

# Extending the scope of reliable offline reinforcement learning

2023-08-21 14:23:49

## Authors

Marnix Suilen

## Abstract

Safe policy improvement (SPI) is an offline reinforcement learning (RL) problem. Specifically, given only a finite data set and the behavior policy that collected data, the SPI problem is to find a new policy that outperforms the behavior policy with a formal guarantee. We extend the scope of SPI in two fundamental ways: 1) We present the first approach to SPI in partially observable environments (POMDPs), and 2) We significantly reduce the amount of data needed to derive said formal guarantee.

## Keywords

reinforcement learning, safe policy improvement, POMDPs, PAC