Core RL Behavior Suite: `bsuite` report

The *Core RL Behavior Suite*, or `bsuite` for short, is a collection of carefully-designed experiments that investigate core capabilities of a reinforcement learning (RL) agent. The aim of the `bsuite` project is to collect clear, informative and scalable problems that capture key issues in the design of efficient and general learning algorithms and study agent behaviour through their performance on these shared benchmarks. This report provides a snapshot of performance on `bsuite2019`, obtained by running the experiments from `github.com/deepmind/bsuite` [5].

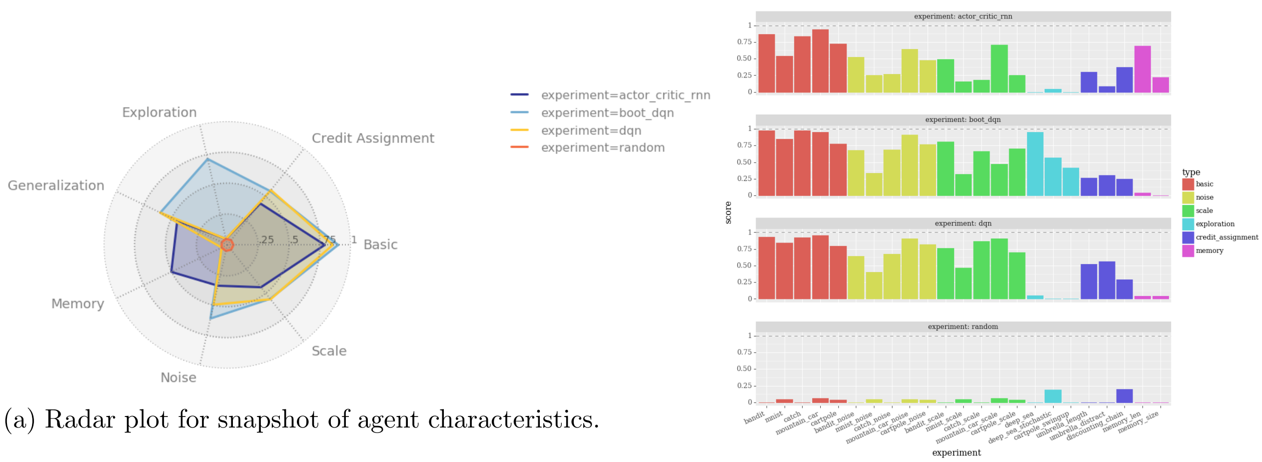
1 Agent definition

In this experiment we compare the performance on on `bsuite2019`of four agents taken from `github.com/deepmind/bsuite/baselines`. The agents that we considered were:

- actor_critic_rnn: an actor critic with recurrent neural network [1].
- boot_dqn: bootstrapped DQN with prior networks [4, 3].
- dqn: a fairly canonical implementation of Deep Q-networks (DQN) [2].
- random: an agent that selects action uniformly at random on each timestep.

2 Summary scores

Each `bsuite` experiment outputs a summary score in [0,1]. We aggregate these scores by according to key experiment type, according to the standard analysis notebook. A detailed analysis of each of these experiments may be found in a notebook hosted on Colaboratory: `ADD-LINK-HERE`.



(a) Radar plot for snapshot of agent characteristics.

(b) Examining score by individual experiment.

Figure 1: Summary output from the `bsuite2019` experiments.

3 Results commentary

All agent perform well on the basic reinforcement learning tasks, however they exhibit significant differences on some of the other dimensions. The recurrent actor-critic agent is the only one to perform better than random on the memory tasks, however it is more sensitive to noise and scaling issues than the other algorithms; As expected, Bootstrapped DQN is particularly effective on the exploration tasks, and it also seems to be more robust to noise, compared to vanilla DQN.

References

[1] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *Proc. of ICML*, 2016.

[2] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[3] Ian Osband, John Aslanides, and Albin Cassirer. Randomized prior functions for deep reinforcement learning. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 8617–8629. Curran Associates, Inc., 2018.

[4] Ian Osband, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. Deep exploration via bootstrapped DQN. In *Advances In Neural Information Processing Systems 29*, pages 4026–4034, 2016.

[5] Ian Osband, Yotam Doron, Matteo Hessel, John Aslanides, Hado Van Hasselt, Eren Sezener, Andre Saraiva, Tor Lattimore, Csaba Szepezvari, Satinder Singh, Benjamin Van Roy, Richard Sutton, and David and Silver. Core RL behaviour suite. 2019.