

Cautious Adaptation For Reinforcement Learning in Safety-Critical Settings

Real-world settings require the Reinforcement Learning (RL) agent to cautious yet optimal behavior. To that end, the work presents a Safety Critical Adaptation (SCA) framework which allows the agent to adapt to scenarios by executing a safe policy. The framework trains the agent in a non-safety critical simulation setting in order to adapt to safety-critical scenarios where failures penalize the agent heavily. The work further proposes a solution based on SCA framework in order to yield risk-averse policies for cautious adaptation. Cautious Adaptation in Reinforcement Learning (CARL) makes use of model-based RL to capture uncertainty and estimate risk followed by planning in sfety-critical settings to avoid catastrophic states. CARL empirically demonstrates risk-averse behavior with fewer failures in comparison to RL baselines.

In most real-world scenarios, agents require safe behavior which would enable risk-averse policies in unforeseen tasks. Such a kind of knowledge transfer results in safe and data efficient adaptation of the agent in the new environment. SCA framework allows safe adaptation in RL by pretraining the agent in a setting where safety is not a concern. Following pretraining, the agent is allowed to adapt to safety-critical settings where catastrophic behavior is heavily penalized. In order to dmeonstrate the suitability of safe adaptation using SCA, the work proposes CARL, which is based on the intuition that agents can estimate risk and reason about uncertainty while learning from a variety of scenarios. CARL is pretrained using an ensemble of probabilistic models on various sandbox environments where safety is not a concern. The framework is combined with PETS but the method is generalizable to any model-based approach. Following pretraining, CARL is made to adapt to safety-critical scenarios where it plans to select actions leading to risk-averse states. Planing is carried out using a novel reward computation scheme which assigns an action score to each planned trajectory. CARL minimizes risk using two variants. Firstly, the agent may seek high rewarding states in order to minimize the risk of low rewards. This is governed using a caution parameter. Secondly, the agent may minimize the risk of reaching catastrophic states. State risk-aversion is based on minimizing the proability of critical states serving as reward penalty.

CARL is evaluated on 4 risk-prone scenarios wherein the agent needs to adapt by seeking risk-averse policies. While CARL depicts improved performance in comparison to baseline methods, it often fails to generalize to Out-of-Domain samples. For instance, CARL presents the maximum number of pole falls and boundary collisions in CartPole and DuckieTown tasks respectively. Sub-optimal risk aversion may be the result of an inaccurate model due to its pretraining on non-critical safety settings. Additionally, the method does not provide intuition on the tradeoff between risk and reward while executng cautious behavior.

CARL exhibits risk-averse policies which enable the agent to restrict visitation of catastrophic states. The work