OTTAWA 2023
64TH WORLD STATISTICS CONGRESS

CPS Paper

Distributional Method for Risk Averse Reinforcement Learning

Author: Dr Ziteng Cheng

Coauthors: Ziteng Cheng, Sebastian Jaimungal, Nick Martin

Submission ID: 660

Reference Number: 660

Presentation File

abstracts/ottawa-2023_132ddd5a9b77c055e337e7daaff72258.pdf

Brief Description

We introduce a distributional method for learning the optimal policy in risk averse Markov decision process with finite state action spaces, latent costs, and stationary dynamics.

We assume sequential observations of states, actions, and costs and assess the performance of a policy using dynamic risk measures constructed from nested Kusuoka-type conditional risk mappings.

For such performance criteria, randomized policies may outperform deterministic policies, therefore, the candidate policies lie in the d-dimensional simplex where d is the cardinality of the action space.

Existing risk averse reinforcement learning methods seldom concern randomized policies, naive extensions to current setting suffer from the curse of dimensionality.

By exploiting certain structures embedded in the corresponding dynamic programming principle, we propose a distributional learning method for seeking the optimal policy.

The conditional distribution of the value function is casted into a specific type of function, which is chosen with in mind the ease of risk averse optimization.

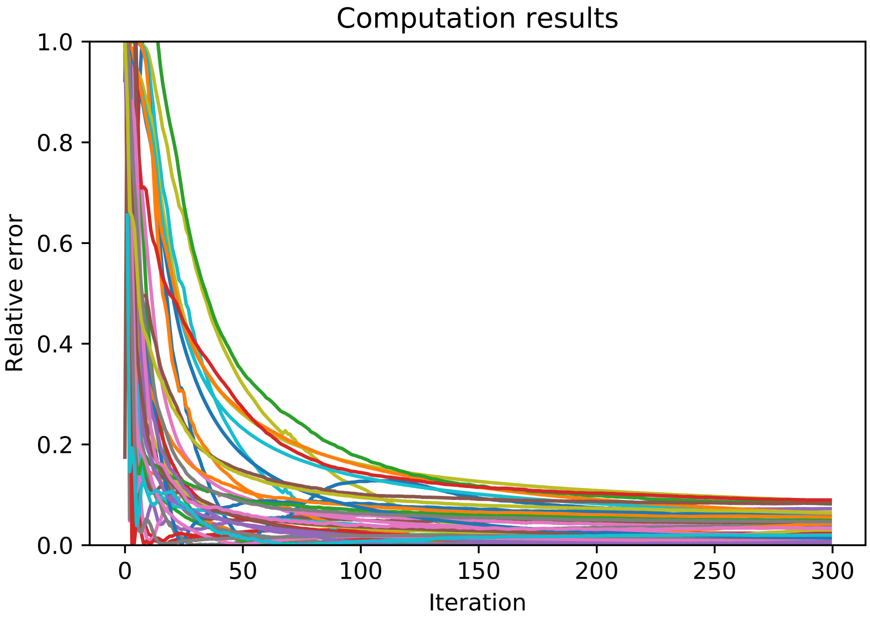
We use a deep neural network to approximate said function, illustrate that the proposed method avoids the curse of dimensionality in the exploration phase, and explore the method’s performance with a wide range of model parameters that are picked randomly.

Abstract

We introduce a distributional method for learning the optimal policy in risk averse Markov decision process with finite state action spaces, latent costs, and stationary dynamics. We assume sequential observations of states, actions, and costs and assess the performance of a policy using dynamic risk measures constructed from nested Kusuoka-type conditional risk mappings. For such performance criteria, randomized policies may outperform deterministic policies, therefore, the candidate policies lie in the d-dimensional simplex where d is the cardinality of the action space. Existing risk averse reinforcement learning methods seldom concern randomized policies, naive extensions to current setting suffer from the curse of dimensionality. By exploiting certain structures embedded in the corresponding dynamic programming principle, we propose a distributional learning method for seeking the optimal policy. The conditional distribution of the risk-to-go is casted into a specific type of function, which is chosen with in mind the ease of risk averse optimization. We use a deep neural network to approximate said function, illustrate that the proposed method mitigates the curse of dimensionality in the exploration phase, and explore the method’s performance with a wide range of model parameters that are picked randomly. This is a joint work with Sebastian Jaimungal and Nick Martin.

Figures/Tables

v


Computation results
1.0
0.8
0.6
0.4
0.2
0.0
Relative error
0
50
100
150
200
250
300
Iteration