IMITATION LEARNING USING THE FORWARD-FORWARD ALGORITHM

Anonymous authors
Paper under double-blind review

ABSTRACT

The forward-forward (FF) algorithm has been recently introduced as a novel approach to training neural networks in a way that approximates the behavior of real neurons. Nevertheless, its application has been limited to visual domains and has not been investigated in the context of imitation learning or reinforcement learning. In this study, we evaluate the FF algorithm in the context of imitation learning. We implement a straightforward imitation model based on the FF algorithm and present a comparative analysis with the backpropagation (BP) model. Our findings indicate that the FF-based model exhibits comparable performance to the BP-based model on larger datasets but demonstrates inferior performance on smaller datasets.

1 INTRODUCTION

Reinforcement Learning (RL) has achieved significant success in diverse domains, including Atari (Mnih et al., 2015), Go (Silver et al., 2017), StarCraft (Vinyals et al., 2019), and protein folding (Jumper et al., 2021), leveraging the flexibility of deep learning in approximating policies and value functions. The stochastic gradient descent approach using backpropagation (BP) is widely employed for training neural networks in deep learning. However, recent research exploring the implementation of BP with real neurons (Lillicrap et al., 2016; 2020; Guerguiev et al., 2017) has revealed ambiguity in whether BP accurately corresponds to real neuron activities. To address this issue, Hinton (2022) proposes a forward-forward algorithm (FF) that replaces the forward and backward passes in BP with two forward passes.

Given this perspective, we consider imitation learning (IL), a problem of learning policies from static datasets, to be an intriguing problem. While FF has shown impressive results in the visual domain, it has not been studied in the context of imitation learning or reinforcement learning.

In this study, we evaluate the efficacy of the FF algorithm in the context of imitation learning and conduct a comparative analysis with the backpropagation (BP) model. We conducted experiments by changing the size of the training dataset in two classic environments. As a result, the FF imitation model performed similarly to the BP model on large datasets but performed worse than BP on small datasets.

2 FORWARD-FORWARD ALGORLITHM

The Forward-Forward (FF) algorithm is a neural network learning algorithm that more closely resembles the learning structure of the human brain than the backpropagation algorithm. The FF algorithm replaces backpropagation's forward and backward passes with two forward passes. These two forward passes work with both positive (i.e., actual) and negative (i.e., generated) data. The network is trained to have high goodness in the positive data and low goodness in the negative data. Therefore, the network distinguishes positive data from negative data by comparing the goodness of the input data. Additionally, positive data include the correct label and negative data include the incorrect label, and the difference between positive and negative data is only the label. The network determines appropriate labels for input data.

3 METHOD

We considered imitation learning as a supervised problem that infers an action based on a given state by treating the state as an input and the action as a label. Therefore, by applying FF, the correct action was inferred from the given state in the same way as the positive data with the correct label was classified from the negative data.

Specifically, we use pairs of correct action and state as positive data and pairs of incorrect action and state pairs as negative data. Regarding generating positive and negative data, the correct action for positive data was determined as the action most likely predicted by an expert-level model in a given state. Conversely, for negative data, a wrong action was selected randomly from the set of actions excluding the correct action. Finally, The action (considered a label) is one-hot encoded and concatenated with the state to be used as positive and negative data.

Our model is trained to output high Goodness for positive data and low Goodness for negative data. During inference, our model computes the goodness for all possible actions and state pairs in a given state and selects the action in the pair with the highest goodness. Goodness is the sum of the square output in every layer.

4 EXPERIMENTS AND DISCUSSION

Our experiments aim to compare the forward-forward (FF) and the BP (BP) algorithms from the perspective of imitation learning. Therefore, Thus, both models were built using a straightforward multi-layer perception model (MLP). We compared models on two standard environments, 'Cartpole' and 'MountainCar' (Brockman et al., 2016). The experience dataset for training was collected from an expert-level Proximal policy optimization (PPO) (Schulman et al., 2017) model. the experience dataset is a total of 100,000 frames, and each data consists of experience tuples of (state, action).

Task	CartPole		MountainCar		
Dataset	1%	5%	10%	50%	100%
FF	422.34 $\pm$152.1	500.0 $\pm$0.0	-121.89 $\pm$27.47	-112.88 $\pm$21.30	-81.18 $\pm$15.20
BP	500.0 $\pm$0.0	500.0 $\pm$0.0	-106.56 $\pm$10.63	-104.64 $\pm$11.30	-80.38 $\pm$14.02
PPO	500 $\pm$0.0		-80.03 $\pm$13.04		

Table 1: Result of performance of Forward-Forward (FF) and Backpropagation (BP). Each value represents the average and standard deviation of the reward. PPO is an expert-level model that generated a dataset used to train FF and BP.

The performance of the FFIL model and the BPIL with various data sizes is compared in the table 1. In the 'CartPole' environment, FFIL and BPIL achieve 500 (maximum) rewards. However, FFIL requires more data than BP to reach the maximum reward. In the 'MountainCar', the experimental results indicate that FFIL, BPIL, and PPO (expert-level model) have comparable performance when trained on the entire dataset. However, as the size of the training dataset decreases, the performance of FFIL shows a considerable decline compared to BPIL.

Our analysis shows that FFIL is more sensitive to dataset size than BPIL. However, the FF paper does not provide insight into the relation between dataset size and performance. Consequently, we cannot ascertain whether this phenomenon is unique to the FFIL or a general characteristic feature of the FF model. One positive aspect is that FFIL can achieve a competitive level of performance with BPIL, given enough data. However, our approach is currently limited to discrete actions. Overcoming this limitation might increase FFIL's applicability to more imitation learning domains.

5 CONCLUSION AND FUTURE WORK

According to our investigation, FFIL performance is more sensitive to the size of the dataset than BPIL. To our knowledge, this is the first study that employs FF in the context of RL. Thus, FF can be applied to RL in a wide variety of research. The development of an online RL model via FF or the implementation of an FFIL model that can infer action in continuous space were our next goals.

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

Jordan Guerguiev, Timothy P Lillicrap, and Blake A Richards. Towards deep learning with segregated dendrites. *Elife*, 6:e22901, 2017.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Timothy P Lillicrap, Daniel Cownden, Douglas B Tweed, and Colin J Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications*, 7(1): 13276, 2016.

Timothy P Lillicrap, Adam Santoro, Luke Marris, Colin J Akerman, and Geoffrey Hinton. Backpropagation and the brain. *Nature Reviews Neuroscience*, 21(6):335–346, 2020.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017.

Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.