Seminar Talk:
Learning in Spiking Neural Networks by Reinforcement of Stochastic Synaptic Transmission

Mark Kröll
Outline

- Purpose of the paper - what is it about?

- Introduction

- Hedonistic Synapse Model

- Training a Multilayer Network (XOR computation)

- Learning dynamical behaviours

- Conclusion

- Discussion
Synaptic transmission

- fundamental for neural computation
- process is surprisingly unreliable

new idea:

- using unreliability for learning purposes
- analogy: genetic replication used for evolution

new concept:

- introducing hedonistic synapses
- hedonistic means ’reward seeking’
- increasing probabilities of release or failure
- responding to a global reward signal
only two actions performed:
  - the synapse either releases neurotransmitter
  - or it fails to release

hedonistic synapse obeys the learning rule
  - the release probability is increased if reward follows release
  - and is decreased if reward follows failure

a synapse is modeled having two states (available and refractory)
  - when the synapse is available, a presynaptic spike stimulates vesicle release
  - there is a transition from available to refractory
  - release probability: \[ p = \frac{1}{1+e^{-q-c}} \]

  ▶ q ... release parameter
  ▶ c ... calcium-like variable to model calcium dynamics
the synapse being in the refractory state recovers with time constant $1/\tau_r$

- while refractory, the synapse cannot release a vesicle

a record of recent actions is needed (eligibility trace)

learning at each synapse:

- $\frac{dq}{dt} = \eta h(t)\overline{e}(t)$
  - is driven by the product of eligibility trace with the global reinforcement signal

for simplicity, short-term plasticity was eliminated from the first example

- $\rightarrow$ calcium-like variable $c$ is held constant
- $\rightarrow$ instantaneous recovery from refractory state ($\tau_r = 0$)
Figure 1: A hedonistic synapse responds to reinforcement by changing its release probability based on recent actions

\[ p = \frac{1}{1 + e^{-q}} \]
\[ q = \log \frac{p}{1-p} \]
Training a multilayer network(1)

consider a network of hedonistic synapses

network was trained to perform XOR computation

- binary variables are encoded in firing rates
- when the input was ”01” or ”10”
  the synapses were rewarded for each output spike
- when the input was ”00” or ”11”
  the synapses were punished for each output spike

200 presentations of each input pattern

results:

- the network learnt to respond to ”01” and ”10”
- and to suppress almost all spiking to ”11”
- a single time-varying reward signal sufficed to train the whole network
Figure 2: A network of hedonistic synapses and integrate and fire neurons learns the XOR function of two binary variables
end result of training → change of release probabilities to increase the reward received by the network

How were the synapses able to determine the changes in release probabilities appropriate for increasing reward?

short answer:

• by comparing two cross correlations
  
  ▶ one between release and reward
  ▶ one between failure and reward

• the difference between the correlations tells the synapse how its actions are causally related to reward

a simple example should illustrate this point
Figure 3: The difference between the two cross correlations provides an appropriate learning signal
Dynamic Synapses(1)

the efficacy of a biological synapse is not static but changes dynamically from spike to spike

a synaptic learning rule must be able to deal with that change

→ short term plasticity is introduced

- the probability of vesicle release to the first spike in the train affects the responses to later spikes
  - if the initial release probability is high → succeeding responses are depressed
  - if the initial release probability is low → facilitation is visible at first, after which depression sets in
Figure 4: Hedonistic learning applied to synapses with short-term facilitation and depression
Results:

- after training, the output neuron became selective to temporal order
- two groups of synapses learnt different dynamical behaviours
- selectivity of response was achieved
  - one group learnt to depress quickly, so that its peak response was immediate
  - the other group learnt to facilitate (by introducing a time lag into its peak response)
Conclusion

learning rules are compatible with features of biological synapses

- driven by presynaptic action potentials and change the membrane conductances of postsynaptic targets
- short-term facilitation and depression

present work is speculative

- however, biological synapses that rely on microscopic randomness for the purpose of optimization could still be correct
- concept of hedonistic synapses is only one realization

so one goal of the paper is to stimulate theorists and experimentalists
Thank you very much for your attention
question:

- There are many sources of randomness in the brain. Why using stochastic vesicle release?

answer:

- It should be made clear that there are many other stochastic processes that could be chosen (f.e. learning could be based on fluctuations due to irregular action potential firing). In this paper stochastic vesicle release was chosen for the following reasons: hedonistic synapses are simple and easy to understand. Second, stochastic vesicle release is a basic and universal property of chemical synaptic transmission. Third, hedonistic synapses are applicable to realistic model neurons.
Discussion (2)

question:

- Do you think learning with hedonistic synapses will scale up to really large networks?

answer:

- that is not realistic, but the difficulty of scale up is not peculiar to hedonistic synapses. It is common desease of all learning methods that use local search to optimize an objective function. Methods as ’backpropagation’ have had impressive successes, but none has scaled up to the challenge of creating artificial intelligence that rivals the human brain in capacity.
question:

- What if the reward signal is delayed in time?

answer:

- A fixed delay in reward has no effect on learning, provided that the same delay is added to the eligibility trace. A variable delay is more problematic: it requires that the time constant of the eligibility trace is made as long as the delay fluctuations → slowdown in learning. In general, one expects temporal delays to slow down hedonistic learning.