Haoming Liang
1430396
haoming3

Eligibility trace is a new parameter introduced to combine Monte Carlo and Temporal Difference methods into TD($\lambda$), in the way of changing the simple return G from either method into the compound return (lambda-return), defined as $G_t^\lambda = (1 - \lambda) \sum_{n=1}^{T-t-1} \lambda^{n-1} G_{t:t+n} + \lambda^{T-t-1} G_t$

where $\lambda \in [0, 1]$. In other words, $\lambda = 0$ stands for TD (0), and $\lambda = 1$ stands for Monte Carlo.

Comparing with Monte Carlo and n-step TD, eligibility trace performs better computationally in an algorithmic sense. Monte Carlo and n-step TD updates state value using future rewards which have not yet determined at the current time step. However, eligibility trace uses the TD error in the current time step and review previous states, which have already been computed, to determine the weight between the two returns.

Thought Question:

Through the process of episodes, the λ parameter behaves fluctuating according to the previous computations. Is there a way to find fixed value which makes the program performs best? Compared with fluctuating λ, how is the performance of a fixed value λ?

Answer:

From the example RMS error analysis in the textbook, different λ combined with different α has different performances, each use of α has its own best λ. However, to find the best fixed value of λ for each specific task is computationally unworthy.