Supplementary Material

1 DETAILS OF THE ENCODER LAYER ALGORITHMIC IMPLEMENTATION

In this supplementary material we analyze the algorithmic formulation of the basic computational units of the Encoder Layer (EL) (i.e. Cortical Columns (CCs) and cell units).

1.1 Proximal dendritic connections

In regards to proximal connections, each neural unit in a CC has the same set of proximal connections to the Distributional Semantic (DS) constraints (i.e. word2vec). Such connections determine a multidimensional space of real numbers. Each CC acquires the statistical distribution on a DS sub-space to which such CC is connected. To that end our modelling approach uses a multidimensional Self Organizing Map (SOM) in each cortical column (Alg. 1).

Algorithm 1 `Plasticity in Proximal Synapses`. Self Organizing Map (SOM) algorithm.

1: given an `input` vector, find the nearest `unit` to such `input` vector in the input space
2: move such `unit` towards the `input` vector in the input space (the magnitude of such movement depends on the learning rate)
3: also move neighbor `units` to the nearest one towards the `input` vector (the magnitude of such movement depends on the learning rate and on a neighborhood measure over the topology of the network of units)

A SOM is an unsupervised clustering algorithm which distributes a continuous multidimensional distribution in a discrete multidimensional distribution of units (Kohonen, 1989, 1982). Hence the algorithm ends up with an array of units of m dimensions in which each unit represents a set of vectors from the continuous distribution in an input space of n dimensions. Generally, $m < n$ in order to reduce the dimensionality in the discrete representation. We added such restriction in our columnar algorithm (Dematties et al., 2019).

In our modelling apporach the SOM algorithm accounts for proximal lateral intra-column interaction, Long-Term Potentiation (LTP) and Long-Term Depression (LTD). Yet more importantly, this algorithm dissociates proximal dendritic inputs from distal dendrites, since it modifies proximal connections following the statistical distribution from DS constraints independently of the activity developed by distal–lateral and apical–dendrites which define the units that fire in such CC. This independence in the plasticity of the proximal dendritic inputs is supported by the property found in cortical tissue by means of which there is dendritic plasticity in the context of partial depolarization of the soma (Reiter and Stryker, 1988)–that is, without an Action Potential (AP).

Spruston (2008) suggested that the mechanisms regulating different dendritic trees differ and that the existence of dendritic domains in pyramidal neurons reflects distinct synaptic inputs, excitability, modulation and plasticity. Hawkins and Ahmad (2016a); George and Hawkins (2009) and Guerguiev et al. (2017) have already developed compelling modeling paradigms around this evidence.

1.2 Distal dendritic connections

Each CC in the EL receives incoming connections from other CCs by means of distal dendritic branches inside the receptive fields–from the same EL and from another Cortical Layer (CL). Each link between such CC and one of its neighboring CCs implies that each cell unit in the CC is linked with a different subset of cell units in the neighbour CC.

Such links account for dendritic branches in neural tissue and we call each connection in a dendrite *potential connection*. Potential connections account for synapses in a dendritic branch. Only a reduced percentage of CCs inside its receptive field are effectively linked to a CC. A cell unit inside its CC ends up with as many dendritic branches as effectively linked CCs inside its receptive field. And each dendritic branch holds a set of *potential connections*.

The term *potential connection* is used to describe a pair of neural units linked by its physical location and dendritic and axonal disposition in cortical tissue. However, an effective connectivity between such neurons will depend upon their sequential pattern of activation which will establish developed synapses between them. If two neural units are linked by means of a distal potential connection such connection will be strengthen only if there is a sequential activation of the cells linked, in two consecutive time steps. If such phenomenon does not repeat itself over time, such synapse will decrease its strength with respect to other synapses in the dendritic branch. A simultaneous activation in both neural units will decrease the strength in such potential connection.

We implemented distal dendritic synaptic plasticity mechanisms by means of Alg. 2. The learning mechanisms implemented on such algorithm simulate neurophysiological phenomena such as Spike-timing dependent plasticity (STDP), and homeostatic regulation plasticity in the synaptic strength regulation in distal dendritic branches. This algorithm presents a completely different synaptic mechanism to the one presented in Alg. 1.

Algorithm 2 `Plasticity in Distal Synapses`. This algorithm accounts for Spike-timing dependent plasticity (STDP) and homeostatic regulation phenomenon in distal dendritic synapses.

```
for every active unit in this cortical column do
    for every dendrite in this active unit do
        increment all the synapses–in this dendrite–potentially connected to units which were active
        in the last time step
    end for
end for
for every active unit in this cortical column do
    for every dendrite in this active unit do
        decrement all the synapses–in this dendrite–potentially connected to units which are active
        in this time step
    end for
end for
if updated step reaches certain value then
    for every unit in this cortical column do
        for every dendrite in this unit do
            if the sum of the synapses in this dendrite is greater than one then
                normalize all synapses in this dendrite
            end if
        end for
    end for
    updated step = 0
end if
updated step++
```

Hence, in this model we segregate the integration of feedforward and feedback information since neurons hold a more complicated morphological organization than the point neurons commonly utilized in prevalent Machine Learning (ML) (Spruston, 2008; Hawkins and Ahmad, 2016a; George and Hawkins, 2009; Guerguiev et al., 2017). Neural cells in cortical tissue have vastly a more complex structure, particularly when compared to artificial single-compartments-units. Different signals could be integrated at distinct dendritic areas. As a matter of fact, areas in cortical tissue receive feedback from other areas which arrives to the distal apical dendrites of pyramidal neurons (Manita et al., 2015; Budd, 1998; Spratling, 2002). Such synapses are far remote–electrotonically speaking–from the basal dendrites where feedforward information arrives (Larkum et al., 1999, 2007, 2009). Spratling (2002); Gris et al. (2017) and Spratling and Johnson (2006) noted that the anatomy of pyramidal neurons may actually provide the segregation of feedforward and feedback information.

Basically, Alg. 2 updates all distal axonal-dendritic synapses, incrementing them always that pre and post-synaptic neural units become active in consecutive time steps, and decrementing them when they become active at the same time step. Occasionally, all the synaptic weights belonging to the same dendrite are normalized every certain number of time steps. This normalization is repeated for all neural units and for all dendritic branches in each neural unit whose sum of synaptic weights is beyond unity.

1.3 Activation Rules in a Cortical Column

First a group of cell units in a CC is partially depolarized by distal connections among such neural units and cell units activated in the previous time step in the EL or in a foreign cortical patch. That is, neural units activated in time step $t = 0$ in the EL or in foreign patches will partially depolarize a set of neural units at time step $t = 1$ in such CC, by means of distal–lateral and apical– dendritic branch synapses established by their learning algorithm (Alg. 2).

Second, afferent proximal connections from DSs will tend to depolarize certain clusters of units in such CC in time step $t = 1$. This *tentative* depolarization is produced by the inputs from the DSs (word2vec inputs) with proximal synapses established by learning in the Alg. 1. A group of neural units is randomly chosen from a discrete distribution whose probabilities are established by the state of excitation in afferent inputs. Namely, the more excited is a cell unit by afferent synapses, the more chance has such unit of being chosen.

If a sufficiently large number of partially depolarized units are in the set of afferently excited units, such partially depolarized units will fire previously in the group. Those units–which fire before–prevent neighboring units in the excited clusters from firing, hyperpolarizing them by means of lateral inhibitory connections in the column.

With the aim of keeping our model simple, we do not implement inhibitory neurons as explicit individual cells, and the function of inhibitory neurons are encoded in Alg. 3. As explained above, cells that fire before prevent other cells in the same CC from firing. This computational hypothesis requires neurophysiological evidence about fast inhibitory mechanisms among neighboring cells sustained by fast inhibitory neurons processing stimulus-related information. Reyes-Puerta et al. (2015) actually showed such properties in the brain and Hawkins and Ahmad (2016a) used them in their bio-inspired model.

Partial depolarization states put cell units in a predictive state generated by the activations produced in the EL–and/or in foreign cortical patches–in previous time steps. That is, lateral and apical activation in previous time steps constitute a context in which current afferent inputs are received.

From the group of units that tend to be depolarized by current afferent inputs from the DSs, only a reduced sub-set of those units are likely to fire in the previous contextual firing history in the EL.

In case there is no context, that is, if not enough units depolarized by afferent inputs are partially depolarized by previous–lateral and apical–activations, almost all units in the afferent excited clusters will be in the same condition. In other words, there are not enough units favoured by its partial depolarization inside the clusters excited by afferent information. Since most of the units are in the same condition, all become active, covering more hypotheses for next inputs. We call this phenomenon as Massive Firing Event (MFE) in our model.

Such activation mechanism is depicted in Alg. 3. In Alg. 3 (Part 1) a ranking is established among neural units–inside a CC–in terms of its afferent excitability, given the afferent inputs (lines 1 and 2). The *number of afferently excited units* refers to the maximum number of units that can be activated by the afferent input in a CC and *minimum number of active units* refers to the number of units that will be active in a CC if a Sparse Distributed Representation (SDR) is achieved as a result of optimal prediction (lines 3 and 4 respectively).

Algorithm 3 `Units activation (Part 1)`. This algorithm establishes the activation rules in a Complex Self-Organizing Map (CSOM) object.

```
1: distances = given an input vector find the euclidean distance each unit has to such input in
   the input space from proximal afferent synapses
2: ranking = sort indexes from the smallest to the largest distances
3: number of afferently excited units = proximal activation percentage*number of units
4: minimum number of active units = (1-sparsity)*number of units
5: if randomness is disabled then
6:    excited units = gets the first number of afferently excited units elements from ranking
7: else
8:    excited units = gets number of afferently excited units random indexes from distances with
      probabilities determined by the relative reciprocal of the distances element values
9: end if
10: for unit = 0 to unit = number of units do
11:    auxiliary = 0
12:    for dendrite = 0 to dendrite = number of distal dendrites do
13:       dendrite accumulator = 0
14:       for active unit = 0 to active unit = number of linked active units do
15:          potential index = find the first coincident index in potential
             connections[dendrite][unit] with linking units[dendrite][active
             unit]
16:          if there exist coincidence then
17:             dendrite accumulator += dynamic synapses[dendrite][unit][potential
                index]
18:          end if
19:       end for
20:       if dendrite accumulator > 100*DISTAL_SYNAPTIC_THRESHOLD then
21:          auxiliary++
22:       end if
23:    end for
24:    total responses[unit] += auxiliary
25: end for
26: updated distances = element wise quotient between distances and total responses
27: updated ranking = sort indexes from the smallest to the largest updated distances
```

If randomness is enabled, *number of afferently excited units* units are chosen at random by means of a discrete distribution whose probabilities are the afferent excitation of each unit. If randomness is disabled, *number of afferently excited units* first units are chosen from the ranking of afferently excited units (lines 5 to 9).

From line 10 to 25 each neural unit accumulates distal–lateral and apical–excitation in order to determine its partial depolarization from units which were active in the previous time step. For each neural unit in a CC, for each distal dendrite in such unit and for each active unit in such distal dendrite the algorithm looks for coincidences between some potential connection in such distal dendrite in the neural unit and the active unit in such distal dendrite. That is, in line 15, the algorithm asks if there is coincidence between some potential connection in this distal dendrite inside the unit and the neural unit activated in the previous time step in the CC linked by such distal dendrite. If there is coincidence, the value of the synaptic weight in such potential connection is accumulated in a dendrite accumulator. After all active units are examined for this dendrite, if the dendrite accumulator is greater than certain threshold, such dendrite is considered active and the total response of the unit is incremented in one.

Segregating neural computation in different compartments in our model transcends the coarse partitioning of broad dendritic domains to determine the computation in individual dendrites. In this regard, layer II/III pyramidal neurons can generate dendritic spikes which is a fundamental property of dendritic excitability and control by inhibition (Larkum et al., 2007). Jadi et al. (2014) showed that dendrites generating local spikes have their own nonlinear input-output curves and also have individual activation thresholds, different from those of the neuron as a whole. There are studies showing that small compartments in the dendritic arms of cortical neurons can each perform complicated operations (Payeur et al., 2019). Furthermore, in a recent study Gidon et al. (2020) showed that individual dendritic compartments could make the computation of the *exclusive OR* which was previously considered unattainable by single-neuron systems.

Each neural unit ends up with an excitation value due to its distal dendrites. The unit distances vector is element- wise divided by distal dendritic excitations vector to get the updated distances and an updated ranking of the units (lines 26 and 27). In this way, units with more distal excitation will decrease its distance more and will be put in a more favorable place in the ranking in order to be activated.

In Alg. 3 (Part 2) the minimum updated distance is found in the group of afferently excited units. Then, a set of units–inside the group of afferently excited units–is identified which have such minimum updated distance. While the number of identified units is less than *minimum number of active units*, the next minimum updated distance is found in the group of afferently excited units and a new set of units–inside the group of afferently excited units–is identified which have such next minimum updated distance. This new set is added to the previous one until the number of units in this accumulative set is greater than or equal to the minimum number of active units.

The functional result of Alg. 3 is that there must be a sufficient amount of–partially and previously depolarized–neural units inside the afferently activated cluster of units in order to get a SDR pattern of activation. Otherwise, the CC will end up with a massive activation pattern, a Massive Firing Event (MFE) in which more than a *minimum number of active units* will be active. In the case of the occurrence of a MFE, the synaptic plasticity is modulated in order to form stronger synapses of those neural units activated during such event.

Each neural unit in a CC establishes its state of partial depolarization based on the contribution from distal dendritic branches from lateral and apical connections. A dendritic branch will contribute to the partial depolarization of the soma in such cell only if such dendritic branch exceeds an activation threshold by means of the contribution from its individual synapses in the context of the patterns of activation in the previous time step.

This mechanism has compelling sequential properties Hawkins and Ahmad (2016b), which have already been applied in the classification of artificially generated sequential data Cui et al. (2016). We apply

Algorithm 3 `Units activation (Part 2)`. This algorithm establishes the activation rules in a CSOM object.

new `distances` = get the updated `distances` elements whose indexes are in excited `units`
new minimum `distance` = get the minimum element from new `distances`
minimum `indexes` = get indexes from updated `distances` vector whose values are equal to new minimum `distance`
apt to be `active` = get the coincident indexes between excited `units` and minimum `indexes`
erase from new `distances` vector, all the elements whose value is equal to new minimum `distance`
while number of elements in apt to be `active` vector < *minimum number of active units* and new `distances` has at least one element **do**
    new minimum `distance` = get the minimum element from new `distances`
    minimum `indexes` = get indexes from updated `distances` vector whose values are equal to new minimum `distances`
    partial apt to be `active` = get the coincident indexes between excited `units` and minimum `indexes`
    incorporate partial apt to be `active` elements into apt to be `active` vector
    erase from new `distances` vector, all the elements whose value is equal to new minimum `distance`
end while
if ENABLE_RANDOM_BEHAVIOUR **then**
    shuffle apt to be `active` vector
end if
for `number` = 0 **to** `number` = number of apt to be `active` elements **do**
    incorporate to `output` the excited `units[`apt to be `active[number]]`
end for
return `output`

such mechanism in our algorithm by adding the contribution of synapses–in a dendritic branch–whose connections are linked with cells that were active in the previous time step in the EL.

REFERENCES

Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265, 1037–1044. doi:10.1098/rspb.1998.0396

Cui, Y., Ahmad, S., and Hawkins, J. (2016). Continuous online sequence learning with an unsupervised neural network model. *Neural Computation* 28, 2474–2504. doi:10.1162/NECO_a_00893. PMID: 27626963

Dematties, D., Rizzi, S., Thiruvathukal, G. K., Wainselboim, A., and Zanutto, B. S. (2019). Phonetic acquisition in cortical dynamics, a computational approach. *PLOS ONE* 14, 1–28. doi:10.1371/journal.pone.0217966

George, D. and Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLOS Computational Biology* 5, 1–26. doi:10.1371/journal.pcbi.1000532

Gidon, A., Zolnik, T. A., Fidzinski, P., Bolduan, F., Papoutsi, A., Poirazi, P., et al. (2020). Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science* 367, 83–87. doi:10.1126/science.aax6239

Gris, K. V., Coutu, J.-P., and Gris, D. (2017). Supervised and unsupervised learning technology in the study of rodent behavior. *Frontiers in Behavioral Neuroscience* 11, 141. doi:10.3389/fnbeh.2017.00141

Guerguiev, J., Lillicrap, T. P., and Richards, B. A. (2017). Towards deep learning with segregated dendrites. *eLife* 6, e22901. doi:10.7554/eLife.22901

Hawkins, J. and Ahmad, S. (2016a). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits* 10, 23. doi:10.3389/fncir.2016.00023

Hawkins, J. and Ahmad, S. (2016b). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits.* 10

Jadi, M. P., Behabadi, B. F., Poleg-Polsky, A., Schiller, J., and Mel, B. W. (2014). An augmented two-layer model captures nonlinear analog spatial integration effects in pyramidal neuron dendrites. *Proceedings of the IEEE* 102, 782–798. doi:10.1109/JPROC.2014.2312671

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics.* 43, 59–69

Kohonen, T. (1989). *Self-organization and Associative Memory: 3rd Edition* (New York, NY, USA: Springer-Verlag New York, Inc.)

Larkum, M. E., Nevian, T., Sandler, M., Polsky, A., and Schiller, J. (2009). Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: A new unifying principle. *Science* 325, 756–760. doi:10.1126/science.1171958

Larkum, M. E., Waters, J., Sakmann, B., and Helmchen, F. (2007). Dendritic spikes in apical dendrites of neocortical layer 2/3 pyramidal neurons. *Journal of Neuroscience* 27, 8999–9008. doi:10.1523/JNEUROSCI.1717-07.2007

Larkum, M. E., Zhu, J., and Sakmann, B. J. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* 398, 338–341

Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., et al. (2015). A top-down cortical circuit for accurate sensory perception. *Neuron* 86, 1304—1316. doi:10.1016/j.neuron.2015.05.006

Payeur, A., Béïque, J.-C., and Naud, R. (2019). Classes of dendritic information processing. *Current Opinion in Neurobiology* 58, 78 – 85. doi:https://doi.org/10.1016/j.conb.2019.07.006. Computational Neuroscience

Reiter, H. and Stryker, M. (1988). Neural plasticity without postsynaptic action potentials: Less-active inputs become dominant when kitten visual cortical cells are pharmacologically inhibited 85, 3623–7

Reyes-Puerta, V., Kim, S., Sun, J.-J., Imbrosci, B., Kilb, W., and Luhmann, H. J. (2015). High stimulus-related information in barrel cortex inhibitory interneurons. *PLOS Computational Biology* 11, 1–32. doi:10.1371/journal.pcbi.1004121

Spratling, M. W. (2002). Cortical region interactions and the functional role of apical dendrites. *Behavioral and Cognitive Neuroscience Reviews* 1, 219–228. doi:10.1177/1534582302001003003. PMID: 17715594

Spratling, M. W. and Johnson, M. H. (2006). A feedback model of perceptual learning and categorization. *Visual Cognition* 13, 129–165. doi:10.1080/13506280500168562

Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience* 9, 206–221