Going Beyond the Point Neuron:
Active Dendrites and Sparse Representations for Continual Learning

Karan Grewal, Jeremy Forest, Subutai Ahmad

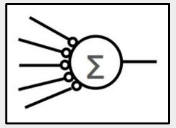
Numenta

Correspondence to
`kgrewal@numenta.com`

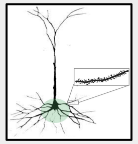
Summary

- The point neuron –used by most artificial neural networks (ANNs)– abstracts away many details of real biological neurons; notably structural complexity.

Point Neuron



Pyramidal Neuron



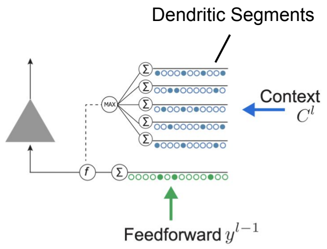
- ANNs suffer from **catastrophic forgetting** when learning tasks in sequence (they are unable to learn new information without erasing what they previously learned) [1,2].
- We hypothesize that the physiology of pyramidal neurons and active dendrites are integral to how the brain learns continually with little forgetting.
- Real neurons have **active dendrites** that act as independent pattern recognizers and can depolarize the cell, making it more likely to generate an action potential [3].
- Coupled with local inhibition, active dendrites lead to the selection of subsets of neurons that train on each task, preventing overwriting of what the network previously learned.
- We demonstrate a functionally unique role for dendrites that differ from the classical view of dendrites as simply adding more layers to an ANN.

Neuron Model

Similar to our previous model [4], we propose **Active Dendrites Networks** by augmenting ANNs with (1) **active dendrites** and (2) **sparse representations.**

1. Active Dendrites

- Each neuron has a set of dendritic segments (i.e., additional weights) which all receive a **context signal** and modulate the activity of the neuron.
- Each segment computes an activation value via a weighted sum of the context signal. A single 'winning' activation is chosen to modulate the neuron.
- The winning activation d acts as a non-linear modifier, or gate, by multiplying the neuron's output by $\sigma(d)$.



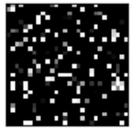
2. Sparse Representations

- We apply a k-**Winner Take All** non-linearity to all hidden layers, where only the top k-ranked activations keep their value.
- k-**winners** incorporates local inhibition and sparsifies the activations of a layer. This reduces the overlap between representations.
- Because dendrite activity modulates the probability of a neuron winning, the context signal impacts exactly which neurons become active leading to context-specific representations

Results On Continual Learning

- In Continual Learning a network is trained on a sequence of tasks (example: tasks 1, task 2 …, task 100).
- We measure accuracy on all tasks at the end of training. ANNs typically perform poorly, and forget older tasks.
- *permutedMNIST* is a benchmark dataset for continual learning; each task requires classifying images of handwritten digits (0-9). In each task, all images have been modified by the same pixel-wise permutation.

"9" from dataset

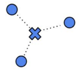


"9" from task A

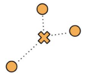
"9" from task B

Computing The Context Signal For *permutedMNIST*

- The ideal context signal would serve as a task identifier, i.e. a context signal unique to each task.
- We use a **task prototype** as the context signal, computed as the mean training example per task.
- At inference time, the closest prototype to each example is chosen as the context signal.



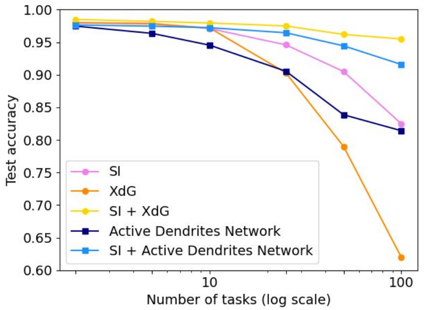
Training samples (o) & prototype (x) for task 1



Training samples (o) & prototype (x) for task 2

Experimental Results

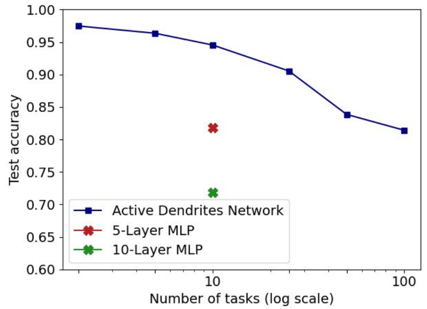
- Active Dendrites Networks achieve 81.4% accuracy when learning 100 tasks in sequence.
- This is far better than standard ANNs, significantly better than **XdG** [1], and competitive with **Synaptic Intelligence** (SI) [2].
- XdG is a model where task-specific sparse subsets of neurons are explicitly stored. Unlike our network, task information is not inferred. In XdG it must be stored separately and provided at test time. This makes the experiment significantly easier since the model does not have to learn to map context to neurons.
- SI is a biologically motivated model of slow and fast synaptic weights. It is a complementary technique.



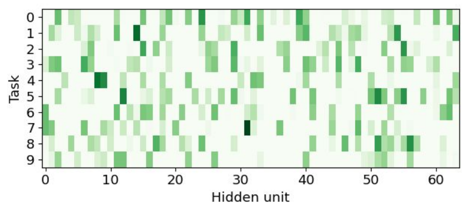
- When augmented with SI, our networks achieve 91.6% accuracy for 100 tasks in sequence. This is slightly lower than XdG + SI. However note that our networks are solving a much harder problem since they must infer task information (unlike XdG).

Dendritic Neuron: Not Just A Multi-Layer Perceptron

- Studies have suggested that a single pyramidal neuron with dendrites can be modeled by a multi-layer ANN [5].
- Here we show a functional advantage of dendrites over standard ANNs (regardless of their depth).
- In continual learning scenarios, a 3-layer network with modulatory active dendrites significantly outperforms 5- and 10-layer standard ANNs without dendrites.

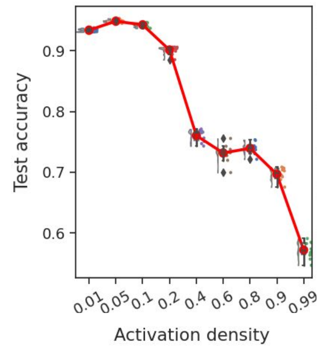


Why Do Dendrites Work Well For Continual Learning?



- Unlike in standard ANNs, active dendrites encourage sparse subsets of neurons to become active.
- *(Left Figure)* Active dendrites select subnetworks of neurons to learn each task, minimizing interference between tasks. We only show 64 neurons in a single layer across tasks.

- Sparse activations of neurons are important for accuracy in continual learning in our network.
- A related method to mitigate catastrophic forgetting based on dendritic function is **Dendritic Gated Networks** (DGNs) [6], however they don't use sparse representations. (We don't compare our results with DGNs since they don't report accuracy on permutedMNIST for >10 tasks.)
- *(Right Figure)* Accuracy as a function of sparsity. Without sparse representations, accuracy on 10 permutedMNIST tasks decreases as representations become more dense.



References

[1] Masse NY, et al.. *PNAS*. 2018, 115(44).
[2] Zenke F, et al.. *ICML*. 2017.
[3] Major et al.. *Ann. Rev. of Neuroscience*. 2013, 36.
[4] Hawkins J, & Ahmad S. *Front. Neural Circ.* 2016, 10(23).
[5] Poirazi P, et al.. *Neuron*. 2003, 37(6).
[6] Sezener E, et al.. *BioRxiv*. 2021.