R4/Sanit Gupta/160100010

August 6, 2020

In this paper, the authors make a case for variants of algorithms that attempt to track the solution instead of trying to converge to it. For non-stationary problems, tracking seems to be the obvious option. On the other hand, for stationary problems, contrary to intuition the authors argue that tracking is the way to go. Tracking gives more weight to the local environment and this ends up helping in some scenarios. I think this was an interesting paper because usually people only care about the convergence of algorithms and in this paper they argue that convergence might not be the right goal with decent empirical evidence.

They test their hypothesis out on two problems.

The first one is the black and white world, a world specifically designed to illustrate the advantages of tracking. In the black and white world, an agent performs a random walk. It can also perform an observe action through which it observes the colour black and white. You are supposed to predict the colour of an observation. Half of its path is white, and the other half is black (these are two contiguous halves). This means having recently observed a certain colour means that it is likely that one is observing the same colour again. A convergent algorithm is unable to capture this, simply predicting either colour with half probability. The tracking algorithm is successful in using these recent observations and outperforms the convergent version (For certain values of the hyperparameter that needs to be set for the tracking algorithm; this hyperparameter can also be set by a meta learning algorithm).

The second one is Go on a 5x5 game board. In spite of the small sounding board size, it has a very large state space ($5 * 10^{10}$ states). In this case, the convergent algorithm is given 250,000 self play episodes to train and the tracking algorithm trains for 10,000 self play episodes (with the current state as initial state) after every time step. Because of how quickly the game usually ends, the tracking agent is extremely unlikely to get more learning time than the convergent algorithm. The tracking algorithm beats the convergent one when both are given extremely simplistic features. Also, making complex features available helps the tracking algorithm more than the convergent agent.

They also adapt the IDBD algorithm to learn the hyperparameter for the black and white world. It is successful in learning reasonable values of the

hyperparameter.

A question I have regarding the paper:

- Should some credit for the victory of the tracking algorithm be attributed to it learning online and from the state it was actually in, unlike the converging algorithm? .