A UNIFIED UNCERTAINTY-AWARE EXPLORATION: COMBINING EPISTEMIC AND ALEATORY UNCERTAINTY

UNIVERSITY OF TORONTO

Parvin Malekzadeh*[1], Ming Hou[2], Konstantinos N. Plataniotis[1]

[1] The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada

[2] Defence Research and Development Canada (DRDC) Toronto Research Centre, Canada

* p.malekzadeh@mail.utoronto.ca

ICASSP 2023.
Paper ID: 1691

Abstract

Exploration is a significant challenge in practical reinforcement learning (RL), and uncertainty-aware exploration that incorporates the quantification of epistemic and aleatory uncertainty has been recognized as an effective exploration strategy. We propose an algorithm that

- ❑ clarifies the theoretical connection between aleatory and epistemic uncertainty;
- ❑ unifies aleatory and epistemic uncertainty estimation; and
- ❑ quantifies the combined effect of both uncertainties.

Experimental results demonstrated that our method achieves substantial improvements in stability and sample efficiency compared to existing frameworks that only consider aleatory uncertainty, epistemic uncertainty, or an additive combination.

Distributional RL

- ❑ The agent-environment interaction is modelled with a Markov Decision Process (MDP):

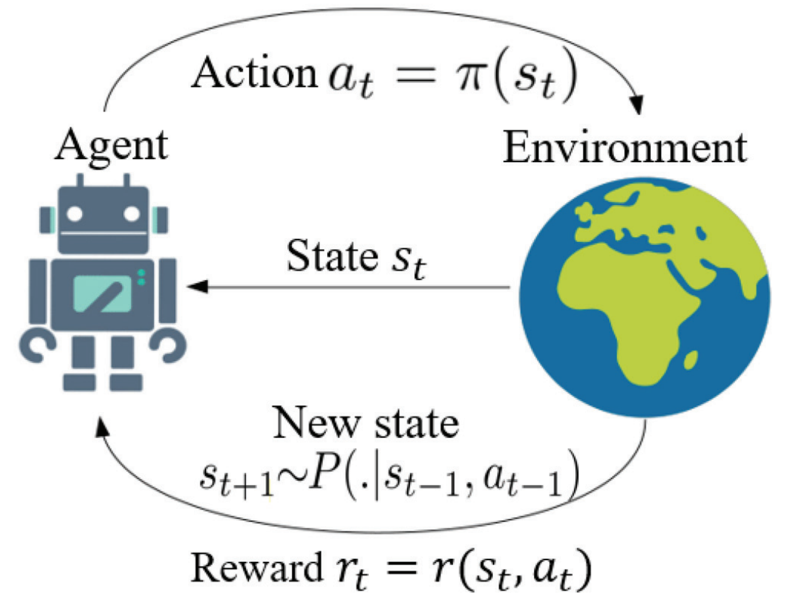


Fig.1: Agent-environment interaction modelled as an MDP.

- ❑ Distributional RL learns $p\big(Z^\pi(\boldsymbol{s}_t, a_t)\big)$, where:

$$Z^\pi(\boldsymbol{s}_t, a_t) = \sum_{k=t}^{\infty} \gamma^{k-t} r(\boldsymbol{s}_k, a_k)$$

- ❑ Distributional Bellman equation:

$$p(Z^\pi_{\boldsymbol{\theta}}(\boldsymbol{s}_t, a_t)) = r_t + \gamma\, p(Z^\pi_{\boldsymbol{\theta}^-}(\boldsymbol{s}_{t+1}, a_{t+1}))$$

- ❑ $p\big(Z^\pi(\boldsymbol{s}_t, a_t)\big)$ captures aleatory uncertainty.

? How can we incorporate epistemic uncertainty about $\boldsymbol{\theta}$?

Proposed Method

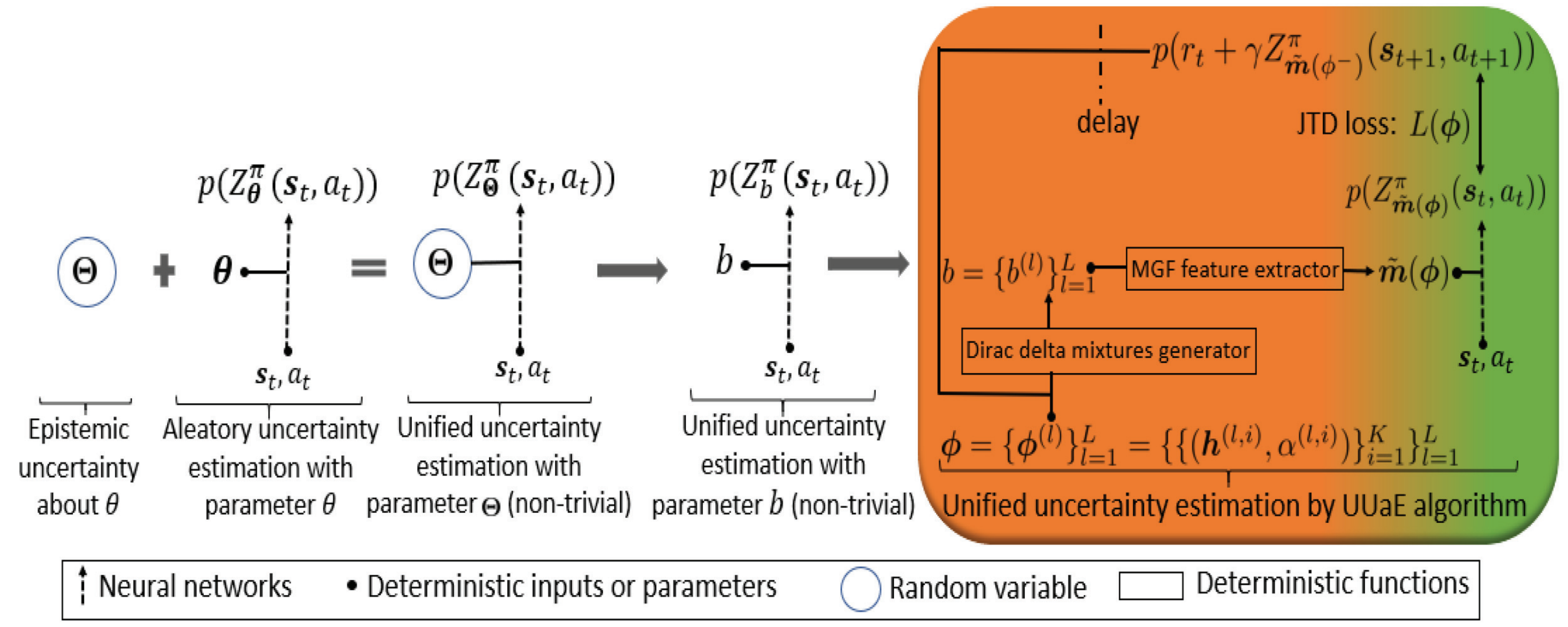


Fig. 2: Steps taken to derive our proposed method. Orange and green shaded areas show learning of epistemic and aleatory uncertainty.

- ❑ **Belief-based distributional RL:** presents epistemic uncertainty about $\boldsymbol{\theta}$ in the form of a belief distribution $b(\boldsymbol{\theta}) = p(\boldsymbol{\Theta} = \boldsymbol{\theta})$ as a mixture of k Dirac delta functions:

$$b(\boldsymbol{\theta}) = \sum_{i=1}^{K} \alpha^i \delta(\boldsymbol{\theta} - \boldsymbol{h}^i) \qquad \boldsymbol{\phi} = \{(\boldsymbol{h}^i, \alpha^i)\}_{i=1}^K$$

Belief-based distributional Bellman equation:

$$p(Z^\pi_b(\boldsymbol{s}_t, a_t)) = r_t + \gamma\, p(Z^\pi_{b^-}(\boldsymbol{s}_{t+1}, a_{t+1}))$$

- ❑ **Unified uncertainty estimation:**
  - Summarizes $b(\boldsymbol{\theta})$ with the feature vector $\tilde{\boldsymbol{m}}(\boldsymbol{\phi})$ consisting of M moments of $\boldsymbol{\Theta}$, achieved from its Moment Generating Function (MGF):

$$p(Z^\pi_{\tilde{\boldsymbol{m}}(\boldsymbol{\phi})}(\boldsymbol{s}_t, a_t)) = r_t + \gamma\, p(Z^\pi_{\tilde{\boldsymbol{m}}(\boldsymbol{\phi}^-)}(\boldsymbol{s}_{t+1}, a_{t+1}))$$

  - Learns $\boldsymbol{\phi}$ using a neural network with the JTD loss function [2]

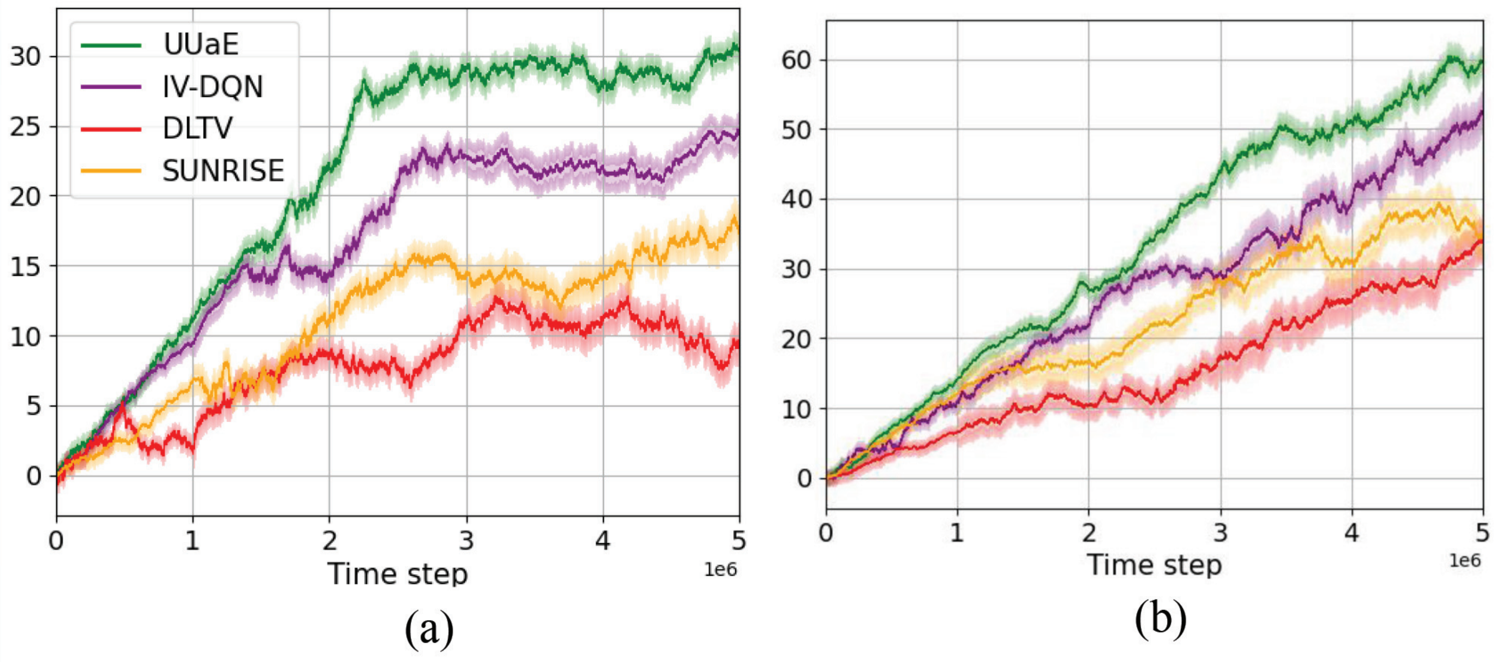
$$L(\boldsymbol{\phi}) = JTD(Z_{\tilde{\boldsymbol{m}}(\boldsymbol{\phi})}(\boldsymbol{s}_t, a_t), r_t + \gamma Z_{\tilde{\boldsymbol{m}}(\boldsymbol{\phi}^-)}(\boldsymbol{s}_{t+1}, a_{t+1}))$$

- ❑ **Composite uncertainty-aware exploration:** chooses an action that rewards high epistemic uncertainty and penalizes high aleatory uncertainty:

$$a_t = \arg\max_{a'} \{\mathbb{E}[Z_{\tilde{\boldsymbol{m}}(\boldsymbol{\phi})}(\boldsymbol{s}_t, a')] - Var(Z_{\tilde{\boldsymbol{m}}(\boldsymbol{\phi})}(\boldsymbol{s}_t, a'))\}$$

Results

- ❑ **Tasks:** two Atari games and an autonomous vehicle driving simulator [3] in a highway.
- ❑ **Baselines:** SUNRISE [4], DLTV [5], and IV-DQN [6], which act based on the epistemic uncertainty, aleatory uncertainty, and their additive formulation.



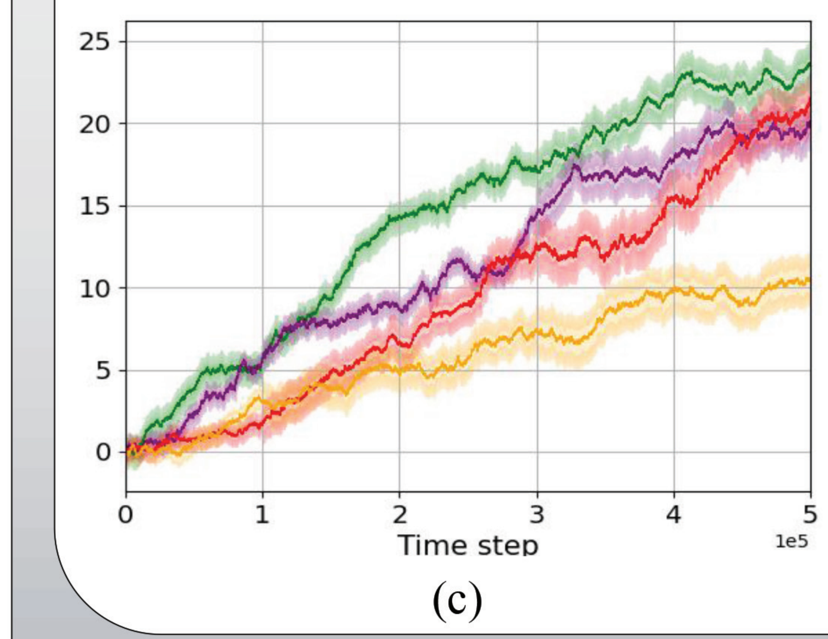


Fig. 3: The cumulative reward for (a): Atari-Asterix, (b): Atari-Seaquest, and (c): Autonomous vehicle driving.

References

[1] W. Dabney, et al., "Distributional reinforcement learning with quantile regression," in AAAI, 2018, vol. 32.
[2] Y. Choi, K. Lee, and S. Oh, "Distributional deep reinforcement learning with a mixture of gaussians," in International Conference on Robotics and Automation, IEEE, 2019, pp. 9791–9797.
[3] E. Leurent, "An environment for autonomous driving decision-making," GitHub, 2018.
[4] K. Lee, et al., "Sunrise: A simple unified framework for ensemble learning in deep reinforcement learning," in PMLR, 2021, pp. 6131–6141.
[5] B. Mavrin, et al., "Distributional reinforcement learning for efficient exploration," in. PMLR, 2019, pp. 4424–4434.
[6] V. Mai, K. Mani, and L. Paull, "Sample efficient deep reinforcement learning via uncertainty estimation," arXiv preprint arXiv:2201.01666, 2022.