Neocortical sparsity, active dendrites, and their relevance in robust sequence learning models

Final Draft

Chase King
chasek22@uw.edu

CSE 599B: "AI and The Brain" — Fall 2020

1 Introduction

Understanding the brain is arguably the greatest scientific problem of our time. The neocortex is the part of the mammilian brain responsible for higher-order brain functions such as learning, sensory perception, sophisticated motor control, decision-making, planning, and language (*Neuroscience. 2nd edition.*, 2001). Experimental neuroscience evidence suggests sparse neuronal activity is used to represent information in the neocortex (Barth & Poulet, 2012). In this paper we will review recent theories and models for sparse representations in the neocortex. We will focus primarily on work published by Numenta[1] that is heavily influenced by recent experimental findings in NMDA spiking and dendritic processing in pyramidal cortical neurons (Ahmad & Hawkins, 2016; Augusto & Gambino, 2019). We will see how "active dendrites", referring to the dynamic abilities of neurons to grow synapses, is a key ingredient in pattern recognition. We will also cover some of the mathematical theory behind sparse representations, and see how sparsity necessarily yields high accuracy even in the presence of noise, which is a key feature of biological intelligence. We will then look at how this theory of sparsity can be applied to neural models and briefly discuss how these models can be used to learning a continuous stream of input data. Finally, we will conclude with discussions on the results of reviewed papers, potential downsides of this model, and possible directions for future work.

2 Background and Related Work

2.1 The Neocortex, Neurons, and Networks

The neocortex is a core component of biological intelligence as we know it. The neocortex is unique to mammals, and the human neocortex is roughly the size and thickness of a dinner napkin and is folded and tightly packed in the outer layers of our brains to increase surface area. The neocortex is not to be confused with the cerebral cortex, which is the part of the brain that encompasses the neocortex. The cerebral cortex consists of roughy 10% of the evolutionarily-older allocortex, and 90% of the more recently-evolved neocortex (*Neuroscience. 2nd edition.*, 2001). In humans the cerebral cortex consists of roughly 16 billion neurons (compared to the roughly 86 billion in the entire human brain) (Herculano-Houzel, 2009). Exact estimates of number of neurons vary, but the neocortex is widely regarded as the center for higher-order cognition and intelligence.

---

[1]Numenta is a neuroscience research institute in Redwood City, CA, specializing in theoretical neuroscience and interested in applying neuroscience principles to machine intelligence.

A neuron is the basic functional unit of the brain. A typical neuron consists of three distinct parts: dendrites, a cell body (soma), and an axon. Dendrites branch out extensively (usually to neurons close by, but occasionally to different parts of the brain or nervous system) and receive inputs from other cells (*Neuroscience. 2nd edition.*, 2001). A synapse is formed when a dendrite (or occasionally the cell body) of one neuron connects to the axon of another. A typical cortical neuron can form tens of thousands of synapses with other neurons, resulting in a highly interconnected structure (Gerstner, Kistler, Naud, & Paninski, 2014). The cell body functions as the information processing unit of the cell. Importantly, signals received at the dendrites travel to the cell body where they can affect the *membrane potential* of a neuron. This refers to the difference of electrical charge inside and outside of the neuron. If the membrane potential exceeds some threshold, an output signal (referred to as an action potential/spike) is generated (Gerstner et al., 2014). The action potential travels down the axon and, at synapses, chemicals called neurotransmitters are released across the synapse and are received by protein channels in the receiving neuron. Neurons can strengthen a particular synapse by adding more protein channels at the site of the synapse.

It is worth mentioning that this is a highly superficial and far from comprehensive summary of neuronal dynamics. There are numerous other biochemical processes that are at play inside neurons but the above information is sufficient for the purposes of this paper.

Let us return to the notion that dendrites can effect the membrane potential when input spikes are received. When neurotransmitters are passed along a synapse and taken in at a receiving neuron's dendrites, they can change the local voltage at that location (Gerstner et al., 2014). Inhibitory synapses can decrease the potential, whereas excitatory synapses can increase the potential. In their theories, Numenta uses the standard neuroscientific division of the dendritic tree into three parts: the proximal zone, the basal zone, and the apical zone which serve the following purposes (Hawkins & Ahmad, 2016). (See Figure 1 below for a diagram of a typical pyramidal neuron.)
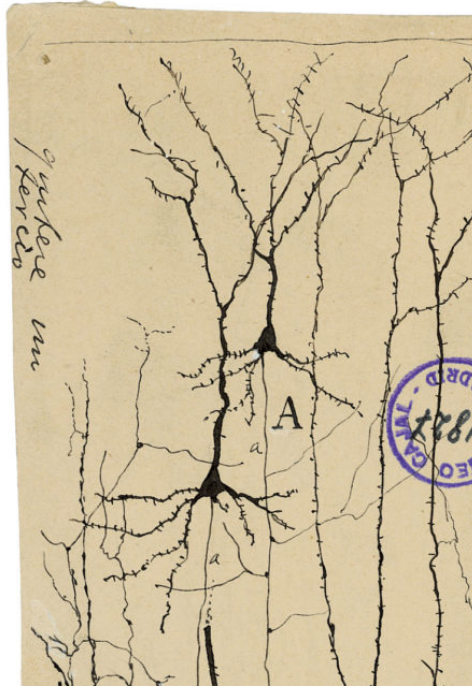
Figure 1: Drawing of pyramidal neurons by Santiago Ramón y Cajal in the late 19th century. Cajal was a pioneer of the field of neuroscience, and his insight that neurons are the discrete unit of computation and that communication between neurons occurs through synapses became known as the neuron doctrine. The cell body is the black blob, the axon is the thin small line extending from the bottom of the cell body (labeled "a"), the proximal dendrites are the branches close to the soma, the basal dendrites are slightly further away from the proximal dendrites, and the apical dendrites extend from the top of the cell body and branch out extensively.

Proximal dendrites have the largest effect at the soma and synaptic activity here is more prone to causing a action potential (Hawkins & Ahmad, 2016). Hence, we can think of proximal synapses as defining the feedforward receptive field of the cell. More specifically, since there are many dendritic segments close to the cell body, the feedforward receptive field can be thought of as a union of feedforward patterns (Hawkins & Ahmad, 2016).

Basal dendrites are farther away from the cell body on the dendritic tree when compared to proximal dendrites. Basal dendrites primarily receive feedforward input from nearby neurons (Park et al., 2019). If a local basal dendritic segment receives enough excitatory synaptic activity, it can cause a dendritic spike to travel down to the cell body (Augusto & Gambino, 2019). (Experimental evidence indicates the coincident activation of 8-20 local synapses is sufficient for generating a dendritic spike (Major, Larkum, & Schiller, 2013).) Also referred to as NMDA spikes, a single one of these pulses is often not enough to lift a cell above its firing threshold, but is still enough to have a longer-lasting depolarization of the cell body (Augusto & Gambino, 2019). Numenta proposes that this depolarization represents a state of prediction that the cell will become active (Hawkins & Ahmad, 2016). What does this have to do with prediction? A slightly depolarized cell will fire slightly earlier than it otherwise would if it receives sufficient feedforward input. By firing earlier, nearby neurons are inhibited which yields sparse patterns of activity when

3

predictions are correct (Hawkins & Ahmad, 2016).[2]

Apical dendrites are the most distal (i.e., farthest away from the cell body), and receive both synaptic input from nearby neurons and more "global" connections with neurons farther away (Park et al., 2019). Typically a single apical dendrite extends from the cell body for several hundred microns before branching out and forming a dendritic tree (this we refer to as the "apical dendrites"). Apical dendrites can also generate NMDA spikes that lead to a temporarily depolarized soma (Augusto & Gambino, 2019). Hence, apical dendrites also recognize patterns and make predictions similar to basal dendrites, but since apical dendrites are connected to neurons farther away, Numenta refers to this prediction as "top-down expectation" (Hawkins & Ahmad, 2016).

2.2   Neuroscientific applications to machine intelligence research

Understanding how the neocortex functions could provide us with useful insights and primitives we can apply to machine intelligence problems. The fields of artificial intelligence and machine learning have been heavily influenced by neuroscience. For example, artificial neural networks used in deep learning models were inspired by the interconnectivity of networks of neurons in the brain (Gerstner et al., 2014). A more recent example is the concept of attention which is also derived from neuroscientific findings (Vaswani et al., 2017).

However, most artificial neural networks use highly simplified artificial neurons when compared with their biological counterparts. For example, artificial neurons in deep neural networks sum together their inputs and pass them through a nonlinearity. However, as summarized above, biological neurons have much more complicated temporal dynamics. Moreover, dendrites don't simply integrate their inputs, as indicated by the complicated spatiotemporal properties of dendritic segments. Another example is fully-connected layers in artificial neural networks. Most cortical pyramidal neurons form on the order of ten thousand synapses with other cells and can dynamically grow or remove synapses with Hebbian learning rules (*Neuroscience. 2nd edition.*, 2001). However, fully-connected layers assume that neurons are connected to *every* neuron in the previous layer, which is not biologically plausible. While these simplifying assumptions work for some specific applications, they do not lend to the incredible generalization properties of the brain. To understand why neurons need active dendrites, Numenta proposes a theory that relies heavily on another neuroscientific principle: sparsity (Hawkins & Ahmad, 2016).

2.3   Sparsity

In the neocortex, sparsity is ubiquitous, found in nearly every subregion of cortex and across all sensory modalities (Barth & Poulet, 2012). (One speculative evolutionary explanation for sparsity relates to power consumption. Neuronal firing is an "expensive" action as we shall see. Evidence suggests that in humans, the brain consumes 20% of the body's energy, so assuming a 2000 calorie daily diet, this equates to roughly 20 watts power consumption (Herculano-Houzel, 2009).) Hence, sparse representations of information appear to be deeply ingrained in biological intelligence, and we cannot truly understand brain function without understanding sparsity.

---

[2]This inter-column inhibition is explained more in the brief article "Quiet Down Now: How Excitatory Neurons Inhibit One Another" (Robinson, 2010).

Sparsity in the brain refers to only a small percentage of neurons being active at any given time, and cells are only physically connected to a relatively small number of cells (*Neuroscience. 2nd edition.*, 2001). On the other hand, sparsity of weights in computational models (e.g., as imposed by L1 regularization) simply zero out weights for particular features. However, the brain does not zero out feature weights. Instead, it finds representations for features using only a small subset of neurons; which neurons are active depends on the input, and can change over time (e.g. through learning).

2.4 Sequence learning

What properties of the brain are so fundamental and necessary for prediction, motor planning, language, and other high-order cognition? Numenta argues that "the most fundamental operation of all neocortical tissue is learning and recalling sequences of patterns" (Hawkins & Ahmad, 2016). Taking inspiration from how brains process information, they note several properties of sequence memory:

1. **Online learning:** Learning must be continuous (compared to the traditional train-inference phase of most models today).

2. **Higher-order[3] predictions:** Correctly predicting complex sequences requires incorporating past contextual information. For example, when you are reading this paper, you likely remember some key information from the first few sections even though you've read other sections in-between.

3. **Multiple simultaneous predictions:** While this may seem counterintuitive, brains continually (and oftentimes subconsciously) make multiple simultaneous predictions. For example, if we hear leaves rustle while taking a hike in the woods, we simultaneously predict that it could be wind, another hiker, or a mountain lion among other possibilities! Therefore our model of sequence memory needs to also have this ability of making multiple simultaneous predictions.

4. **Local learning rules:** Instead of having a global objective function (like most models today use), the learning rules for our sequence memory must be local[4] to each individual neuron.

5. **Robustness:** Brains are incredibly robust to noise, variance in inputs, and neuronal failure. For example, if you saw a pixelated or blurry image of a dog you've never seen before, you would instantly be able to recognize a dog. Hence our sequence memory should exhibit similar robustness properties.

We will return to these properties when we discuss models of neurons in section 5.1.

3 Mathematical Foundations of Sparse Representations

In Ahmad and Hawkins (2016), Numenta examines some interesting properties of Sparse Distributed Sepresentations (SDRs), which are the primary representation method used in their Hierarchical Temporal Memory (HTM) systems. As alluded to before, representations are sparse because only a small percentage of neurons are active at any given time,

---

[3]The term "higher-order" refers to higher-order Markov chains which have the property of holding information for many time steps.

[4]Local in both space and time.

and distributed because information is encoded across a set of active neurons and not just a single neuron. Here we review some of the most important properties of SDRs given in Ahmad and Hawkins (2016) before seeing how they can be applied to machine intelligence systems.

SDRs. We make the simplifying assumption that if we take a small enough time bin, e.g. milliseconds, we can observe sets of neurons that are active in each bin. Thus, we can represent the instantaneous activity of a population of n neurons by a binary vector $x \in \{0, 1\}^n$ where a $x_i = 1$ if an only if neuron i is active. These binary vectors are highly sparse (i.e., a small number of components are 1). We use $|x|$ to denote the number of "on" (1) bits in x (i.e., $|x| = \sum_{i=1}^{n} x_i = \|x\|_1$). When we refer to many SDRs with the same number of on bits we will use the variable w to refer to the number of on bits.

Overlap and matching. The overlap between two SDR vectors is the number of bits that are in the on (1) state in the same location. Since these are binary vectors as explained above, we can compute the overlap between two SDRs simply by taking the dot product $(x \cdot y)$. For an example[5], consider the following where $n = 16$ and $w = 4$:

$$x = \texttt{0010001000110000}$$
$$y = \texttt{0001001000100001}$$
$$\text{overlap}(x, y) = x \cdot y = 2$$

since there are two places where x and y both have a 1. We say two SDR vectors have a match if their overlap exceeds some threshold θ. That is, there is a match between x and y if $x \cdot y \geq \theta$.

SDR representation space. Given a fixed n and w, the number of unique SDR encodings is n choose w:

$$\binom{n}{w} = \frac{n!}{w!\,(n-w)!} \tag{1}$$

since we can choose w places in the n-dimensional vector to place a one. Note that this is much smaller than the number of possible recordings when we don't enforce sparsity, which is 2^n since at every position we can place either a zero or a one. However, with values of $n = 2048$ and $w = 40$ the number of possible SDR encodings is 2.37×10^{84}, still an astronomically large number. The inverse of (1) is the probability that two SDRs chosen uniformly at random are equal, which is essentially zero for any reasonable choice of n and w.

Inexact SDR matching. We now observe how many SDRs of size n have exactly b bits of overlap with some SDR x with $|x|$ on-bits. Assuming $b \leq |x|$ and $b \leq w$, this number can be expressed as

$$\text{overlapsize}_x(n, w, b) = \binom{|x|}{b} \times \binom{n - |x|}{w - b} \tag{2}$$

where the first term indicates the number of ways to choose b overlapping bits and the second term indicates the number of ways to choose the remaining $w - b$ on-bits in the places of x that have a zero. Keeping this fact in mind, we turn to how we can make our matching system more tolerant to noise. Clearly, exact matching (i.e., $\theta = w$) is infeasable since only one slight perturbance will cause our system to not recognize the match. Moreover, every sensory experience your brain has ever was novel since sensory inputs never repeat exactly; exact matching is clearly not how the brain solves this problem. If, for ex-

---

[5]This is a highly trivial example and in reality n is much larger.

ample, we set $\theta = w/2$, random noise can change 50% of the on-bits and still match two SDRs. However, there is a tradeoff between robustness to noise and generating more false positives (i.e., a match when there shouldn't be one). Formally, given a SDR encoding x and another random SDR y, the probability of a false match (i.e., $x \cdot y \geq \theta$ can be expressed

$$\text{FP}_w(n, w, \theta) = \frac{\sum_{b=\theta}^{w} \text{overlapsize}_x(n, w, b)}{\binom{n}{w}} \tag{3}$$

where the numerator indicates the total number of SDRs that match at least θ bits with w and the denominator indicates the number of total patterns from (1). This probability is very low; see Figure 2 below for a visual.
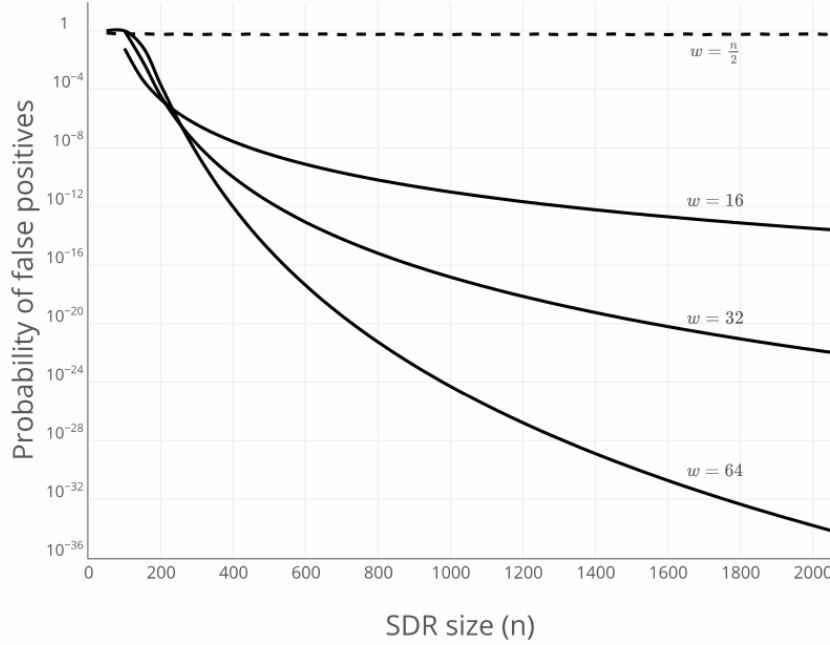

Figure 2: False positive probabilities for SDRs corresponding to equation (3). The SDR size n is on the x-axis and the false positive rate is on the y-axis. Each solid black curve corresponds to a different number of on bits w. As the SDR size increases, the false positive probability becomes essentially zero. The dashed black line at the top indicates the false positive probability of SDRs when $w = n/2$ (i.e., half the bits are 1). In thse case, the error rate is around 50% which implies sparsity is a necessary component for robustness with SDRs. Figure from Hawkins et al. (2016).

4 Sparse representations in single neurons

Now that we have sufficient background on SDRs, we turn to the question of how biological neurons operate on such representations. We will primarily cover information in Ahmad and Hawkins (2016) while also discussing some relevant information from Hawkins and Ahmad (2016). Note that in this section, we will focus on static analysis (i.e., no learning) and specifically the theory behind active dendrites.

4.1 Methods

First, one important aspect of this model is that the instantaneous activity of presynaptic cells is either on (1) or off (0); this enables simpler mathematical analysis that is in-line with the background on SDRs we covered above. The proposed model treats a neuron as having a set of independent dendritic segments (Hawkins & Ahmad, 2016). Each segment receives some input and decides whether or not to fire a dendritic spike depending on whether it recognized some learned pattern. Each dendritic segment of a neuron has a certain region of potential connections that could be made with other cells; this region is a subset of all possible connections. The number of actual connections the segment makes is a small percentage of the number of possible connections in the aforementioned region.

Mathematically, let n represent the number of possible connections a dendritic segment could make. We can thus represent a dendritic segment with a binary vector $d \in \{0, 1\}^n$, where $d_i = 1$ indicates the segment is connected to presynaptic cell i in the region, and the sum of elements in d indicates the number of connections the synapse has made (i.e., $|d|$). Experimental neuroscience findings indicate the number of dendritic segment connections on typical cortical pyramidal neurons is between 20 and 300, whereas the number of possible connections n is taken to be much larger (Major et al., 2013). Likewise we can represent the presynaptic activity at time t in the region of possible connections with the binary vector $a_t \in \{0, 1\}^n$. Let $|a_t|$ denote the number of active cells. As discussed above a dendritic spike fires when the match between a_t and d exceeds some threshold θ.

4.2 Results

Other typical computational models assume full connectivitiy between a cell and its set of possible inputs (e.g. "fully connected" layers in artificial neural networks, where every neuron in one layer is connected to every neuron in the next layer). However, the theory discussed here assumes very sparse connectivity. How, then, can we assume our model dendritic segment can reliably detect a pattern when it is not connected to very many cells?

Recall the tradeoff between robustness to noise (i.e., a dendritic segment can still detect patterns even with noisy neuronal firing) and generating false positives (i.e., a segment detects a pattern when there is none). Note that a high overlap threshold θ makes a segment less robust to noise but more confident in its predictions, and vice-versa. By equation (3), the probability that any random presynaptic activity pattern generates a match is simply the probability of a false positive:

$$P(a_t^T d \geq \theta) = \frac{\sum_{b=\theta}^{|d|} \text{overlapsize}_d(n, |a_t|, b)}{\binom{n}{|a_t|}} \tag{4}$$

While this equation appears quite messy, note the denominator increases much more than the numerator. Hence, by increasing n and $|d|$ we can make the probability of a false positive very small (refer to Figure 2 for false positive rates). Likewise, the probability of a false negative can be decreased in a similar fashion.

We now take a step back and look at populations of dendritic segments. Suppose a neuron has m independent dendritic segments that represent m patterns (assume that each segment detects a single pattern), and each segment has $|d|$ synaptic connections and a matching threshold of θ. A false positive occurs when any of the segments falsely detect

a different pattern. Let $d_i \in \{0,1\}^n$ denote the ith dendritic segment connections, for $1 \leq i \leq m$. Then given some presynaptic input $a \in \{0,1\}^n$ representing a new pattern, we have

$$P(a \text{ matched}) = 1 - P(a \text{ not matched}) = 1 - \prod_{i=1}^{m} P(a^T d_i < \theta)$$

and for high-dimensional sparse vectors this probability is very small.

Thus, we have shown that robust recognition of patterns is possible if presynaptic activity is sparse. Intuitively, if the number of active neurons at any time is much smaller than the population size (i.e., sparse activity), and a dendritic region is connected to a small number of neurons that happen to be some of the active ones, then the chances are quite high that these neurons represent a specific pattern. The mathematical analysis above shows that pattern recognition can be almost perfect even with a large amount of noise and sparse synaptic connections.

5 Sparse representations in networks of neurons

Now that we understand how a single neuron can operate using sparse representations, we turn to the following questions:

- How can we represent the salient properties of neurons (discussed above) in a computational model?

- How can networks of these model neurons learn sequences of patterns?

Potential answers for these questions correspond to work done in Hawkins and Ahmad (2016).

5.1 Methods

Numenta refers to their model neuron as a "HTM neuron" where HTM is Hierarchical Temporal Memory, a term used to describe their models of neocortex (Hawkins & Ahmad, 2016). HTM neurons are modeled essentially identical to the dendritic properties discussed above: HTM neurons have a set of dendritic segments (referred to as "coincidence detectors"), each with their own synapses, and they are in three groups corresponding to the proximal, basal, and apical dendrites of a cortical pyramidal cell (Hawkins & Ahmad, 2016). In this paper, HTM neurons have 128 coincidence detectors with up to 40 synaptic connections on each dendritic segment. See Figure 3 below for a comparison of (A) typical artificial neurons, (B) biological pyramidal neurons, and (C) HTM neurons.
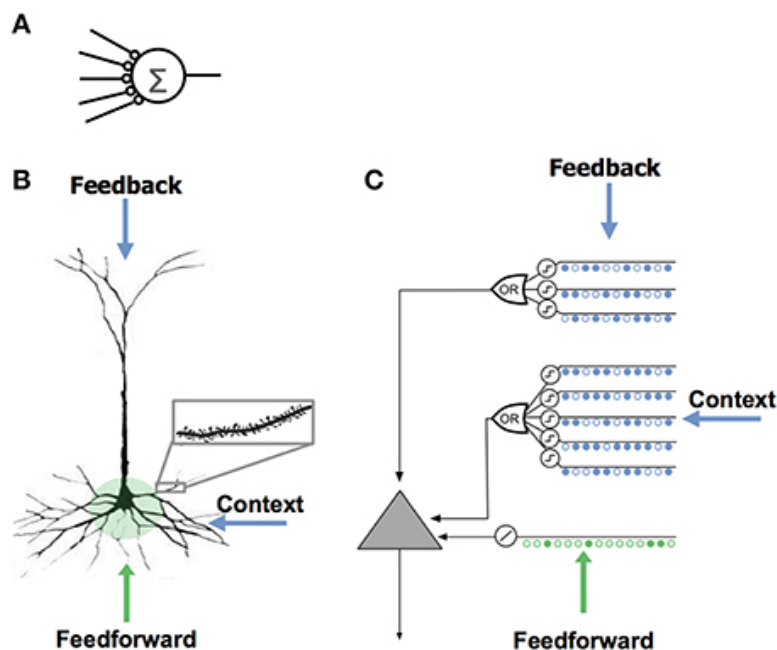
Figure 3: Comparison of different neuron models (Hawkins & Ahmad, 2016). (A) is the typical model used in most ANNs. (B) is a diagram of a typical neocortical pyramidal neuron. Proximal dendrites are highlighted in green and receive feedforward input that can lead directly to (somatic) action potential. The distal dendrites receive contextual input and apical dendrites receive feedback; both these can trigger NMDA spikes and help depolarize the cell if enough spatially-close synapses are activated together. (C) is the computational model of (B), referred to as an HTM neuron.

Recall the properties of sequence memory as listed in section 2.4. We will now cover the basics of using HTM neurons to model sequence learning. (For a more in-depth review refer to (Hawkins & Ahmad, 2016).)

Higher-order sequence memory requires both a representation of the input to the network at a certain time step, and an input for the temporal context. For example, when reading this paragraph, your brain receives visual input from the words on as you read, but there is also an internal representation of the previous sentences and paragraphs enabling you to build on concepts. Figure 4 below, from (Hawkins & Ahmad, 2016), and its caption illustrate basic example, and provide information how cells in mini-columns[6] learn high-order sequences.

---

[6]Without going into too much depth, the cortex is typically divided into 6 layers, and mini-columns span all layers (*Neuroscience. 2nd edition.*, 2001). All the neurons in a mini-column (on the order of 100) share the same feedforward receptive field properties (Hawkins & Ahmad, 2016).
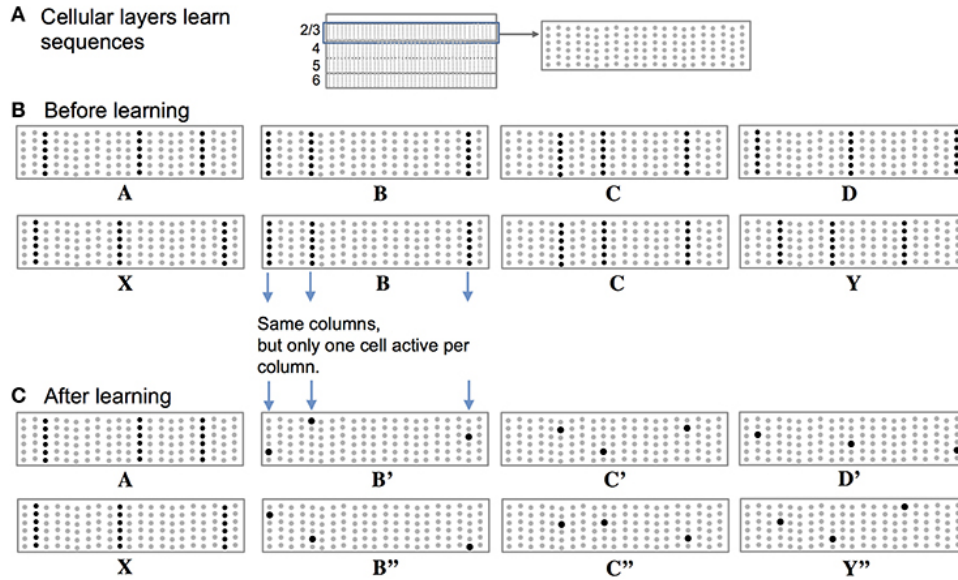
Figure 4: Consider the sequences ABCD and XBCY. In order to correctly predict the fourth letter after receiving the first three, our sequence learning system must have different representations of BC. In (b), all cells in a column become active prior to learning. In (c), only particular cells in columns are active after learning. This essentially serves to encode the past information; for example the B' encoding indicates A was active before and the B'' encoding indicates X was active before. (Panels greatly simplified for clarity.) Figure from Hawkins and Ahmad (2016).

Note how the minicolumn activity exhibited in Figure 4 makes sense from a theoretical perspective. All cells in a minicolumn share the same feedforward properties, so if an input is novel (i.e., hasn't been learned), all cells will become active. However, in the context of a learned sequence of inputs, one or more of the cells in a minicolumn will be depolarized as a result of prediction from the previous input.[7] Hence, the depolarized cells will reach threshold first, thus generating an action potential before other cells and inhibiting nearby cells. This leads to sparse patterns of activity that are unique to a particular element of a particular sequence (Hawkins & Ahmad, 2016).[8]

Now we turn to the synaptic learning rule. For a model neuron to recognize a certain pattern of activity, a set of synapses in a dendritic segment need to connect to a subset of cells that are active when the pattern is recognized (Hawkins & Ahmad, 2016). Hence our model neuron needs to have the ability to "grow" new synapses. This is achieved by maintaining a set of potential synapses in each dendritic segment, where each synapse has a "permanence" value representing the stage of growth of a synapse between the dendritic segment and some other neuron (0 represents no growth and 1 represents a fully-formed synapse) (Hawkins & Ahmad, 2016). The permanence value is altered dynamically using a Hebbian-like learning rule, where increased used leads to stronger permanence and permanence decays over time. Moreover, the model treats all synapses with a permanence exceeding some threshold (e.g., 0.3) as having the same weight, and all others as not being

---

[7]Specifically, the basal synapses learn the transitions between patterns of inputs, whereas the proximal synapses receive the input to the network. Since the basal synapses are essentially a step ahead of the network, predicting the next input from the current input, they are more depolarized than the inactive cells, so upon more feedforward input at the next time step, the cells with basal input reach action potential threshold first.

[8]Perhaps referring to this network as modeling "sequence memory" is not the best term since there is no representation of dynamic-length "sequences." Instead, this network learns transitions between inputs. The nuances of various sequences are encoded in the sparse neuron activity at each time step.

formed at all. This enables for binary-based computation but the permanence is still used internally to model how repeated patterns result in strengthened synapses, and one-time patterns are typically caused by noise.

5.2   HTM vs. other temporal models

Both these models are good at representing information across many time steps, and hence both are good at learning sequences. However, there are also several differences. First, HTM neurons involve more complicated processing of inputs (e.g., dendritic segments as discussed above). Second, HTM models do not train using backpropagation (as used by RNNs); instead, they use a more simple Hebbian-style learning rule. This essentially combines the train/inference phases of typical RNNs and other DL models and reduces computational cost. Third, HTMs are based around sparse neuron activations and sparse neuron connectivity, whereas typical RNNs or other models involve neurons that aren't sparsely activated and don't have sparse connectivity. Finally, HTMs are based on binary representations whereas most other models involve real-valued representations. These properties enable HTM neurons to excel at modeling continual data streams with minimal computational cost. The sparsity properties also enable impressive noise robustness and fault tolerance properties. Finally, there is minimal hyperparameter tuning whereas typical RNN models require tuning of many hyperparameters.

5.3   Results

In Hawkins and Ahmad (2016), a network of HTM neurons (with 2048 minicolumns, each containing 32 HTM neurons) was used to model synthetic data containing high-order sequences in addition to random noise (where the maximum possible average prediction accuracy is 50%). Each HTM neuron had 128 basal dendritic segments, each segment with up to 40 synapses.[9] Figure 5 below shows the accuracy of the network.
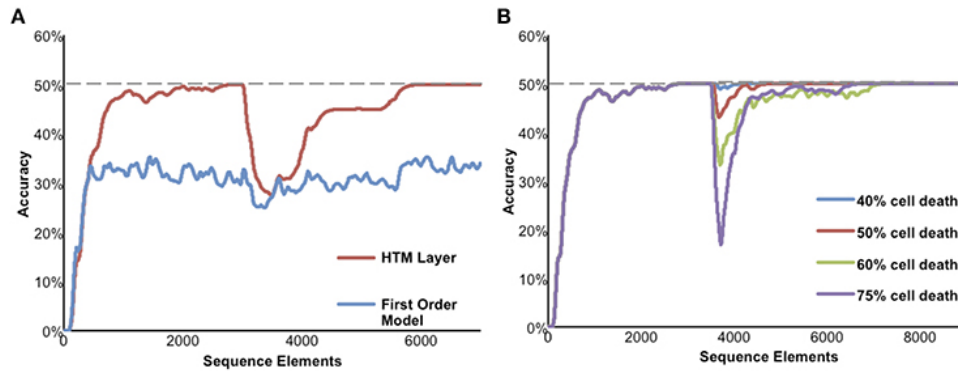

Figure 5: Simulation results of HTM network from (Hawkins & Ahmad, 2016). (A) online learning starting from scratch. After 3000 samples the sequences in the data stream were changed so the model has to recover and learn the new sequences. The blue line represents the performance of a first-order model. (B) After about 3500 samples, a random selection of neurons were inactivated. Note how at 40% cell death the network's performance is hardly affected, and even with 75% cell death the network is able to return to previous accuracy levels after about 2500 more samples.

---

[9]No apical synapses were included because this network was designed to illustate basic properties of sequence memory.

The embedded sequences in the synthetic stream are six elements long (e.g. `XABCDE`) and require high-order representations to fully disambiguate the patterns and achieve the maximal accuracy of 50%.

6   Conclusion

In this paper, we have seen how sparsity in the brain enables it to generalize more broadly and detect patterns even with significant amounts of noise. We can use the theory of SDRs to design robust sequence learning algorithms, such as the HTM model neuron discussed in this paper, that have similar sparsity and noise-robustness properties as the brain. Results of basic experiments with HTM neurons (i.e., those discussed in section 5.3) indicate significant ability to generalize and rapidly adapt to novel patterns. However, the accuracy comparison in Figure 5 only compares the HTM network to a "first-order model" (which was just another HTM network with only one cell per column) (Hawkins & Ahmad, 2016). It would be interesting to make additional comparisons to other popular models capable of learning high-order sequences (such as RNNs and hidden Markov models) to gain a better understanding of how HTM networks compare in practice.

One limitation of these models is that they are naturally suited to continuous streaming data. However, sparsity is a high-level concept that can be applied universally, so another interesting line of work would be to apply it to typical deep neural network models. Dropout, for example, can encourage more sparse models by "killing" a percentage of the neurons and is an effective tool for regularization (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). However, dropout is added on top of traditional dense layers as an augmentation to a typical dense network, so it would be interesting to take a different approach and incorporate sparseness into the layers themselves. Numenta has research into building sparse CNNs, for example, and it would be interesting to further explore this work and apply it to novel problems other than just MNIST (Ahmad & Scheinkman, 2019).

As for this paper, the mathematical analysis of sparsity assumes neural activity is (1) sampled from uniform random distributions, and (2) decorrelated. For the former, clearly the sensory inputs our brain receives is not sample uniformly at random. For example there are patterns of sensory inputs brains process frequently, and other patterns brains never process. Hence extending this analysis to other distributions of neural activity is an interesting topic for future work. For the latter, the topic of neural correlation and distributions of individual spiking neurons in cortex is debated, but due to the inherent robustness of sparse reprsentations, neural activity that is not completely uncorrelated still yields high fault tolerance (Ahmad & Hawkins, 2016).

Additionally, our assumptions of binary synapses is also up for debate. However, if for example we were to change synapses from a single bit to 3 bits (representing 8 possible synaptic weights) without changing sparsity, the noise robustness of our system would still remain the same. This is because robustness relies on the sparse representations. Theoretically, adding more bits of resolution to synapses would only expand the space of possible representations, and this presents another possible direction for future investigation.

The understanding of active dendrites has not been fully established in neuroscience literature, but establishing theories for their purpose can enable future hypothesis-driven experimental work. Nonetheless, sparsity is a core component of neocortical function so this line of work is a promising start to the future of brain-inspired intelligent algorithms.

References

Ahmad, S., & Hawkins, J. (2016). *How do neurons operate on sparse distributed representations? a mathematical theory of sparsity, neurons and active dendrites.*

Ahmad, S., & Scheinkman, L. (2019). *How can we be so dense? the benefits of using highly sparse representations.*

Augusto, E., & Gambino, F. (2019). Can nmda spikes dictate computations of local networks and behavior? *Frontiers in Molecular Neuroscience*, *12*.

Barth, A., & Poulet, J. (2012, 05). Experimental evidence for sparse firing in the neocortex. *Trends in neurosciences*, *35*, 345-55. doi: 10.1016/j.tins.2012.03.008

Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. USA: Cambridge University Press.

Hawkins, J., & Ahmad, S. (2016). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits*, *10*, 23. Retrieved from https://www.frontiersin.org/article/10.3389/fncir.2016.00023 doi: 10.3389/fncir.2016.00023

Hawkins, J., Ahmad, S., Purdy, S., & Lavin, A. (2016). *Biological and machine intelligence (bami)*. Retrieved from https://numenta.com/resources/biological-and-machine-intelligence/ (Initial online release 0.4)

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, *3*, 31. Retrieved from https://www.frontiersin.org/article/10.3389/neuro.09.031.2009 doi: 10.3389/neuro.09.031.2009

Major, G., Larkum, M., & Schiller, J. (2013, 07). Active properties of neocortical pyramidal neuron dendrites. *Annual review of neuroscience*, *36*, 1-24. doi: 10.1146/annurev-neuro-062111-150343

Neuroscience. 2nd edition. (2001). Sunderland (MA): Sinauer Associates.

Park, J., Papoutsi, A., Ash, R. T., Marin, M. A., Poirazi, P., & Smirnakis, S. M. (2019). Contribution of apical and basal dendrites to orientation encoding in mouse v1 l2/3 pyramidal neurons. *Nature Communications*, *10*(1), 5372. Retrieved from https://doi.org/10.1038/s41467-019-13029-0 doi: 10.1038/s41467-019-13029-0

Robinson, R. (2010, September). Quiet down now: how excitatory neurons inhibit one another. *PLoS biology*, *8*(9). Retrieved from https://europepmc.org/articles/PMC2935453 doi: 10.1371/journal.pbio.1000474

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929-1958. Retrieved from http://jmlr.org/papers/v15/srivastava14a.html

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). *Attention is all you need.*