Reinforcement Learning Progress

Today, OpenAI released a new result. We used PPO (Proximal Policy Optimization), a general reinforcement learning algorithm invented by OpenAI, to train a team of 5 agents to play Dota and beat semi-pros.

This is the game that to me feels closest to the real world and complex decision making (combining strategy, tactics, coordinating, and real-time action) of any game AI had made real progress against so far.

The agents we train consistently outperform two-week old agents with a win rate of 90-95%. We did this without training on human-played games—we did design the reward functions, of course, but the algorithm figured out how to play by training against itself.

This is a big deal because it shows that deep reinforcement learning can solve extremely hard problems whenever you can throw enough computing scale and a really good simulated environment that captures the problem you're solving. We hope to use this same approach to solve very different problems soon. It's easy to imagine this being applied to environments that look increasingly like the real world.

There are many problems in the world that are far too complex to hand-code solutions for. I expect this to be a large branch of machine learning, and an important step on the road towards general intelligence.