Reward based Hebbian Learning in Direct Feedback Alignment (Student Abstract)

Ashlesha Akella, [1] **Sai Kalyan Ranga Singanamalla,** [1] **Chin-Teng Lin,** [1, 2]
[1]School of Computer Science, University of Technology Sydney, Australia
[2]Center for Artificial Intelligence, University of Technology Sydney, Australia
{ashlesha.akella,saikalyanranga.singanamalla} @student.uts.edu.au, chin-teng.lin@uts.edu.au

Abstract

Imparting biological realism during the learning process is gaining attention towards producing computationally efficient algorithms without compromising the performance. Feedback alignment and mirror neuron concept are two such approaches where the feedback weight remains static in the former and update via Hebbian learning in the later. Though these approaches have proven to work efficiently for supervised learning, it remained unknown if the same can be applicable to reinforcement learning applications. Therefore, this study introduces RHebb-DFA where the reward-based Hebbian learning is used to update feedback weights in direct feedback alignment mode. This approach is validated on various Atari games and obtained equivalent performance in comparison with DDQN.

Introduction

Gradient descent based error backpropagation (BP) is the widely used mechanism for training neural networks and was proven to be an efficient optimization technique for varied practical applications. BP assumes that a neuron in a given layer has knowledge of all of its downstream synaptic connectivity for precise gradient estimation in addition to the presence of symmetric feedback weights. However, the human brain is not known to accommodate such symmetric feedback connections for the weight transport problem. For this reason, recent studies explored alternatives to BP to achieve biological realism. For example, (Lillicrap et al. 2016) introduced feedback alignment (FA) and has shown that it is not necessary for the feedback weights need not be symmetric to feed-forward weights and can still produce efficient learning with random weights. Building on this, (Nøkland 2016) developed direct feedback alignment (DFA) in which the feedback connection arrives from the output layer to hidden layers earlier in the pathway for error-driven learning. This approach has proven it applicability to spiking neural networks too (Lee et al. 2020). In all these studies, the feedback weights are chosen randomly and remain static throughout training and only allow the feedforward weights to update. Recently, (Akrout et al. 2019) devised an approach to update these feedback weights in FA by adopting a mirror neuron concept in which a network identical to the forward network is established for a backward pass. The feedback weights are updated via Hebbian learning, with the activity from the forward network, and has shown the performance is equivalent to BP. These advancements raise the question of its compatibility in the Reinforcement Learning (RL) domain. Though reward-based learning is the crux of RL, the synaptic weight update still rely on BP. However, Neuroscientific studies have shown the existence of reward-modulated Hebbian learning in adjusting synaptic weights using a reward in addition to pre and post-synaptic activity. Therefore, this work explores the role of DFA using reward-modulated Hebbian learning (referred to as RHebb-DFA) for updating feedback weights in RL application. Hebb-DFA is validated with different Atari games and compared to feed-forward based Double Deep Q Network (DDQN) for baseline comparison.

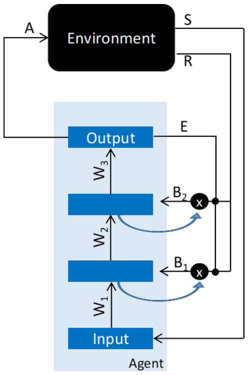

Figure 1: RHebb-DFA architecture

.

RHebb-DFA

In RL, an agent observes the current state s of the environment and chooses an action a and executes it on the environment. As a consequence, it receives a reward r and the environment transits to a new state $s^{/}$. Deep Q-Learning is an algorithm where a multi-layered neural network $Q(s, .; \theta)$ (θ are the parameters) is used to estimate values of each action given a state, and the value of an action is defined as the expected sum of future rewards. DDQN (Van Hasselt, Guez, and Silver 2015) is an off-policy learning algorithm which

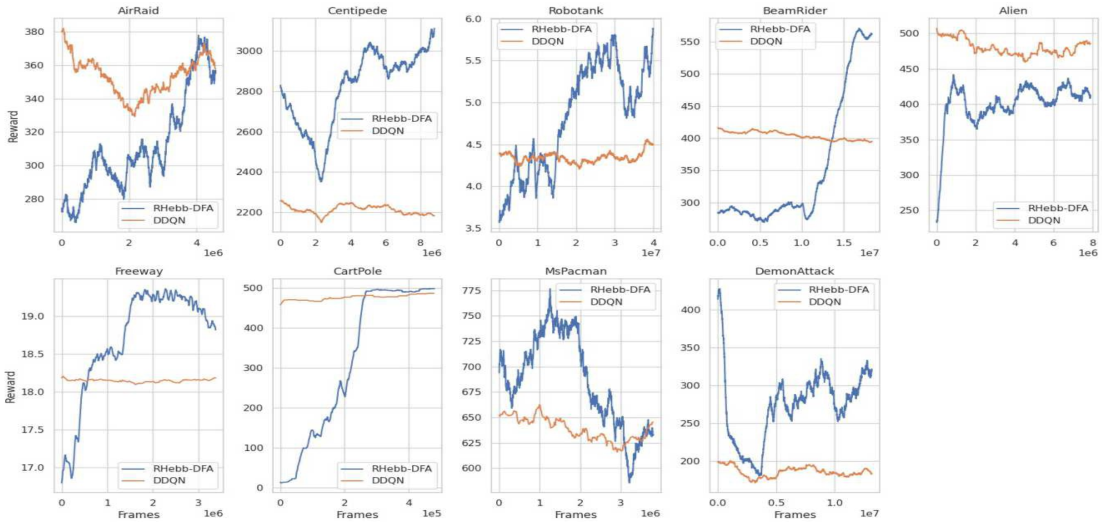

Figure 2: Performance of RHebb-DFA in comparison to DDQN on various Atari games

.

learns to estimates the value of an action given a state by minimizing the following error

$$E = \mathrm{E}_{s,a,r,s^{/}\sim\rho(.)}\ \left(y_i - Q_{actual}\,(s,a;\theta)\right)^2 \qquad (1)$$

where y_i is called the TD (temporal difference) target and is defined by $y_i = r + \gamma \max_{a^/} Q_{target}\ (s^/,\ a^/;\ \theta)$. Q_{actual} and Q_{target} are two Q-networks used in Double DQN, where gradients are computed for Q_{actual} and the weights are copied to Q_{target} for every few episodes. In this work we used RHebb-DFA to train the Q-network by propagating the error E using direct feedback alignment.

For simplicity, assumes the Q-network has only two hidden layers, where W_i denotes the weights that connects the layer below to a hidden layer i, as shown in Figure 1 and B_i is a weight matrix that connects the output layer directly to each hidden layer i. B_i is initialized randomly with appropriate dimensions and are trained using reward-modulated hebbian learning for every training epoch prior to the training of w_i as dictated by Equation 2.

$$\Delta B_i = E * r * Q_{actual}(s) \qquad (2)$$

where E is the error as defined in Equation 1, r is the reward agent received at state s and action a and $Q_{actual}(s)$ is a vector of action values for a state s. After updating B, ΔW is computed using Equation 3.

$$\Delta W_i = -(E * B_i * f^/(Q_{actual}(s))) \quad f^/(O_i) * x_i \qquad (3)$$

where $f^/$ is the derivative of activation function f, x_i and O_i are the input and output of layer i respectively. $Q_{actual}(s)$ is the final layer output value. The work validated RHebb-DFA on few Atari games as shown in Fig. 2. The algorithmic pseudocode, maximum scores obtained for RHebb-DFA on each of these environments and source code is provided in supplementary.

Discussion

In this preliminary study, we leveraged the properties of reward-based Hebbian learning for updating feedback weights in DFA. The proposed method was performed comparable to DDQN and better on a few Atari games. Since DFA is not compatible with the convolution layer(Akrout et al. 2019), RHebb-DFA was tested only for networks with feedforward layers. One of the current limitations of this approach is that it is sensitive to reward and resulted in large fluctuations during the training process. Therefore, the future direction of this work is to stabilize the training along with proving convergence from a mathematical standpoint.

References

Akrout, M.; Wilson, C.; Humphreys, P.; Lillicrap, T.; and Tweed, D. B. 2019. Deep learning without weight transport. In *Advances in neural information processing systems*, 976–984.

Lee, J.; Zhang, R.; Zhang, W.; Liu, Y.; and Li, P. 2020. Spike-train level direct feedback alignment: sidestepping backpropagation for on-chip training of spiking neural nets. *Frontiers in Neuroscience* 14.

Lillicrap, T. P.; Cownden, D.; Tweed, D. B.; and Akerman, C. J. 2016. Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications* 7(1): 1–10.

Nøkland, A. 2016. Direct feedback alignment provides learning in deep neural networks. In *Advances in neural information processing systems*, 1037–1045.

Van Hasselt, H.; Guez, A.; and Silver, D. 2015. Deep reinforcement learning with double q-learning. *arXiv preprint arXiv:1509.06461* .