TBMI26 – Computer Assignment Reports  
Reinforcement Learning

Deadline – Mars 12 2018

Author/-s:

In order to pass the assignment you will need to answer the following questions and upload the document to LISAM. If you meet the deadline we correct the report within one week after the deadline. Otherwise we give no guarantees when we have time.

1. **Define the V- and Q-function given an arbitrary policy as well as a given optimal policy (See lectures/classes).**

Q-function:

V-function:

1. **Define a learning rule for the Q-function (Theory, see lectures/classes).**

In the beginning we explore more (we use a large ɛ) and then we lower the ɛ and exploit instead.

We can also change the learning rate.

1. **Describe your implementation, especially how you hinder the robot from exiting through the borders of a world.**

We initialize a look-up table for the size of the world (10 \*15) and then depth 4 ( dimension that describes the action). The we let the robot run and learn it’s way. For every 1000th epoch we lover the epsilon so that the robot in the beginning explore a lot and later on it exploits instead.

To avoid the robot bumping into the wall we use a while loop. If the action is invalid a new action is choosen until the action is valid. See following code:

while act.isvalid ~= 1

[a, oa] = chooseaction(look\_up, pos\_prev(1), pos\_prev(2), [1 2 3 4], [1 1 1 1], eps);

act = gwaction(a)

end

1. **Describe the differences between the worlds explored by the robot. Any surprises?**

The obstacles are different in the different worlds. The robot hade some issues with the obstacles and keep going back and forth next to the obstacle.

1. **For each world: Plot the V-function, i.e. how do you get to the goal from each position.**
2. **For each world: describe the key observations you have made with respect to parameter choices. Provide documentation of the parameters you have used for each figure! A good rule is to provide each figure with a caption. Plot policies and the V-function for appropriate worlds to the extent you find appropriate in order to explain what you have done and learned during the assignment.**
3. **What would happen if we where to only use Dijkstra's shortest path finding algorithm in the ''Suddenly Irritating blob'' world? What about in the static ''Irritating blob'' world?**

Then the look-up table would be initialized from us in the beginning (the rewards would already be there), which mean that the robot doesn’t learn by itself, thus the robot can not be better than the teacher.

1. **Include an in-depth description of the to/from HG worlds (world 3 and 4). What happens on the way from HG? How and why can this problem be solved with Q-learning? Which path does the robot prefer, and why?**
2. **Can you think of any application where reinforcement learning could be of practical use? A hint is to use the Internet.**
3. **How does the different parameters () influence learning and appearance of the Q- and V-functions?**