R8/Sanit Gupta/160100010

August 6, 2020

In this paper, the authors consider the problem of making the Sony Aibo robot walk as fast as possible. Locomotion of robots having legs is a complex control problem. Some time before this paper was published, hand tuning of parameters was the standard approach. Even though this approach worked well enough, there were issues including it being too time consuming. All the training was done on real robots and their total training amounted to just 1000 field traversals and 3 hours. Achieving a speed better than both the best hand tuned and learned approaches in such a short time is extremely impressive.

Given that evaluation of a parameter vector is very expensive, efficiency was a major concern for the authors. The authors' approach was to start with an initial vector and to perform gradient descent. They proposed an interesting method to estimate the gradient. First,they would evaluate a number of randomly generated parameter vectors in the proximity of the initial vector. These random parameter vectors would be generated in such a manner that it would likely that we'd get a reasonable estimate of the partial derivative of the vector in each direction. Next, these vectors would be evaluated and using the derivatives estimated a step would be made to change the gait so as to increase the speed of the robot.

There were some issues I had with this paper. First of all, it seems to me that this is just a blackbox optimization problem where they have used gradient descent (using some heuristic methods to estimate the gradient) to find the optimal parameters. I don't think was a reinforcement learning problem. Clearly, their approach works very well and they are able to achieve extremely high speeds but I think they have given mislabelled their solution techniques. In my opinion, this was not a reinforcement learning problem or a reinforcement learning paper.

Also, they start with 'the hand-tuned UT Austin Villa walk as a starting point for learning.' Would having a random starting point not increase the sample complexity a lot? Is that something one should be concerned about? Does this not qualify the claim they made about there being no 'human intervention'?