

Continual Learning Overview: Summary

Motivation. Standard machine learning assumes a closed-world setting with fixed data distributions, fixed label spaces, and one-shot training, whereas real-world agents encounter sequential data, distributional shifts, and new concepts, and naive adaptation in this setting leads to *catastrophic forgetting*, motivating continual learning methods that balance *plasticity* (learning new tasks) with *stability* (retaining old knowledge).

Problem Settings. Continual learning includes multiple formulations depending on task structure and supervision, including Instance-Incremental Learning (IIL), Domain-Incremental Learning (DIL), Task-Incremental Learning (TIL), Class-Incremental Learning (CIL), as well as more realistic settings such as Task-Free Continual Learning, Online Continual Learning, Blurred Boundary Continual Learning, and Continual Pre-training where data arrives sequentially without clear task boundaries.

Biological Inspiration: Complementary Learning Systems. The Complementary Learning Systems (CLS) hypothesis posits two interacting memory systems—a fast, episodic system that minimizes interference and a slow, distributed system that supports generalization—where memory consolidation occurs through interleaved replay of new and old experiences, inspiring replay-based and regularization-based continual learning algorithms.

Approach Taxonomy. Continual learning methods can be grouped into five interacting families: *regularization-based methods* that constrain important parameters using measures such as the Fisher Information Matrix (e.g., EWC), *replay-based methods* that approximate past data distributions via stored or generated samples, *optimization-based methods* that constrain gradient updates to avoid increasing loss on previous tasks (e.g., GEM), *representation-based methods* that learn stable and transferable features through robust representations and flat loss landscapes, and

architecture-based methods that modify model structure over time via parameter isolation, expansion, or routing.

Evaluation Metrics. Continual learning performance is evaluated using an accuracy matrix measuring performance on each task after learning subsequent tasks, from which metrics capturing overall performance (Average Accuracy), memory stability (Forgetting Measure, Backward Transfer), and learning plasticity (Intransigence, Forward Transfer) are derived, with no single metric fully characterizing behavior.

Key Takeaway. Continual learning is fundamentally about controlling learning trajectories under non-stationary data distributions, and effective solutions typically combine replay, regularization, optimization control, representation learning, and architectural design to enable knowledge accumulation without destructive interference.