Rainbow: Combining Improvements in Deep Reinforcement Learning

Manan Tomar : ED14B023

Abstract

This work (Hessel et al., 2017) examines various independent extensions to the DQN (Mnih et al., 2015) algorithm and proposes combining these to extract their individual benefits. To this end, an algorithm called Rainbow-DQN is introduced which is shown to perform better than all other variants.

1. Approach

There are six extensions to DQN considered here :

- **Double DQN** (Van Hasselt et al., 2016) uses two separate networks in the maximization of the target, one for selecting the action and the other for evaluating the Q value for the selected action. This is done so as to avoid overestimation of Q values observed in the original DQN setup.
- **Prioritized Replay** samples some transitions more frequently than others where the sampling probability is proportional to the TD error
- **Multistep Learning** uses a multi step return instead of the standard single step used in DQN. The Q value in the target is then computed for the $t + n^{th}$ state where n defines the multi step.
- **Distributional RL** uses a distribution over the return instead of only using the expected value, i.e. the state value function. A variant of the Bellman equation is proposed which applies to distributions instead of scalars and it is shown that minimizing the KL divergence between the target distribution and the current approximating distribution results in learning optimally.
- **Noisy Nets** introduces a noisy term to a linear layer of the Q network. This allows for efficient exploration initially, with the network learning to ignore this term slowly as more gradient steps are performed.
- **Dueling DQN** decouples the Q value estimation by estimating the state value function $V(s)$ and the advantage function $A(s, a)$ separately and then aggregating these to produce Q. This allows for learning faster as high rewarding states are quickly learnt irrespective of the action taken.

The combined approach here adds each of the above step by step, first introducing a multi step version to the distributional case, then having two networks for evaluating and selecting actions to incorporate Double DQN, using prioritized replay with sampling probabilities proportional to the KL divergence from the distributional setting and replacing all linear layers by a noisy version. Finally, to add the dueling architecture, separate streams for V and A are added and softmax is taken over their aggregation to produce the return distribution.

2. Experiments

Rainbow-DQN is evaluated on all 57 Atari games and the normalized average scores against human performance are reported. Ablation studies are also conducted by removing one of the six components from the combined version to study the effect each one has on the overall performance. It is observed that the multi step return component's absence hurts the final performance the most.

The paper concludes that in general rainbow DQN can further be combined with a lot more other methods as well, using concepts from Heirarchical RL, exploration based methods such as count based strategies, intrinsic motivation and bootstrapped DQN, making use of auxiliary tasks, etc. being some of such methods.

References

Hessel, Matteo, Modayil, Joseph, Van Hasselt, Hado, Schaul, Tom, Ostrovski, Georg, Dabney, Will, Horgan, Dan, Piot, Bilal, Azar, Mohammad, and Silver, David. Rainbow: Combining improvements in deep reinforcement learning. *arXiv preprint arXiv:1710.02298*, 2017.

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

Van Hasselt, Hado, Guez, Arthur, and Silver, David. Deep reinforcement learning with double q-learning. In *AAAI*, volume 2, pp. 5. Phoenix, AZ, 2016.