Regularization and Feature Selection in Least-Squares Temporal Difference Learning

Tanay Dixit (EE19B123) & Vibhhu Sharma (EE19B128)

The paper "Regularization and Feature Selection in Least-Squares Temporal Difference Learning" by J. Zico Kolter and Andrew Ng aims to improve linear value function approximation in Reinforcement Learning problems with very large state spaces. In this context, it proposes to use L1 regularization as a regularization framework for the Least Squares Temporal Difference(LSTD) algorithm. The paper introduces the problem statement and background in an elaborate manner that is also easy to understand before getting into the minutiae of the algorithm. It outlines the need for regularization and the LSTD method before explaining how to incorporate each type of regularization and their corresponding benefits and drawbacks.

Reproducibility: The paper provided pseudo-code for the algorithm and follows it up with a step-by-step explanation. For experimentation, the paper clearly defines the parameters and state and action spaces of the example problems it considered:one of them being the famous "mountain car" problem. Thus, these results can be reproduced. The use of a classic problem as a demonstration example makes it more familiar and approachable to the reader.

Strengths:

- The paper makes a significant contribution in the domain of reinforcement learning with large state spaces. It succeeds in developing an algorithm based on Least Angles Regression(LARS) that uses L1 regularization to produce sparse solutions.
- Their method prevents overfitting and gives a higher discounted reward as the number of irrelevant features increase.
- The authors obtained a computational complexity that is linear in the total number of basis functions, in contrast to the original complexity of standard-LSTD which is at least quadratic.

Novelty: Prior to this paper, there had been work done in the domain of regularization and feature selection in Reinforcement Learning. One paper utilized l2 regularization for temporal difference based policy iteration algorithms. Another paper utilized L1-regularization, however they do not consider fixed points of a Bellman backup, and hence the solution loses all interpretation as a fixed point, which is very important for the Temporal Difference solution. The authors improve upon these shortcomings and thus provide a significant contribution to the field.

Areas of improvement:

- The authors claim that one can ensure that `A` is a P-matrix even when sampling off-policy by adding some amount of l2-regularization. However, this decreases the sparsity of the solution and brings with it the same disadvantages seen with typical l2-regularization.
- LARS-TD is not optimal for the sequential nature of policy improvement, because each new policy requires us to restart.
- The authors have not compared the performance of the algorithm with alternate methods such as those based on the addition of new features in a greedy manner(Tropp et. al, 2004[1]). These algorithms have also been proven to work well and a performance comparison would be helpful.

We believe that we have understood the contributions of the authors fairly well. Based on our understanding, the paper was presented in a manner that was accessible to new practitioners in the field without compromising on the objective of producing a significant result.

[1] http://users.cms.caltech.edu/ jtropp/papers/Tro04-Greed-Good.pdf