Review-22

The Act Of Remembering: A Study In Partially Observable Reinforcement Learning

Learning memoryless policies hinders Reinforcement Learning (RL) agents to demonstrate long-term suitable performance in partially-observable settings. Lack of information obtained from the environment necessitates the need for external memory. The work presents a study of why and when memoryless agents fail in partially-observable settings. To this end, the work throws light on theoretical aspects of meory utility and proposes a novel scheme which provides the agent with external memory. With sufficient expressivity of external memory, agent policies converge to globally optimal solutions and demonstrate suitable performance on long-term partially-observable scenarios.

Requirement of external memory in partially-observable setting sis highlighted upon examining the lack of retrieval of long-term information by the agent. The proposed schemes (Ok and OAk) present the agent with external memory which results in a meory-augmented environment. In contrast to using conventional memroy-based methods, Ok and OAk memor modules combine the memory information with state of the agent and expand its action space respectively. This allows the agent to learn an efficient memoryless policy which only does not rely on internal memory mechanisms. Ok and OAk memory modules are a generalization of k-order memory schemes which are buffers of fixed k-size bits. While k-order memories do not allow storage of information beyond k-steps, Ok memories address this problem by allowing the agent to select whether to push the experience in memory or not.

In comparison to contemporary memory modules, Ok and OAk memories demonstrate suitable performance on a suite of partially-observable tasks. Additionally, generalization of memory modules to variable RL schemes such as multi-step actor-critic, TD methods, Sarsa(λ) and DDQN demonstrate their potential for practical utilization. However, the study presents two main caveats with respect to its analysis. Experiments demonstrate the suitability of Ok and OAk for various different values of k steps. The study does not provide insights on a suitable value of k or a general selection criterion which would aid in further understanding when Ok and OAk memories may not be sufficient. Lastly, realizing memory modules is sufficient for small-scale navigation tasks but may not necessarily scale to large-scale partially-observable settings. The work does not throw light on large-scale scenarios consisting of significant partial-observability. A consistent evaluation scheme for assessing the agent with variable degrees of observability may yield insights into the hindrances faced by agent.