World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:19, No:08, 2025

A Fully Interpretable Deep Reinforcement Learning-Based Motion Control for Legged Robots

Authors : Haodong Huang, Zida Zhao, Shilong Sun, Chiyao Li, Wenfu Xu

Abstract : The control methods for legged robots based on deep reinforcement learning have seen widespread application; however, the inherent black-box nature of neural networks presents challenges in understanding the decision-making motives of the robots. To address this issue, we propose a fully interpretable deep reinforcement learning training method to elucidate the underlying principles of legged robot motion. We incorporate the dynamics of legged robots into the policy, where observations serve as inputs and actions as outputs of the dynamics model. By embedding the dynamics equations within the multi-layer perceptron (MLP) computation process and making the parameters trainable, we enhance interpretability. Additionally, Bayesian optimization is introduced to train these parameters. We validate the proposed fully interpretable motion control algorithm on a legged robot, opening new research avenues for motion control and learning algorithms for legged robots within the deep learning framework.

Keywords : deep reinforcement learning, interpretation, motion control, legged robots
Conference Title : ICLR 2025 : International Conference on Learning Representations
Conference Location : Sydney, Australia
Conference Dates : August 30-31, 2025