Aproaches on Q learning agent for snake with fixed obstacles

Introduction

The project has the goal of implementing a working agent for the game of snake with fixed obstacle position and variable map size using reinforcement learning i.e Q learning. The first agent is implemented using tabular q-learning with experience learning and epsilon-greedy algorithm built completely in numpy, while the second will be using the pythorch library and will be implemented using deep Q-learning with experience learning.

States

I have tried different ways to model the states including: relative position of snake head and tail to apple, snake's head possition in the grid and snake's head positon relative to obstacles and the food. Finally, the best approach was to use a 11 dimension boolean array storing information about the relative position of the head to the food, relative position of head to the obstacles and border of the map and direction of motion. Thus, the tabular version will have 2^11 * 3 = 6144 qtable entries which is acceptable memory-wise.

Rewards

Different type of rewards seem to converge to the same result in the same time. For the deep q learning approach I used a big reward for eating the apple and a negative reward for hitting an obstacle. For the tabular approach, I used positive rewards for getting closer to the apple, negative rewards for going further away and big positive rewards for eating the apple and negative ones for hitting the walls or itself. The agent also recieves a negative reward if the snake has made more than 100 * len(snake) moves without any result.

Actions

I chose to model only 3 actions, continuing the direction of motion and going left/right w.r.t the direction of motion.

Parameters

I chose the discount factor to be 0.9 as we want long term results, learning rate = 0.1 which then turns to 0.05 after 300 matches in the tabular agent.

Results

The deep learning approach got the best results by far, obtaining a mean score of 10 after only 100 played games and reaching a maximum score of 25 after 300 games, while the tabular q learning aproach needs over 300 games to start getting visible result and a mean score of 10.