Linear Least-Squares Algorithms for Temporal Learning

Soheil

1 Main Ideas

Temporal Difference learning algorithms were developed to address episodic MDPs. In TD learning, the value function approximation stages are done in each time step. At the end of each episode (trial), then, a policy evaluation is performed. However, when the state-action size of our problem is enormous, tabular representation of the problem (in P(transition probability) and R(rewards)) becomes our main issue. For instance, take the well-known inverted pendulum as an example. We are dealing with a continuous problem, with presumably infinite number of states. We can discretize the state into a set of for instance 360*100 states (to have one degree resolution for the θ and a resolution of 0.1 for the $\dot{\theta}$ between -5 to +5). If we only consider a resolution of 0.1 for the motor (from -5v to +5v as an input) that controls the position of the pendulum, the state-action space will get gigantic (36000*100), and dealing with this problem will become practically very hard. This is only happening in a small problem. However, it is good example that fairly reveals a big issue we have in our previous MDP solvers. Storing a unique value for each state-action pair is not practical for large state spaces.

This problem is addressed by considering different types of function approximations for the state (or action) value function. In particular, if the number of actions are small, we can use linear approximation for the value function:

$$V(s) = \phi_s^\top \theta$$

A general approach to the use of linear approximation in solving a MDP is elaborated in the paper in Figure 1. A least-square approach, which is the main contribution of this paper, is also explained in details in figure 2. This approach, however, is computational intensive, and has $O(m^3)$ complexity. Instead of a LS TD algorithm, we can use the recursive version of it, which is introduced in formula 13 to 15.

The only challenge in the way of implementing this approach is to how to define the basis function for a given MDP.

2 Questions

1. How is LSTD relate to TD (temporal difference)? And for extra credit: When will TD and LSTD solutions be the same?

In small problems with small number of states and actions, the tabular representation would be enough to solve a MDP. However, as soon as we are getting more states, the problem starts to get exponentially hard to be solved using standard TD methods, in which we estimate the value (or action) function in each state (or state-action pair). There it comes LSTD, that approximates the value function with a linear combination.