Fisher-Guided Selective Forgetting (FGSF) For Deep Reinforcement Learning

Massimiliano Falzari and Matthia Sabatelli

University of Groningen

Abstract. We present a novel algorithm suitable for Deep Reinforcement Learning (DRL) problems that leverages information geometry to implement strategic and selective forgetting. Our method aims to tackle DRL's Primacy Bias and enhance adaptability and robustness within the sequential decision-making framework. We empirically show that by including a selective forgetting mechanism, implemented by leveraging the Fisher Information Matrix, one can obtain faster and more robust learning compared to traditional DRL methods that are only focused on learning. Our experiments, performed on the popular DeepMind Control Suite benchmark, strengthen the idea - already present in the literature - that forgetting is a fundamental part of learning, particularly in situations with non-stationary targets.

Keywords: Deep Reinforcement Learning · Fisher Information Matrix · Information Geometry · Primacy Bias · Selective Forgetting

1 Overcoming The Primacy Bias with The Fisher Information Matrix

The Primacy Bias (PB) is one of the many potential reasons why Deep Reinforcement Learning (DRL) agents struggle to adapt to new experiences and tend to overfit to early training data [5]. While in DRL this phenomenon has been introduced relatively recently, the PB appears to be a more fundamental issue within the Machine Learning community and has arguably been addressed with different names (e.g., Plasticity Loss, Capacity Loss) [3,1,2] This widespread occurrence suggests that the PB, alongside its instances, might be a fundamental issue of the general learning dynamics of neural networks, rather than being a DRL-specific problem. In this paper, we present a novel, computationally efficient, regularization method that addresses the PB through selective forgetting, significantly improving DRL agents' adaptability. Our approach uses a Fisher Information Matrix (FIM)-based mechanism to partially forget information from previous learning experiences while at the same time preserving the learning progress. This procedure builds upon previous work targeting Machine Unlearning [4] and is adapted to consider DRL's sequential and not-stationarity nature. It also provides a theoretical framework grounded in information geometry for analyzing the learning dynamics of DRL agents that allows us to characterize the PB under the lens of the FIM. Specifically, we show that the PB occurs when a rapid increase and decrease in the magnitude of the FIM's trace appears during the early training phase as depicted in the first row of Figure 1. To overcome this behavior, we design an algorithm that implements selective forgetting by adding noise based on the FIM as follows: $w := w + \lambda\epsilon$, where w represents the weights of the neural network, λ is a regularization constant that controls the magnitude of the injected noise, and ϵ is sampled from $\mathcal{N}(0, F^{-\frac{1}{2}})$ with F being the FIM calculated on the sampled batch of trajectories that is used at the weight optimization step. It can be shown that by performing this update, the parameter distribution obtained is closer to the parameter distribution of a model that is trained only on the specific batch of trajectories on which the FIM was calculated, hence forgetting all previously encountered trajectories.

We evaluate our method using the Soft-Actor Critic algorithm on the popular DeepMind Control Suite benchmark. Specifically, we show that for cases in which the PB appears, our method regularizes the trace of the FIM (first two plots of the first row of Fig. 1) and also outperforms the reset method proposed in [5], in terms of reward (first two plots of the second row of Fig. 1). For situations where the PB is not present (last plot of Fig. 1), and the reset method is actually detrimental, it performs just as well as the baseline.

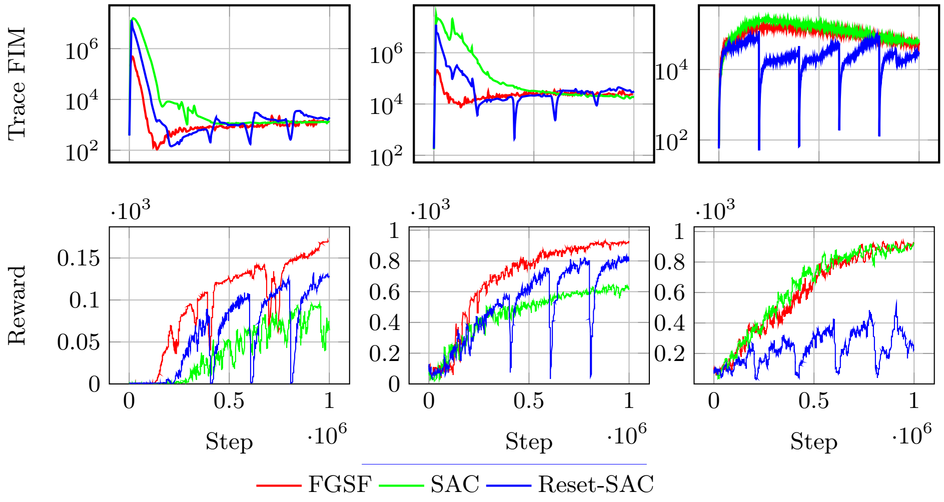

Fig. 1: Performance comparison of FGSF, SAC, and Reset-SAC across three environments: `humanoid-run`, `quadruped-run`, and `finger_turn-hard`. The upper row shows the trace of the FIM during training on a logarithmic scale, while the lower row displays the reward curves. In the first two environments, a clear increase and decrease of the FIM's trace is observed. This pattern is notably absent in the third environment.

References

1. Alessandro Achille, Matteo Rovere, and Stefano Soatto. Critical learning periods in deep networks. In *International Conference on Learning Representations*, 2018.
2. Hongjoon Ahn, Jinu Hyeon, Youngmin Oh, Bosun Hwang, and Taesup Moon. Catastrophic negative transfer: An overlooked problem in continual reinforcement learning.
3. Shibhansh Dohare, J Fernando Hernandez-Garcia, Parash Rahman, A Rupam Mahmood, and Richard S Sutton. Maintaining plasticity in deep continual learning. *arXiv preprint arXiv:2306.13812*, 2023.
4. Aditya Golatkar, Alessandro Achille, and Stefano Soatto. Eternal sunshine of the spotless net: Selective forgetting in deep networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9304–9312, 2020.
5. Evgenii Nikishin, Max Schwarzer, Pierluca D'Oro, Pierre-Luc Bacon, and Aaron Courville. The primacy bias in deep reinforcement learning. In *International conference on machine learning*, pages 16828–16847. PMLR, 2022.