Self-Attention Meta-Learner for Continual Learning
Ghada Sokar\textsuperscript{1}, Decebal Constantin Mocanu\textsuperscript{1,2}, Mykola Pechenizkiy\textsuperscript{1}
\textsuperscript{1}Eindhoven University of Technology, \textsuperscript{2}University of Twente, The Netherlands
g.a.z.n.sokar@tue.nl

Introduction

- \textbf{Continual learning} aims to build agents that can learn a set of tasks \textit{sequentially}, acquiring new knowledge from each task without \textit{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 \textit{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

- 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 \textit{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

- 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.

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.