When to Learn and When to Stop: Quitting at the Optimal Time (Student Abstract)

Diana Vins[1,2], **Jean Erik Delanois**[2], **Maxim Bazhenov**[2]
[1] School of Biological Sciences, University of California, San Diego
[2] Department of Medicine, University of California, San Diego
9450 Gilman Dr, La Jolla, 92092-0100
dvins@ucsd.edu, jdelanois@ucsd.edu, mbazhenov@ucsd.edu

Abstract

Artificial neural networks (ANNs) struggle with continual learning, sacrificing performance on previously learned tasks to acquire new task knowledge. Here we propose a new approach allowing to mitigate catastrophic forgetting during continuous task learning. Typically a new task is trained until it reaches maximal performance, causing complete catastrophic forgetting of the previous tasks. In our new approach, termed Optimal Stopping (OS), network training on each new task continues only while the mean validation accuracy across all the tasks (current and previous) increases. The stopping criterion creates an explicit balance: lower performance on new tasks is accepted in exchange for preserving knowledge of previous tasks, resulting in higher overall network performance. The overall performance is further improved when OS is combined with Sleep Replay Consolidation (SRC), wherein the network converts to a Spiking Neural Network (SNN) and undergoes unsupervised learning modulated by Hebbian plasticity. During the SRC, the network spontaneously replays activation patterns from previous tasks, helping to maintain and restore prior task performance. This combined approach offers a promising avenue for enhancing the robustness and longevity of learned representations in continual learning models, achieving over twice the mean accuracy of baseline continuous learning while maintaining stable performance across tasks.

Introduction

Traditional machine learning models, which are typically trained in isolation on a fixed dataset, struggle to retain prior knowledge when exposed to new information. As the demand for more adaptable and lifelong learning systems grows, there is a need to develop methods that can effectively balance the learning of new tasks while preserving the performance of old ones.

One existing method to mitigate CF is Sleep Replay Consolidation (SRC) algorithm (Tadros et al. 2022) inspired by the role of biological sleep in memory and learning (Klinzing, Niethard, and Born 2019). With SRC approach, each new task training phase is followed by a period of "sleep" - noise driven spontaneous replay of previously learned activation patters with local unsupervised synaptic plasticity - wherein previous tasks performance is recovered (Tadros et al. 2022).

In this new study, we propose an extension to SRC and potentially many other continual learning strategies. Our new method, Optimal Stopping (OS), dynamically determines when to halt training based on overall network performance. OS monitors the mean validation accuracy across all tasks after each training epoch. When this metric begins to decline, indicating that catastrophic forgetting is outpacing new learning, training is halted and the network parameters are reverted to their previous state. This approach is conceptually similar to early stopping for preventing overfitting (Caruana, Lawrence, and Giles 2000), but focuses on preserving cross-task performance rather than generalization within a single task. In combination with SRC, OS shows further performance improvements, suggesting that the timing of SRC application is important for enabling effective memory consolidation through replay.

Methods

Optimal Stopping (OS): The implementation of OS uses a simple yet effective criterion based on mean task performance. During standard backpropagation training, the model's validation accuracy is tracked for all tasks (trained and untrained). After each epoch, if the average performance across all tasks decreases compared to the previous epoch, the model's parameters are reverted to their previous state and training is halted. This ensures new knowledge is only integrated when it provides a net benefit to the overall network performance.

Sleep Replay Consolidation (SRC): SRC is implemented, as previously (Tadros et al. 2022), by mapping the trained ANN to a spiking neural network (SNN) with the same architecture. The SNN's activity is driven by Poisson distributed spike trains whose frequencies are set based on the average input values observed in the training datasets. Local unsupervised Hebbian-type plasticity modifies weights, i.e. synaptic strengths, during the "sleep" phase. A synapse is increased if presynaptic activation was followed by postsynaptic activation and reduced if postsynaptic activation occurred without presynaptic activation. After the sleep phase is completed, the SNN's updated parameters are mapped back to an ANN where standard ANN testing and training can resume.

Experimental setup: The model employed here is a fully connected, 4-layer ANN. MNIST was split into five tasks,

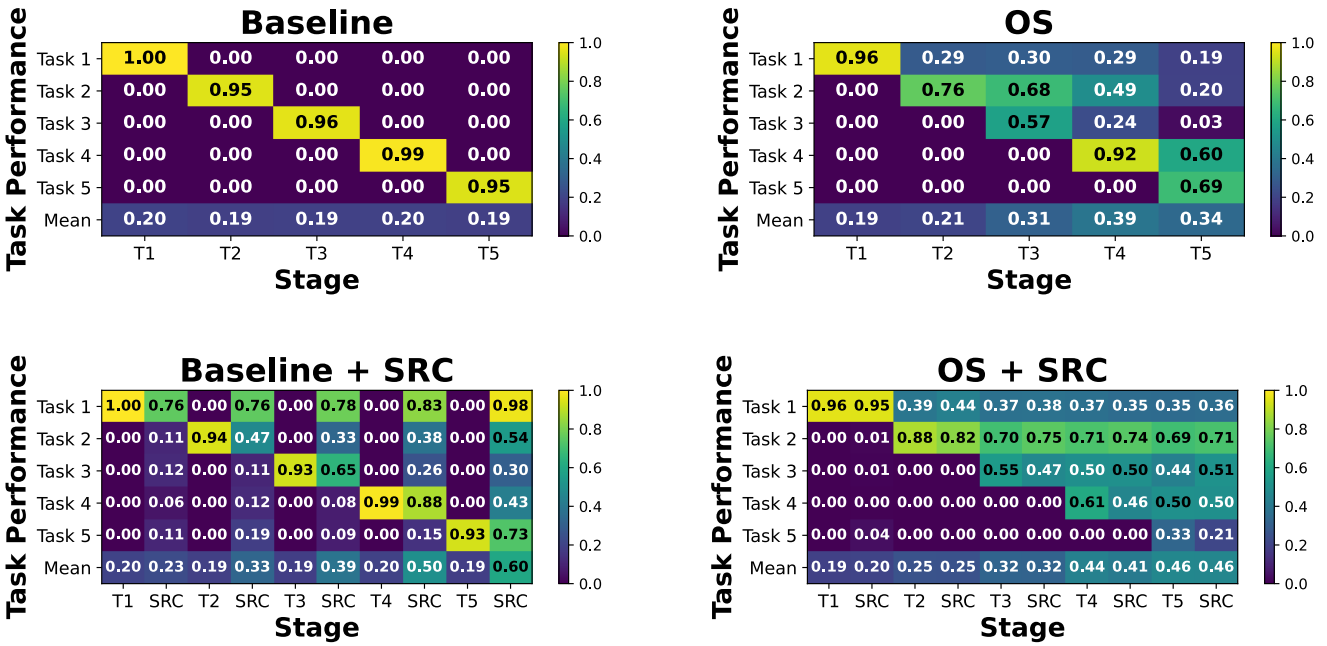

Figure 1: Performance for Baseline, Baseline + SRC, OS, and OS + SRC. Horizontal axis denotes training stage while vertical axis specifics specific task performance, color indicates average accuracy across 5 trials. Complete catastrophic forgetting can be seen in the baseline model (upper left) while OS is able to introduce a degree of new task performance without losing previous task accuracy, resulting in overall improved mean accuracy (upper right). The addition of SRC to both methods further increases accuracy (bottom left and bottom right)

each classifying two digits, and was trained sequentially. We compared four experimental conditions: (1) Baseline: training each task for a fixed number of epochs, (2) OS: using optimal stopping to determine training duration, (3) Baseline+SRC: each task is trained with fixed epochs and followed by a sleep phase as completed in (Tadros et al. 2022), and (4) OS+SRC: each task is trained to optimal stopping and sleep is applied after each training stage. Five trials with different random seeds were run for each condition. Detailed model hyperparameters are provided in the supplement.

Results

When the Baseline model is analyzed, catastrophic forgetting is clearly observed (Figure 1 upper left). Following the predefined training duration, each new task achieves optimal performance while all previously learned tasks drop to zero accuracy, resulting in a consistent mean performance around 20%.

When the OS training paradigm is introduced, notably higher mean accuracy can be achieved (Figure 1 upper right). Here we can observe a clear balance between old and new task performance. Under the OS regime, the model is constrained to only train the new task so long as the mean accuracy of current and previous tasks is rising. Such limited training of a new task results in lower new task performance, however, the network is able to retain performance on previous tasks resulting in a marked improvement in mean performance across all tasks from about 20% to 35% (Compare Figure 1, Baseline Mean T5 vs OS Mean T5). When OS is combined with SRC, this pattern is exaggerated resulting in around 45% accuracy, over doubling Baseline performance. Our results were consistent across multiple random task orders (see supplement).

For comparison with published work, we implemented the standard SRC protocol following (Tadros et al. 2022). Our implementation achieved better mean performance (∼60%) than originally reported (48%). However, SRC alone also revealed high accuracy variability, with mean performance oscillating between peaks after sleep phases and troughs during training. In contrast, OS+SRC enables a smooth integration of new information, suggesting more stable memory acquisition. While current OS+SRC performance does not exceed standard SRC, it may offer advantages when the ANN needs to learn continuously never loosing performance on the older tasks.

Acknowledgments

Supported by NSF (grants 2323241 and 2223839).

References

Caruana, R.; Lawrence, S.; and Giles, C. 2000. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. *Advances in neural information processing systems*, 13.

Klinzing, J. G.; Niethard, N.; and Born, J. 2019. Mechanisms of systems memory consolidation during sleep. *Nature neuroscience*, 22(10): 1598–1610.

Tadros, T.; Krishnan, G. P.; Ramyaa, R.; and Bazhenov, M. 2022. Sleep-like unsupervised replay reduces catastrophic forgetting in artificial neural networks. *Nature communications*, 13(1): 7742.