International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V2I1P110
Eureka Vision Publication | Volume 2, Issue 1, 84-87, 2021

Original Article

SYNAPSE-Fed: A Bio-Inspired Framework for Continual, Secure, and Explainable AI

Mohan Siva Krishna Konakanchi
Independent Researcher, USA.

Abstract - Artificial intelligence systems deployed in the real world must be capable of continual learningacquiring new knowledge and skills over time without catastrophically forgetting previously learned information. This challenge is particularly acute in decentralized settings where data is streamed and owned by different entities. Drawing inspiration from the brain's mech- anisms of neuroplasticity, we introduce SYNAPSE-Fed (Synaptic Network Adaptation for Perpetual Secure Explainable Federated Learning), a novel framework designed for robust continual learning in a federated environment. At its core, SYNAPSE- Fed features a meta-learning algorithm that models synaptic consolidation to mitigate catastrophic forgetting. By identifying and protecting network parameters crucial for past tasks, our algorithm preserves existing knowledge while adapting to new data streams. To ensure the integrity of this learning process across distributed silos, we embed our algorithm within a trust- aware federated learning protocol. A dynamic trust metric evaluates each participant's contribution based on performance, consistency, and their ability to balance knowledge stability with plasticity, ensuring accountability. Finally, we address the critical need for transparency by introducing a quantitative frame- work to optimize the trade-off between the continual learner's performance and its explainability. We demonstrate through experiments on continual learning benchmarks that SYNAPSE- Fed significantly outperforms existing methods in preventing catastrophic forgetting and shows high resilience in a federated setting with heterogeneous participants.

Keywords - Continual Learning, Meta-Learning, Federated Learning, Explainable AI (XAI), Neuroplasticity, Catastrophic Forgetting.

1. Introduction

Modern AI systems, particularly those based on deep neural networks, have achieved remarkable success but operate on a fundamentally flawed assumption: that the data distribution is stationary. When trained on a sequence of tasks, these networks exhibit a phenomenon known as **catastrophic forgetting**, where learning a new task drastically degrades performance on previously learned ones [1]. This limitation severely hinders the deployment of AI in dynamic, real-world environments that require systems to learn continuously and adapt to new information over their lifetime.

The human brain, in contrast, is an exemplary continual learner. It seamlessly acquires new skills and knowledge with- out overwriting the old. This ability is underpinned by complex biological mechanisms like **synaptic plasticity**, where the strength of synaptic connections is selectively modified to consolidate long-term memories while forming new ones [2]. Inspired by this, the field of Continual Learning (CL) seeks to develop algorithms that enable AI to learn sequentially.

Our work addresses three interconnected challenges in de- veloping practical continual learning systems. First, we need more effective algorithms to combat catastrophic forgetting. We draw inspiration from synaptic consolidation to propose a novel **bio-inspired meta-learning algorithm**. By meta- learning a parameter-importance mapping, our model learns *how* to protect critical knowledge while efficiently adapting to new tasks.

Second, real-world continual learning often happens in a decentralized manner. For instance, a fleet of robots or a consortium of hospitals continually learn from their unique, local data streams. **Federated Learning (FL)** is the natural paradigm for such scenarios, but it requires a mechanism to ensure the integrity and accountability of the learning process across diverse and potentially unreliable participants (silos) [3].

Third, as continual learning systems make evolving deci- sions over time, their behavior must be **explainable** to engender trust and allow for human oversight. A model that silently modifies its knowledge base without transparency is unacceptable in high-stakes applications. This necessitates a formal method to manage the trade-off between the model's adaptive performance and its scrutability.

To this end, we introduce **SYNAPSE-Fed**, a unified framework that synergistically addresses these three chal- lenges. Our primary contributions are:

- A novel, bio-inspired meta-learning algorithm that mim- ics synaptic consolidation to effectively mitigate catas-

trophic forgetting in a sequential task environment.

- A **Trust-Aware Federated Learning** protocol specif- ically designed for continual learning, which assesses client reliability based on their contribution to both knowledge stability (low forgetting) and forward transfer (fast adaptation).
- A quantitative framework to analyze and optimize the inherent **trade-off between continual learning perfor- mance and model explainability**, providing a princi- pled approach for deploying adaptable yet transparent AI.
- Extensive empirical validation on benchmark CL datasets, demonstrating SYNAPSE-Fed's superior per- formance and robustness compared to state-of-the-art baselines.

2. Related Work

2.1. Continual Learning (CL)

Approaches to mitigate catastrophic forgetting are typically categorized into three families.

2.2. Meta-Learning

Meta-learning, or "learning to learn," aims to train a model on a distribution of tasks so that it can adapt to a new task with very few examples [7]. Model-Agnostic Meta-Learning (MAML) is a popular algorithm that learns a parameter initialization that is sensitive to fine-tuning on new tasks.

2.3. Federated Learning for Continual Learning

Combining FL and CL is a natural yet challenging direction. The continual arrival of data at each client creates a dual non-IID challenge: data is heterogeneous across clients and non- stationary over time for each client. Some works have started to explore this "Federated Continual Learning" setting.

3. The Synapse-Fed Framework

SYNAPSE-Fed integrates a bio-inspired continual learner into a secure federated network with an explainability-control module.

3.1. Bio-Inspired Continual Meta-Learning

Let us consider a sequence of tasks $T_1, T_2, ..., T_N$. After training on task T_i, the model must learn T_{i+1} without degrading its performance on tasks $\{T_1, ..., T_i\}$.

Inspired by synaptic consolidation, we associate each model parameter ω_j with an "importance" weight ε_j. When learning a new task, changes to parameters with high importance are penalized. The key novelty of our approach is that we **meta-learn** the function that computes these importance weights.

Let the model parameters be ω. The learning objective for a new task T_{new} is:

$$L(\omega) = L_{new}(\omega) + \vartheta \sum_{j} \Lambda_{\omega}(j)(\omega_j \rightarrow \omega^{\rightarrow})^2$$

Where L_{new} is the standard loss for the new task, $\omega^{\rightarrow}$ are the parameters after learning the previous task, ϑ is a reg- ularization strength, and $\Lambda_{\omega}(j)$ is a meta-learned importance function, parameterized by ϖ. This function Λ_{ω} is a small neural network that takes parameter indices or properties as input and outputs their importance.

The meta-training process involves simulating a continual learning scenario on a set of meta-training tasks. The meta-objective is to find the parameters ϖ for the importance net- work that minimize forgetting across the entire sequence. This allows the model to learn a data-driven, dynamic regularization strategy that is more effective than the static heuristics used in methods like EWC.

3.2. Trust-Aware Federated Continual Learning

We deploy our continual learner in a federated setting with K silos. Each silo continually receives new data. The central server orchestrates the training and maintains the global model's parameters ω and the meta-learned importance estimator's parameters ϖ.

At each communication round ϱ , the server computes a trust score for each silo k. Our trust metric T_k is designed for the CL context:

$$T_k = w_1 S_{perf} + w_2 S_{stability} + w_3 S_{plasticity}$$

- Performance (S_{perf}): Standard performance of the client's updated model on a global validation set con- taining samples from all tasks seen so far.
- Stability ($S_{stability}$): Measures how little the client's update degrades performance on *old* tasks. This is backward

transfer, or ”forgetting.” A lower forgetting score is better.

- Plasticity ($S_{plasticity}$): Measures how well the client's update improves performance on the *most recent* task. This is forward transfer, or adaptability.

The server uses these trust scores to perform a weighted aggregation of the client updates for both the main model ω and the importance estimator ϖ. This ensures that clients who contribute to a stable and adaptable global model are given more influence.

3.3. Explainability-Performance Optimization

To ensure transparency, we create a hybrid system. The final decision is arbitrated between our high-performance SYNAPSE-Fed continual learner (M_{CL}) and a static, inter- pretable baseline model (M_{INT}), such as a rule-based system. A gating mechanism, controlled by a scrutability parameter ϑ_{exp}, manages this arbitration. When the continual learner's uncertainty on a given input is high (e.g., low softmax confidence, or high predictive entropy), the system can be configured to rely on the simpler, more predictable M_{INT}. We define **Explainability (X)** as the percentage of decisions that are routed through or can be validated by M_{INT}. By varying ϑ_{exp}, we can trace a Pareto frontier of models, plotting CL performance (average accuracy across all tasks) against X. This allows an operator to deploy a model that meets a required level of transparency.

4. Experimental Setup

4.1. Datasets and Tasks

We use two standard continual learning benchmarks:

- Permuted MNIST: A sequence of 10 tasks, where each task is the MNIST classification problem but with a fixed random permutation of the input pixels.
- Split CIFAR-10: The CIFAR-10 dataset is split into 5 sequential tasks, each containing 2 classes (e.g., Task 1: airplane/automobile, Task 2: bird/cat, etc.).

4.2. Federated Scenario

We simulate a federation of $K = 20$ clients. The data for each task is distributed in a non-IID fashion across the clients. To test robustness, we designate 4 clients as ”forget- ful,” meaning their local training process uses a much lower regularization parameter ω, causing them to contribute updates that exhibit high catastrophic forgetting.

4.3. Baselines

We compare SYNAPSE-Fed against:

- Fine-tuning (FT): A simple SGD baseline that shows catastrophic forgetting.
- EWC (Elastic Weight Consolidation): A classic regularization-based CL method.

5. Results and Discussion

5.1. Continual Learning Performance

We measure the average classification accuracy across all tasks after the model has been trained on the full sequence.

Table 1: Average Accuracy (%) On Continual Learning Benchmarks

Method	Permuted MNIST	Split CIFAR-10
Fine-tuning (FT)	21.5	19.8
EWC	74.2	55.1
MAML	68.9	49.3
FedAvg+EWC	72.5	53.2
SYNAPSE-Fed	91.3	68.7

The results in Table I show that fine-tuning leads to near-total catastrophic forgetting. EWC provides a strong baseline, but its performance is limited by its static parameter importance heuristic. SYNAPSE-Fed signif- icantly outperforms all baselines. Its ability to meta- learn an importance mapping allows it to form a more effective protection against forgetting, leading to substan- tially higher final accuracy. The trust-aware aggregation in SYNAPSE-Fed also proves superior to the simple averaging in FedAvg+EWC, especially in our scenario with ”forgetful” clients.

5.2. Trust Mechanism Robustness

During the federated training, we observed that the trust scores for the 4 ”forgetful” clients were consistently lower than those for the normal clients. Their low ”Stabil- ity” score ($S_{stability}$) led to their updates being down- weighted, thereby protecting the global model from their tendency to forget. This demonstrates that our CL-specific trust metric is effective at identifying and isolating unre- liable participants.

5.3. Explainability-Performance Trade-off

Plotting the Pareto frontier for SYNAPSE-Fed on Split CIFAR-10 revealed a graceful trade-off. The top- performing model (68.7% accuracy) had an explainability score of $X = 0$. However, a model with $X = 0.75$ (75% of uncertain decisions validated by an interpretable model) could still achieve an accuracy of 64.1%. This al- lows a practitioner to gain a large amount of transparency for a small, measurable cost in performance.

6. Conclusion

SYNAPSE-Fed provides a novel, integrated solution to the challenges of continual learning in decentralized, real- world settings. By drawing inspiration from the principles of synaptic plasticity and framing the problem in a meta-learning context, it effectively mitigates catastrophic forgetting. The integration with a trust-aware federated learning protocol ensures that this continual learning can proceed securely and robustly across multiple data silos. Finally, the framework's ability to quantify the explainability-performance trade-off provides a crucial tool for deploying AI systems that are not only adap- tive and intelligent but also transparent and trustworthy. Future work will explore more complex biological mech- anisms, such as memory replay and structural plasticity, to further enhance the continual learning capabilities of AI.

References

[1] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," in *The psychology of learning and motivation*, vol. 24, pp. 109-165, Elsevier, 1989.

[2] W. C. Abraham and M. F. Bear, "Metaplasticity: the plasticity of synaptic plasticity," *Trends in Neurosciences*, vol. 19, no. 4, pp. 126-130, 1996.

[3] P. Kairouz et al., "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1-210, 2019.

[4] G. M. van de Ven and A. S. Tolias, "Three scenarios for continual learning," *arXiv preprint arXiv:1904.07734*, 2019.

[5] S.-W. Lee et al., "Overcoming catastrophic forgetting by incre- mental moment matching," in *Proc. Advances in Neural Informa- tion Processing Systems (NeurIPS)*, 2017.

[6] J. Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks," in *Proc. National Academy of Sciences (PNAS)*, vol. 114, no. 13, pp. 3521-3526, 2017.

[7] J. Schmidhuber, "Meta-learning," in *Encyclopedia of Machine Learning*, pp. 664-667, Springer, 2011.

[8] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. on Machine Learning (ICML)*, 2017.

[9] S. M. A. M. H. M. M. S. J. P. P. Yoon, "Federated continual learning with plastic connections for glioma segmentation," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[10] A. Rusu et al., "Progressive neural networks," *arXiv preprint arXiv:1606.04671*, 2016.

[11] M. D. Riemer et al., "Learning to learn without forgetting by maximizing transfer and minimizing interference," in *Proc. Int. Conf. on Learning Representations (ICLR)*, 2019.

[12] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. on Artificial Intelligence and Statistics (AISTATS)*, 2017.

[13] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[14] A. Krizhevsky, "Learning multiple layers of features from tiny images," University of Toronto, Tech. Rep., 2009.

[15] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G. Yang, "XAIExplainable artificial intelligence," *Science Robotics*, vol. 4, no. 37, 2019.

[16] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated opti- mization in heterogeneous networks," in *Proc. Conf. on Machine Learning and Systems (MLSys)*, 2019.

[17] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, 2015.

[18] D. Silver et al., "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354-359, 2017.

[19] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206-215, 2019.

[20] D. P. Kingma and J. Ba,"Adam: A method for stochastic optimization," in *Proc. Int. Conf. on Learning Representations (ICLR)*, 2015.