19D070052 – Sheel Shah

a. What is the main difference between POMCP on the one hand, and traditional POMDP planning algorithms such as value iteration and policy iteration on the other? On a related note, describe the properties of tasks on which POMCP would be a better choice than the traditional algorithms, and vice versa.

=> POMDP planning algorithms update beliefs and use classic MDP planning algorithms to solve for the belief-MDP using some compact representation. The representation grows exponentially after each iteration. POMCP uses Monte-Carlo sampling, and hence can handle the blowing up of dimension. This is done for both planning as well as belief updates, which were done by actual calculation in PI/VI (belief updates were done by the formulas [Bayes' theorem] seen in last class, and the MDP had to be solved in every iteration, but POMCP approximates both these steps)

b. Provide a summary of POMCP, indicating the main steps that it performs.

=> 1. Action selection is done via MCTS (based on UCB based exploration-exploitation). The search tree nodes are not states, but histories, since POMDPs do not have state information.

2. "we approximate the belief state using an unweighted particle filter, and use a Monte-Carlo procedure" as in the paper. After every interaction (action + observation), MC simulation updates the particles of the filter, where each simulation is begun by a random state chosen as per the current belief.

3. The algorithm is in essence partially observable MCTS, combined with usage of the simulations to update the belief state as well.