Latent Property State Abstraction

John Burden[1], Sajjad Kamali Siahroudi[2], Daniel Kudenko[2]

[1]CSER, University of Cambridge,
[2]L3S Research Centre, Hannover

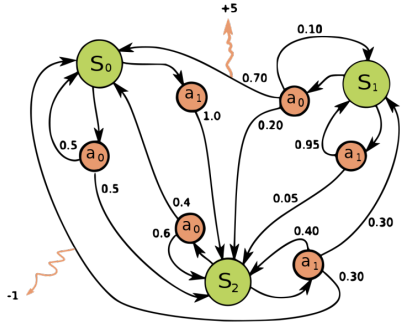
2020

Summary and Motivation

- Reinforcement Learning suffers from a high sample-complexity, often making learning slow.
- Domain knowledge of the environment can be used to improve this.
- Abstractions of the environment can be used to identify broad domain knowledge.
- Abstractions are usually given by a human domain expert; our approach opts to replace this with a method for an agent to learn its own abstraction through experience with the environment.

Abstract Markov Decision Processes

- An Abstract Markov Decision Process embodies an abstraction of a Markov Decision process; capturing the environment and its dynamics at a broader level. $\mathcal{A} = (S_{\mathcal{A}}, A_{\mathcal{A}}, R_{\mathcal{A}}, P_{\mathcal{A}})$
- We focus on state abstractions and build our AMDPs around those.
- We therefore require a state-abstraction function $Z : S_{\mathcal{M}} \rightarrow S_{\mathcal{A}}$, where $\mathcal{M}$ is the original MDP and $\mathcal{A}$ is the AMDP.



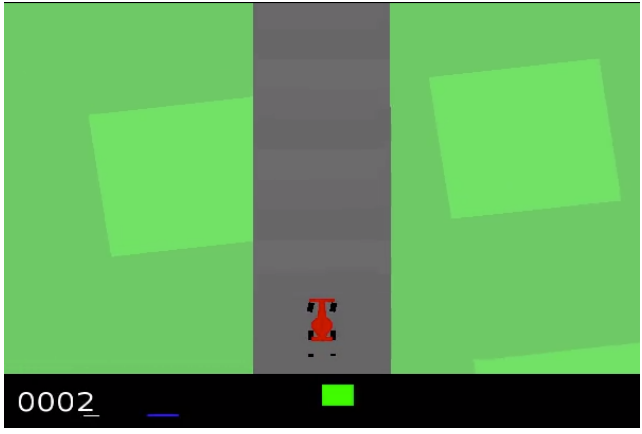
Credit: MistWiz - Wikipedia

Reward Shaping

- Reward Shaping is a mechanism for encouraging desired behaviour in an agent and involves giving the agent an additional reward.
- If the additional reward given is the difference between two potential functions of the states, then the optimal policy is guaranteed to remain unchanged. Additional reward $F(s, s') = \gamma\phi(s') - \phi(s)$ [5]
- AMDPs can be used for reward shaping and the optimal potential function is $\phi(s) = V(s)$ for $s \in S_{\mathcal{A}}$ [4].

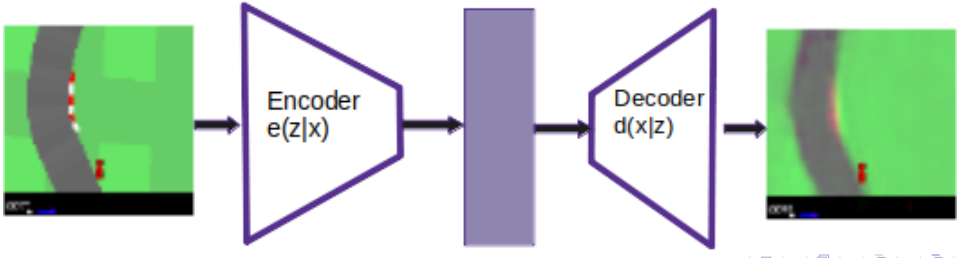
Car Racing Environment

- We use a version of the OpenAI Car Racing environment [1] where the action space has been discretised to 20 actions.



Learning Latent State Spaces

- Variational Auto-encoders [3] can learn to reconstruct pixel-based input through lower-dimensional bottlenecks.
- The original state-space can go from a $96x96x3$ pixel representation to a 128 dimension vector whilst retaining most of the information required to reproduce the image.
- This process can be done automatically with random rollouts of the environment.
- We can leverage the encoder portion of the network as a state abstraction function forming an abstract state-space.

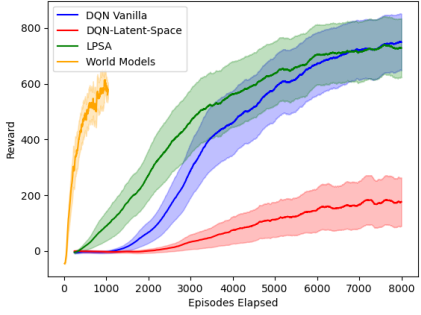


Latent Property State Abstraction

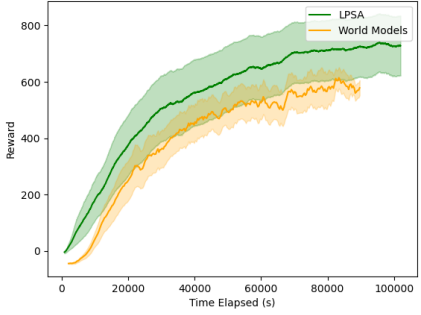
- We can use this state abstraction to form an AMDP, retaining the original actions and viewing the reward and transition functions through this abstract lens.
- This AMDP has a much smaller state-space than the original pixel-based environment.
- Interaction with this AMDP can be done model-free using the original environment.
- We can use learned abstract state values for reward shaping to improve the learning rate.

Results

- We augment DQN with our reward shaping approach and evaluate performance within the Car Racing domain. We compare our approach against DQN, the purely abstract agent, as well as World Models [2].
- Comparison against very different algorithms is difficult.
- Pre-processing time not shown so we can focus on shaping efficacy.



(a) Episodes



(b) Time

Conclusion

- Latent state-spaces can be learned from auto-encoders and used as abstract state-spaces.
- From this we can completely automate the reward shaping process.
- The learnt reward shaping functions can improve performance of existing algorithms.

References

[1] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba.
Openai gym, 2016.

[2] David Ha and Jürgen Schmidhuber.
World models.
CoRR, abs/1803.10122, 2018.

[3] Diederik P. Kingma and Max Welling.
An introduction to variational autoencoders.
CoRR, abs/1906.02691, 2019.

[4] Bhaskara Marthi.
Automatic shaping and decomposition of reward functions.
In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, pages 601–608, New York, NY, USA, 2007. ACM.

[5] Andrew Y. Ng, Daishi Harada, and Stuart J. Russell.
Policy invariance under reward transformations: Theory and application to reward shaping.
In *Proceedings of the Sixteenth International Conference on Machine Learning*, ICML '99, pages 278–287, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.