Information Processing and the Brain CW 2

Kheeran Naidu

December 2019

Introduction

For CW2, I have decided to implement and explore the behaviour of the classical *Reinforcement Learning* algorithm using *Temporal Difference Learning*. Reinforcement learning was developed from an area of psychology of animal lerning under the name trial-and-error learning [1] and the area of the optimal control problem - using value functions and dynamic programming - in the form of Markov Decision Processes [2][3]. Temporal-difference (TD) methods for reinforcement learning were proposed as a model of classical (or Pavlovian) conditioning in 1987 [4], and refined to the TD learning rule in 1990 [5]. Put simply, reinforcement learning is the computational method for goal-oriented learning and the TD learning rule is one which is based on the difference in the current state value and next state value.

Question 1

The reinforcement learning algorithm is based on the formal framework of Markov decision processes.

[[’’ ’’ ’’ ’’ ’’ ’’ ’’ ’’] [’’ ’•’ ’•’ ’•’ ’•’ ’•’ ’’ ’’] [’’ ’•’ ’E’ ’’ ’’ ’•’ ’’ ’’] [’’ ’’ ’•’ ’’ ’’ ’•’ ’•’ ’’] [’’ ’’ ’•’ ’’ ’’ ’’ ’’ ’’] [’’ ’•’ ’’ ’’ ’’ ’’ ’’ ’’]]

References

[1] Robert Sessions Woodworth. "Experimental Psychology. New York: Holt, 1938". In: *Department of Psychology Dartmouth College Hanover, New Hampshire* (1937).

[2] R. E. Bellman. "A markov decision process. journal of Mathematical Mechanics". In: (1957).

[3] R. E. Bellman. "Dynamic programming, princeton univ". In: *Prese Princeton, 1957* (1957).

[4] Richard S Sutton and Andrew G Barto. "A temporal-difference model of classical conditioning". In: *Proceedings of the ninth annual conference of the cognitive science society*. Seattle, WA. 1987, pp. 355–378.

[5] Richard S Sutton and Andrew G Barto. "Time-derivative models of pavlovian reinforcement." In: (1990).