Dopamine: A framework for flexible Reinforcement Learning research

Pablo Samuel Castro
Google Brain
psc@google.com

Subhodeep Moitra
Google Brain
smoitra@google.com

Carles Gelada
Google Brain
cgel@google.com

Saurabh Kumar
Google Brain
kumasaurabh@google.com

Marc G. Bellemare
Google Brain
bellemare@google.com

Abstract

We introduce Dopamine, a compact and flexible framework for speculative Reinforcement Learning research. Our framework targets research involving radical changes from established baselines while emphasizing simplicity, robustness, and reproducibility.

1 Introduction

Over the past decade there has been a continual growth in Reinforcement Learning (RL) research, resulting in many results with a large variety of agent types and environments. Many of these advances have come out of radical departures from the established algorithms of the time. Some notable examples are the use of replay memories Mnih et al. (2015), large-scale distributed training Espeholt et al. (2018), and distributional methods Bellemare et al. (2017). Because of the significant departure from traditional methods, the authors of these improvements typically implement their algorithms separately from any existing framework, which becomes a problem for reusability and reproducibility. It is unfortunately too common for these "original-author" implementations to never be open-sourced, resulting in multiple re-implementations of the same algorithm.

We introduce Dopamine, a new framework for the kind of speculative RL research that has driven these radical discoveries. By focusing on a specified type of RL research, our framework is compact and simple to modify, allowing researchers to try out radical new ideas without having to commit to a larger (and likely more complicated) framework. In addition, we provide state-of-the-art algorithms and ready-to-go results for easy comparison with new algorithms. In this paper we discuss some of the design choices we made when building this framework and the resulting qualities they produce.

2 Compactness and simplicity

Our framework focuses on value-based methods evaluated on the Atari games of the Arcade Learning Environment (ALE) Bellemare et al. (2013). The ALE is a mature and well understood benchmark that has become a standard for new RL algorithms. The 60 Atari games it supports provide a rich and diverse set of environments that pose a variety of challenges to agents. Indeed, they often help highlight the strenghts and weaknesses of different algorithms relative to each other.

In addition, we focus our framework on value-based agents. These design choices allow us to provide a codebase that is compact and easy to understand. Further, value-based agents demand fewer computational resources than typical policy-based methods. Dopamine provides stable implementations

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

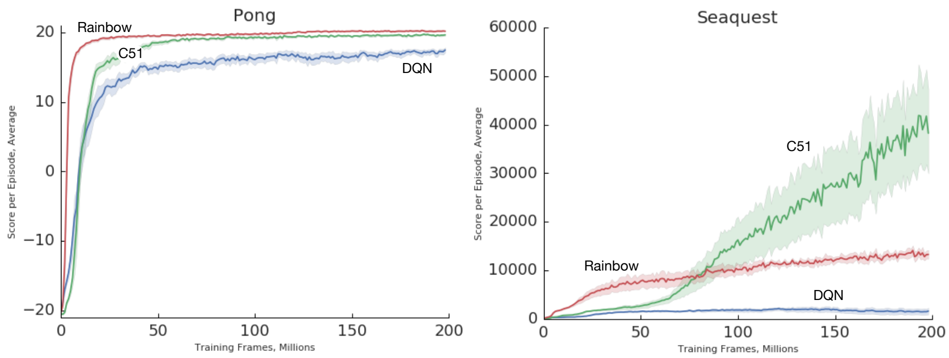

Figure 1: Baseline agents on SpaceInvaders and Seaquest

of DQN Mnih et al. (2015), C51 Bellemare et al. (2017), a simplified version of Rainbow Hessel et al. (2018), and Implicit Quantile Networks Dabney et al. (2018). Rainbow and Implicit Quantile Networks are current state-of-the-art methods.

Our implementation consists of only around 3000 lines of code spread over 11 python files (not including tests). This includes extensive inline documentation that help clarify the logic of the different components. We have written the code in a way that makes it very easy for both newcomers and experienced researchers to understand the inner workings, and begin modifying to try out new ideas. Nonetheless, our framework also makes it easy to train one of the provided agents with just a few lines of Python code.

In addition to exporting standard RL training statistics to Tensorboard, Dopmaine comes with a set of colabs that make it easy to plot the results against any of the provided baselines.

3 Reproducibility

We are particularly sensitive to the importance of reproducibility in RL research, as has been argued in Islam et al. (2017). The following features of Dopamine help us satisfy this important requirement.

- We have followed the recommendations given by Machado et al. (2018) on standardizing evaluation on the ALE.
- For each agent provided, we include the hyperparameter configurations that match those used in the original papers.
- We provide an "updated" set of configuration files for each agent that unifies the hyperparameter settings amongst them. This enables apples-to-apples comparison between the different agents.
- We provide the full training data comparing all provided agents on all 60 games from the ALE for facilitating the benchmarking of new agents. These are available as Python pickle files (for agents trained with Dopamine) and as JSON data files (for comparison with agents trained in other frameworks). Figure 1 plots the returns of our provided agents for two Atari games.

4 Conclusion

Some of the most significant advances in Reinforcement Learning have come as a result of dramatic departures from established methods. Unfortunately many of these are not accompanied with open-sourced author-approved implementations which is problematic for reproducibility and for promoting further improvements on new algorithms. Dopamine aims to fill this gap by providing a compact and flexible framework. We believe our design choices make Dopamine a strong starting point for newcomers and experienced researchers alike; our hope is that Dopamine's flexibility and ease-of-use will empower researchers to try out new ideas - both incremental and radical - very quickly, and help in the advance of scientific research.

References

Bellemare, M. G., Dabney, W., and Munos, R. (2017). A distributional perspective on reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 449–458.

Bellemare, M. G., Naddaf, Y., Veness, J., and Bowling, M. (2013). The arcade learning environment: An evaluation platform for general agents. *J. Artif. Int. Res.*, 47(1):253–279.

Dabney, W., Ostrovski, G., Silver, D., and Munos, R. (2018). Implicit quantile networks for distributional reinforcement learning. In *Proceedings of the 35th International Conference on International Conference on Machine Learning*, ICML'18.

Espeholt, L., Soyer, H., Munos, R., Simonyan, K., Mnih, V., Ward, T., Doron, Y., Firoiu, V., Harley, T., Dunning, I., Legg, S., and Kavukcuoglu, K. (2018). IMPALA: Scalable distributed deep-RL with importance weighted actor-learner architectures. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1406–1415, Stockholmsmässan, Stockholm Sweden. PMLR.

Hessel, M., Modayil, J., van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., Horgan, D., Piot, B., Azar, M. G., and Silver, D. (2018). Rainbow: Combining improvements in deep reinforcement learning. In *AAAI*. AAAI Press.

Islam, R., Henderson, P., Gomrokchi, M., and Precup, D. (2017). Reproducibility of benchmarked deep reinforcement learning tasks for continuous control. *CoRR*, abs/1708.04133.

Machado, M. C., Bellemare, M. G., Talvitie, E., Veness, J., Hausknecht, M. J., and Bowling, M. (2018). Revisiting the arcade learning environment: Evaluation protocols and open problems for general agents. *J. Artif. Intell. Res.*, 61:523–562.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.