Overcoming forgetting in continual learning with an ensemble of experts

Continual learning and exemplar-free class incremental learning. Continual learning (CL) is an important field in machine learning aimed at developing systems capable of learning from a continuous data stream, similar to human learning. One critical aspect of CL is exemplar-free class incremental learning (EFCIL), which focuses on teaching models new concepts without retaining past data samples. This approach is especially vital in situations with privacy constraints or limited storage capacity. However, EFCIL faces a significant challenge known as catastrophic forgetting, where models forget previously learned information upon encountering new tasks (depicted in Fig. 1).

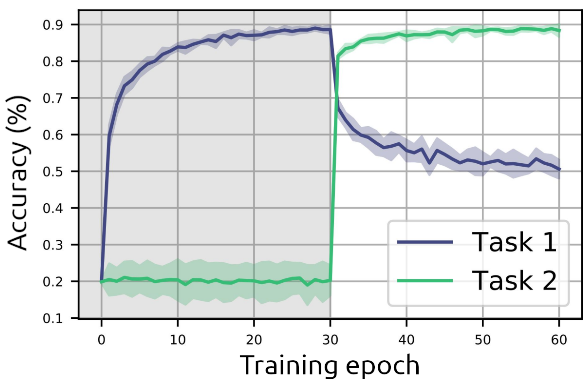

Figure 1: Visualization of catastrophic forgetting when training an artificial neural network. Once it starts to learn a new task (epoch 30), its accuracy in solving the first task drastically drops, showcasing catastrophic forgetting.

Scientific goal of the project. This project aims to address catastrophic forgetting in EFCIL by developing an advanced ensemble method that leverages the strengths of multiple neural networks, referred to as experts. Similar to humans, experts can specialize in different tasks and cooperate to solve them better. Additionally, they can help each other to reduce forgetting or explain complex tasks, thus improving the learning process. Traditional single-model approaches to EFCIL often suffer from semantic drift, where the model's internal representation of past classes changes unfavorably during new learning tasks, leading to forgetting. In this project, we seek to utilize the knowledge of other experts to help currently trained expert alleviate its semantic drift and forgetting.

Research questions. Alleviating semantic drift: How can we mitigate semantic drift within an ensemble of experts? The hypothesis is that when training a single expert, leveraging knowledge from other experts in the ensemble can better maintain its integrity of past class representations. That will reduce the expert's semantic drift and decrease its forgetting. Ensembled knowledge distillation: How can we improve knowledge distillation techniques for ensembles? Knowledge distillation is one of the best ways to reduce forgetting. The project proposes a novel method to distill knowledge across experts, ensuring that the model retains information about previous tasks more robustly, potentially outperforming existing logit and feature distillation techniques.

Significance and impact of the project. Current CL approaches often rely on storing past exemplars, creating challenges in memory usage, privacy, and scalability. This project aims to pioneer more efficient, scalable, and privacy-preserving incremental learning techniques by developing novel ensemble methods that eliminate the need for stored exemplars, potentially setting new benchmarks in the field. By integrating ensemble techniques with CL, the project expects to enhance model robustness and performance in dynamic environments, crucial for applications like autonomous systems and real-time analytics. The findings will have significant implications for industries such as healthcare and finance, enabling continual learning without compromising data privacy. By addressing core challenges like semantic drift and catastrophic forgetting through innovative ensemble methods, the project aspires to significantly advance the state-of-the-art in continual learning. By reducing the need for data storage and ensuring compliance with privacy regulations, this project aligns with both practical and ethical considerations in modern AI applications, contributing valuable insights and inspiring further research in the academic community.

Work plan. This project will last 12 months and is divided into three phases. We will finalize it by creating a submission for an international machine learning conference. Firstly, we will verify research hypotheses regarding overcoming semantic drift in an EFCIL ensemble method. Secondly, we will focus on the ensembled knowledge distillation. In the last phase, we will propose and implement a novel ensemble method based on the conclusion derived from the previous steps. We will evaluate it against state-of-the-art methods and perform its ablation. Our research findings, models, and code will be published at a prestigious machine learning conference and made available online in line with the principles of open science.