Paper Review:
Trust Region Policy Optimization

Summary:

The paper, “Trust Region Policy Optimization”, suggests saying that there is an iterative procedure that has the purpose of optimizing policies, that should guarantee a consistent improvement. TRPO uses a trust region approach to ensure that the policy update step is conservative and doesn't lead to a significant decrease in performance. The TRPO algorithm uses a trust region construct to ensure that policy updates do stray too far from the active policy, hence being able to improve efficiency and ensure that the policy converges to a localized optimum.

Contributions:

The authors observe and report on the effectiveness of TRPO on several benchmark tasks and discuss its performance against other algorithms imposed on agents. The authors also supply the mathematical formulas and theories needed for users to expand on the experiment. The algorithm requires the solution of a large optimization problem at each iteration, which can be computationally expensive and time-consuming. The paper states that there was “monotonic improvement” for an algorithm when frequently optimizing a “local approximation to the expected return of the policy.”

Strengths and Weaknesses:

Some of this paper’s strengths would include how it is explaining mathematical concepts surrounding the TRPO in a rich manner. This article has a weakness in that it does not seem to use a sufficient breakdown for terms that are for newer readers exploring the artificial intelligence technology field who also happened to stumble upon DRL. In TRPOs the algorithm requires the solution of a large optimization problem at each iterative series, which can be computationally expensive and time-consuming. Another weakness is how the computational requirements for training the models are high, which may limit their practical applicability, as well as accessibility.

Experimental Validity:

This experiment has traits that would suggest that it is valid as it does seem to generate results expressed in charts and supported by formulas and theories. This experiment was very interesting as it seems to explore a possible solution for improving algorithms. Overall, the experiment is beyond being considered anything short of valid.

Additions to Paper:

To extend on this paper, computational efficiency seemed to be a big factor in determining this methods approach as feasible. While the proposed algorithms achieve cutting edge performance, the computational minimums for training the models are high. Someone could explore methods to improve the computational efficiency of these algorithms.