Multi-Objective RL With Multi-Agent Bidding

Anonymous authors
Paper under double-blind review

Keywords: Multi-Objective RL, Multi-Agent RL

Summary

The summary appears on the cover page. Although it can be identical to the abstract, it does not have to be. One might choose to omit the stated contributions in the Summary, given that they will be stated in the box below. The original abstract may also be extended to two paragraphs. The authors should ensure that the contents of the cover page fit entirely on a single page. The cover page does **not** count towards the 8–12 page limit.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Contribution(s)

1. Provide a succinct but precise list of the contribution(s) of the paper. Use contextual notes to avoid implications of contributions more significant than intended and to clarify and situate the contribution relative to prior work (see the examples below). If there is no additional context, enter "None". Try to keep each contribution to a single sentence, although multiple sentences are allowed when necessary. If using complete sentences, include punctuation. If using a single sentence fragment, you may omit the concluding period. A single contribution can be sufficient, and there is no limit on the number of contributions. Submissions will be judged mostly on the contributions claimed on their cover pages and the evidence provided to support them. Major contributions should not be claimed in the main text if they do not appear on the cover page. Overclaiming can lead to a submission being rejected, so it is important to have well-scoped contribution statements on the cover page.
   Context: None
2. The submission template for submissions to RLJ/RLC 2025
   Context: Built from previous RLC/RLJ, ICLR, and TMLR submission templates
3. *[Example of one contribution and corresponding contextual note for the paper "Policy gradient methods for reinforcement learning with function approximation" (Sutton et al., 2000).]*
   This paper presents an expression for the policy gradient when using function approximation to represent the action-value function.
   Context: Prior work established expressions for the policy gradient without function approximation (Williams, 1992).

Multi-Objective RL With Multi-Agent Bidding

Anonymous authors
Paper under double-blind review

Abstract

1 Introduction

2 Multi-Agent Bidding Mechanism for Multi-Objective RL

Definition 1 (MO-MDP). *A multi-objective Markov decision process (MO-MDP) with $m \in \mathbb{Z}_{>0}$ objectives is specified by a tuple $\mathcal{M} = (S, A, T, R, \mu_0)$, where*

- *S is the set of states,*
- *A is the set of actions,*
- *$T : S \times A \to \mathbb{D}(S)$ is the transition function mapping a state-action pair to a distribution over states,*
- *$R : S \times A \times S \to \mathbb{R}^m$ is the reward function with each output component corresponding to the different objectives, and*
- *$\mu_0 \in \mathbb{D}(S)$ is the initial state distribution.*

A *policy* in an MO-MDP $\mathcal{M}$ is a function $\pi : (S \times A)^* \times S \to \mathbb{D}(A)$ that maps a history of state-action pairs and the current state to a distribution over actions.

Definition 2 (MAB-MDP). *Let $\mathcal{M} = (S, A, T, R, \mu_0)$ be an MOMDP with m objectives and let $b \in \mathbb{Z}_{>0}$ be the bid upper bound. Also define $B = \{0, \ldots, b\}$ to be the range of bids. We can define the corresponding multi-agent bidding Markov decision process (MAB-MDP) as a tuple $\mathcal{B} = (S, \hat{A}, \hat{T}, \hat{R}, \mu_0)$ where*

- *$\hat{A} = (A \times B)^m$ represents the action space of the m agents in which each agent selects an action from A and a bid from B.*

References

R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems 12*, pp. 1057–1063, 2000.

R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992.