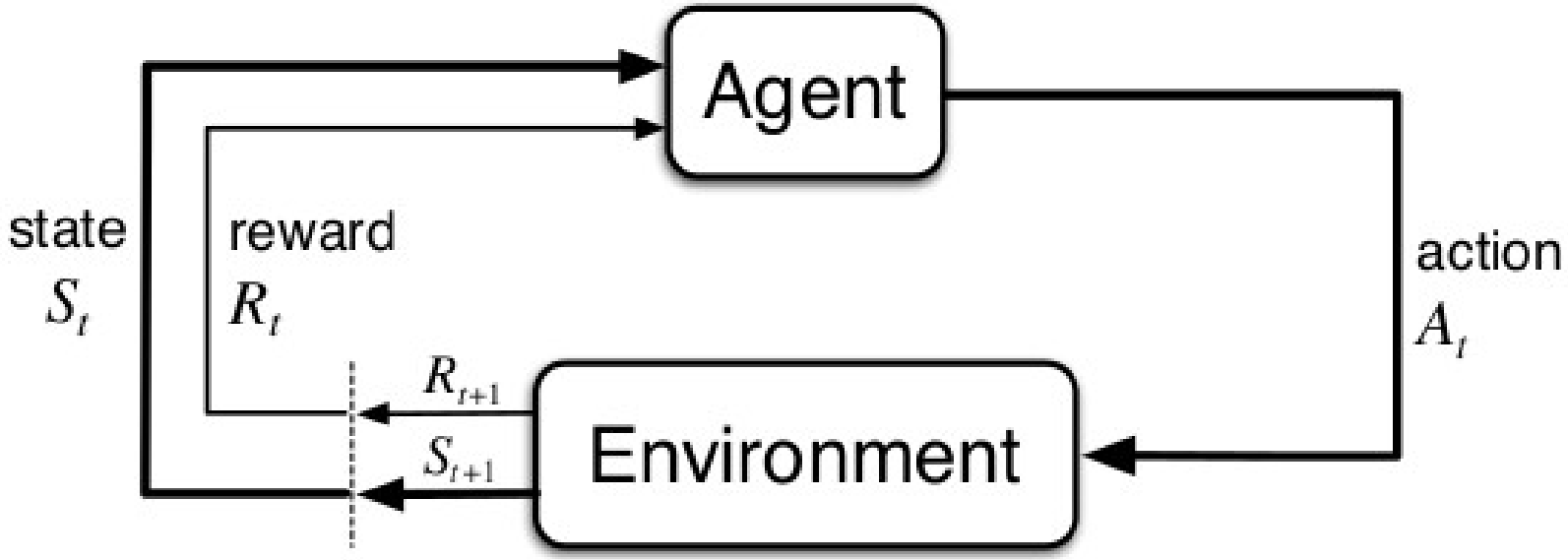
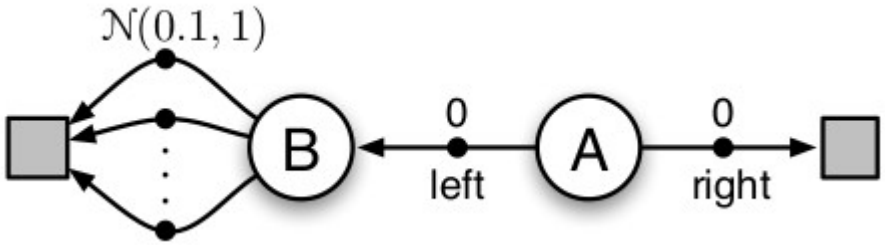
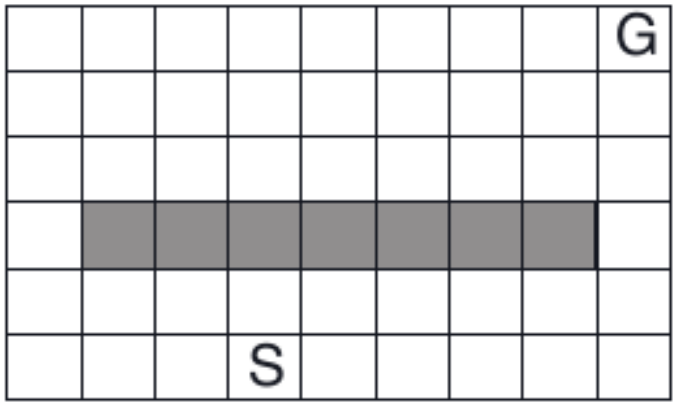


Figure from Sutton and Barto, 2018

Exploration





Figures from Sutton and Barto, 2018

Intrinsic Motivation

A way to provide agents with a denser reward signal (more feedback).

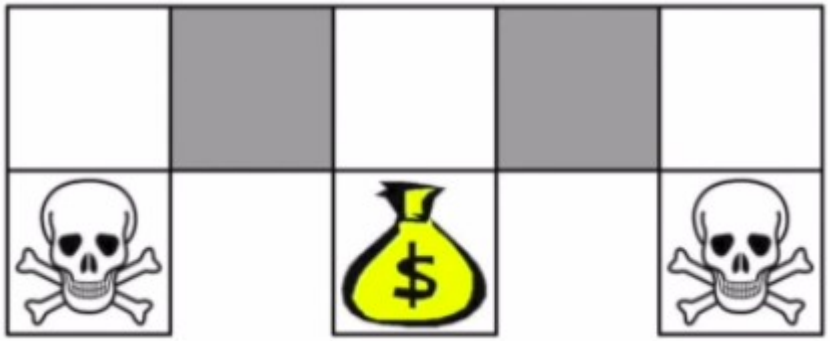
Curiosity- Rewarding agents for having new experiences.

Random Network Distillation (RND) – Curiosity algorithm.

RND: arxiv.org/abs/1810.12894

Aliased States

Due to partial observability

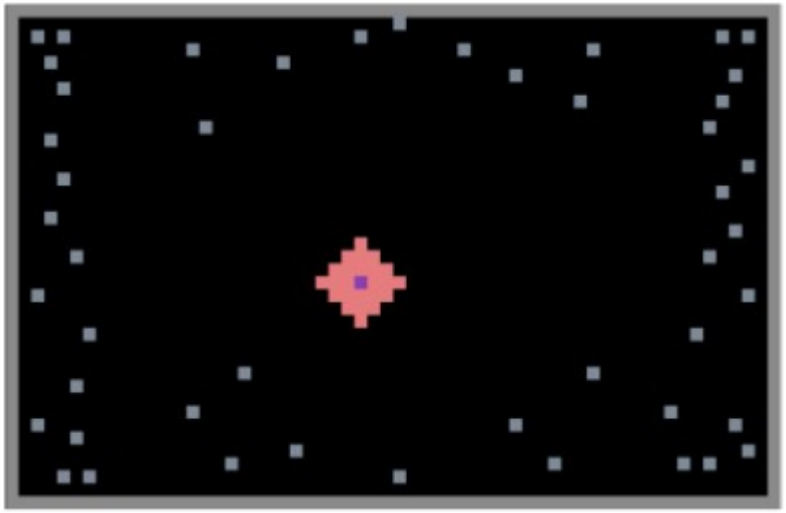


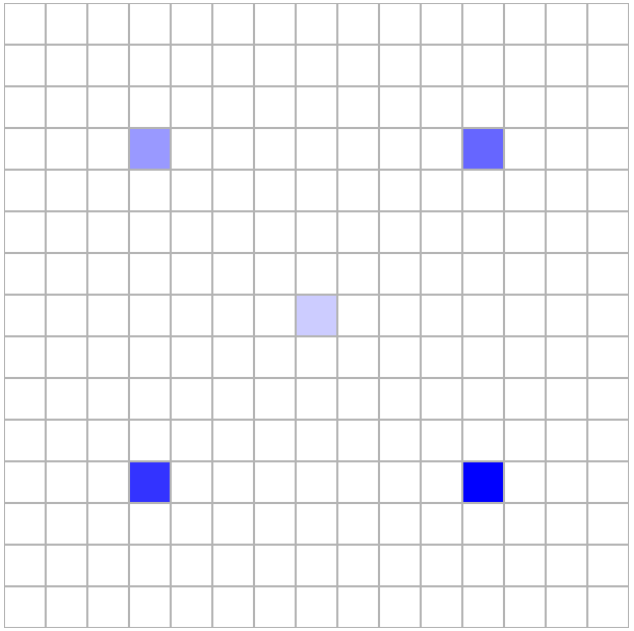
Currently solved with:

- **Memory**
- **Redefining “state” in MDP**

Figure from lecture by David Silver

Experiments





Top left figure: arxiv.org/abs/1812.07019v1

Results

Random variable is final return (Normal)

	Feature	Reward	Feature + Curiosity	Reward + Curiosity
sample mean	3.03	3.49	2.51	3.75
sample variance	1.76	1.71	1.17	2.22

Random variable is success rate (Bernoulli)

	Feature	Reward	Feature + Curiosity	Reward + Curiosity
p	18/40	28/40	8/40	29/40

Discussion

- **Reward trails were significantly better than feature trails**
- **Curiosity did not significantly improve upon either method**
- **Both unaliasing methods performed better than pure curiosity, which failed every time**

Future Work

- **Hyperparameter search**
- **Try other approaches**
- **Compare to RNN memory methods**
- **Test in multi-agent scenarios**