Brain-inspired synaptic rule for adaptive continual learning in deep neural networks

Suhee Cho (suhee.cho@kaist.ac.kr)
Department of Brain and Cognitive Sciences,
Korea Advanced Institute of Science and Technology,
Daejeon, 34141, Republic of Korea

Hyeonsu Lee (hslee9305@kaist.ac.kr)
Department of Bio and Brain Engineering,
Korea Advanced Institute of Science and Technology,
Daejeon, 34141, Republic of Korea

Se-Bum Paik (sbpaik@kaist.ac.kr)
Department of Brain and Cognitive Sciences,
Korea Advanced Institute of Science and Technology,
Daejeon, 34141, Republic of Korea

Abstract

While deep neural networks (DNNs) outperform humans in various tasks, they still encounter challenges in continual learning due to catastrophic forgetting. To tackle this challenge, we propose a model emulating the brain's ability to memorize sequential information. We specifically targeted the serial position effect and the Hebb repetition effect, which illustrate the working memory's capacity to retain sequential information. Inspired by synapses within the working memory system, we designed synapses with various flexibilities and randomly distributed them within the network, training the network incrementally to learn classes. As a result, our model successfully replicates the serial position effect by effectively memorizing items learned both earlier and later in the sequence. It further reproduces the Hebb repetition effect, enhancing memory through repetitive learning. Consequently, our model adaptively allocates memory resources to sequentially presented information, suggesting a potential pathway for adaptive continual learning in DNNs.

Keywords: continual learning; incremental learning; catastrophic forgetting; sequential working memory

Introduction

Recent advances in deep neural network (DNN) models have demonstrated remarkable performance across various tasks, often surpassing human capabilities (Barui, Sanyal, Rajmohan, Malik, & Dudani, 2022). However, DNN models face a significant challenge with continuous streams of information (Fig. 1a), particularly catastrophic forgetting—where the network abruptly loses memory of prior items upon learning new ones (Fig. 1b, blue) (McCloskey & Cohen, 1989; French, 1999). This highlights the need for DNNs to allocate their memory resources more evenly across the continual information streams, ultimately facilitating the retention of prior knowledge alongside the incorporation of new inputs.

In contrast, the human brain dynamically allocates memory resources. Key cognitive phenomena such as the "serial position effect" and the "Hebb repetition effect" observed in human working memory underscore this adaptability. The serial position effect illustrates the brain's tendency to better memorize items positioned at the sequence's beginning and end (Fig 1b, red) (Murdock, 1962; Avons & Mason, 1999), while the Hebb repetition effect shows the enhancement of memory for sequential inputs through the repetitive presentation of inputs (Fig. 1c) (Hebb, 1961; Burgess & Hitch, 1999). These phenomena highlight the brain's adaptive and robust continual learning capabilities, suggesting the importance of integrating such mechanisms into machine intelligence.

Motivated by these cognitive insights, we propose a DNN model with mixed synaptic flexibilities to emulate the serial position and Hebb repetition effects. Our aim is to bridge the gap between artificial and biological intelligence by endowing DNNs with adaptive learning capabilities akin to the human brain.

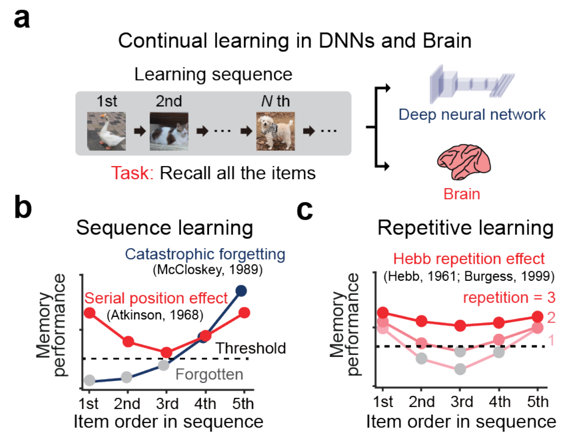

Figure 1: (a) Continual learning in DNNs and the brain. (b) Serial position effect in the brain and catastrophic forgetting in DNNs. (c) Hebb repetition effect in the brain.

Our model

Synapse design

We designed a set of synapses with various flexibilities, adopting the characteristics of the labile long-term potentiation (LTP) found in the brain (Fig. 2a) (Pradier et al., 2018). Labile LTP can be either stably maintained or unstably discharged depending on conditions. To emulate this dynamic profile, we introduced the concept of "synapse flexibility" (Fig. 2b). A synapse with low flexibility (stable) gets fixed as its weight deviates significantly from its initial values (Fig. 2b, brown). Conversely, a synapse with high flexibility (unstable) allows continuous weight adjustments throughout training (Fig. 2b, blue). We hypothesized that stable synapses contribute to retaining the memory of early items and unstable synapses facilitate learning new items and thereby combining the two can generate the serial position effect (Fig. 2c).

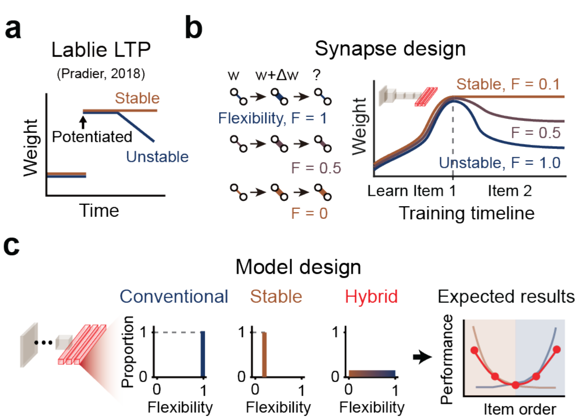

Figure 2: (a) Labile LTP observed in the brain. (b) Example weight profiles of stable and unstable synapses. (c) Three network models with different synaptic compositions.

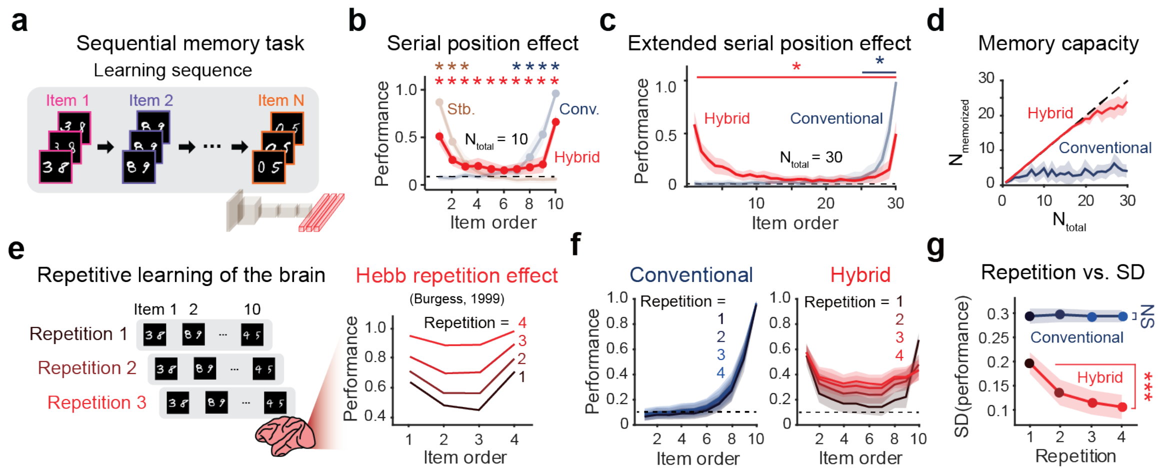

Figure 3: (a) Continual learning task to train deep neural network models. (b) Memory performance of items within the sequence in three different model networks; $*P < 0.05$, one-sample t-test. (c) Results with the extended length of the learning sequence; one-sample t-test, $*P < 0.05$. (d) Number of memorized items for various sequence lengths; $N_{memorized}$ is calculated using cross-validation (one-sample t-test, $P < 0.05$). (e) Repetitive sequential learning of the brain. (f) Repetitive learning results in the conventional and the hybrid model. (g) Variation in standard deviation across items in relation to repetition frequency; paired t-test, $***P < 0.001$; NS, $P = 0.98$

Network design

We constructed three distinct models with varying synaptic compositions within their fully-connected layers: a conventional model featuring fully unstable synapses, a stable-only model with synapses having low flexibility, and a hybrid model with a uniform distribution of flexibility values ranging from 0 to 1 (Fig. 2c). Each model was trained to sequentially memorize and classify distinct handwritten numbers.

Results

The serial position effect emerges from the coexistence of stable and unstable synapses

We trained each model with the ten-item sequence and analyzed the memory performance (Fig. 3a). The conventional model recalled the latest four items with performance exceeding chance levels while forgetting earlier ones — a characteristic of catastrophic forgetting (Fig. 3b, blue). In contrast, the stable-only model retained memory of the first three items but struggled with learning subsequent items (Fig. 3b, brown). Remarkably, the hybrid model exhibited superior memory performance for both early and recent items, showcasing the serial position effect by memorizing all items in the sequence (Fig. 3b, red). Moreover, the serial position effect manifested by the hybrid model remained robust across varying sequence lengths. Even when we expanded the sequence to include 30 items, the hybrid model memorized all items above chance levels, whereas the conventional model only retained the memory of the last five items (Fig. 3c). This result exhibits the hybrid model's flexible utilization of memory resources: with an increasing number of items to memorize, the hybrid model adjusts the resource allocation to accommodate the new information (Fig. 3d).

The model adaptively improves the memory performance through repetitive review

Despite the enhanced memory retention for items at the sequence's beginning and end in our model, items positioned in the middle exhibited comparatively lower memory performance. To address this, we proposed leveraging repetitive reviewing of the item sequence, drawing inspiration from the Hebb repetition effect (Fig. 3e).

We trained the networks in four repetitive cycles of training, using an identical learning sequence each time. The results showed that the conventional DNN's performance lacked repetition dependency (Fig. 3f, left). However, the hybrid model demonstrated substantial memory enhancement, particularly for items in the middle of the sequence, replicating the Hebb repetition effect (Fig. 3f, right). Additionally, repetitive learning reduced the performance disparity between items within the sequence in the hybrid model, leading to a decrease in the standard deviation across item performances (Fig. 3g). Consequently, the hybrid model was capable of uniformly memorizing items through repetitive learning, unlike the conventional model which overly prioritizes the recent items.

Conclusion

In this study, we demonstrated that integrating synapses with diverse flexibilities in DNNs leads to the emergence of the serial position effect and the Hebb repetition effect, crucial aspects of sequential working memory. Our model offers new insights into developing more resilient continual learning in DNNs, proposing potential breakthroughs in overcoming catastrophic forgetting.

References

Avons, S. E., & Mason, A. (1999, February). Effects of visual similarity on serial report and item recognition. *Q. J. Exp. Psychol. A*, *52*(1), 217–240.

Barui, S., Sanyal, P., Rajmohan, K. S., Malik, A., & Dudani, S. (2022, Sep 30). Perception without preconception: comparison between the human and machine learner in recognition of tissues from histological sections. *Scientific Reports*, *12*(1), 16420. doi: 10.1038/s41598-022-20012-1

Burgess, N., & Hitch, G. J. (1999, July). Memory for serial order: A network model of the phonological loop and its timing. *Psychol. Rev.*, *106*(3), 551–581.

French, R. M. (1999, April). Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.*, *3*(4), 128–135.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal..

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), (Vol. 24, p. 109-165). Academic Press. doi: https://doi.org/10.1016/S0079-7421(08)60536-8

Murdock, B. B. (1962, November). The serial position effect of free recall. *J. Exp. Psychol.*, *64*(5), 482–488.

Pradier, B., Lanning, K., Taljan, K. T., Feuille, C. J., Nagy, M. A., & Kauer, J. A. (2018, May). Persistent but labile synaptic plasticity at excitatory synapses. *The Journal of Neuroscience*, *38*(25), 5750–5758. doi: 10.1523/jneurosci.2772-17.2018