Pavlovian and Instrumental Conditioning in Spiking Neural Networks

Kira Radinsky and **Michael Goldish**
Technion–Israel Institute of Technology
Haifa 32000, Israel
{kirar , mgoldish}@tx.technion.ac.il

Abstract

In this study we have examined several neural network models in order to describe learning tasks, such as Pavlovian conditioning and instrumental conditioning. This study emphasizes the importance of precise firing patterns in brain dynamics and suggests how a global reinforcement signal in the form of DA can selectively influence the right synapses at the right time.

Introduction

In this work we have investigated how Pavlovian and instrumental conditioning occur in spiking neural networks. To be more precise, how does the brain know what firing patterns of what neurons are responsible for the rewards it later receives. We have created conditions in which those patterns no longer exist and the entire network is stimulated. In this work we base our method on the methods described in (Izhikevich 2007). In his work, the model is based on a model network of cortical spiking neurons with spike-timing–dependent plasticity (STDP) modulated by dopamine.

Related Work

In this study we use two spiking neuron models:

1. **Leaky Integrate-and-Fire (LIF)**

   In the integrate-and-fire neuron model the state of the neuron is characterized by its membrane potential. The membrane potential receives excitatory or inhibitory contributions by synaptic inputs that arrive from other neurons by their associated synapses. These inputs, that are each weighted by their respective synaptic strength, are modeled either as injected current (current synapse models in which summation is linear) or as a change in the membrane conductance (conductance synapse models in which summation of the synaptic input is nonlinear, i.e., the amplitude depends upon the value of the membrane potential). Conductance synapse models take account of the change in amplitude of the excitatory and inhibitory inputs, which depends upon the difference between the membrane potential and the corresponding reversal potential. In this model we neglect the structure of the neuron associated with the dendrites. The neuron is leaky since the summed contributions to the membrane potential decay with a characteristic time constant (the membrane time constant). When the membrane potential reaches a (fixed) threshold, an output spike is generated – the integrate-and-fire mechanism. After the membrane potential crosses the threshold it is reset to its resting value. In our implementation of the model, we used a synaptic depression mechanism to simulation the refractoriness of the neuron.

   For a more detailed description of this model we refer the reader to (Burkitt 2006).

2. **The Izhikevich model**

   The Izhikevich neuron model was developed as an efficient, powerful alternative to the integrate and fire model. The model uses two variables, a variable representing voltage potential and another representing membrane recovery (activation of potassium currents and inactivation of sodium currents).

   This is a spiking neuron model, so when the voltage passes a threshold value a spiking event occurs, and the voltage and recovery variable are reset.

   We use v to represent activation (voltage potential) and u to represent the recovery variable. Voltage is computed by integrating the following two differential equations using Euler's method:

   $$\dot{v} = 0.04v^2 + 5v + 140 - u + I$$
   $$\dot{u} = a(bv - u)$$

   I is the total input current – a weighted sum of the inputs from other neurons; a and b are abstract parameters of the model. When the voltage exceeds a threshold value, which is preset at 30mv, both v and u are reset, as follows:

   $$v \longleftarrow c$$
   $$u \longleftarrow u + d$$

   Thus, the model has four parameters:

   (a) a describes the time scale of the recovery variable u,

   (b) b describes the sensitivity of the recovery variable u to the subthreshold fluctuations of the membrane potential v,

   (c) c describes the after-spike reset value of the membrane potential v and

(d) d describes the after-spike increase of the recovery variable u.

According to Izhikivech, "The model can exhibit firing patterns of all known types of cortical neurons with [a suitable] choice of parameters" (Izhikevich 2004).

The Algorithm

Following (Izhikevich 2007), we describe the state of each synapse using two variables:

1. s – synaptic strength/weight
2. c – synaptic tag ("eligibility trace")

which we update according to the following equations:

$$\dot{c} = -c/\tau_c + STDP(\tau)\delta(t - t_{pre/post})$$
$$\dot{s} = cd$$

Here d describes the extracellular concentration of DA, $\delta(t)$ is the Dirac delta function that step-increases the variable c, and $\tau = t_{post} - t_{pre}$ is the interspike interval.

The algorithm is described in figure 1. For a more detailed description of it refer to (Izhikevich 2007) (section "Materials and Methods").

Experimental Evaluation

Multiple stimuli

We used a network of 1000 neurons (with 100,000 synaptic interconnections), of which we selected 100 random groups consisting of 50 neurons each (the groups may overlap): $s_1, s_2, ..., s_{100}$. Each group represents a stimulus. Delivering a stimulus s_i to the network means manually injecting current into all 50 neurons in group s_i. During training, we select one of the 100 predefined groups at random, and stimulate it. Then after an interval of 100-300ms we repeat this with a different group. This process continues until 1000 stimuli have been delivered. Whenever group s_1 is stimulated, we deliver a reward in the form of extracellular DA, with a delay of up to 1s. The experiment was conducted using two models:

1. The LIF model.
2. The Izhikevich model.

Prior to training, stimulating any group triggers a small response (a small number of spikes following the stimulus). Throughout the experiment, the synapses emanating from group s_1 are strengthened and eventually stimulating s_1 triggers a very strong response. The synapses emanating from other groups are strengthened as well, but to a lesser extent.

Pavlovian conditioning

We used a network of 1000 neurons (with 100,000 synaptic interconnections), and randomly chose 3 groups of neurons (A, B, C), each consisting of 50 neurons. We manually fully connected group B to group C with synaptic strength 1, and group A to group C with synaptic strength 0. During training we stimulate group A and B together and deliver a reward with a delay of up to 1s. Prior to training, stimulating group B triggers a strong response in group C, whereas stimulating group A triggers no response in C. During training the connections from group A to group C are reinforced and we want to show that eventually stimulation of group A results in a response in group C. This experiment was inspired by Pavlov's original experiment: group A represents the conditioned stimulus, group B represents the unconditioned stimulus and group C represents the unconditioned response.

Instrumental conditioning

We used a network of 1000 neurons (with 100,000 synaptic interconnections), and randomly chose 3 non-overlapping groups of neurons, each consisting of 50 neurons. We denote the groups by S, A and B. Group S represents the input stimulus to the network. A and B are two groups of motor neurons that give rise to 2 motor responses of the network. During training, we stimulate group S and count the number of spikes fired by neurons in A and B during a 20 ms time window after the stimulation. We denote these counts as $|A|$ and $|B|$ respectively. If $|A| > |B|$, we deliver a reward (in the form of extracellular DA) with a delay of up to 1 s. If $|A| < |B|$, no reward is delivered. The experiment consists of 1000 such trials separated by 10 s. Prior to training, stimulation of group S results in equal activity in A and B, on average. The purpose of training is to make group A fire more than group B in response to stimulus S. One might think of neurons in groups A and B as projecting to 2 motor areas that innervate 2 antagonistic muscles; to produce a noticeable movement, one group has to fire more spikes than the other group.

In another version of the experiment, we chose to reward the network only when its activity met a stronger requirement: $|A| > 2|B|$. In yet another version, we delivered negative rewards when $|B| > 2|A|$, in addition to positive rewards when $|A| > 2|B|$.

For specific details regarding parameter value selection please refer to the appendix.

Instrumental conditioning – Learning Agent Demo

As a final presentation of the Instrumental experiment we have presented a learning agent interacting with the environment. The agent motion function is such as a spring, i.e. it moves right and left returning to its center. In this stage the agent is a neural network, which receives a positive reward when reaching the left wall, and negative reward when reaching the right wall. After a training period we expect to see more motions of the agent toward the left wall.

Results

In this section we discuss the results of our experiments.

The "multiple stimuli" experiment has been fully reproduced as discussed in the work of Izhikevich 2007. It is evident that the network exhibits a significantly stronger response to stimulus s_1 than to other stimuli. However, when trying to reproduce those results with the LIF model, the results were not as good. Although the response of the network after stimuli s_1 was a bit stronger, we have not seen evident results that the network responds substantially different after stimuli s_1 compared to other stimuli. We rather

```
Procedure SIMULATION($chart$)
  While (Time Permits)
    $\{STDP(neuron) = 1 | neuron \in fired\}$
    INJECTCURRENT(S) ;update the current according to s
    Foreach $n \in fired$
      $preNeurons = \{neuron \in neurons | neuron \to n\}$
      $postNeurons = \{neuron \in neurons | n \to neuron\}$
      $c_{preNeurons,neuron} = c_{preNeurons,neuron} + 1.0 * STDP(preNeurons)$
      $c_{neuron,postNeurons} = c_{neuron,postNeurons} - 1.0 * STDP(postNeurons)$
    $c = c - dt * (c/\tau_c)$
    $s = s + dt * (c * DA)$
    $STDP = STDP - dt * (STDP/\tau_{stdp})$ ; stdp decay for all neurons
    $DA = DA + dt * (-DA/\tau_d + DA_t)$ ; dopamine decay for all neurons
```

Figure 1: Simulation

Figure 2: Multiple stimuli with LIF model. Each point on the graph represents the percentage of firing neurons (lower graph) after each of the stimuli (upper graph)

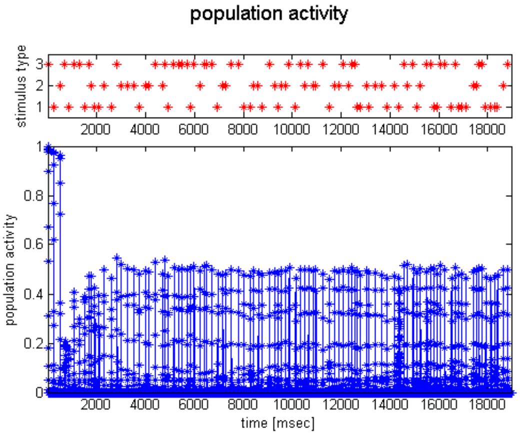
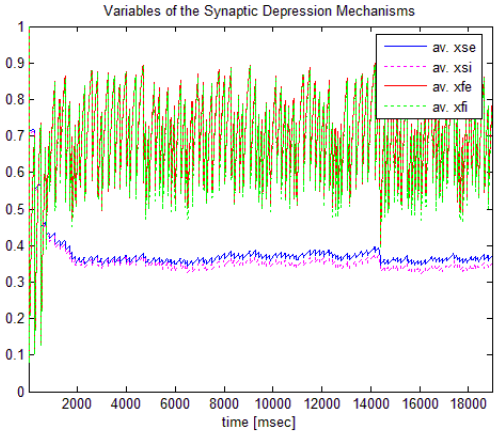


Figure 3: The depression parameters in the multiple stimuli with LIF model.



observed that the network responds with an excessive activity after all stimuli. The results can be seen in 2, with their corresponding depression parameter behavior 3.

In the Pavlovian conditioning experiment (see figure 4 and 5), we show that during training the connections from group A to group C are reinforced and we can observe that eventually stimulation of group A results in a response in group C. The figure shows that synaptic strength of the neurons connecting group A and C. The two figures show the results after a period of 5 minutes training. Almost all connections in the network have been stronger in the end of the training.

The instrumental conditioning experiment yielded results that are not entirely consistent with those in (Izhikevich 2007). When we used a relatively weak reward requirement such as $|A| > |B|$, each run of the experiment (consisting of 1000 trials) ended with a different result: sometimes synaptic connections to group A were stronger, and sometimes connections to group B were stronger. In figure 7 we show the results for when A exceeded B. Using a stronger requirement, such as $|A| > 2|B|$, yielded consistent results (see figure 6), especially with the addition of negative reinforcements (when the opposite requirement $|B| > 2|A|$ was met). Each point on the graph represents a spike count during a 20 ms window following the stimulus. It is evident that spike counts in group A are significantly higher than those in group B after a period of training.

Conclusions

We have shown how STDP modulated by DA may give rise to reinforcement learning with delayed reward. The experiments described in (Izhikevich 2007) have been reproduced successfully, except for the instrumental conditioning experiment, which we could only reproduce using a stronger reinforcement requirement (i.e. $|A| > 2|B|$).

One result which is evident in all our experiments is that although specific firing patterns result in reward, all firing patterns are reinforced during training. While we may expect only certain synapses to be strengthened, an overall increase in synaptic strength is apparent throughout the entire

Figure 4: Pavalovian Experiment. The synaptic strength between A and C, after 3 minutes of training.

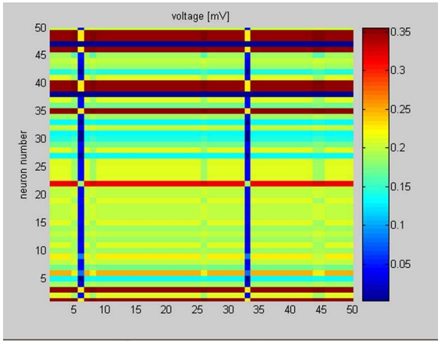
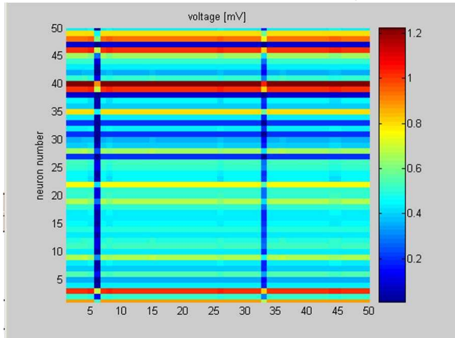


Figure 5: Pavalovian Experiment. The synaptic strength between A and C, after 5 minutes of training.



network. The firing patterns leading to reward, however, are reinforced more than others. In (Izhikevich 2007) it is suggested that random post-then-pre coincidental firings should lead to an overall decrease in synaptic strength over time, thereby leaving only the relevant synapses potentiated. We did not witness this phenomenon in our experiments, probably due to the rarity of coincidental firings. This gives rise to the question whether this suggested mechanism alone can account for the specificity in synaptic potentiation which we expect to find in biological neural networks.

Appendix

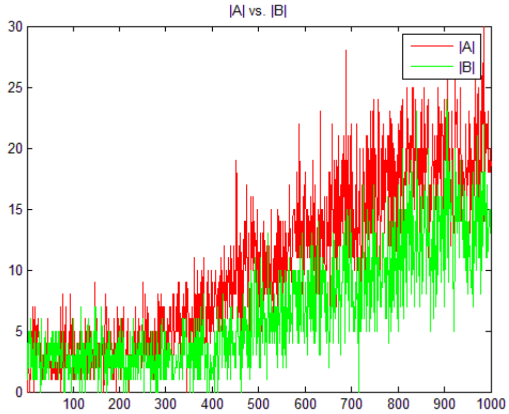
In our experiments we use the networks built with the following parameters:

1. $\tau_{stdp} = 0.02171$
2. $\tau_c = 1$
3. $\tau_d = 0.2$
4. $DA_t = 0.01$
5. excitatory neurons: 80 percent and inhibitory neurons 20 percent
6. $dt = 1e^{-3}$

Specific LIF model parameters:

1. $noise_std = 0.2$
2. $\tau_m = 10e^{-3}$

Figure 6: Instrumental with reward when $A > 2B$. Each point on the graph represents a spike count during a 20 ms window following the stimulus. The red graph is the response of the network (the number of spikes in the network) after stimuli A, and the green is after stimuli B.



3. $tau_{se} = 1$ s – slow synaptic depression mechanism, e – excitatory population
4. $tau_{si} = 1$ s – slow synaptic depression mechanism, i – inhibitory population
5. $Use = 0.05$
6. $Usi = 0.05$
7. $tau_{fe} = 200e^{-3}$ f – fast synaptic depression mechanism, e – excitatory population
8. $tau_{fi} = 200e^{-3}$
9. $Ufe = 0.25$
10. $Ufi = 0.25$

Specific IZHIKEVICH model parameters:

1. $a(excitatory) = 0.02; a(inhibitory) = 0.1$
2. $a(excitatory) = 8; a(inhibitory) = 2$
3. $maximalsynapticstrength = 4$ The synaptic weight can't be more than this

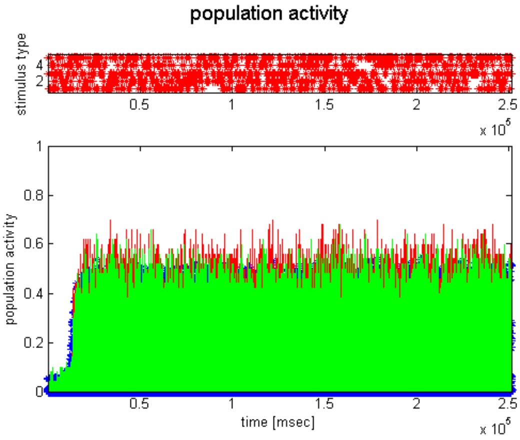
References

Burkitt, A. N. 2006. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biological Cybernectics.*

Izhikevich, E. 2004. Which model to use for cortical spiking neurons. *IEEE Transactions on Neural Networks.*

Izhikevich, E. M. 2007. Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cerebral Cortex.*

Figure 7: Instrumental with reward when $A > B$. Each point on the graph represents a spike count. The red graph is the response of the network (the number of spikes in the network) after stimuli A, and the green is after stimuli B.


population activity
stimulus type
x 10^5
population activity
time [msec]
x 10^5