Revisiting Rainbow: Promoting more insightful and inclusive deep reinforcement learning research

Anonymous Author(s)
Affiliation
Address
email

Since the introduction of DQN [Mnih et al., 2015] reinforcement learning has witnessed a dramatic increase in research papers [Henderson et al., 2018]. New methods are typically evaluated on a set of environments that have now become standard, such as the Atari 2600 games made available in the Arcade Learning Environment (ALE) [Bellemare et al., 2012]. While these benchmarks have helped to evaluate new methods in a standardized manner, they have also implicitly established a minimum amount of computing power in order to be recognized as valid scientific contributions. Although classic reinforcement learning tasks such as CartPole, Acrobot, and grid worlds have not gone away, they are now used mostly for evaluating theoretical contributions; indeed, it is quite difficult to have a paper proposing a new reinforcement learning method accepted at one of the major machine learning conferences unless it includes experiments with one of the benchmarks just mentioned. Furthermore, at a time when efforts such as Black in AI and LatinX in AI are helping bring people from underrepresented (and typically underprivileged) segments of society into the research community, these newcomers are faced with enormous computational hurdles to overcome if they wish to be an integral part of said community.

It thus behooves the reinforcement learning research community to incorporate a certain degree of flexibility and creativity when proposing and evaluating new research. This paper is partly a position paper, partly an empirical evaluation. We argue for a need to change the status-quo in evaluating and proposing new research to avoid exacerbating the barriers to entry for newcomers from underprivileged communities. We complement this argument by revisiting the Rainbow algorithm [Hessel et al., 2018] on a set of small- and medium-sized tasks. This allows us to conduct a "counterfactual" analysis, and investigate whether there is scientific value in exploring empirical research in reinforcement learning when restricting oneself to small- to mid-scale environments.

Revisiting Rainbow

We follow a similar process as Hessel et al. [2018] in evaluating the various algorithmic variants mentioned above: we investigate the effect of adding each on top of the original DQN agent as well as the effect of dropping each from the final Rainbow agent, sweeping over learning rates for each. Our implementation is based on the Dopamine framework [Castro et al., 2018]. We perform our empirical evaluation on small-scale environments which are all available as part of the OpenAI Gym library [Brockman et al., 2016] . In order to strengthen the Rainbow Connection, we also ran a set of experiments on the MinAtar environment [Young and Tian, 2019].

In Figure 1 we evaluate both the *addition* of each algorithmic component to DQN, as well as their *removal* from the full Rainbow agent. What we find is that, in aggregate, the addition of each of these algorithms does improve learning over the base DQN. Nevertheless, for most algorithms the gains are not consistent throughout all environments; a finding which is consistent with the results observed by Hessel et al. [2018] in Figure 4. However, while Hessel et al. [2018] found prioritized replay and multi-step to be the most impactful additions, in these environments the gains from these additions are more tempered.

If we focus on the effect of removing components from the full Rainbow agent, what seems to hurt performance the most is the removal of distributional RL. However, it is rather interesting to see

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

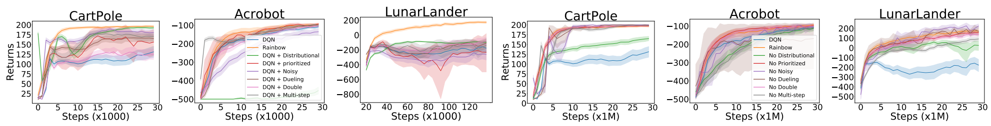

Figure 1: Comparison of the different algorithmic components on the small environments using the optimal hyper-parameters, averaged over 100 independent runs (shaded areas show 95% confidence intervals). Top row explores adding on top of DQN, bottom row explores removing from Rainbow.

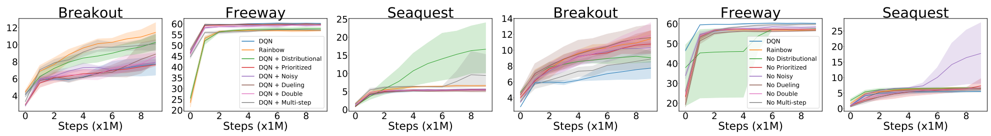

Figure 2: Comparison of the different algorithmic components on the MinAtar games, averaged over 5 independent runs (shaded areas show 95% confidence intervals). Top row explores adding on top of DQN, bottom row explores removing from Rainbow.

that when distributional RL is added to DQN *without any of the other components*, only adding distributional RL to DQN the gains can sometimes be minimal (see LunarLander), and can sometimes have a large negative effect on learning (Acrobot).

In MinAtar games, Figure 2 , we still find distributional RL to be the most significant of the additions, but it does not seem to require being coupled with another algorithm to produce improvements. This begs the question as to whether the use of distributional RL with convolutional layers avoids the pitfalls sometimes observed in the classic control experiments. However, the results from Seaquest suggest that the combination with one of the algorithms is hurting performance; based on the ablation results from Rainbow, it appears that noisy networks have a detrimental effect on performance.

Beyond the Rainbow

Running on small-scale experiments enables us to run more expansive research. As an example, we question the choice of the Huber loss, which is what is usually used to train DQN agents as it is meant to be less sensitive to outliers. We trained our agents using the MSE loss and found the surprising result that on all environments considered using the MSE instead of the Huber loss yielded improvements, and in all but one environment (Freeway) the improvements were significant, sometimes even surpassing the performance of the full Rainbow agent (Figure 3)..

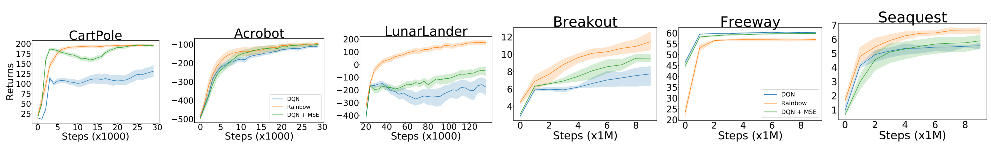

Figure 3: Evaluation of the use of the mean-squared error loss, instead of the Huber loss, in DQN.

Conclusion

On a limited computational budget we were able to reproduce, at a high-level, the findings of Hessel et al. [2018] and uncover new and interesting phenomena. Evidently it is much easier to revisit something than to discover it in the first place. However, our intent with this work was to argue for the relevance and significance of empirical research on small- and medium-scale environments. We believe that these less computationally intensive environments lend themselves well to a more critical and thorough analysis of the performance, behaviours, and intricacies of new algorithms.

We are by no means calling for less emphasis to be placed on large-scale benchmarks. We are simply urging researchers to consider smaller-scale environments as a valuable tool in their investigations, and reviewers to avoid dismissing empirical work that focuses on these smaller problems. By doing so, we believe, we will get both a clearer picture of the research landscape and will reduce the barriers for newcomers from diverse, and often underprivileged, communities. These two points can only help make our community and our scientific advances stronger.

References

Marc G. Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, Vol. 47:253–279, 2012. cite arxiv:1207.4708.

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

Pablo Samuel Castro, Subhodeep Moitra, Carles Gelada, Saurabh Kumar, and Marc G. Bellemare. Dopamine: A Research Framework for Deep Reinforcement Learning. 2018. URL `http://arxiv.org/abs/1812.06110`.

Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. Deep reinforcement learning that matters. In *Proceedings of the Thirthy-Second AAAI Conference On Artificial Intelligence (AAAI), 2018*, 2018.

Matteo Hessel, Joseph Modayil, Hado van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. Rainbow: Combining Improvements in Deep Reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015.

Ian Osband, Yotam Doron, Matteo Hessel, John Aslanides, Eren Sezener, Andre Saraiva, Katrina McKinney, Tor Lattimore, Csaba Szepesvári, Satinder Singh, Benjamin Van Roy, Richard Sutton, David Silver, and Hado van Hasselt. Behaviour suite for reinforcement learning. In *International Conference on Learning Representations*, 2020. URL `https://openreview.net/forum?id=rygf-kSYwH`.

Kenny Young and Tian Tian. Minatar: An atari-inspired testbed for thorough and reproducible reinforcement learning experiments. *arXiv preprint arXiv:1903.03176*, 2019.