Assignment 2: Morris Water Maze Compuational Model

COMP596: Brain-Inspired AI

Solim LeGris

March 29, 2021

1 Part I - Understanding the Foster, Morris and Dayan Paper

1.1 Global consistency

1.1.1 Describe the global consistency problem.

A navigational system such as the hippocampus must solve the problem of computing coordinates that are globally consistent while only having access to local and relative information. This means that it must compute coordinates that are independent of its relative position in space.

1.1.2 What type of navigational system suffers from the global consistency problem? Why is it a problem for them?

A model that assumes that the place cells in the hippocampus become associated with metric coordinates for locations within environments suffers from the global consistency problem. Such models run into trouble because the only information they have available is relative, notably self-motion. Supposing that a path to the goal is first computed by the place cells, all the stored information will be relative to the initial position of the animal. Starting from a new location at some other time, it will have trouble finding its way since the coordinates it has learnt are only consistent relative to the location from which they were initially encoded.

1.2 Distal Reward Problem

1.2.1 Describe the distal reward problem.

A navigational system, in an attempt to reach a goal location, must solve the problem of encoding appropriate action sequences for which the reinforcement signal is distant. Notably, the signal may only exist once the goal location is reached.

1.2.2 What type of navigational system suffers from the distal reward problem? Why is it a problem for them?

A model that assumes that place cells provide the ideal representation for reward-based learning will suffer from this problem since any given place cell usually encodes a shorter length of space than the distance that separates it from its goal. As a result, most place cells cannot directly be associated with appropriate actions to reach a goal location since the reinforcement signal that would allow this is distant from the area of space any given place cell encodes.

1.3 The Computational Model

1.3.1 What is the role of the actor in the Foster et al. (2000) model? Specifically, what is the actor learning to encode?

The actor's role is to provide actions (e.g. directions) to the animal such that they will bring him closer to the goal location. The actor encodes weighted place cell activity thereby enabling it to represent the direction in which the animal should move to get closer to the goal location.

1.3.2 What is the role of the critic in the Foster et al. (2000) model? Specifically, what is the critic learning to encode?

The critic's role is to provide a value of particular locations to the animal by criticizing actions taken by the animal. The critic encodes weighted place cell activity thereby enabling it to represent the value of any particular location in which the animal is within an environment.

1.3.3 What is the error signal used in TD learning? Give the equation for it and explain all the terms/variables in it.

The error signal used for TD learning is the difference between the critic's output at time $t+1$ and at time t added to the reward at time t as follows $\delta_t = R_t + \gamma C(p_{t+1}) - C(p_t)$. R_t is the reward given at time t where $R_t = 0$ unless the animal has reached the goal in which case $R_t = 1$. $C(p_{t+1})$ and $C(p_t)$ are the critic's outputs at times $t+1$ and t. γ is the discounting factor and scales how much the output of the critic at time $t+1$ is considered in the error signal.

1.3.4 What is the role of place cells in the Foster et al. (2000) model? Specifically, what do they encode?

Place cells in the model encode overlapping locations of the environment in which the mouse navigates. They provide a representation that will allow the actor-critic system to learn the appropriate path to the goal.

1.3.5 What sorts of information can place cells not encode that a navigational system might want?

Pace cells cannot encode spatial or navigational quantities such as distance to a goal, direction to a goal or next action to take to get to a goal.

1.4 Discussion

1.4.1 Does TD learning with an actor critic system solve the distal reward problem? Explain why or why not?

The actor-critic system may solve the distal reward problem by learning a value function over place cell activity as mentioned in Foster et al (2000). This value function is approximated by the critic and allows it to criticize the actor's actions leading the system to learn the value of locations such that it will lead it to its goal. The distal reward problem occurs when the reinforcement signal is too distant for a system to use it to associate values to actions and locations that will lead it to find its goal. The actor acts as an intermediate between the actor and the reinforcement signal. Once the value function has been accurately approximated by the critic such that $C(p) \approx V(p)$, then the critic can selectively reinforce actions that will lead the system to the goal.

1.4.2 According to Foster et al. (2000), what information must be available to a rat if it is going to solve the global consistency problem with TD learning? Is it realistic that a rat possess this information?

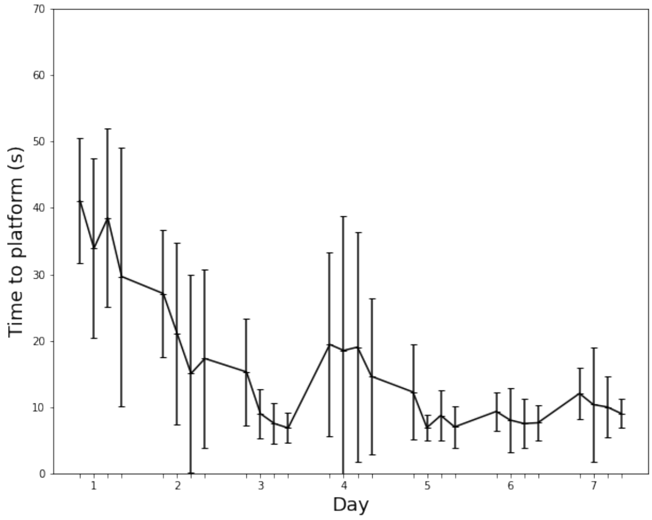
In order for a rat to solve the global consistency problem, it must have access to an allocentric representation of the environment which is a representation that is independent of the rat's own position and point of origin and which can be computed from self-motion estimates, directly available to the animal. Foster et al. (2000) provide evidence from the scientific literature that animals do in fact have access to instantaneous estimates of self-motion. Although I do not believe that this information is unavailable to the animal, I am not convinced that the implementation of self-motion estimates as defined in the paper mirrors how real rats do it.

2 Part II: Implementing the pure TD algorithm

2.1 See code.

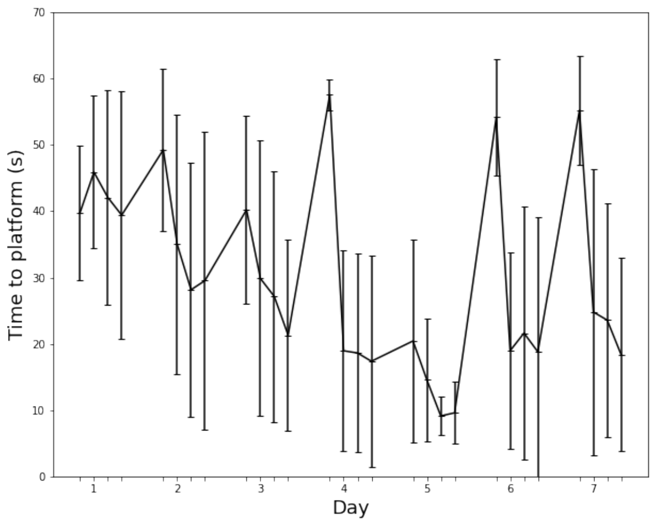
2.2 With a single location does the model successfully reduce latency over trials?

As can be seen in the plot below, the mouse reaches the platform faster as it learns its position.



2.3 What happens to the model when the flag is toggled to a multi-platform case?

As seen in the plot below, the model cannot use the information learnt on previous trials effectively and there is some interference causing the latency to increase drastically everyday before the new location is learnt. This is contrary to what is seen in real mice performing this task. As mentioned in part I, the actor critic model solves the distal reward problem since it learns to a approximate a value function over locations. Nonetheless, when the platform is moved it faces the global consistency problem. Since the representations that it uses to find the goal are egocentric, there is no way for it to find the goal using the information learned on the trials of the day before. This information therefore interferes with learning the new location of the platform causing the increased latency at the first trial of the day and the gradual decrease in latency.

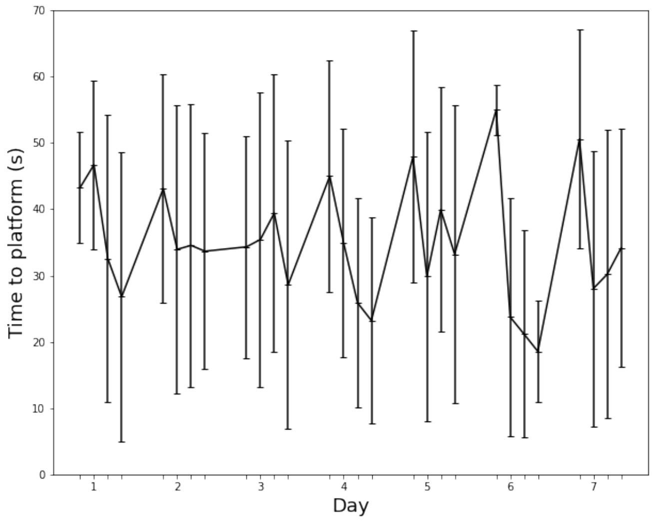


3 Part III: Implementing the combined coordinate and TD algorithm

3.1 See code.

3.2 With multiple platform location, does the model successfully reduce latency over trials?

The model reduces the escape latency over trials but fails to do so better than the pure TD model for some reason that I could not decipher. This is possibly because the hyper-parameters that I chose are not optimal and as such further optimization is required. Additionally, I was not able to assure that overflow is handled correctly since the source of this problem was difficult to determine. This might also cause the model to perform sub-optimally.



3.3 What is your opinion of the coordinate model? Does it seem like a good strategy for an AI?

The coordinate model as presented by Foster et al. has a few issues. For one, it is unclear how exactly the model must compute self-motion as it is not defined accurately in the paper. In my implementation, the model uses the angle of the action taken and computes x and y coordinates using step size as the norm. Putting self-motion computation considerations aside, the model itself is very simplistic. The environment in which it operates has no obstacles and is 2-dimensional. Furthermore, the model does not makes use of sensory inputs from a visual sensory system or other sensory information which should be crucial to navigate any environment. Sensory information could for example increase the ability of the model to generate allocentric maps through the use of landmarks. Using this information, an AI model could learn the spatial relations between objects in its environment and its goal, inevitably leading it to reach its goal more efficiently and more flexibly.