Biologically Inspired Sleep Algorithm for Reducing Catastrophic Forgetting in Neural Networks (Student Abstract)

Timothy Tadros,[1] **Giri Krishnan,**[1] **Ramyaa Ramyaa,**[2] **Maxim Bazhenov**[1]
[1]Department of Medicine, University of California, San Diego
9500 Gilman Drive, La Jolla, California 92092, +18585343377
[2]Department of Computer Science, New Mexico Tech
{tttadros, gkrishnan, mbazhenov}@ucsd.edu, ramyaa.ramyaa@nmt.edu

Abstract

Artificial neural networks (ANNs) are known to suffer from catastrophic forgetting: when learning multiple tasks, they perform well on the most recently learned task while failing to perform on previously learned tasks. In biological networks, sleep is known to play a role in memory consolidation and incremental learning. Motivated by the processes that are known to be involved in sleep generation in biological networks, we developed an algorithm that implements a sleep-like phase in ANNs. In an incremental learning framework, we demonstrate that sleep is able to recover older tasks that were otherwise forgotten. We show that sleep creates unique representations of each class of inputs and neurons that were relevant to previous tasks fire during sleep, simulating replay of previously learned memories.

Introduction

Although ANNs have equaled and even surpassed human performance on various tasks, they suffer from a problem known as catastrophic forgetting (McClelland, McNaughton, and O'reilly 1995). While humans can continuously learn from new information, ANNs perform well on new tasks at the expense of older tasks. In biological networks, sleep has been hypothesized to play an important role in memory consolidation and generalization of knowledge (Rasch and Born 2013). During sleep, neurons are spontaneously active without external input and generate complex patterns of synchronized oscillatory activity across brain regions. Previously experienced or learned activity is believed to be replayed during sleep (Ji and Wilson 2007). This replay of recently learned memories along with relevant old memories is thought to be the critical mechanism that results in memory consolidation. Biophysical modelling of sleep has shown that sleep can recover memories lost due to catastrophic forgetting by replaying old memories and decreasing representational overlap among interfering memories (Gonzalez et al. 2019). In this study, we implemented the main mechanisms behind sleep activity to reduce catastrophic forgetting.

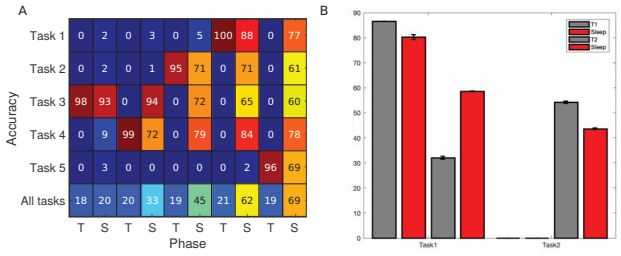

Figure 1: Sleep reduces catastrophic forgetting for MNIST (A) and CUB-200 (B) datasets. In A, each task is a pair of MNIST digits (T = training, S = sleep). In B, each task is half the CUB-200 dataset (training-grey, sleep-red) (left group = first half, right group = second half).

Sleep Algorithm

While most ANNs are modelled after simplified brain dynamics, spiking neural networks (SNNs) seek to more closely model temporal brain dynamics. The key advantage of using SNNs for solving catastrophic forgetting is that spiking activity or replay related to the older tasks may still occur in SNNs, even if the information seems to have been forgotten based purely on classification performance. To use this SNN advantage, we utilized an existing algorithm for converting an ANN to SNN (Diehl et al. 2015) and stimulated the SNN with noisy versions of the average input observed during normal training while updating weights during the sleep phase (see pseudocode). We utilized a partial version of spike-timing dependent plasticity (STDP): (a) if a pre-synaptic spike induces a post-synaptic spike, then the weight between these neurons is increased; (b) if a post-synaptic spike occurs, but the pre-synaptic neuron did not cause this spike, then the weight is decreased. During sleep, inputs must be converted into spike-trains in order to propagate activity from the input layer to the hidden layers of the network. We converted inputs (real-valued pixel intensities or features) to spike trains by computing a Poisson-distributed spike train, such that inputs with higher values (i.e. brighter pixels) will spike more than inputs with lower values. We utilized the average image seen so far (from all

procedure SLEEP($nn, I, scales$) ▷ I is input

  for $t \leftarrow 1$ to Ts **do** ▷ Ts - duration of sleep

    $S(1) \leftarrow$ Convert input I to Poisson-distributed spike train

    for $l \leftarrow 2$ to n **do** ▷ n - number of layers

      $v(l,t) \leftarrow v(l,t-1) + (scales(l-1)\mathbf{W}(l)S(l-1))$ ▷ W(l) - weights

      $S(l) \leftarrow v(l,t) > threshold(l)$ ▷ Propagate spikes

      $\mathbf{W}(l,l-1) \leftarrow \begin{cases} \mathbf{W}(l,l-1) + inc & \text{if } S(l) == 1 \,\&\, S(l-1) == 1 \\ \mathbf{W}(l,l-1) - dec & \text{if } S(l) == 1 \,\&\, S(l-1) == 0 \end{cases}$ ▷ Apply STDP rule

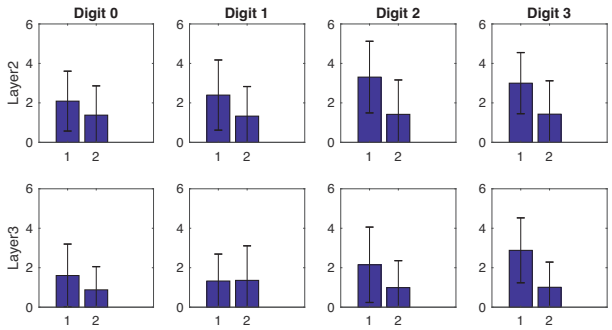

Figure 2: Firing rate of digit specific neurons (left-most bar in each subplot) is greater than firing rate of random neurons (right-most bar) during sleep. Digit-specificity is defined by picking 100 neurons with highest firing rates for each digit after running each digit through the SNN.

previous tasks) as SNN input in order to simulate noisy replay during sleep without revealing information about previous tasks. Next, activity was propagated through the network as spikes and the STDP rule was applied. To simulate sleep, with periods of up-states where activity is elevated and a neuron sequence is thought to be replayed, weights were up-scaled to induce high firing rates during sleep.

Results

We tested our algorithm on two datasets: MNIST and CUB-200 Res-Net50 features. The MNIST dataset consists of 70,000 28x28 greyscale images of handwritten digits, with 60,000 in the training set and 10,000 in the testing set. CUB-200 is a high resolution dataset of images of birds with 200 bird species, with few (30) images per class. For MNIST, we analyzed an incremental learning task where the ANN must learn pairs of digits sequentially. For CUB-200, we used the same incremental learning framework, splitting the dataset into 2 tasks (Task 1 = bird species 1-100, Task 2 = 101-200). Figure 1 shows that sleep is able to recover performance on tasks that were otherwise forgotten.

In biological networks, sleep consolidates memories by replaying previously learned information and reducing representational overlap. Using spiking information in the SNN, we verified that previously learned task information in the ANN is reactivated during sleep (Fig. 2). This alters the weights in a manner such that task representation becomes more distinct, allowing the network to respond to digits of previously learned tasks more accurately (Fig. 3).

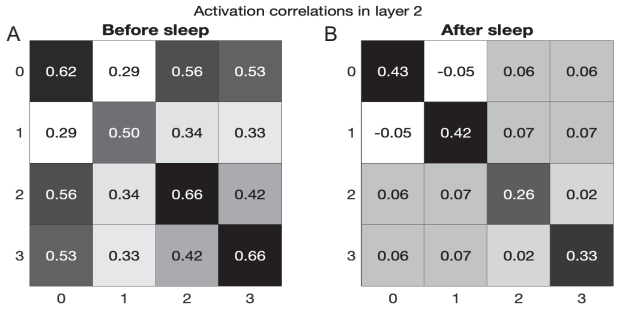

Figure 3: Correlation of activations (computed after learning task 2) before (A) and after sleep (B) in layer 2 for each pair of digits.

Conclusion

We developed a sleep-like algorithm which reduces catastrophic forgetting in an incremental learning setting by reducing representational overlap and replaying previously learned information without storing this information.

Acknowledgments: This work is supported by the L2M program from DARPA/MTO (HR0011-18-2-0021).

References

Diehl, P. U.; Neil, D.; Binas, J.; Cook, M.; Liu, S.-C.; and Pfeiffer, M. 2015. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In *2015 International Joint Conference on Neural Networks (IJCNN)*, 1–8. IEEE.

Gonzalez, O. C.; Sokolov, Y.; Krishnan, G.; and Bazhenov, M. 2019. Can sleep protect memories from catastrophic forgetting? *BioRxiv* 569038.

Ji, D., and Wilson, M. A. 2007. Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience* 10(1):100.

McClelland, J. L.; McNaughton, B. L.; and O'reilly, R. C. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review* 102(3):419.

Rasch, B., and Born, J. 2013. About sleep's role in memory. *Physiological reviews* 93(2):681–766.