Exercise 5: Eye movements

Objective: The goal of this assignment is to learn to model a reinforcement learning agent and its way of collecting visual information to improve its state representation. The agent that you are modeling is the same as in the exercises 3 and 4. You should solve this assignment individually.

Components: There is only a written component. This is an "open book" exercise, thus you can consult any book, website, etc, during this time. Please do not forget to introduce your references in the section "Bibliography" at the end of the exercise.

Submission: Please solve the assignment directly here and convert it to pdf. Change the name of the file to MI210_Exercise5_Surname_Name.pdf, where Surname and Name is your surname and name. Send me the pdf per email to daniela.pamplona@ensta.fr. The subject of the email must be MI210_Exercise5. Please do not forget to write your name in the header of your report. You should submit until 13H of May 4, 2021

Recommendation:

The book "Reinforcement Learning: An Introduction" from Richard S. Sutton and Andrew G. Barto and the class slides are good points to start when you need help. The slides are available in the Moodle of the class, the book is available here:

https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf

Good luck!

Surname: WANG **Name: Yu**

Group number: E **Engineering system: Domestic robot**

Q1) Can humans move their eyes independently? If yes, give examples. If no, why not?

Generally, it's impossible for humans to move their independently. There is a complex structure at the base of the brain dedicated to making sure that the two eyes move together to focus on the object of interest. If they are not looking at the same thing, the person will be confused.

But independent eye movements may be caused by diseases and abnormalities. For example, conjugate gaze paralysis or strabismus can cause one eye to not fully "follow" the direction of the second eye.

Q2) For the engineering system that you are working on, describe a visually guided task that can be modeled as a reinforcement learning problem and write in detail what would be the set of states, the set of actions, rewards and policy (Please choose something different from obstacle avoidance/ litter collection and the ε-greedy policy). Justify your proposal and point its possible benefits and drawbacks.

In the precedent exercises, my group and I were working on a domestic robot. This robot was designed to do all the cleaning things in the house and help and keep company to people in the house. More specifically, it is important that the robot has to be able to move around the house or the apartment and it also needs indoor classification to recognize objects in the house.

Task:

Find the target, judge whether it needs to be cleaned, clean the object or find the next target.

Set of states:

Moving forward or waiting on a position.

Set of actions:

Identify the object, clean the object, move forward, turn left, turn right, stay still

Rewards:

+10 If the target is in the center of the image, the reward decreases as the object approaches the border of the image. If it is a boundary, it is +0; if the object is missing, it is -2.

Policy:

Greedy policy. The goal is to maximize the Rewards by making sure each action tends to a reward of +10.

The object will not move. If we know its location, we can move forward directly according to its location. As we approach the object, we can determine whether it needs to be cleaned. If we need it, we clean it, if we don't need it, we choose to go to the next target object.

However, one possible problem is that the fish may be willing to hide from our robot. In this case, our robot may be too slow to keep up with the changes in the direction of the fish. In addition, if a fish is surrounded by many other fish that look like him, our robot may not be able to follow the good fish.

Q3) Would you penalize late rewards? What would be the value of \gamma? Why?

We will penalize late reward (gamma != 1). Because at every moment, we want to make sure that in the future the object will be in the center of the camera. Therefore, we need to make sure that the reward at the next step will be high. If we penalized late rewards, the algorithm would try to maximize only the current reward. And we know that the object is still. With late rewards penalized, our robot would be stable, and the recording would have a good quality.

Q4) Design a new policy to control the agent's camera. Please choose something different from minimizing the cost of uncertainty, as defined in the class Explain what would be the advantages and disadvantages of such policy. Justify your proposal and point its possible benefits and drawbacks.

A new policy could be to make sure that the quality of the recording is the best possible. In order to do so, the algorithm could calculate the cost of an action regarding the quality of the video. For example, it is not useful to follow an object that is clean. If it is possible, it could then stop the mission and start following a new object.

The benefits are that our robot don't lose time on getting unusable data and it can use this time to get more data on other species. However, it will spend more time on seeking new objects and it might lead to significantly less data.