Robust Adversarial Reinforcement Learning, Pinto et al

Summary

A reinforcement learning method relies on massive amount of training data. For real-world tasks, the process of collection of data is prohibitively expensive and in some cases dangerous. For transfer of learning policies that are trained on simulated data, the gap between simulation and the real-world often results in a failed transfer of the learning policy. Moreover, explicit definition of environmental parameters of the simulation for collection of training data or modeling errors such as different values of friction etc., would increase the training set by a huge amount. This issue of a lack of robustness and generalization of RL policies, is an important field of research in AI, and has been addressed in this paper.

Approach and Strength

The paper, uses an approach, similar to that in a two-player zero-sum discounted game, or Generative Adversarial Network. The protagonist learns a policy to maximize the reward while another agent, the adversary, learns a policy to minimize the reward function of the protagonist. The reward function can be visualized as a function of stability of the agent, for example, the stability of a walker in a simulation. The algorithm optimizes the the agents in an alternating procedure, that is, the protagonist learns its policy, using TRPO, while holding the adversary's policy fixed and vice versa. This circumvents the issue of solving for the exact equilibrium solution at each iteration, if the protagonist and adversary's policy were trained simultaneously.

The learned policy using RARL, outperforms the policy learned using TRPO, even when no noise or disturbance is added to the simulation, which is a major strength of the paper. There is a significant improvement between the baseline policy and RARL, when test parameters such as friction, mass etc., are changed or introduced as disturbances in the test environment. The baseline policy fails to generalize in these cases.

Weakness

While introducing disturbances or varying the environmental parameters, the noise is introduced as an adversarial noise. The paper does not present results with random noise in the test environment.

The paper also does not present real-world experimental results, which is contrary to the discussion of better transfer of simulation policies to real-world.

Future Work and Applications

The application of RARL can be extended to a multi-agent setting, where a single, or multiple adversarial agents, could work against a multi-robot system, to simulate disturbances in an environment or to model the failure of components in the robots.

The RARL algorithm could be applied to learn a driving policy for modeling the behaviour of an autonomous vehicle in a traffic environment. For example, this could be a trajectory planning problem for a vehicle merging into a heavy traffic lane, on an highway, where the adversaries are the neighboring vehicles.