

Synaptogenesis: Constraining Synaptic Plasticity Based on a Distance Rule

Jordi-Ysard Puigbò¹(✉), Joeri van Wijngaarden¹, Sock Ching Low¹,
and Paul F.M.J. Verschure^{1,2}

¹ Laboratory of Synthetic, Perceptive, Emotive and Cognitive Science (SPECS),
DTIC, Universitat Pompeu Fabra (UPF), Barcelona, Spain

jordiysard.puigbo@upf.edu

² Catalan Research Institute and Advanced Studies (ICREA), Barcelona, Spain

Abstract. Neural models, artificial or biologically grounded, have been used for understanding the nature of learning mechanisms as well as for applied tasks. The study of such learning systems has been typically centered on the identification or extraction of the most relevant features that will help to solve a task. Recently, convolutional networks, deep architectures and huge reservoirs have shown impressive results in tasks ranging from speech recognition to visual classification or emotion perception. With the accumulated momentum of such large-scale architectures, the importance of imposing sparsity on the networks to differentiate contexts has been rising. We present a biologically grounded system that imposes physical and local constraints to these architectures in the form of synaptogenesis, or synapse generation. This method guarantees sparsity and promotes the acquisition of experience-relevant, topologically-organized and more diverse features.

Keywords: Machine learning · Connections · Biologically constrained

1 Introduction

Typically, artificial networks of learning nodes, or neurons, have been conceived as all-to-all connected. There are two reasons to begin with this approach. First, at a low scale, biological neurons are strongly connected with most of their neighbors, small networks being approximated in this way. Second, the flow of information among plastic networks is usually computed through dot products. This leads to fully disconnected networks becoming fully connected with the slightest amount of noise. This becomes equivalent to having all to all connections between all nodes, regardless of some weights trending towards 0. These two assumptions have been implicitly used for several years. Still the exponential increase in computational power is limited by this approach, as the complexity in the data increases, affecting the stability of the network.

1.1 Sparseness in Artificial Networks

State of the art approaches to neural computation, which has been gaining momentum, has discovered the benefits of sparsity and local connectivity patterns:

Convolutional Networks (ConvNets) approaches the problem by defining local sets of weights (or kernels) shared by all or part of the other neurons. These kernels are limited in size, limiting the effect of one neuron to the others, and are convolved over the network. This implies that neurons are locally connected to just a small fraction of other neurons, dramatically reducing the amount of connections to below 5% of all possible connections and, by extension, the computational power needed.

Reservoir Computing (RC) considers pools of units with complex temporal dynamics, that are randomly connected. The idea is that a sufficiently large and complex pool would contain potentially useful features. With it, one could use simple, shallow classifiers that read from the pool (or reservoir) and learn combinations of such features for a range of different tasks. One of the principal requirements for the convergence of these networks is having a spectral radius smaller than one [1], what is strongly influenced by the degree of sparsity of the network.

Deep architectures, combined with other techniques, are currently a trend that is used under the lemma of reusing features from previous layers, indirectly increasing the sparsity of the whole network. This occurs because the number of connections existing in the layered architecture could be reconverted into a shallow network which would have lost most of the initial connections.

Sparsity, then, is a feature present in State of the Art machine learning, typically imposed through architectural constraints. Still, the techniques are usually designed artificially, occasionally with inspiration from biology.

1.2 Sparseness in Biological Systems

The connections between and within areas of the brain have been widely studied in order to understand what makes it so unique. One of the most outstanding regions of the brain in this sense is the neocortex. The neocortex is the larger extension of neurons in the primate brain. Over its long extension, rich, functional heterogeneity at large scale levels conflicts with apparently strong structural homogeneity among cortical areas at a neuronal level. This introduces a dichotomy that is present also at the level of connectivity: long range, inter-areal connectivity matrices seem to be very dense [2] with over 90% of possible connections existing. This does not mean that the connections are evenly distributed in terms of weight. Additionally, at the neuronal level, neurons strongly exhibit lots of short range, local connections to their neighbors, and the longer the distance, the lower the probability of connecting 2 neurons, which results in

very sparse networks as they scale up. Kennedy et al. has proposed that cortical connections must follow a distance rule that determines their level of neighboring connectivity [3]. We therefore present a model of how such a distance rule can generate new network topologies driven by external activity patterns.

1.3 Synaptogenesis

Connections are functionally critical for neurons: how neurons connect to each other determines the way the neural network would operate. Synaptogenesis occurs not only in neurons grown during the embryonic and neonatal stages of life, but also in adult-grown neurons. In the case of adult neurogenesis, it is necessary for newly-grown neurons to not only create synapses with older neurons but also do so in a manner that would not disrupt the preexisting network. However, adult neurogenesis in mammals is relatively uncommon and occurs mainly in specific regions of the brain, like the olfactory bulb [4] and the dentate gyrus [5]. Regardless of when the neurons exhibiting synaptogenesis are grown, synapses' proliferation and survival are of scientific interest as they offer insight on how the brain processes stimuli.

Synaptogenesis has been shown to be dependent on activity of the neuron [6] and genetic traits. In addition, the distances between neurons could also play a significant role in synaptogenesis. Particularly in the cortex, most connections within the area are local, with approximately 80% of connections in the V1, V2 and V4 stemming from intra-areal sources [2], with 95% of these intrinsic connections arising from within 1.9 mm. Such evidence suggest that the brain's neural network is composed of clusters of densely connected neurons which are then connected to each other by sparse, long-range connections [3].

Moreover, tone directionality and frequency tuning are characteristics that identify receptive fields in primary auditory cortex (A1) [7]. As receptive fields are shaped by the connections between neurons, these phenomena should also be reflected by synaptogenesis. While it is also possible that neurogenesis could contribute to their formation, the low rate of occurrence of adult neurogenesis in the cortex suggests that synaptogenesis is a plausible mechanism for the early formation of receptive fields. Other plastic mechanisms at the level of neuron receptors might then have a more important role on their later fine tuning. This paper attempts to propose a model of synaptogenesis that can describe the structure and function of the cortex, in particular A1.

2 Methods

We propose a model of a cortical layer that uses Izhikevich neurons with spike time dependent plasticity (STDP) to update connections and a distance rule to model how connections are formed during development.

2.1 Spiking Cortical Dynamics

Cortical dynamics were modeled using python scripts. The network consisted of 800 excitatory and 200 inhibitory Izhikevich neurons in total, that together

represent part of the primary auditory cortex [8]. The neuron’s membrane potential v and membrane recovery variable u follow two differential equations:

$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + I \quad (1)$$

$$\frac{du}{dt} = a(bv - u) \quad (2)$$

With $a = .02/.10$ for E/I, $b = .2$, after-spike membrane reset $c = -65$ and after-spike recovery reset $d = 2.0$. Both populations had a rectangular shape with a ratio of 2:1 (*i.e.*, 40×20) with a randomly initialized connectivity matrix (see Table 1 for initialization values). The network was first trained using real-world auditory signals, consisting of 7 different songs converted to input signals with the multi-taper fast Fourier Transformation (FFT), that gives the music’s power spectrum over time. Songs covered different genres, including 70’s and 80’s pop, rock and metal, and lasted approximately 34 min. The FFT’s spanned 100 frequencies in the range of 150–650 Hz, and were normalized at each time point such that the maximum input value was always 1. This input was then mapped to the 800 excitatory neurons in a 1:10 ratio with small overlap using Gaussian smearing over both the x- and y-axis. Simulations ran at a temporal resolution of 1 ms, and the connectivity matrix was updated at each iteration based on Synaptic Time-Dependent Plasticity (STDP) principles. These principles constitute temporal asymmetric Hebbian learning, where synapse strength increases as a pre-synaptic action potential is followed by a post-synaptic spike and decreases vice versa. This was implemented on a population level through two variables $M(t)$ and $P(t)$ that either increase or decrease synaptic weights based on the order of spikes:

$$\tau_+ \frac{dP}{dt} = P + a_+ \quad \text{and} \quad \tau_- \frac{dM}{dt} = -M + a_- \quad (3)$$

This implementation was based on [9], with τ_+ and τ_- being the time constants of synaptic potentiation and depression respectively and a_+ and a_- their amplitude. Where weights W_{ij} are updated for each spiking neuron:

$$W_{ij} = W_{ij} + M_{ij} + P_{ij} \quad (4)$$

Alternatively an artificial network was trained using the same setup but using a different learning rule. A rate version of STDP was extracted from [10] and used as a learning rule for the rate approximation of the same system as:

$$W_{ij} = W_{ij} + \eta(W_M - W_{ij})(W_m - W_{ij})x_i x_j + W_e \epsilon (W_M - W_{ij})W_{ij} \quad (5)$$

Table 1. Model parameter values

Neuron	a	b	c	d
E	.02	.20	-65	2.0
I	.10	.20	-65	2.0

Where $W_e \epsilon$ corresponds to the integral of the STDP rule, W_M and W_m are the maximum and minimum thresholds for the weights, making the learning rule bi-stable and with a homeostatic decay. The rule is in function of the spiking rates of the presynaptic (x_i) and postsynaptic (x_j) neurons.

2.2 Synaptogenesis

In order to model synaptogenesis, a distance matrix has been computed in order to identify the position of each neuron of the space. In order to do this, the neurons have been distributed along two axis and given a normalized distance of 1 unit. From that, the probability of forming a connection from a neuron i to a neuron j given their distance d_{ij} is:

$$P(C_{ij}|d_{ij}) = \frac{1}{1 + kd_{ij}^2} \quad (6)$$

Where k is a scaling constant that defines the range connections will reach. The exponential has been chosen squared standing for the distribution gradient a typical fluid will suffer on a 2D medium (see Fig. 1 for reference). This rule was extended in two ways:

- The probability of a connection between neuron i and j being created is influenced by nearby existing connections to neurons k . With the probability distribution in Eq. 6, scaled by the weight between the origin neuron i and the neighbor neuron k , in order to promote clusters of specialized neurons and rich-club effects. The computation is then equivalent to a dot product like:

$$P(C_{new}|W, P) = WP \cdot P^T \quad (7)$$

Where C corresponds to the Boolean connectivity matrix, W is the weights matrix and P is the probability matrix obtained from Eq. 6 and shown in Fig. 1a.

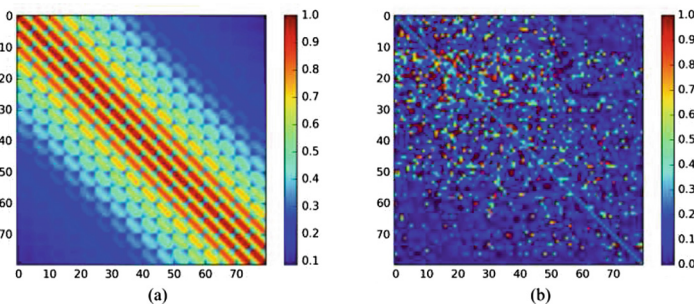


Fig. 1. (a) Sample distribution of probabilities extracted from Eq. 6. The network was initialized with size 20×40 neurons and considered a 2D layer of interconnected excitatory neurons. Parameters: k : 0.1, network size: 10×5 . (b) Connectivity matrix W of the model using synaptogenesis. One can observe the weights organized in clusters, spatially concentrated around the low frequencies (neurons 10–20).

- The connection probability was then scaled by the plasticity rule. It increases the weight as defined by the STDP rule above, promoting the formation of only relevant connections between the input and cortical layers and filtering out random ones.

3 Results

3.1 Synaptogenesis Creates Sparse Networks

In order to understand the sparsity of the receptive fields, we trained our STDP and rate networks for around 30 min of real music. We observed higher degrees of sparsity in the model using synaptogenesis as compared to the model without (Fig. 2).

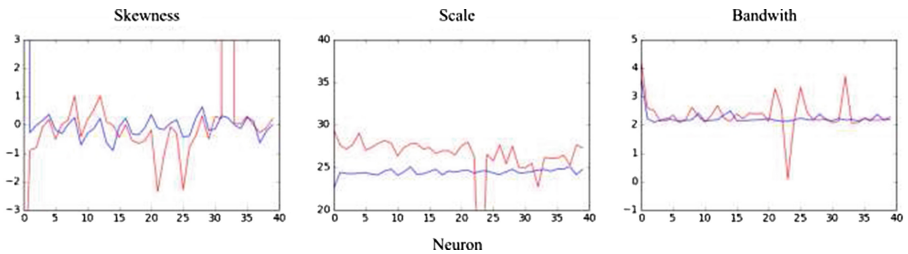


Fig. 2. Synaptogenesis (red) produces more variety (higher standard deviation) in skewness, bandwidth and scale than raw STDP (blue). Y axis show the fitting value of the data in a skewed Gaussian distribution. This data was generated using the rate based model. (Color figure online)

3.2 Synaptogenesis Converges to Richer Receptive Fields

We aimed to reproduce the data observed in [7], who found a high variability in the cortex, in terms of skewness, scale and bandwidth, and which corresponds to the three main parameters describing a skewed Gaussian distribution. We selected 10 evenly distributed frequencies and tested the network trained in the previous experiment for 10 trials. We then computed the rate of subpopulations of the network by summing the number of spikes in bins of 25 neurons, selected accordingly from the 2 dimensional pool. We extracted the spectral receptive fields for each population of neurons, as shown in Fig. 3. In order to extract the receptive fields we used the same methodology typically used in the study of the auditory cortex [7]. This showed slight differences between the receptive fields of the different neurons, where using synaptogenesis usually led to more dissimilar receptive fields among the neurons of the population. The minimal differences observed in Fig. 3 are attributed to the use of too strong inputs during testing, and the task of producing more realistic background noise and auditory input in order to show more relevant differences is left out of the scope of

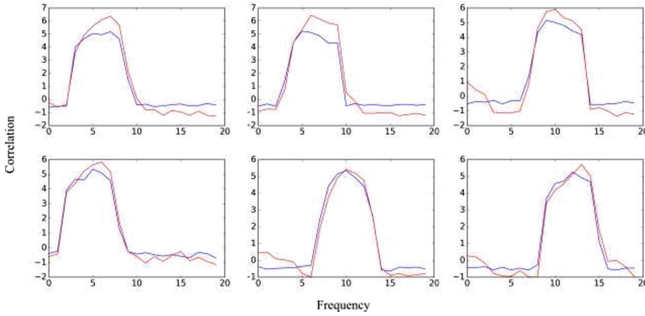


Fig. 3. Examples of the variety of receptive fields found. Top row shows three sample neurons with receptive fields that show greater variability in skewness and scale, whereas the bottom row shows neurons that have similar receptive fields for both approaches. The data was generated using the rate based model. Blue: STDP, red: synaptogenesis (Color figure online)

this paper. Finally, in order to test the relevance of these small differences, we fit a skewed Gaussian curve to the receptive fields. We observed increased variability (standard deviation) in the measures of skewness, bandwidth and scale of the model with synaptogenesis, relative to the model without synaptogenesis. We then conclude that additional work should show significant differences on the trends observed.

4 Conclusions

We have proposed a model of synapse generation, or synaptogenesis, based on a distance rule. This rule promotes the formation of a richer family of receptive fields, specializing neurons for variations in bandwidth, skewness and intensity. We have shown this variations comparing plasticity rules for rate and spiking neurons, with and without the synaptogenesis process.

We have shown that this process leads to sparser networks, a characteristic highly valued in state of the art artificial neural networks. Nonetheless, the capacity of this process to filter out redundant information and keep just relevant connections has yet to be shown. A big improvement to this model would involve the addition of apoptosis, or neural death, what would help prune connections that have become irrelevant. Still, the processes underlying apoptosis are mostly unknown and good measures to guide the pruning are still under debate.

We have proposed this experimental setup as a potential substrate of a single cortical layer. In this sense, the layer has a realistic ratio of excitatory and inhibitory neurons. Moreover, our model has been trained on auditory data, allowing the generation of a richer variety of features which is already observed in the auditory cortex of the ferret [7]. Next steps include completing the cortical model

with several layers and a better set of neuron types. Moreover, the input was mathematically modeled as observations in the A1 of the ferret (described in [11]), but in order to account for a real model of the cortex, the input should be filtered through attentional processes mainly driven by thalamo-cortical connections.

References

1. Schrauwen, B., Verstraeten, D., Van Campenhout, J.: An overview of reservoir computing: theory, applications and implementations. In: 15th European Symposium on Artificial Neural Networks, pp. 471–482 (2007)
2. Markov, N., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M., Giroud, P., Ercey-Ravasz, M., Pilaz, L., et al.: Weight consistency specifies regularities of macaque cortical networks. *Cereb. Cortex* **21**(6), 1254–1272 (2011)
3. Ercey-Ravasz, M., Markov, N.T., Lamy, C., Van Essen, D.C., Knoblauch, K., Toroczkai, Z., Kennedy, H.: A predictive network model of cerebral cortical connectivity based on a distance rule. *Neuron* **80**(1), 184–197 (2013)
4. Luskin, M.B.: Restricted proliferation and migration of postnatally generated neurons derived from the forebrain subventricular zone. *Neuron* **11**(1), 173–189 (1993)
5. Kuhn, H.G., Dickinson-Anson, H., Gage, F.H.: Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *J. Neurosci.* **16**(6), 2027–2033 (1996)
6. Kelsch, W., Sim, S., Lois, C.: Watching synaptogenesis in the adult brain. *Ann. Rev. Neurosci.* **33**, 131–149 (2010)
7. Wang, K., Shamma, S.A., Byrne, W.J.: Noise robustness in the auditory representation of speech signals. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1993, vol. 2, pp. 335–338. IEEE (1993)
8. Izhikevich, E.M., et al.: Simple model of spiking neurons. *IEEE Trans. Neural Netw.* **14**(6), 1569–1572 (2003)
9. Song, S., Miller, K.D., Abbott, L.F.: Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* **3**(9), 919–926 (2000)
10. Abbott, L., Gerstner, W.: Homeostasis and learning through spike-timing dependent plasticity. In: Summer School in Neurophysics, no. LCN-PRESENTATION-2007-001 (2003)
11. Wang, K., Shamma, S.: Self-normalization and noise-robustness in early auditory representations. *IEEE Trans. Speech Audio Process.* **2**(3), 421–435 (1994)