Homework #4
Q-Learning

Problem

Description

In this homework you will have the complete RL experience. You will work towards implementing and evaluating the Q-learning algorithm along with a few modifications (eligibility traces) on a simple domain. Q-learning is a fundamental RL algorithm and has been successfully used to solve a variety of decision-making problems. As such, working with it will be a very useful experience. For this homework, you will have to think carefully about algorithm implementation, associated parameters, exploration, evaluation metrics and performance improvements. By the end you should have a deeper understanding of the algorithm, how it interacts with the domain and the overall RL pipeline.

The domain you will be tackling is called Taxi. It is a discrete MDP which has been used for RL research in the past. This will also be your first opportunity to become familiar with the OpenAI Gym environment (https://gym.openai.com/). This is a cool and unique platform where users can test their RL algorithms over a selection of domains, comparing and sharing implementations and results. We will be implementing and testing our algorithm through OpenAI Gym which offers the Taxi environment.

The Taxi problem was introduced in Dietterich(2000). It is a grid-based domain where the goal of the agent is to pick up a passenger at one location and drop them off in another. There are 4 fixed locations, each assigned a different letter. The agent has 6 actions; 4 for movement, 1 for pickup, and 1 for dropoff. The domain has a discrete state space and all of the transitions are deterministic.

Procedure

Implement a basic version of the Q-learning algorithm and use it to solve the taxi domain. The agent should explore the MDP, collect data to the learn the optimal policy and the optimal Q-value function. (Be mindful of how you handle terminal states, typically if St is a terminal state, V(St+1) = 0). Use γ = 0.90.

You can evaluate your agent offline or by uploading your experiment file to the OpenAI server using a GitHub account. The latter will generate a learning curve (reward/steps vs episodes) which is indicative of performance. The OpenAI server will indicate if your implementation has

solved the domain. Note that all evaluations uploaded to the OpenAI server are publicly accessible. **Please do not upload your code online as a gist writeup**.

Examples

Below are the optimal Q values for 5 (state, action) pairs of the Taxi domain. Remember that all states and actions are 0-indexed.

- Q(462, 4) = -11.374402515
- Q(398, 3) = 4.348907
- Q(253, 0) = -0.5856821173
- Q(377, 1) = 9.683
- Q(83, 5) = -12.8232660372

Resources

The concepts explored in this homework are covered by:

- Lectures
  - Convergence
  - Exploring Exploration
- Readings
  - Asmuth-Littman-Zinkov-2008.pdf
  - littman-1996.pdf (chapters 1-2)

Submission Details

Due Date: **Jun 25 11:59 pm (AOE)**

You will be evaluated based on optimality of results. This will be assessed by your algorithm's optimal Q-values for 10 specific state-action pairs (remember to use γ = 0.90). You will submit your results to 10 problems selected for you on the rldm website.

Optionally, you might want to, with the same implementation, solve the environment under OpenAI's criteria.

Were you able to solve the environment and get the optimal Q values with the same implementation?

Dietterich, T. G. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. Journal of Artificial Intelligence Research, 13, 227–303.