Learning Anticipation through Priming in Spatio-Temporal Neural Networks

Nooraini Yusoff[1] and André Grüning[2]

[1] School of Computing, UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia
nooraini@uum.edu.my

[2] Department of Computing, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK
a.gruning@surrey.ac.uk

Abstract. In this paper, we propose a reward-based learning model inspired by the findings from a behavioural study and biologically realistic properties of spatio-temporal neural networks. The model simulates the cognitive priming effect in stimulus-stimulus-response association. Synaptic plasticity is dependent on a global reward signal that enhances the synaptic changes derived from spike-timing dependent plasticity (STDP) process. We show that by priming a network with a cue stimulus can facilitate the response to a later stimulus. The network can be trained to associate a stimulus pair (with an inter-stimulus interval) to a response, as well as to recognise the temporal sequence of the stimulus presentation.

Keywords: Reward-based learning, Spiking neural networks, Priming effect, Associative learning

1 Introduction

In 1999, Erickson and Desimone published a behavioural experiment on visual discrimination tasks with primates [1]. Well-known as the GO/NO-GO experiment, for each trial the subjects were shown a visual image, namely a predictor, proceeded by another image, namely a choice, with a delay, whilst the neuronal activity in those primates was observed following some conditional performance rules. The performance was measured through actions by the subjects that required them to release or not to release a bar followed by a reward, i.e. a reward was applied if the subjects indicated if the choice was the correct match of the predictor by releasing the bar, or the subjects did not release the bar if the stimuli were unmatched. From their experiments, after a number of learning trials, recordings from cells in the associative cortex of the monkeys had shown persistently increasing activity in the brain when presented with a stimulus. The activity was indicating not only response to the shown stimulus but also the stimulus that the subjects were expecting to be seen, one that had been associated to the shown stimulus. Hence, the recordings from cortical neurons have concluded two types of task-related activity in the brain namely retrospective - related to previously shown stimulus and prospective - related to a stimulus that the brain expects to appear [1], [5].

We perceive their finding as the effect of priming in memory recall. The cognitive behaviour of priming effect shows signs of influence of previous information on the perception of subsequent information [2], [6]. The effect is a result of spread activation mechanism in the brain in which a recently probed stimulus invokes its associated information, consequently strengthening the retrieval of information of a later proceeded stimulus when both are related. In many circumstances, depending on the type of a cue, negative or positive, we can observe the effective use of priming practiced in our daily life. For example, the priming effects of television food advertising on eating behaviour, political campaigns on poll counts, and landmarks on route following.

For this study, inspired by the behavioural experiment findings in [1], we propose a reward-based learning scheme for mapping delayed stimulus pair to a target response. The learning is goal-directed with only minimal information on the target response in order to maximise reward. Furthermore, we use a generic recurrent neural network with random and sparse connectivity that simulates spiking behaviours of the real cortical neurons. The strength of our work could be ascribed to the implementation of the learning protocol as suggested in behavioural study by [1], with integration of biologically realistic properties of spiking neural network found in [10]. We demonstrate that, in such environment with rich and realistic dynamics, a network can be trained to exhibit the benefits of priming effect in the brain. This is shown through the effect of stimulus cueing that facilitates recognition to a target response. The network can be trained not just to associate stimulus-stimulus to a response but also to recognise the temporal sequence of the stimuli.

2 Neural Network Model

For learning experiments, we simulate a recurrent neural network consisting of 1000 neurons with 80% excitatory and 20% inhibitory spiking neurons, as proposed in [10]. The connectivity between neurons is random and sparse with probability of $p = 0.1$ (no self-feedback). Each excitatory neuron is randomly connected to 100 neurons, and each inhibitory neuron is randomly connected to 100 excitatory neurons only. The synaptic transmission delay is set randomly between 1 to 20 ms. Synaptic weights are initialised with 1.0 and -1.0 mV for excitatory and inhibitory weights, respectively. In our model, learning only affects the connections between excitatory to excitatory neurons, and excitatory to inhibitory neurons, whilst the rest are not updated (i.e. not plastic). The range of modifiable weights (i.e. excitatory synapses) is $0 \leq w \leq 4.0$ mV.

The excitatory population is divided into subpopulations of neurons namely m stimulus groups (S), non-selective neurons (NS) and n response groups (R). Each group S composed of 50 neurons represents a certain stimulus, meanwhile neurons in the NS group are assumed to be not selective to any stimulus. Each response group consists of 100 excitatory neurons. In our model the inhibitory subpopulation IH only acts as a global inhibition. The dynamic properties of a neuron are based on Izhikevich spiking neuron model [9]. In this study, to exhibit the spiking behaviours of pyramidal cells and interneurons in the cortical network, all excitatory neurons are regular spiking type neurons and all inhibitory are fast spiking neurons (detailed description can be found in [3]).

3 Synaptic Plasticity Rules

We implement a reward-based learning to associate a stimulus pair to a target response. In this study we refer the stimulus pair as the predictor-choice pair denoted by (S_i, S_j). The predictor S_i is a cue for the response of the later stimulus, the choice S_j. The network is given a positive reward for a correct response to a pair, or otherwise negatively rewarded for an incorrect response. The reward signal acts as a reinforcement signal that consolidates the synaptic changes derived from a spike-timing dependent plasticity (STDP) function, as in 1.

$$\Delta w_{stdp} = \begin{cases} A_+ e^{\frac{-\Delta t}{\tau+}} & \text{if} \quad \Delta t \geq 0 \\ A_- e^{\frac{\Delta t}{\tau-}} & \text{if} \quad \Delta t < 0 \end{cases} \tag{1}$$

where the spike timing-dependent synaptic change $\Delta_w stdp$ is determined by the difference in firing times between a postsynaptic neuron and its presynaptic ($\Delta t = t_{post} - t_{pre}$). The synaptic potentiation is applied for $\Delta t \geq 0$ and synaptic depreciation is applied for otherwise. The magnitude of synaptic change is given by $A_+ e^{\frac{-\Delta t}{\tau+}}$ (for potentiantion) and $A_- e^{\frac{\Delta t}{\tau-}}$ (for depreciation), where A is the maximal change when the Δt is approaching 0, and τ is the time constant (in ms). For our STDP curve, $\tau_+ = \tau_- = 20$ ms, $A_+ = 0.1$, and $A_- = 0.15$ (following [8]).

In learning with modulated spike-timing dependent plasticity, the synaptic change is dependent on a reward signal, $r(t)$ and an eligibility trace, $z(t)$. $r(t)$ is determined based on a reward policy that counts the number of neuron firings (F) of response groups within an interval of 20 ms from the onset of a choice (see 2).

$$r(t) = \begin{cases} r(t-1) + 0.5 & \text{if} \quad F_i \geq 2F_j \\ 1 - F_j/F_i & \text{if} \quad F_j < F_i < 2F_j \\ -0.1 & \text{if} \quad F_i < F_j \end{cases} \tag{2}$$

From 2, F_i and F_j are the number of firings of a target response group, and non-target group, respectively. The type of reward (i.e. strong positive, weak positive and negative) determines the rate of the signal. The eligibility trace, $z(t)$ is the summation of $\Delta_w stdp$. Therefore, the synaptic change (3) is read as [7], [8]:

$$\Delta w(t) = [\alpha + r(t)]\, z(t) \tag{3}$$

where α is the activity-independent increase of synaptic weight.

3.1 Learning Protocol

Every learning simulation runs for 20 mins. A network is given a set of pair-response mappings $(S_i, S_j) \to R_k$, with different pairing strategies depending on the task. The purpose of learning is to train the network to associate a predictor-choice pair (S_i, S_j) to a target response, R_k. For current stimulation, the network is supplied with a superthreshold current of 20 pA that could immediately trigger any stimulated neuron to fire. In our experiment, a network is set with a background activity, that we randomly stimulate a neuron for every ms.

For learning initialisation, we start up a network with only background activity for 100 ms. As mentioned earlier, the background activity is implemented via stimulation to an arbitrary neuron with 20-pA current for every ms. Within this interval, the network is only at an asynchronous state with absence of intensified currents to target groups. At time $t = 0$, the membrane potential for each neuron is set to $v =$ -60 mV, just above the resting potential ($v =$ -65 mV). This is to assume some activity prior to learning as well as to facilitate neuron activation.

After the initialisation phase, we begin a learning trial by presenting the predictor stimulus, S_i. For this purpose, at time $t = t_n$, we stimulate all neurons in S_i (with 1-ms pulse 20-pA current). After an inter-stimulus interval (ISI), at time t $= t_{n+ISI}$, we stimulate all neurons with the same amount of current to the paired choice stimulus, S_j. We choose an optimal ISI from a range of 10 to 50 ms based on a preliminary experiment. The presentation of learning pairs is done randomly with uniform distribution. From the onset of the choice stimulus, we observe the activation in the response groups within a 20-ms interval. The winner of response groups is the one with the highest number of neuron activations. As described in Section 3, depending on the firing rate, the network is rewarded with strong positive, weak positive or negative reinforcement.

We then test the trained network with the same stimulus presentation settings by recalling the learned pairs, unlearned pairs, and noisy stimuli. To generate the noisy stimuli, we vary the number of neurons for random stimulation in learned groups with probability of less than 1.0. The testing result shows the average percentage of performance over a number of trials, i.e. performance = (number of correct recalls/number of trials)*100.

4 Simulation Results

We began training a network with exclusive stimulus groups in which each neuron was a member of one group only. From the population of 800 excitatory neurons, we selected 8 non-overlapping stimulus groups of 50 excitatory neurons each, and 2 exclusive groups of 100 excitatory neurons each were selected as response groups, R_m, i.e. $R_0 = A$ and $R_1 = B$, where A and B are the group labels. The learning set is as follows: $Pair - Response = \{(S_0, S_1) \to A, (S_2, S_3) \to B, (S_4, S_5) \to A, (S_6, S_7) \to B\}$.

We stimulated all 50 neurons in a predictor group S_i proceeded by a stimulation to all neurons in its choice S_j. Initially, the ISI between the predictor and the choice was fixed to 10 ms as the average of synaptic transmission delays, 1 to 20 ms. For testing (i.e. probe trial), at this stage of experiment, we as well stimulated all neurons in the learned groups prior to investigating error tolerance in a tested stimulus group as opposed to stimulate a fraction $p < 1.0$ of them.

The correct mapping of stimulus pair to target response for training and testing, respectively, were achieved at 94.08% and 99.9%. As a result of learning, the averaged number of spikes for target group is 9.98, when compared with the non-reinforced group with 7.18 and the negatively rewarded group with 3.15.

4.1 Inter-stimulus Interval (ISI)

In the following experiment, we studied how the delay between a predictor and a choice in a pair influenced the association to a reinforced response group. We have trained the network with a set of ISIs in between 10 to 50 ms. The network learned to associate a stimulus pair to its target response when the ISI $\leq$ 20 ms with the averaged performance achieved for training and testing were 82.2% and 91.07%, respectively. For ISI $>$ 20 ms, the average performance was below the chance level and the invariance of spike counts in the target, non-target and control groups was low indicating only random activity in those groups.

The optimal performance was achieved when the choice was delayed 15 ms after the onset of the predictor. When the stimulation delay was 10 ms, we found only small variance ($p < 0.1$) of the averaged performance between the recognition of learned stimulus pairs (S_i, S_j) and unlearned stimulus pairs (S_j, S_i), i.e. 99.9% (for learned stimulus pair) and 98.1% (for unlearned stimulus pair). This is to show that shorter ISI could cause dominance of a predictor over its choice in a learning pair. For example a network trained with $(S_0, S_1) \to A$ for ISI = 10, when tested with an unlearned pair of (S_1, S_0) would also respond to A with high probability. The network only strongly associated the predictor to its pair response, $S_0 \to A$, hence resulting in less effect of the choice, S_1. The discrepancy in testing performance between learned and unlearned pairs was higher ($p < 0.1$), when the delay between stimuli was greater than 10 ms (and delay $\leq$ 20 ms). The error rate of the recall to unlearned stimulus pairs increased from 7.24% (ISI $\leq$ 10 ms) to 17.11% (ISI $>$ 10 ms) implying the network had learned the sequence of presentation of stimuli.

We then further analysed the performance of learning with ISI = 15 ms and ISI = 20 ms. There was a trade-off between the increase in ISI and response rate. An increase in ISI decreased the activation of neurons in the target groups. When trained with ISI = 20 ms, variance of averaged spike counts in the target, non-target and control (non-reinforced response) groups was low giving the ratio of averaged activations of target to control, and target to non-target, 3.60: 2.51 and 3.60:2.08, respectively. On the other hand, networks trained with ISI = 15 ms was observed with averaged activations of 6.71: 4.38 (target:control) and 6.71:2.44 (target: non-target). Hence, the optimal delay between predictor and choice has been found at 15 ms. In a learning trial, by priming a network with a predictor delayed by ISI = 15 ms, can still facilitate the response to its choice.

4.2 Probe Trials

After the network had been trained with $Pair - Response = \{(S_0, S_1) \rightarrow A, (S_2, S_3) \rightarrow B, (S_4, S_5) \rightarrow A, (S_6, S_7) \rightarrow B\}$, we ran a number of probe trials for different recall tasks.

a) Recalls with random selectivity of the neurons in the stimulus groups
During training, at stimulation times t_n and t_{n+ISI}, all neurons (N=50) in the predictor and choice groups were supplied with a superthreshold current of 20 pA. To test for noise robustness in a response recall, for every learned stimulus pair, we performed response recalls by randomly activating neurons in the predictor and choice groups. We tested a trained network with selectivity of neurons to be stimulated, p_n, from 0.5 to 1.0. The averaged recall performance over 100 probe trials showed that each stimulus group required minimal activations of 70% (35 out of 50) of neurons at minimum of 65.48% of correct recalls. In other words, a network with random synaptic connectivity of 0.1, tolerated maximal distortion probability of 0.3.

b) Recalls with only choice groups (neutral condition), congruent pairs, and incongruent pairs
We ran a series of probe trials to see the effect of priming in response recalls. A trained network was probed with 3 conditions of stimuli namely neutral-the network was only presented with learned choices without their predictor, $\{S_1, S_3, S_5, S_7\}$, congruent-learned paired stimuli (predictor-choice), $\{(S_0, S_1), (S_2, S_3), (S_4, S_5), (S_6, S_7)\}$, and incongruent-predictor and choice with conflicting responses, $\{(S_0, S_3), (S_2, S_1), (S_4, S_7), (S_6, S_5)\}$. For trials with congruent and incongruent conditions, the ISI was 15 ms. In response to a single stimulus (neutral condition-with choice only), the averaged performance over 100 trials was 53.93%. When presented with congruent pairs, the percentage of correct recalls achieved 95.85%, meanwhile averaged correct recalls when responding to incongruent pairs decreased to 42.28%. This indicates a facilitation effect when a choice is preceded with its correct paired predictor. Priming the network with a predictor acting as a cue to its choice gives advantage in recalling the response. Meanwhile high competition or interference exists when the network is probed with predictor-choice having conflicting target responses.

4.3 Learning with Non-exclusive Stimulus Groups

In the previous experiments we trained a network with exclusive paired predictor-choices. The network only associated different stimulus groups with their target response, A or B. In the following experiments, we investigated the learning performance with different pairing strategies. We manipulated the sequence of predictor-choice and exclusivity of the predictor and/or choice in a learning trial. Here we present four conditions of learning experiments with different sets of non-exclusive stimulus groups; non-exclusive predictor (I), non-exclusive choice (II), orthogonally identical learning pairs (III), and asymmetrically different learning pairs (IV). For each case, we trained the network with stimulus pairs having correlated spike patterns as the studied pairs, and with exclusive stimulus groups as the controls. Table 1 shows the probe trial results for the studied pairs.

Table 1. Correct recall to target response in probe trials

Condition	*Pair – Response*	*Correct*(%)
I	$\{(\mathbf{S_0},\mathbf{S_1}) \rightarrow \mathbf{A}, (\mathbf{S_0},\mathbf{S_2}) \rightarrow \mathbf{B}, (S_3,S_5) \rightarrow A, (S_4,S_6) \rightarrow B\}$	69.10%
II	$\{(\mathbf{S_0},\mathbf{S_1}) \rightarrow \mathbf{A}, (\mathbf{S_2},\mathbf{S_1}) \rightarrow \mathbf{B}, (S_3,S_5) \rightarrow A, (S_4,S_6) \rightarrow B\}$	68.30%
III	$\{(\mathbf{S_0},\mathbf{S_1}) \rightarrow \mathbf{A}, (\mathbf{S_1},\mathbf{S_0}) \rightarrow \mathbf{B}, (S_3,S_5) \rightarrow A, (S_4,S_6) \rightarrow B\}$	71.90%
IV	$\{(\mathbf{S_0},\mathbf{S_1}) \rightarrow \mathbf{A}, (\mathbf{S_1},\mathbf{S_2}) \rightarrow \mathbf{B}, (S_3,S_5) \rightarrow A, (S_4,S_6) \rightarrow B\}$	78.05%

* percentage shows the performance of the studied pairs (in bold)

Generally, our algorithm can be applied to train a network to learn temporal sequences particularly for pair associate tasks consisting of learning pairs with non-exclusive predictor or choice, those are orthogonally identical and asymmetrically different. However, correlation of spike patterns between two paired stimuli decreases response discrimination rate. The averaged correct recall to target for studied pairs is lower than the controls with exclusive neuron membership. From those tested conditions, we have found that a network trained with non-exclusive choice suffers more interference compared to other conditions. This is the result of delay between stimuli allowing small depression due to random spikes (noise) and the absence of reward in the period of ISI. In such cases, the contribution from the predictor decreases. On the other hand, the ISI of 15 ms can sufficiently give temporal difference between two orthogonally identical pairs.

5 Conclusion

In this paper, we show that, in a stochastic and noisy environment, a network can be trained to perform a stimulus-stimulus-response association. We use a generic neural network with realistic properties in Izhikevich spiking neuron model. By cueing the network with a prime stimulus known as the predictor, it can facilitate the response to a later stimulus, the choice, even both are separated by a temporal delay. For this, we support the behavioural experiment finding that concludes the priming effect that could benefit a response processing [1]. As a result of learning, we show that the network can learn stimulus-stimulus association as well as the anticipation to the response of the following stimulus. Hence, this could lead to a range of applications, for example an agent can be trained to associate stimuli for visual recognition task, path tracking, and multimodal person authentication e.g. audio-visual. Unlike other existing supervised gradient-descent based learning approaches, our model does not require any spike template for learning target. Learning supervision is only dependent on a global reinforcement signal determined by a reward policy with minimal assumption about the network dynamics. The network activity is adjusted to maximise the reward signal rate, and as a result of learning through a series of reward-actions, the target reinforced synapses are strengthened.

For the work under progress, we are improving the model performance in learning with higher response competition. In our current model, the connectivity between neurons is random and sparse. Without any inhibition mechanism or anatomical constraint, the model could perform well for learning with exclusive stimulus groups. Nevertheless, for learning with high correlation in spike patterns, the model performance decreases for learning with non-exclusive stimulus groups. In addition, with the current network connectivity, it may lead to undesired causal firings. For learning with competing target responses, A and B, in which reinforcing the synapses $S_i \rightarrow A$ could also lead to triggering of synapses $A \rightarrow B$. Some initial results (not discussed in this paper) have shown some lights on the practicality of our to be proposed lateral inhibition mechanism in improving performance for more competitive learning environment.

Acknowledgments. This research has been funded by the Ministry of Higher Education (Malaysia) and partially supported by EPSRC grant EP/I027831/1.

References

1. Erickson, C. A., Desimone, R.: Responses of Macaque Perirhinal Neurons during and after Visual Stimulus Association Learning. Journal of Neuroscience 19(23), 10404–10416 (1999).
2. Filippova, M. G. Does Unconscious Information Affect Cognitive Activity?: A Study Using Experimental Priming. The Spanish Journal of Psychology 14(1), 20–36 (2011).
3. Izhikevich, E. M.: Which model to use for cortical spiking neurons. IEEE Trans. Neural Networks 15(5), 1063–1070 (2004).
4. Tulving, E., Schacter, D. L., Stark, H. A.: Priming Effects in Word Fragment Completion are independent of Recognition Memory. Journal of Experimental Psychology: Learning, Memory and Cognition 8(4), 336–342 (1982)
5. Mongillo, G., Amit, D.J., Brunel, N.: Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network. European Journal of Neuroscience 18, 2011–2024 (2003)
6. Schacter, D. L. Priming and multiple memory systems: Perceptual mechanisms of implicit memory. Journal of Cognitive Neuroscience 4(3), 244-256 (1992).
7. Florian, R.V.: Reinforcement learning through modulation of spike-timing dependent synaptic plasticity. Neural Comput. 6, 1468–1502 (2007)
8. Izhikevich, E.M.: Solving the distal reward problem through linkage of STDP and dopamine signaling. Cereb Cortex 17 2443–2452 (2007)
9. Izhikevich, E.M. Simple Model of Spiking Neurons. IEEE Trans. Neural Networks 14(6), 1569–1572 (2003)
10. Izhikevich, E.M.: Polychronization: Computation with Spikes. Neural Computation 18, 245–282 (2006)