中国科学院材料力学行为和设计重点实验室

CAS Key Laboratory of Mechanical Behavior and Design of Materials

Uncertainty in Bayesian Reinforcement Learning for Robot Manipulation Tasks with Sparse Rewards

Li Zheng, Yanghong Li, Yahao Wang, Guangrui Bai, Haiyang He and Erbao Dong

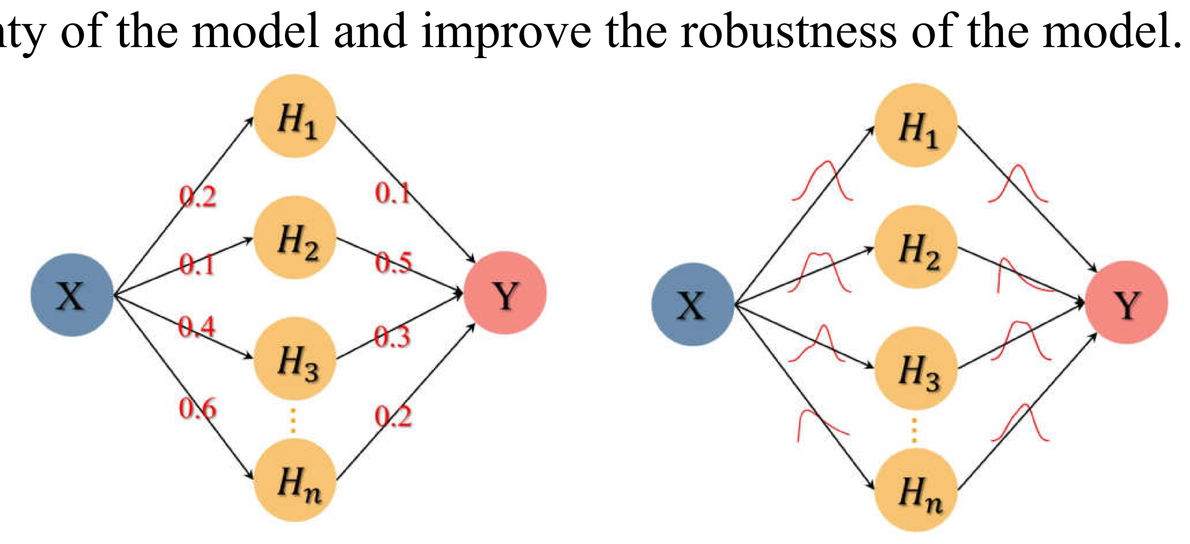
University of Science and Technology of China

INTRODUCTION

This study aims to explore the application of Bayesian deep reinforcement learning (BDRL) in robot manipulation tasks with sparse rewards, focusing on addressing the uncertainty in complex and sparsely rewarded environments. Conventional deep reinforcement learning (DRL) algorithms still face significant challenges in the context of robot manipulation tasks. To address this issue, this paper proposes a general algorithm framework called BDRL that combines reinforcement learning algorithms with Bayesian networks to quantify the model uncertainty, aleatoric uncertainty in neural networks, and uncertainty in the reward function. The effectiveness and generality of the proposed algorithm are validated through simulation experiments on multiple sets of different sparsely rewarded tasks, employing various advanced DRL algorithms. The research results demonstrate that the DRL algorithm based on the Bayesian network mechanism significantly improves the convergence speed of the algorithms in sparse reward tasks by accurately estimating the model uncertainty.

PRELIMINARIES

Due to the deterministic nature of the model weights and biases in a deep neural network, it cannot provide probabilistic explanations for the model. BNNs models can be used to obtain uncertainty estimates for deep neural network models. Bayesian probability theory provides a solid mathematical framework for obtaining model uncertainty. The weights and biases of the neural network model are assigned as prior distributions *(a-b)*. By solving the model and obtaining the posterior distribution, we can capture the uncertainty of the model and improve the robustness of the model.



(a) (b)

METHOD

1. The study defined reward function uncertainty, which complements the aleatoric and model uncertainties, to evaluate the uncertainty related to data incompleteness and inaccurate definition of reward functions.

➢ Model Uncertainty

$$\mathbb{U}^{(i)}_{\text{model}} = \int \{\boldsymbol{p}(a \mid s_i, \boldsymbol{\omega}) - \mathbb{E}[\boldsymbol{p}(a \mid s_i)]\}^{\otimes 2} q^*_\theta(\boldsymbol{\omega}) d\boldsymbol{\omega}$$
$$\approx \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{p}(a \mid s_i, \hat{\omega}_t) - \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{p}(a \mid s_i, \hat{\omega}_t)\right]\right)^{\otimes 2}$$
$$= \frac{1}{T}\sum_{t=1}^{T} D^2\left(\hat{a}_{i,t}, \frac{1}{T}\sum_{t=1}^{T}\hat{a}_{i,t}\right)$$

➢ Aleatoric uncertainty

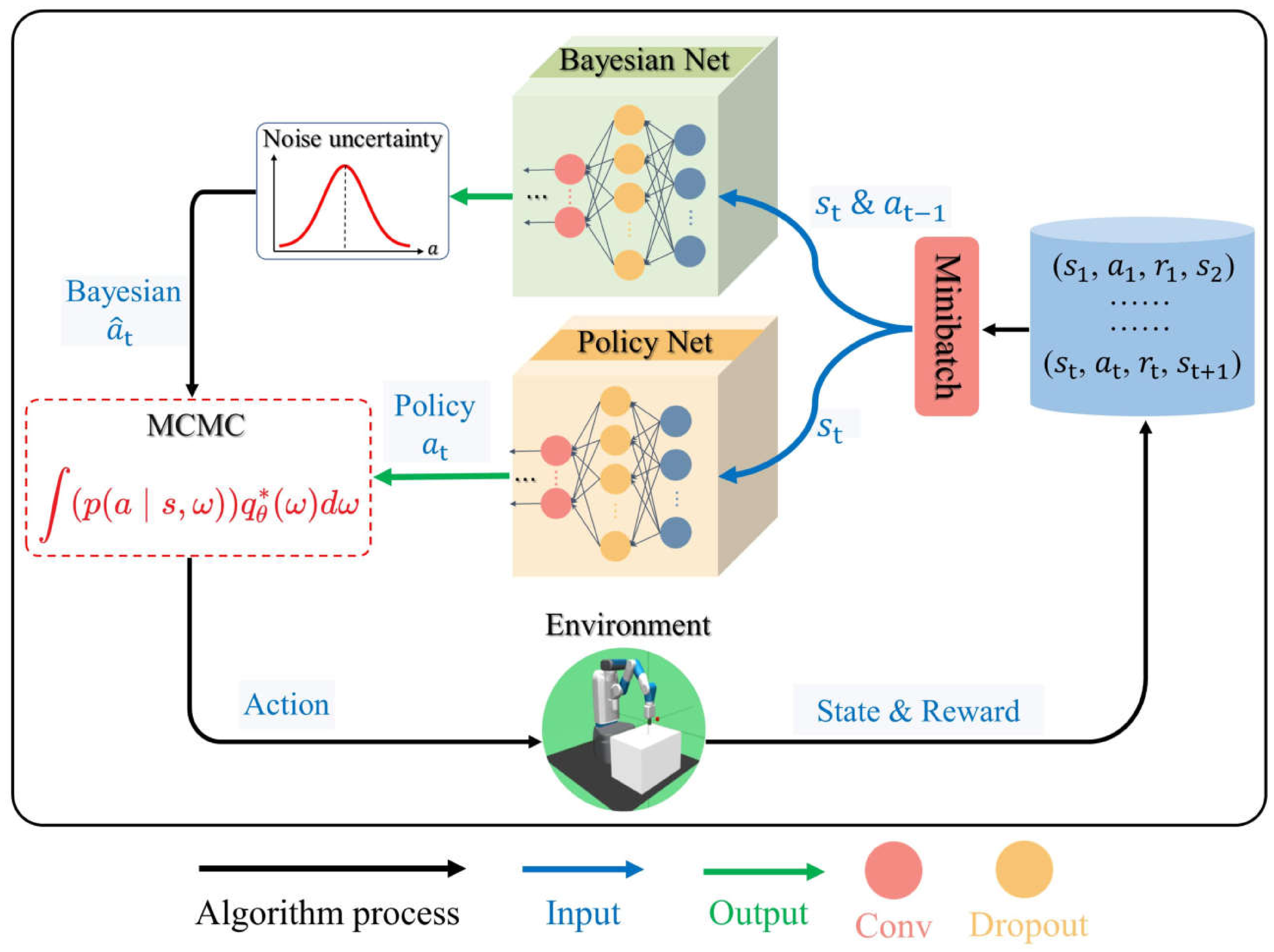
$$\mathbb{U}^{(i)}_{aleatoric} = \int (\sigma_i^2(\omega)) q^*_\theta(\omega) d\omega \approx \frac{1}{T}\sum_{t=1}^{T}(\sigma_{i,t})^2$$

➢ Reward Functional uncertainty

$$E(a \mid s) = \mathbb{E}_{a\sim\mathcal{N}(0,\sigma^2)}[r(a,s)]$$

$$\mathbb{U}^{(i)}_{reward} = D_{KL}(E \parallel R) = \sum_{i=1}^{n} E(i) log \frac{E(i)}{R(i)}$$

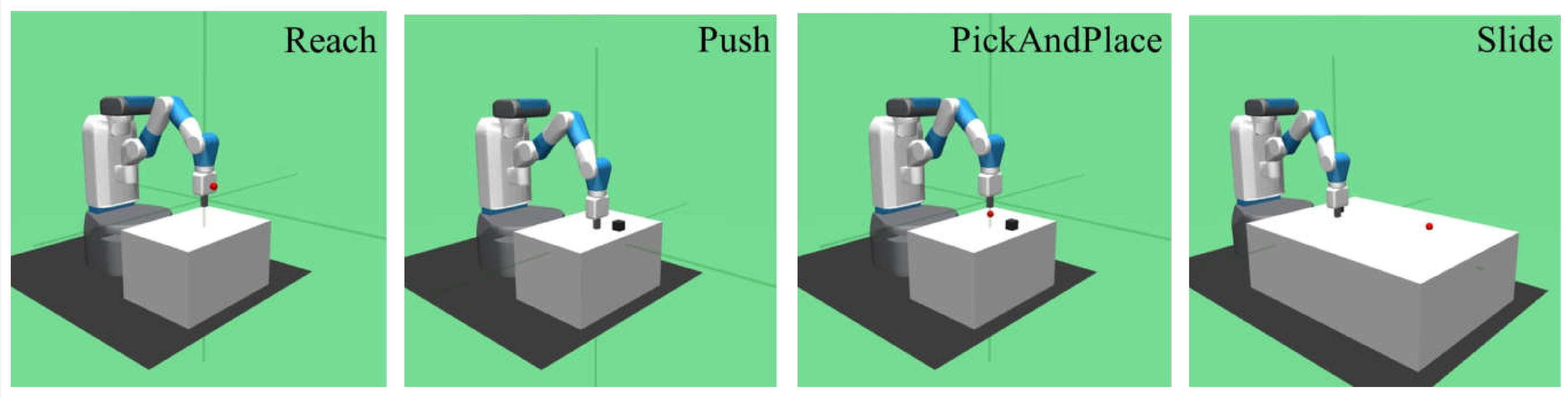
2. The study proposed a Bayesian deep reinforcement learning framework that uses uncertainty estimation to accelerate the training process of traditional DRL algorithms in complex environments and improve their performance.



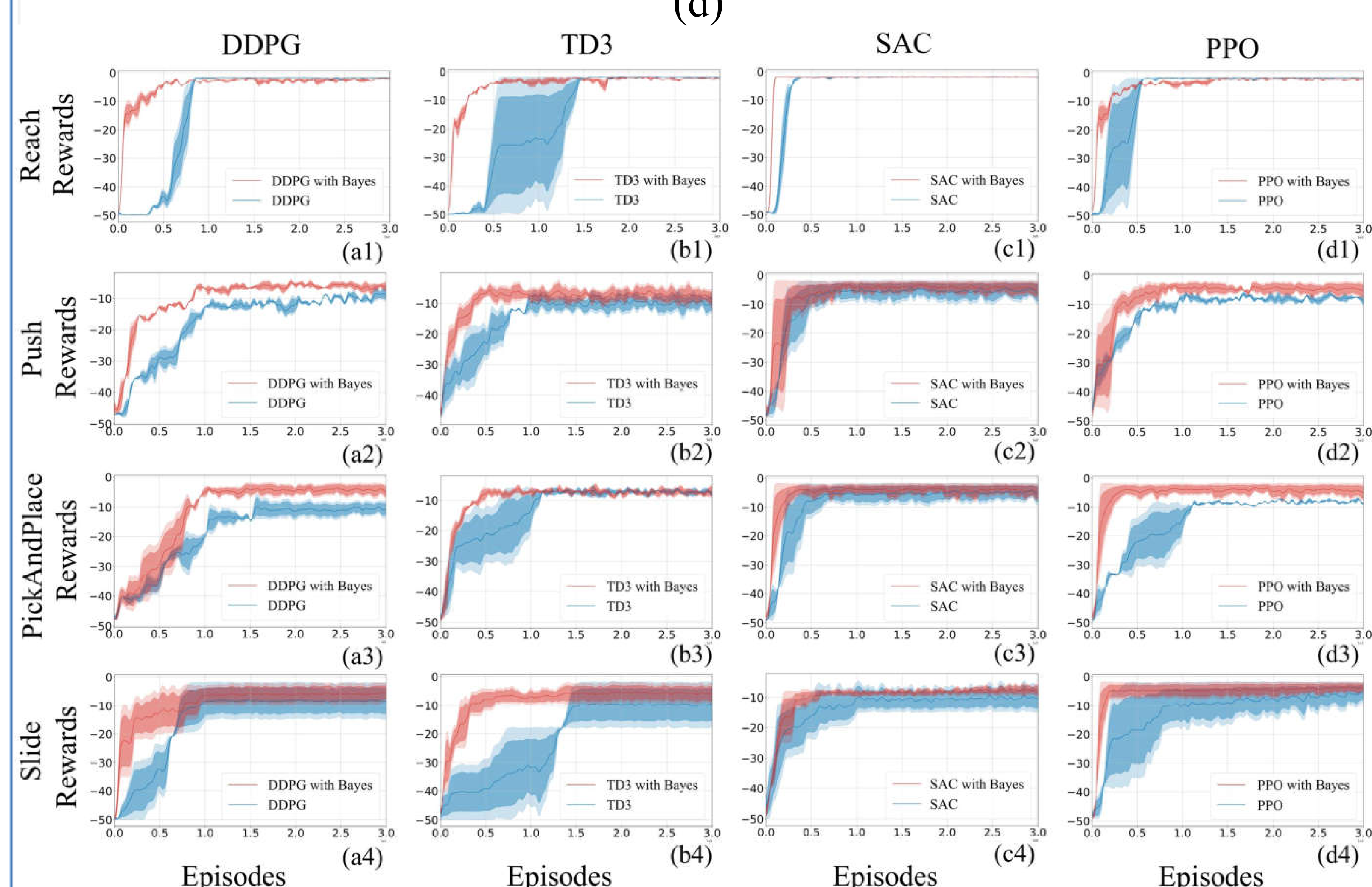
(c)

EXPERIMENTS AND RESULTS

To validate the universality of the proposed BDRL framework, we conducted experimental comparisons and validations between the Bayesian-based DRL algorithm and the conventional DRL algorithm in various reward-sparse environments.



(d)



(e)

CONCLUSIONS

The study proposed DRL algorithm based on the Bayesian network mechanism demonstrates superior performance compared to conventional DRL algorithms when facing different reward-sparse tasks. Through a combination of comparative experiments on four reward-sparse environments and evaluation validation experiments, we have verified the algorithm's advantages in handling dynamic environmental changes and unknown factors. The BDRL algorithm not only exhibits excellent performance in experiments but also holds promising potential for wide-ranging applications.

USTC-LMBD