Accelerate Reinforcement Learning with ProhibitiveBoundaries

Zongqiang Pang,Liping Bai *Member, IEEE,*

***Abstract*—As human children pick up new skills, the parents rarely just let them try things out themselves. Grown ups would guard toddlers from falling as they try to figure out how to walk; training wheels are installed on bicycles to assist teenagers learn how to balance the apperatus. This kind of failure assistance can be expressed as prohibition boundaries with punishment in the reinforcement learning context. The purpose of such a boundary is to prevent termination of training episodes while provide negative feedbacks. In this paper, we demostrate out idea of prohibition boundary through a series of experiments with OpenAI Gym environment. The reinforcement learning agents are based on tensorforce, which is the latest implementation of nearly all the RL agents written in PyTorch. As the data suggest, our method can accelerate learning considerably if the parameters for prohibition boundaries are set correctly.**

***Index Terms*—Reinforcement Learning, Assisted Reinforcement Leraning, Reward Engineering**

I. Introduction

REINFORCEMENT learning is the process of methotically extracting information from observations to gradually bound the policy distribution, either directly through policy gradient methods or via scaffolding measurements, maximizing the expected reward along a trajectory. Considerable progress has been made in applying reinforcement learning to robotics. Pierter Abeel... EHT..ADD PAPRES HERE. However, one of the problems facing reinforcement learning based control is how to gather enough real world data when things are expensive. Imagine every piece of information is collected with crashed pricy helicoptors. Simmulation is clearly one way to circumvent it, PAPERS. Yet, there are still residual dynamics that evade simulation and can only be collected via physical experience, with potential monatary loss.

In this paper, we suggest an easy to implement methods to provide training assitance to reinforcement learning agent. Training assistance is a common occurance in athletic practice. For instance, figure skaters learn how to spin 4.5 resolutions on protective wires before they do so on ice; gynmnists practice their cartwheels on forms before they do so on the balance beam. Without such protective schemes, it is doubtful that there would be any athelets still alive. The same is true for children, without those helping hands from parents and training wheels, most children would be severely injued as they learn how to walk or bike.

Nanjing Unversity of Posts and Telecommunications, College of Automation & College of Artificial Intelligence, Nanjing, Jiangsu,210000 China email:zqpang@njupt.edu.cn

II. Where to Approximate

III. System Modelling

IV. Policy Gradient

V. Approximate Dynamic Programming

Control theoriests are familiar with Approximate Dynamic Programming, it

A. Value Iteration and Policy Iteration

B. Maximum Entropy Reinforcement Learning

VI. Other Ways to Utilize Increased Computational Power

A. Multiagent Formulation

B. Random Shoot