Double Q-learning and Deep Q-learning with a Working Memory Concept

David Ludwig - dwl2x@mtmail.mtsu.edu
Advisor: Dr. Joshua Phillips

Abstract

Working memory is a part of our brain's memory system that temporarily retains a small amount of information which we use to accomplish tasks. Using a framework called the Working Memory Toolkit, we can easily model working memory based on cognitive neuroscience models using reinforcement learning (RL). The toolkit implements RL traditionally on a single-layer neural network, and the conceptual information encoded as holographic reduced representation (HRR) vectors can get quite large. In more complicated problems, learning can also be unstable and tricky to converge due to "randomness" in the environment, requiring additional effort to solve. By replacing the single-layer neural network with a multi-layer neural network using deep Q-learning, we show that the size of the HRR vectors can be reduced while retaining a reliably learning ability. The gathered evidence also suggests that double Q-learning reduces the noise in the learned Q-functions from exploratory actions. By incorporating combinations of these algorithms into the WMtk, the memory usage required for the tasks can be reduced while still learning and further stabilizing the desired Q-function; thus, leading to potential advancements in working memory modeling.

Introduction

- The Working Memory Toolkit (WMtk) is a framework that allows easy integration of working memory into artificial intelligence (AI) agents [3].
- Q-learning is a popular temporal-difference learning (TD-learning) algorithm that is specifically designed to learn the Q-function, the function that calculates the expected reward given a state [5, 7].
- Double Q-learning is a variation to the Q-learning TD-learning algorithm that aims to reduce the learning overestimation that traditional Q-learning is prone to do in some stochastic environments [2].
- Deep Q-learning is another variation of the Q-learning algorithm that implements it on a multi-layer neural network. The deep neural network is capable of learning more sophisticated functions than single-layer neural networks [6].

Specific Aims

- Implement and integrate double Q-learning and deep Q-learning into the Working Memory Toolkit model
- Show that deep Q-learning can reduce the size of the HRR vectors required for reliable learning
- Show that the integration of double Q-learning can stabilize learning by reducing noise and increase learning reliability in stochastic environments

Methods

- As this is a standalone project implemented outside of the actual working memory toolkit, a small reinforcement learning framework was constructed to allow quick and easy manipulation of the provided Q-learning algorithm. Everything was implemented using Python and Keras (with Tensorflow backend) in a Jupyter Notebook.
- In order to test the new models, a simple 1D maze problem was built up. On each episode, the agent is placed at a random position in the maze and must navigate to the specified goal position by moving either left or right. A epsilon-greedy approach was used for exploratory actions. This task is illustrated in Figure 1.

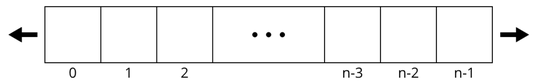

Figure 1: Maze problem diagram

- To incorporate the usage of working memory, the agent is presented with a signal on the first step of an episode which indicates what the goal position is. The agent must learn to hold onto this signal in memory so that it may find the goal. This gives the agent two additional actions that it must consider when making a move: **store** the signal that is currently in the environment (could be no signal at all), or **protect** what is currently in memory.
- To test the learning ability of the deep Q-learning model, the size of the HRR was simply reduced such that the deep Q-learning model retained a reliable learning ability. Assuming our predictions are correct, using this HRR size with the traditional model should result in unreliable learning behavior or even complete failure to learn the function.
- Double Q-learning should result in more stable learning and reduce the effects of noise, especially in more stochastic environments. By adding "randomness" to the reward schedule, this effect can be simulated and compared.

Intermediate Results & Conclusions

- The implementation of deep Q-learning into the working memory model resulted in an immediate improvement to the ability/reliability of learning when working with smaller HRR vectors. The deep Q-learning implementation used included a single hidden layer equivalent to half the size of the HRR vectors. This additional layer granted the neural network the ability to learn a more complex function, thus allowing it to distinguish the HRR vectors more easily. Figure 2. Shows a comparison of the results of Q-learning vs. deep Q-learning after 10,000 episodes with the same learning parameters. In testing, deep Q-learning managed to reliably learn the optimal Q-function with the smaller HRR vectors while standard Q-learning either struggled greatly or failed entirely. In practice, it was found that the HRR size could be reduced by nearly fifty percent, while retaining a reliable learning ability. This could allow the inclusion of additional conceptual information within the same memory space.
- While inconclusive, evidence suggests that the inclusion of double Q-learning reduces noise in the learning process, and therefore learns the Q-function more reliably. Due to time constraints within this project, support for this claim has not yet been properly gathered. It was recognized during experimentation however that with double Q-learning the noise from exploratory actions was reduced, allowing the agent to "learn" the optimal function in fewer episodes, whereas traditional Q-learning required either additional episodes or an annealing schedule for epsilon to "learn" the function.

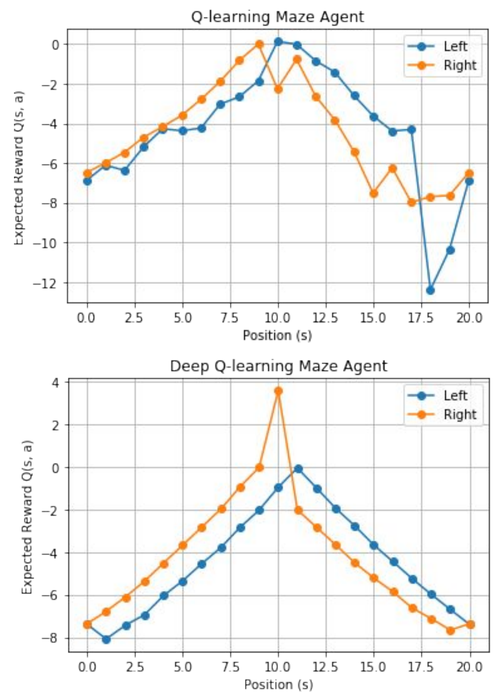

Figure 2: Standard Q-learning vs. deep Q-learning with smaller HRR vectors
Maze Size: 20; HRR Size: 64, learn rate: 0.05, epsilon: 0.2, discount: 0.95

Background

- To encode conceptual information for the neural networks, the WMtk uses HRR vectors to automate the encoding process[1]. Every single information concept is represented by a unique HRR vector. Concepts can be combined through the means of circular convolution to produce a new unique HRR vector that is orthogonal to the original vectors [4], allowing the neural network to learn and make predictions on that combined concept. This idea allows a single neural network to learn many tasks, and adapt to new tasks without changing the structure of the net. This model has been been labeled as "N-task."
- While HRR vectors are very powerful in that sense, they grow in size very quickly. Even in problems that require a small number of state-spaces require a large HRR vector size to retain orthogonality after the circular convolutions [1, 8]. One approach to mitigate this issue is to incorporate the deep Q-learning algorithm into the WMtk. The WMtk traditionally uses a single-layer critic network to evaluate the actions the agent can take. Deep Q-learning implements Q-learning on a multi-layer neural network [6]. By replacing Q-learning with deep Q-learning, the agent could retain the learning ability and reliability while reducing the size of the HRR vectors.
- Another variation of Q-learning is double Q-learning. Double Q-learning aims to solve the issues of overestimation that can occur in certain stochastic environments using traditional Q-learning. These overestimations can cause often cause difficulties when learning more complicated tasks, or when there is randomness in the reward schedule [2]. Double Q-learning works by incorporating a second Q-function. Rather than updating itself, a Q-function that makes a prediction updates the expected value on the other Q-function [2]. This behavior could stabilize and reduce noise that occurs in traditional Q-learning.

References

1. DuBois, G. M. and Phillips, J. L. Working memory concept encoding using holographic reduced representations. In Proceedings of the 28th Modern Artificial Intelligence and Cognitive Science Conference, 2017.
2. Hasselt, H. V. Double q-learning. In Lafferty, J. D., Williams, C. K. I., Shawe-Taylor,J., Zemel, R. S., and Culotta, A., editors,Advances in Neural Information Processing Systems 23, pages 2613–2621. Curran Associates, Inc., 2010.
3. Phillips, J. and Noelle, D. Working memory for robots: Inspirations from computational neuroscience. 01 2006.
4. Plate, T. A. Holographic reduced representations.IEEE Transactions on Neural Networks, 6(3):623–641, May 1995.
5. Sutton, R. S. Learning to predict by the methods of temporal differences.Machine Learning, 3(1):9–44, Aug 1988.
6. van Hasselt, H., Guez, A., and Silver, D. Deep reinforcement learning with double q-learning. CoRR, abs/1509.06461, 2015.
7. Watkins, C. J. C. H. and Dayan, P. Q-learning. Machine Learning, 8(3):279–292, May 1992.
8. Williams, A. S. and Phillips, J. L. Multilayer context reasoning in a neurobiologically inspired working memory model for cognitive robots. In Proceedings of the 40th Annual Meeting of the Cognitive Science Society, 2018.