Asynchronous Training

A disadvantage of the policy gradient methods presented in the previous section is that performing the rollout operations requires evaluating agent behavior in some (likely simulated) environment. Most simulators are complicated pieces of software that can't be run on the GPU. As a result, taking a single learning step will require running long CPU-bound calculations. This can lead to unreasonably slow training.

Asynchronous reinforcement learning methods seek to speed up this process by using multiple asynchronous CPU threads to perform rollouts independently. These worker threads will perform rollouts, estimate gradient updates to the policy locally, and then periodically synchronize with the global set of parameters. Empirically, asynchronous training appears to significantly speed up reinforcement learning and allows for fairly sophisticated policies to be learned on laptops. (Without GPUs! The majority of computational power is used on rollouts, so gradient update steps are often not the rate limiting aspect of reinforcement learning training.) The most popular algorithm for asynchronous reinforcement learning currently is the asynchronous actor advantage critic (A3C) algorithm.

CPU or GPU?

GPUs are necessary for most large deep learning applications, but reinforcement learning currently appears to be an exception to this general rule. The reliance of reinforcement learning algorithms to perform many rollouts seems to currently bias reinforcement learning implementations toward multicore CPU systems. It's likely that in specific applications, individual simulators can be ported to work more quickly on GPUs, but CPU-based simulations will likely continue to dominate for the near future.

Limits of Reinforcement Learning

The framework of Markov decision processes is immensely general. For example, behavioral scientists routinely use Markov decision processes to understand and model human decision making. The mathematical generality of this framework has spurred scientists to posit that solving reinforcement learning might spur the creation of artificial general intelligences (AGIs). The stunning success of AlphaGo against Lee Sedol amplified this belief, and indeed research groups such as OpenAI and DeepMind aiming to build AGIs focus much of their efforts on developing new reinforcement learning techniques.

Nonetheless, there are major weaknesses to reinforcement learning as it currently exists. Careful benchmarking work has shown that reinforcement learning techniques are very susceptible to choice of hyperparameters (even by the standards of deep learning, which is already much finickier than other techniques like random forests).