My masters

Simen Storesund

January 2, 2024

1   Introduction

Introduction One of the baffling developments in robotics and computer navigation has been how seemingly simple animals perform superiorly to any human-made construct in path finding and understanding the underlying structure in new and familiar environments. Although remarkable feats of navigation are found in nature, such as the continental flights of trekking birds or salmon returning to their home rivers after spending their lives roaming the ocean, robots struggle with the more ordinary tasks of making their way through simple environments such as mazes or obstacle courses. This type of navigation is critical for expanding upon navigation technologies such as self-driving vehicles or appliances such as robotic vacuum cleaners or lawn mowers.

Meanwhile, neuroscientific approaches have made good progress into what appears to be the navigational system in rodents. It has long been known that neurons that respond significantly to single locations in the environment, called place cells, are found in the rodent hippocampus, a brain area which is instrumental in the formation of new episodic memories. These cells exhibit a series of interesting features - during sleep, sequences of place cells that were activated during awake time are activated in a sped-up manner, a phenomenon coined replay. During navigation, an array of place cells activate, ranging from those that were just activated in past positions, to the current position and including place cells that predict future behavior. This activity, called preplay, seems to indicate that the place cell system is either directly involved with, or sits downstream of, navigational systems that compute future positions in a goal-oriented way.

It is suggested that a network of place cells either in CA1 or CA3 can underlie a cognitive map of the environment, which can encompass all locations an animal has previously visited. This is the kind of map that might underlie navigation systems, which can help path planning by producing sequences of place cells leading from some start location to a goal location. Computer algorithms have long been able to find shortest paths in graph networks consisting of connected nodes. Typically, each node in a graph network represents one location, and is connected to its neighboring nodes by some weight, which can represent distance or effort required to get from the node to the neighbor. Dijkstra's algorithm

finds the shortest path from a start to a goal by effectively branching outwards from the start node, and records, for every node it gets to, the second to last node it would get there by. Then, when it finds the goal neuron, it will iterate backwards over the second-to-last nodes from the goal to the start, and thus gets the inverse of the sequence. The brain is probably unable to utilize such a method, because of the demand it places on separate random-access memory structures that are biologically implausible.

Instead, multiple approaches have been attempted that use reinforcement learning to find paths in an unknown environment. In most of these approaches, the place cell network is seen as a recurrently connected neural network, in which neighboring place cells are connected by some random, unbiased weight. The goal is to introduce a learning rule that lets random activity expanding from the start eventually converge to a single sequence effectively from start to goal. The criteria for such a learning rule is that it can be proven to converge to some path, given enough time, and that it works under biologically plausible rules. Within the framework of reinforcement learning, using eligibility traces has been a popular approach, which implies that each neuron or synapse has some hidden trace, whose strength typically increases after neuron activation and decays over some time scale. The degree of learning in the neuron depends on some global reward signal, which indicates how well the system performs, along with this trace, solving the credit assignment problem. Some approaches here involve simply trying a series of single sequences, based on a monte carlo random walk, and in which each path is graded based on how fast the goal is reached. These methods can be proven to converge, but they require a lot of iterations. For the kind of sequencing animals perform, rate of convergence is also critical. Ideally, sequences should be found quickly, but also unlearned, so the network can form other sequences without being overfitted to past sequences. For example, if a mouse considers whether to navigate to one place in a maze, which involves a left-turn in a T-intersection, and then changes its mind to a goal in which going right is faster, the navigation network should flexibly adapt to avoid a bias to paths involving a left-turn.

Another interesting question in such a network is which variable is learnt over. It is typically assumed that the learning variable is the synapse strength, or weight, between neurons. This is possible both for long-term and short-term learning. Another possibility, which has been tried already, is that the changing variable is activation delay, which changes the time it takes from a place cell receiving an input until producing an output. One model designed specifically to address the question of navigation in the context of place cells is the TSS model. This model assumes that transition information, which would connect nodes in a Dijkstra-like network, is stored in a separate network, and shows that cells in this network would be optimized by firing in grid-like patterns, at least in square or circular environments. In addition, to optimally accelerate the rate of convergence, grid cells should scale by a factor of the square root of two. As such, the TSS model makes a compelling argument connecting optimal navigation and neuroscientific findings. Under this model, place cells don't directly activate neighboring place cells during mental navigation. Instead, they act

through mediating grid cells, which connect them to surrounding place cells. As such, for navigational tasks, the TSS model assumes that place cell - to grid cell weights and grid-cell to place cell weights were already learnt during exploration. Activation time is then a compelling learning variable during sequence-extraction. Instead of finding sequences from start to goal by trying random sequences, the TSS-model instead posits an outwards propagation of place cell activity from start, until the goal is found. This can be thought of as trying all possible paths from the start simultaneously, but it complicates the credit assignment problem. To enable learning, reaching the goal must provide information to some place cells whether they contributed to reaching the goal or not. The goal of this work is to implement a reinforcement-learning approach that unites the computational TSS model with state of the art learning rules using eligibility traces. Additionally, it seeks to answer how the multiple grid scales can be used to stabilize and accelerate retrieval of place cell sequences, and includes concrete predictions for how a network would function under this assumption.