REWARD-MODULATED SPIKE-TIMING DEPENDENT PLASTICITY

August 28, 2023

Amin Zeinali
University of Tehran
Faculty of Mathematics, Statistics, and Computer Sciences

1 Abstract

Reward-modulated spike-timing-dependent plasticity (R-STDP) has been proposed as a learning rule that enables spiking neural networks to solve the distal reward problem through dopaminergic reinforcement signals. We implemented an R-STDP rule in a simulated spiking neural network to investigate how reward-based modulation of synaptic plasticity can train desired target neurons. The network consists of interconnected excitatory and inhibitory neuron groups, with R-STDP on excitatory-to-excitatory synapses that reach target neurons. Reward signals are provided by a payoff module and control dopamine concentration, which multiplicatively modulates R-STDP. Simulations demonstrate successful propagation of reward to distal synapses and selective strengthening of connections to frequently rewarded targets. The model provides a framework for investigating reward processing in spiking networks and validates the ability of dopamine-modulated plasticity to implement reinforcement learning for value-based decision making.

2 Introduction

The notebook associated with the report implements a spiking neural network in PyTorch using the PymoNNtorch library. Key components include leaky integrate-and-fire neuron models, STDP and R-STDP synaptic plasticity rules, a reward/punishment behavior module, and dopamine concentration modeling. Simulations demonstrate how R-STDP allows propagating reward signals to distant synapses and can train target neurons through reinforcement learning.

3 Methods

3.1 Network Structure

The network consists of interconnected excitatory and inhibitory neuron groups. The excitatory population receives random input currents generated from Normal distribution, while the inhibitory population receives no external inputs.

The learning rules implemented are:

- STDP for excitatory-to-inhibitory connections.
- R-STDP for excitatory-to-excitatory connections.

- Anti-Hebbian STDP for inhibitory-to-excitatory connections.

The R-STDP rule incorporates a reward-modulated learning term based on dopamine concentration. Dopamine concentration is updated at each timestep based on reward signals provided by the payoff module.

The network is simulated for alternating epochs of rewarding one of two target excitatory neuron groups. R-STDP drives the selective strengthening of synapses to the rewarded target group.

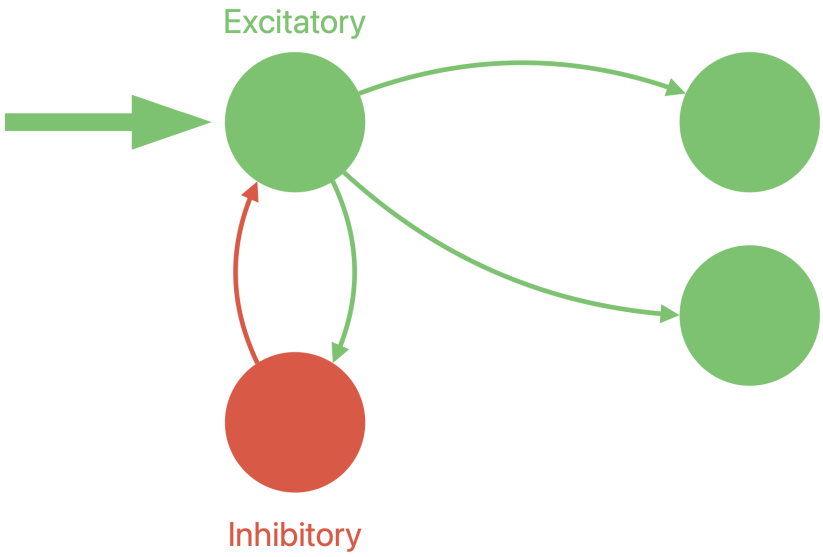

Figure 1: The structure of the network used in simulations.

3.2 R-STDP

According to the figure above, in this section, we have two target excitatory populations. The goal is that each target excitatory population learns a specific signal which is determined by the user.

Reward-modulated STDP (R-STDP) like any other reinforcement learning technique needs a reward/punishment function or as we call it payoff function. In this case, two target populations are made, namely p_0 and p_1. Assume that p_0 is supposed to learn the signal s_0. When s_0 is given to the network the payoff function will compute the following quantity:

$$DA(t) = \frac{n_0 - n_1}{N} C$$

n_i $i \in \{0,1\}$ is the number of neurons in p_i which have fired a spike and N is the total number of neurons in each population. C is a constant that is determined by the user.

After computing the payoff function, we calculate the new level of dopamine using the following equation:

$$\frac{\mathrm{d}d}{\mathrm{d}t} = -\frac{d}{\tau_d} + DA(t)$$

In the equation above, τ_d is the decay constant of dopamine. To complete the algorithm, we need an eligibility trace to keep track of the STDP events, representing the activation of an enzyme important for plasticity. In the following equations, c is the eligibility trace, τ_c is the decay constant of c, $t_{pre/post}$ is the spike time of pre- or post-synaptic neuron and w is the synaptic strength.

$$\frac{\mathrm{d}c}{\mathrm{d}t} = -\frac{c}{\tau_c} + STDP(\tau)\delta(t - t_{pre/post})$$
$$w = w + cd$$

4 Results

We use raster plots to examine the activity of the output neuron group. The signals are illustrated in two different colors, orange and purple.

As you can see in figure 2, each signal has been given to the network 25 times, in a total of 5000 iterations. It is clear that when a signal reaches the network, initially most of the neurons in both of the populations fire spikes and the dopamine level fluctuates around 0. Over time, R-STDP weakens synaptic strengths that lower the dopamine level. Consequently, the fire rate of the desired population rises. While for the other neuron group, it falls. This causes an increase in the level of dopamine in the entire network which then remains relatively stable as long as the input signal has not been changed.

Figure 2: Raster plots of neuron groups during the simulation.

5 References

E. M. Izhikevich, "Solving the Distal Reward Problem through Link-age of STDP and Dopamine Signaling," Cerebral Cortex, vol.17,pp.2443–2452, 01 2007