LEARNING A DYNAMICAL ATTRACTOR BY MAXIMIZING MULTIVARIATE INFORMATION

ABSTRACT

It has been previously proposed that the human perceptual system can be viewed as a statistical inference engine whose function is to infer the probable causes of sensory input. From the neuroethological point of view, the evolution of such a system should be guided by the stimuli to which it is exposed. Here we present a recurrent spiking neural network model capable of efficiently maximizing information about a stimulus via spike timing dependent plasticity mechanisms. We show that information gradients can be estimated by the quality factor of the network response and that such gradients "carve out" the basins of a dynamical attractor. Interestingly, resulting networks exhibit properties consistent with previous work examining dynamical properties of neural networks well-suited for computation, such as criticality.

1 INTRODUCTION

It is thought that neural networks that lie at a critical point between order and chaos maximize basic processing properties such as sensitivity, dynamic range, correlation length, information transfer, and susceptibility. Multiple signatures of criticality have been proposed in the literature, one being power-law distributions of neuronal avalanches following a branching process similar to that seen in physical theories of avalanches. Others have searched for information theoretic signatures of criticality. However, it has recently been demonstrated that criticality is not a universal determinant of computational power; rather, critical networks are only well-suited for more complex tasks such as those that require integration of stimulus information over long timescales [1]. It has also been shown that a set of readout neurons trained to extract information from a pool of recurrently connected "reservoir" neurons can extract meaningful information from the reservoir. While such a combination of information theory and biologically realistic learning rules is a pioneering approach, the highly specific tasks that readout neurons are trained to accomplish degrades the interpretability of the evolution of synaptic weights and the generalization bounds of the readout given the statistical structure of the stimulus.

To begin to answer this question, we write down the matrix mechanics that describe a simulation of N spiking neurons with evolving synaptic weights and fixed connectivity. By manipulating connectivity matrices as suggested in [2], we search for systematic methods for training sub-critical, critical, and super-critical networks. Synaptic weights evolve according to a spike timing dependent plasticity (STDP) rule which will allow for adaptation to a stimulus with well-defined statistical structure in an unsupervised fashion. Criticality of trained networks will be assessed according to established measures e.g., the branching parameter, autocorrelation time, and avalanche size distributions. With these different types of networks in hand, we examine the dynamics of synaptic weights and the relationship between criticality and information transfer to a set of readout neurons.

2 METHODS

2.1 NETWORK MODEL

Networks are recurrently connected and contain N neurons each of which is excitatory with probability p_e and inhibitory with probability p_i. Synapses are defined mathematically as a pair of matrices $\Phi = (\mathbf{A}, \mathbf{B})$ where $\mathbf{A}$ is a recurrent weight matrix and $\mathbf{B}$ is an input weight matrix. The raw connectivity of the network is defined by two binary matrices: $\tilde{\mathbf{A}} \in \mathbb{F}_2^{n\times n}$ and $\tilde{\mathbf{B}} \in \mathbb{F}_2^{n\times k}$ where k is the total number of inputs. The full model is initialized as

$$\mathbf{\Phi} = (\mathbf{A}_0 * \tilde{\mathbf{A}}, \mathbf{B}_0 * \tilde{\mathbf{B}})$$

where $*$ represents an element-wise product. Since the chemical synapse is unidirectional, we must have that if $\tilde{a}_{ij} \neq 0$ then $\tilde{a}_{ji} = 0$ or vice versa. In other words, if the weight $a_{ij} \neq 0$, then neuron j is the postsynaptic cell and if $a_{ji} \neq 0$ then neuron i is the postsynaptic cell. One obvious consequence of this is that rows of $\mathbf{A}$ are inputs and columns of $\mathbf{A}$ are outputs.

First, we will partition $\tilde{\mathbf{A}}$ into four partitions: excitatatory-excitatory, excitatory-inhibitory, inhibitory-excitatory, and inhibitory-inhibitory. The size of these partitions are determined by the above probabilities. For a neuron i in the excitatory partition, the probability of a connection to a neuron in the inhibitory partition j i.e. the probability $\tilde{a}_{ij} \neq 0$ is $p_{e \to i}$. A similar definition follows for the inverse $p_{i \to e}$. The excitatory-excitatory and inhibitory-inhibitory probabilities are $p_{e \to e}$ and $p_{i \to i}$, respectively.

However, we need to prevent bi-directional connections while maintaining p_{ee}. For example, we delete redundancies between the upper and lower triangles of the excitatory-excitatory paritition. Deletion occurs with equal probability for either triangle which will remove $(Np_e)^2 p_{ee}^2$ connections on average. To account for this, we adjust probabilities according to the following equations

$$\begin{aligned} p'_{ee} &= p_{ee} + p_{ee}^2 \\ p'_{ii} &= p_{ii} + p_{ii}^2 \\ p'_{ie} &= p_{ie} + p_{ie} \cdot p_{ei} \\ p'_{ei} &= p_{ei} + p_{ie} \cdot p_{ei} \end{aligned}$$

The input connectivity matrix $\tilde{\mathbf{B}}$ is simpler. We only need to divide $\tilde{\mathbf{B}}$ into two partitions for excitatory inputs and inhibitory inputs. Inputs are, by definition, unidirectional so we can just set two probabilities for excitatory inputs and inhibitory inputs: q_e and q_i.

We will start by looking for constraints on $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ that correspond to sub-critical, critical, and super-critical networks. In [9] it was suggested that the input strength relative to recurrence determines network criticality. We pursue this line of reasoning, looking for matrices that satisfy a particular degree of recurrence.

Apriori it is difficult to say how the ratio of each p and q affects network criticality although recent results have shown that stronger recurrence relative to the input can increase criticality. We term the family of ratios here *recurrence ratios* and we will analyze the sensitivity of computational properties of the network to these values.

We initialize $\mathbf{A}_0$ from $\mathbf{B}_0$ by drawing connection weights from a uniform distribution $\mathbf{A}_0 \sim U^{n \times n}[0, g_{max}]$.

2.2 A GENERAL MODEL

Let $\mathbf{V}[I]$ be a vector recording the voltage per neuron for the duration of the simulation. We adopt the following model for the time evolution of $\mathbf{V}[I]$:

$$\begin{aligned} \tau_m \frac{d\mathbf{V}[i]}{dt} &= \sum_n \mathbf{I}_n[i] \\ &= \sum_n g_n(\mathbf{V}[i] - E_n) \end{aligned}$$

where $\mathbf{I}_n[J]$ is a vector of currents of class n e.g., excitatory current and E_n is the reversal potential. Using our definition of the conductance matrices in the preceding section, we have

$$\tau_m \frac{d\mathbf{V}[I]}{dt} = (\mathbf{V}[I] - E) \cdot \sum_j \mathbf{A}[I, j] + (\mathbf{V}[I] - E_{in}) \sum_k \mathbf{B}[I, k]) \tag{1}$$

However, since $\mathbf{A}$ and $\mathbf{B}$ are not static matrices, we would need more information to write down a solution to (1). It turns out that we cannot write down a general solution since the evolution $\mathbf{A}$ and $\mathbf{B}$ is non-deterministic. To even write the update equations we need to constrain their evolution

We additionally define the *observable state* matrix $\mathbf{Z}$ which is drawn at each time step from a distribution p

$$p(\mathbf{Z}[i,t] = 1|\mathbf{V}[I]) = \sigma(\mathbf{V}[I])$$

where σ is a sigmoidal activation function.

2.2.1 SYNAPTIC DYNAMICS

We begin with a naive version of synaptic plasticity, reserving the more sophisticated discussion until we have define the relevant model components. In the naive model, we consider two mechanisms by which synaptic conductances can change (i) release of neurotransmitter into the synaptic cleft and (ii) long-term potentiation and depression via spike timing. The former is a rather simple update to $\mathbf{A}$ and $\mathbf{B}$ computed from $\mathbf{Z}$

$$\tau_A \frac{d\mathbf{A}}{dt} = -\mathbf{Z}\mathbf{A}$$
$$\tau_B \frac{d\mathbf{B}}{dt} = -\mathbf{Z}\mathbf{B}$$

where the initial conditions will, in general, depend on the baselines set by synaptic plasticity. For the latter criterion, we look for a Hebbian update to the conductance matrices $\mathbf{A}$ and $\mathbf{B}$ which are also derived from $\mathbf{Z}$, ignoring sub-threshold activity.

For potentiation, consider the block $\mathbf{Z}[i, t-\tau]$ where $\tau > 0$ and the vector describing recurrent inputs to neuron i: $\tilde{\mathbf{A}}[i, J]$. We use this vector to zero out spikes irrelevant to neuron i by taking an element-wise product for each $t' \in [t-\tau, t]$

$$\mathbf{U}[i, J, t] = \sum_{t-\tau}^{t} F\left(\mathbf{Z}[i, t-\tau] * \tilde{\mathbf{A}}[i, J]\right)$$

Similarly, for depression we use the vector describing recurrent outputs of neuron i

$$\mathbf{U}[i, J, t] = \sum_{t-\tau}^{t} G\left(\mathbf{Z}[i, t-\tau] * \tilde{\mathbf{A}}[I, j]\right)$$

Computing this for all i and suppressing indices allows us to write the abstract update

$$\Delta\mathbf{A} = \lambda\mathbf{U} + \eta(t)$$

where $\eta(t) \sim P^{n\times n}$ and $P = \mathcal{N}(\mu, \sigma^2)$ is white-noise.

2.2.2 INFORMATION MAXIMIZATION

The above equations in their general form are unlikely to be able to permit any meaningful form of learning. To rectify this, we can make use of recent postulates on efficient coding, where synaptic plasticity is the learning rule by which a network finds a solution to a rate-distortion problem [x]. Simply put, we would like to know how a network can estimate information gradients and how gradient ascent can be achieved via synaptic plasticity. In other words how does a network estimate the gradient of the following Lagrangian

$$\mathcal{L} = I(\tilde{R}; S) - \gamma(\tilde{R}; R)$$

where R, $\tilde{R}$ and S satisy the Markov chain $S \to R \to \tilde{R}$. Maximization of the above Lagrangian provides an estimate of distribution $P(\tilde{R}|R)$. In the past decade, Williams et al. developed a method that exhaustively decomposes the Shannon information in a multivariate system in terms of the redundancy between synergies of subsets of the sources [8].

2.2.3 UPDATE EQUATIONS

$$\mathbf{1} : \mathbf{V} \to \mathbb{F}_2^{n \times t}$$

Now we can write a stochastic update for the entire system

$$\mathbf{V}[I,t] = \frac{\Delta t}{\tau_m}\left((\mathbf{V}[I,t-1] - E) \cdot \sum_j \mathbf{A}[I,j,t-1] + (\mathbf{V}[I,t-1] - E_{in}) \cdot \sum_k \mathbf{B}[I,k,t-1]\right)$$

3 THE STIMULUS

We would expect that that statistical structure of the stimulus would be causally related to the time evolution of the weight matrix $\mathbf{A}(t)$. Let the stimulus be white-noise $s \sim Q^t$ where $Q = \mathcal{N}(\mu, \sigma^2)$. We will parameterize a Poisson spike generator with an estimate of the spike rate $r(t)$ for this stimulus. To do that, we can use to generate an estimate of the spike rate using the Volterra expansion

$$r(t) = r_0 + \int_0^\infty D(\tau)s(t-\tau)d\tau$$

where we estimate $D(\tau)$ according to Adelson and Bergen (1985)

$$D(\tau) = \alpha \exp(-\alpha\tau)\left(\frac{(\alpha\tau)^5}{5!} - \frac{(\alpha\tau)^7}{7!}\right)$$

and we take $\alpha = 1/15$

REFERENCES

[1] Buesing et al. A spiking neuron as information bottleneck

[2] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities

[3] Maass et al. Edge of chaos and predicting computational performance for neural circuit models

[4] Cramer et al. Control of criticality and computation in spiking neuromorphic networks with plasticity

[5] Zierenberg et al. Homeostatic plasticity and external input shape neural network dynamics

[6] Beggs et al. Universal critical dynamics in high-resolution neuronal avalanche data

[7] Chalk et al. (2019) Toward a unified theory of efficient, predictive, and sparse coding *PNAS*

[8] Williams et al. (2010) Nonnegative decomposition of multivariate information *arXiv*

[9] Bellec et al. (2020) A solution to the learning dilemma for recurrent networks for spiking neurons *arXiv*