Self-Attention Meta-Learner for Continual Learning

Ghada Sokar[1], Decebal Constantin Mocanu[1,2], Mykola Pechenizkiy[1]

[1] Eindhoven University of Technology, [2] University of Twente, The Netherlands

g.a.z.n.sokar@tue.nl

Introduction

- **Continual learning** aims to build agents that can learn a set of tasks ***sequentially***, acquiring new knowledge from each task without ***forgetting*** the previous learned ones.
- We propose self-attention meta-learner (SAM) that build *prior* knowledge necessary for CL and select the *relevant* knowledge for each task from the past experience.

Contributions

- We show the effectiveness of the ***selective transfer*** performed by SAM in increasing the performance and reducing forgetting by allowing for a selective update for the weights.
- We show the importance of having a prior generic knowledge in increasing the forward transfer especially when the tasks are dissimilar.
- We address the task agnostic scenario where the task identity is not available during inference. We also assume that the previous data is not available.

Proposed Method

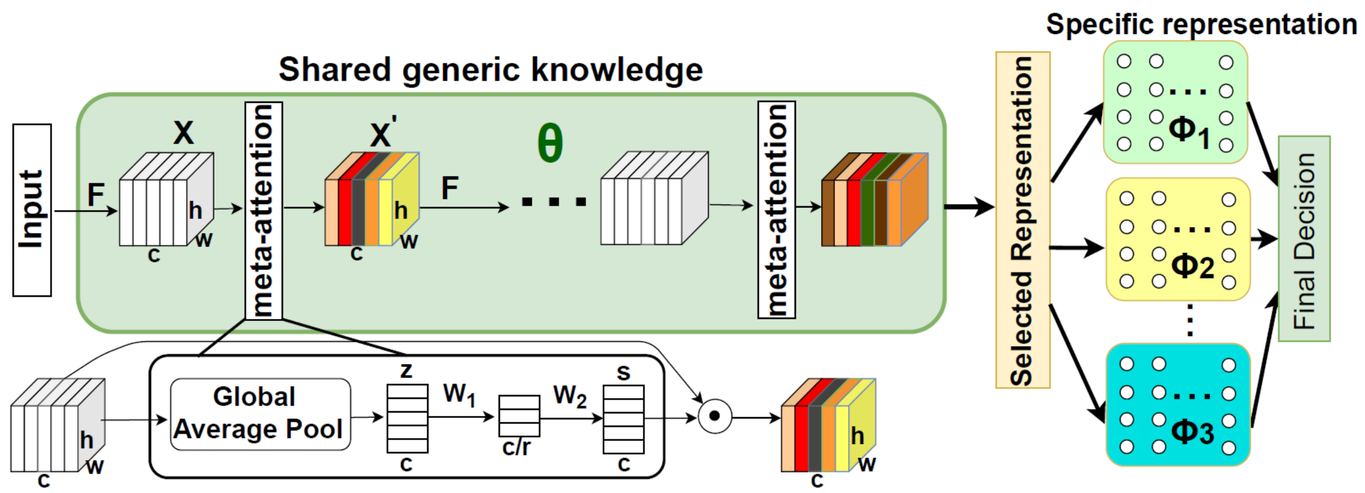

Figure 1: An Overview of our proposed method (SAM).

- The network consists of a shared sub-network and specific sub-networks.
- The shared sub-network learns a prior generic knowledge using optimization-based meta-learning algorithm MAML [1]. It also incorporates a self-attention mechanism meta-trained that learns to boost the relevant features to the input task from each layer in the shared subnetwork.
- Each task in the continual sequence builds a specific branch on the top of the ***selected relevant*** presentation from the shared subnetwork.
- The final decision module is responsible for deciding the predicted class from all the seen classes so far.

Results

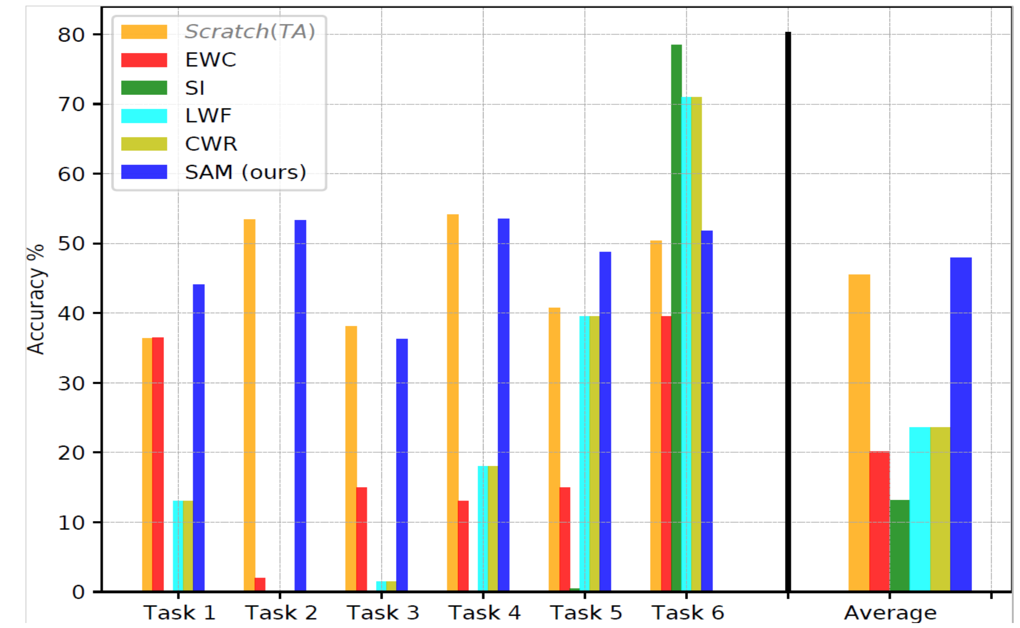

Figure 2: The accuracy of each task of split CIFAR-10/100 as well as the average accuracy across all tasks.

Table 1: Enhancing existing CL strategies by SAM. "Standard" represents the original form of the methods.

	Split MNIST		Split CIFAR-10/100	
Method	Standard	SAM	Standard	SAM
Fine-tuning	19.86 ± 0.04	**53.87** ± 1.73	12.24 ± 0.05	**25.45** ± 1.76
SI [2]	19.99 ± 0.06	**67.32** ± 0.43	13.39 ± 0.04	**42.92** ± 1.01
MER [3]	32.66 ± 2.33	**50.04** ± 1.85	-	-

Conclusion

- We propose the self-attention meta-learner (SAM) for the continual learning paradigm.
- We show that our approach outperforms the state-of-the-art methods in the class-incremental-setting.
- We illustrate the performance boost achieved by the popular existing methods when they are integrated in our framework.

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic metalearning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 1126–1135.
[2] Friedemann Zenke, Ben Poole, and Surya Ganguli. 2017. Continual learning through synaptic intelligence. In Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 3987–3995.
[3] Riemer, Matthew, Ignacio Cases, Robert Ajemian, Miao Liu, Irina Rish, Yuhai Tu, and Gerald Tesauro. Learning to Learn without Forgetting by Maximizing Transfer and Minimizing Interference. In International Conference on Learning Representations. 2018.