Optimization in Markov Decision Programming

For Udacity AIND Planning Project Research Review

In this report, we will be discussing about the three major optimization techniques used in Markov Decision Programming. Planning has long history in the field of AI from classical planning of Theorem proving to Problem Solving which was first introduced in a paper by Nilsson and Fikes in 1971 and it was called STRIPS. Then PDDL, Planning graphs & GraphPlan were introduced. But this classical model had many disadvantages like Uncertainty and Multi, competing objectives. So to counter the uncertainty in planning models these Markov Decision Process was introduced.

MDPs generalized the view of classical planning models by introducing the general objective functions called **rewards** which allowed trade offs to be made between transition probabilities and general solution concepts called **policies**.

MDPs can be solved by linear programming or dynamic programming. The algorithm has the following two kinds of steps, which are repeated in some order for all the states until no further changes take place. Their order depends on the variant of the algorithm; one can also do them for all states at once or state by state, and more often to some states than others. As long as no state is permanently excluded from either of the steps, the algorithm will eventually arrive at the correct solution. These variants of the algorithm are the optimization method which helps in determining the optimal policies.

Value Iteration (Bellman 1957)
In this method, the markov property of the MDPs allows exploitation of the dynamic programming principle for optimal policy construction and removes the need to enumerate large number of possible policies. In this the policy function is not used instead the value of the policy function is calculated in the Value function whenever it is needed. In simple terms, we calculate the value of Value function for a state from a guessed value of the Value function for an initial state. And then we iterate it for all states and get the value of Value function for all states until the value function converges.

For a note, Optimal value function in this method is unique but the optimal policy is not. Also, it only works well in finite planning horizon i.e. where the stage of termination is known. Please refer to the references for better understanding.

Policy Iteration (Howard 1960)
In this method, given a fixed policy this can compute its value exactly. Unlike the value iteration function this always provides the optimal policy. This method works well for the infinite planning horizon where the state of termination is unknown. It also works for all three objective functions of infinite planning horizon. This method converges faster(generally), is intuitive, flexible and gives exact value of the optimal policy. Please refer to the references for better understanding.

Prioritized sweeping (Moore & Atkeson 1993)
In this method, the focus is on the “Important” states which are defined using a priority metric and all the predecessors are lined accordingly. Thus, if a state is important then all its predecessors will close to the top of priority queue and if a state is not important then search will focus on other parts of state space.

This method is comparatively new to other methods and gives efficient prediction and control of stochastic markov systems. It uses the Temporal differencing and q-learning which have fast real time performance. It uses all the previous experiences being a memory based method and prioritizes the important states and guides the search.

Most effective optimization method

For most effective method, it largely depends on the kind of the planning problem and the criteria of optimality of that problem i.e. if a planning problem has its criteria of optimality as finite planning horizon, then Value Iteration is the best Method. Whereas if the criteria of optimality is infinite planning horizon then Policy Iteration seems to be the one. But Prioritized Sweeping and Linear Programming can also be used for infinite planning horizon.

References

[1] Dynamic Programming and Markov Processes by Dimitri Bertsekas
[2] Dynamic Programming and Markov Decision Processes by Anders Kristensen
[3] A Markovian Decision Process by Richard Bellman
[4] Slides on Markov Decision Process by Craig Boutilier
[5] Markov Decision Process Wikipedia Page
[6] Prioritized Sweeping by Andrew Moore and Chris Atkeson