Towards Robust, Efficient, and Practical Decision-Making: From Reward-Maximizing Deep Reinforcement Learning to Reward-Matching GFlowNets

Ling Pan
Hong Kong University of Science and Technology
lingpan@ust.hk

The standard RL paradigm enables agents to learn optimal policies by maximizing expected long-term rewards, achieving success in various domains. However, the singular focus on optimizing reward proxies $(\arg\max \hat{R}(x))$ and learning the *reward-maximizing* policy often proves insufficient for real-world applications where diverse solution sets are critical. For example, in molecular generation, imperfect reward functions in simulations necessitate multiple high-reward candidates. The Generative Flow Network (GFlowNet) (Bengio et al. 2021), represents a paradigm shift by learning *reward-matching* policies $(\pi(x) \propto \hat{R}(x))$. This enables the discovery of diverse, high-quality solutions while addressing critical challenges in traditional RL, particularly in scientific discovery applications where simulation uncertainties demand multiple promising candidates. Despite widespread success in various domains, sequential decision-making still faces significant challenges, particularly with deep neural networks and in multi-agent settings where agent interactions introduce additional complexity. We target three fundamental problems in sequential decision-making paradigms: (1) how can agents learn efficiently for decision-making? (2) how can we ensure a convergent and robust learning behavior of agents? and (3) how to successfully apply them in important practical applications?

Efficient Training. Sequential decision-making systems often suffer from low sample efficiency in learning optimal policies. We improve learning efficiency across core dimensions: *state* transition dynamics, *action* selection (Pan et al. 2023b), *reward* (credit assignment) (Pan et al. 2023a), *policy*, and the *agent* learning paradigm (Pan et al. 2022, 2024).

Robust Learning Behavior. The reliability of decision-making systems in real-world applications is often compromised by estimation uncertainties and environmental stochasticity. We investigate how to enhance its learning robustness, considering more robust value estimates for overestimation reduction (Pan, Cai, and Huang 2020; Pan et al. 2021) and developing robust approaches for extending GFlowNets to stochastic environments (Pan et al. 2023c; Zhang et al. 2024) with inherent uncertainty..

Practical Applications. Intelligent decison-making algorithms demonstrate strong practical impact to critical real-world challenges. We develop the first large-scale dockless bike-sharing system rebalancing approach (Pan et al. 2019) in computational sustainability, and demonstrat promising results in molecular generation and combinatorial optimization (Zhang et al. 2023), where these advances can lay the groundwork for other important applications in computational sustainability, healthcare, and autonomous systems.

References

Bengio, E.; Jain, M.; Korablyov, M.; Precup, D.; and Bengio, Y. 2021. Flow network based generative models for non-iterative diverse candidate generation. In *NeurIPS*.

Pan, L.; Cai, Q.; Fang, Z.; Tang, P.; and Huang, L. 2019. A deep reinforcement learning framework for rebalancing dockless bike sharing systems. In *AAAI*.

Pan, L.; Cai, Q.; and Huang, L. 2020. Softmax Deep Double Deterministic Policy Gradients. In *NeurIPS*.

Pan, L.; Huang, L.; Ma, T.; and Xu, H. 2022. Plan better amid conservatism: Offline multi-agent reinforcement learning with actor rectification. In *ICML*.

Pan, L.; Jain, M.; Madan, K.; and Bengio, Y. 2024. Pre-Training and Fine-Tuning Generative Flow Networks. In *The Twelfth International Conference on Learning Representations (ICLR)*.

Pan, L.; Malkin, N.; Zhang, D.; and Bengio, Y. 2023a. Better Training of GFlowNets with Local Credit and Incomplete Trajectories. In *ICML*.

Pan, L.; Rashid, T.; Peng, B.; Huang, L.; and Whiteson, S. 2021. Regularized Softmax Deep Multi-Agent Q-Learning. In *NeurIPS*.

Pan, L.; Zhang, D.; Courville, A.; Huang, L.; and Bengio, Y. 2023b. Generative Augmented Flow Networks. In *ICLR*.

Pan, L.; Zhang, D.; Jain, M.; Huang, L.; and Bengio, Y. 2023c. Stochastic Generative Flow Networks. In *UAI*.

Zhang, D.; Dai, H.; Malkin, N.; Courville, A.; Bengio, Y.; and Pan, L. 2023. Let the Flows Tell: Solving Graph Combinatorial Optimization Problems with GFlowNets. In *NeurIPS*.

Zhang, D.; Pan, L.; Chen, R. T.; Courville, A.; and Bengio, Y. 2024. Distributional GFlowNets with Quantile Flows. *TMLR*.