Report : Reinforcement Learning

Shmuel Naaman

In your report, mention what you see in the agent's behavior. Does it eventually make it to the target location?

The agent eventually make it to the target location, the number of steps that it required increase exponentially as the distance between the start point and the end point increase.

Identify and update state:

The most interesting states is the 'Next waypoint', this state provide with the cab the prefer direction that should be taken if possible.

The next important states are the 'Red-Green light' that restricts the possibilities for the cab. Following the 'Red-Green light' rules will make sure that we do not loose points on the way to the destination.

Another important state is the 'oncoming' that fine tune the 'Red-Green light' and provide us some additional possibilities and might reduce the number of steps needed to arrive at destiny at some cases. The reason that this parameter is not so important is that not many cars are on the road.

What changes do you notice in the agent's behavior?

As we start to use the Q table the number of steps required to arrive at destination reduce considerably. In addition when you are looking at the cab it seems that it chooses the best possible direction when possible. As the number of runs increase, it seems that also the number step to destination decrease.

Report what changes you made to your basic implementation of Q-Learning to achieve the final version of the agent. How well does it perform?

I did not had to change the setting, for the learning rate 'Alpha' = 0.5, and the discount 'Gamma=0.01'. However when choosing direction, if the self state was all zeros, the agent chooses the 'Next waypoint'. In addition to that I set a random function that once in awhile (~ 7 runs) the agent chooses random direction. This is an attempt to prevent the agent to go into a local minimum.

Does your agent get close to finding an optimal policy, i.e. reach the destination in the minimum possible time, and not incur any penalties?

Yes the agent performs well and reaches the destination in the minimum possible time, and not incurs any penalties. When I set the (n_trials=100) the reward_Total = 2460

Ideal or optimal policy would be

1) Sense the environment (see what changes naturally occur in the environment)
2) Take an action - get a reward
3) Sense the environment (see what changes the action has on the environment)
4) Update the Q-table
5) Repeat