Continuous (Machine) Learning - A Comprehensive Overview

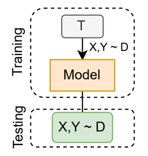
by Anaïs Lavorel, Nadia Kabachi and Elhadj Benkhelifa

Continuous (Machine) Learning for Artificial Neural Network

1. Knowledge Acquisition

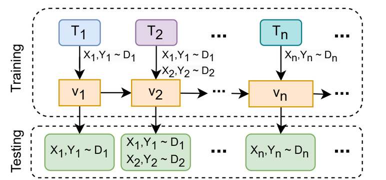
1.A. Data Availability

1.A.I. Offline Learning



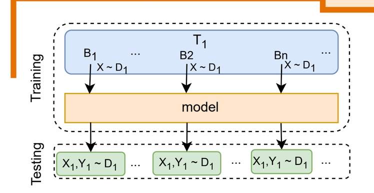
Data were fully collected before training occurred [1].

1.A.II. Continual Learning (CL)



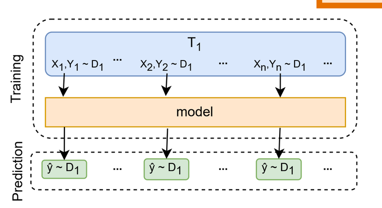
Datasets are acquired over time to encompass new tasks, classes, and domains, prompting model expansion for their support. **Continual learning enables the acquisition of new knowledge while mitigating critical forgetting of prior tasks** [1,2].

1.A.III. Online Learning (OL)



Data are received incrementally, enabling the **models to adjust swiftly and effectively manage data drift** [1].

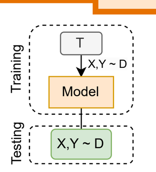
1.A.IV. Stream Learning



Data are **ingested and analyzed sequentially, without any storage**. Models rely on unsupervised learning, generally employing simplified techniques to support analysis and adapt to data drift with reduced delais [3].

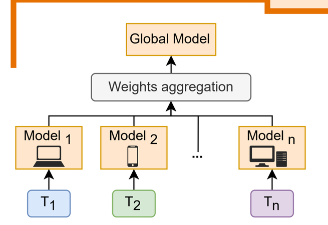
1.B. Ressources Accessibility

1.B.I. Centralized Learning



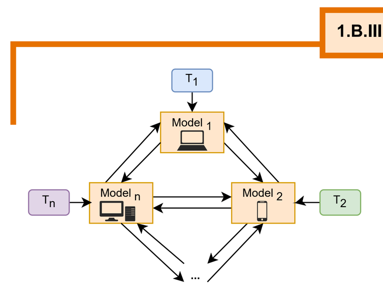
Once all data has been aggregated in a central location, **training is conducted on an individual device** [1].

1.B.II. Federated Learning



Replicas of the model are dispatched to each designated "worker" device to undergo **training on a segmented portion of the dataset,** thereby accelerating the training process before the **aggregation of weights into a unified global model** [4].

1.B.III. Distributed Learning



Each **participant device receives a replica of the model** for localized data training. Upon achieving model stability, participants contribute their respective weights to collaboratively refine the overall solution [4].

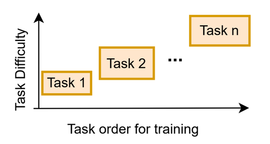
1.C. Learning Mechanism

1.C.I. Hand-Craft

The **architecture** of the Artificial Neural Network is **established prior to the training**, with the intention of modifying only the weights durring the experiment [1].

1.C.II. Curriculum Learning

The efficiency of learning was improved by **refining the sequence in which tasks were presented** [5].

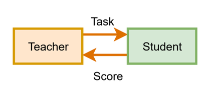


1.C.II.a. Hand-crafted Curriculum

An expert chose the order in which the task will be encountered [5].

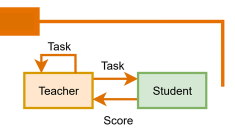
1.C.II.b. Teacher-Student based

The student model **improved its learning by receiving tasks selected by the teacher model** [5].



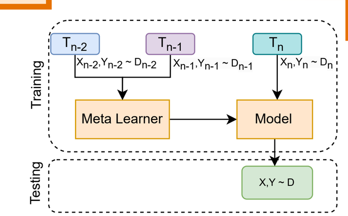
1.C.II.c. Self-Play

The Teacher-Student paradigm was re-examined through the implementation of **competitive tasks.** Within this structure, the 'Teacher' is assigned progressively complex tasks that align with the student's learning progression [5].



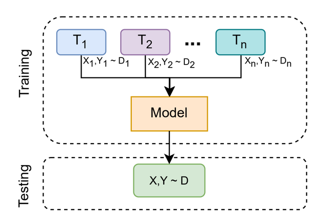
1.C.III. Meta Learning

An AutoML approach that **leverages dataset metadata through iterative learning to enhance performance**, employing metric-based, model-based, and optimization-based strategies [1].



1.C.IV. Multi-Task Learning

Models trained concurrently on a variety of discrete tasks [1].

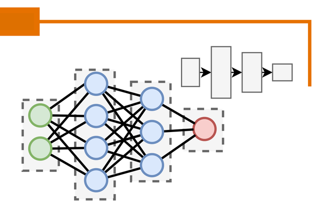


1.C.V. Modular Learning

Sub-units are defined in the model's architecture [6].

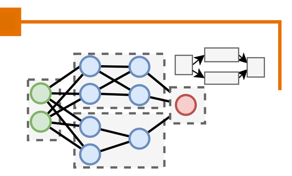
1.C.V.a. Layer Modular Network

It mirrors a conventional Artificial Neural Network, with its **neurons organized into distinct layers** *[6]*.



1.C.V.b. Path Modular Network

Within the Artificial Neural Network subcategory, **specific pathways are allocated for individual tasks** *[6]*.



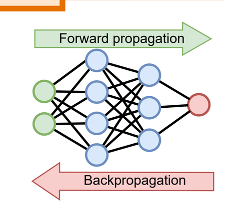
1.C.VI. NeuroEvolution

The model is developed during training, abstaining from layer dependencies, and replicates the principles of **genetic evolution** observed in a population [7].

1.D. Learning Assessment

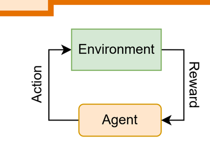
1.D.I. Backpropagation

Weights in the Artificial Neural Network are adjusted based on **δ(y_pred,y)**, which represents the difference between the predicted output (y_pred) and the target value (y). This adjustment is achieved by **propagating the error backward through the network** [7].



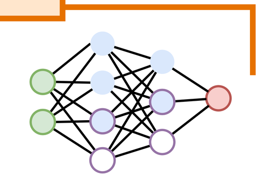
1.D.II. Reinforcement Learning

Specify the goals and the **reward** mechanism that require optimization. The model will **evaluate its architecture using the received reward** signal rather than the δ(y_pred,y) [7].



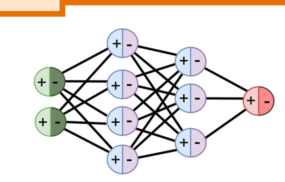
1.D.III. Hebbian Learning

Review the weights using the principle that neurons **firing together wire together**. ΔWij= n×xi×yi, where *xi* is the activation of the pre-synaptic neuron, *yi* is one of the post-synaptic neurons, and *n* is the learning rate [7].



1.D.IV. Predictive coding

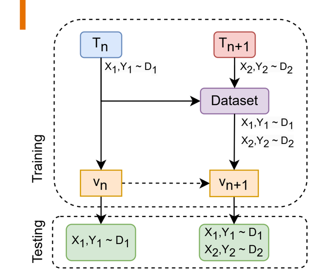
To emulate the brain's unsupervised learning, the network consistently predicts subsequent inputs. This is accomplished by **employing positive and negative neurons to simulate simultaneous forward and backward activations** [8].



2. Knowledge Retention

2.A. Replay-based

Expand the training dataset with stored / generated examples from prior task to maintain prior knowledge retention [1, 9].



2.A.I. Real Data Replay

Retain essential data for reincorporation into the training dataset [1,9].

2.A.II. Real Feature Replay

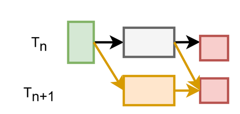
Utilize a feature extraction model to **retain and replay salient features from prior tasks,** analogous to data replay mechanisms [1,9].

2.A.III. Generative Replay

A generative model is employed to create synthetic data, which is then used to augment the training dataset for a separate model, expanding its learning resources [1,9].

2.B. Lock-based

2.B.I. Lock & Expand

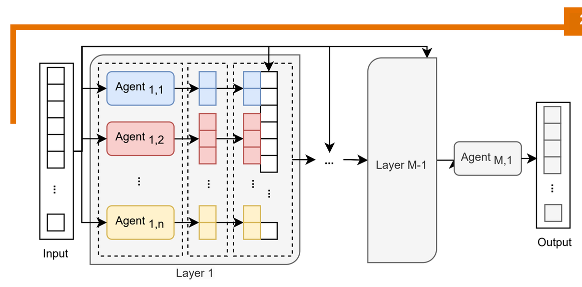


To accommodate a novel task Tn+1, the model introduces **a new pathway and output, thereby leveraging Path Modular Network behavior** to preserve prior knowledge while permitting access to the existing subnetwork to minimize redundancy [1,9].

2.B.II. Multi-Agent based

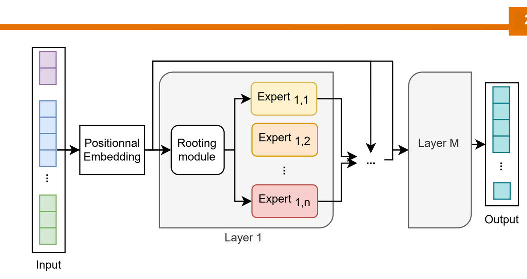
The model architecture contains a **collection of distinct agents** [10].

2.B.II.a. Mixture of Agent (MOA)



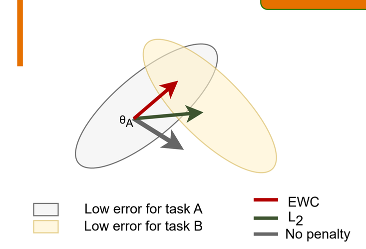
A model featuring a **collection of agents that work collaboratively** to accomplish each task [10].

2.B.II.b. Mixture of Expert (MOE)



The architecture of a model consists of a **routing mechanism and a set of expert models**, wherein the routing system directs specific experts to execute a given prediction [10].

2.C. Regularisation-based



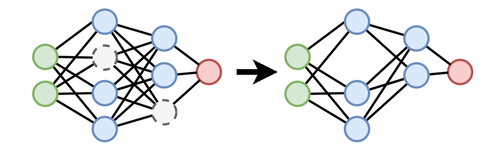
Inspired by neuroplasticity and neuromodulation, **synapse/neurone relevance is calculated for the previous task**. This allows freezing or restricting their plasticity to retain knowledge [9].

3. Knowledge Transfert

3.A. Model Compression

3.A.I. Prunning

To enhance efficiency, elements within the structure exhibiting negligible influence on task performance, indicated by **near-zero neuronal or synaptic weights, are eliminated** [11].



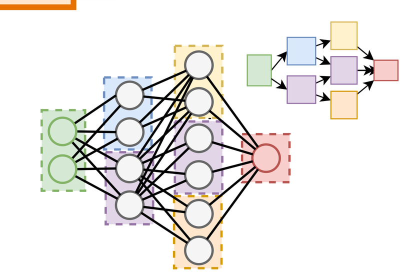
3.A.II. Quantisation

The model's **precision is diminished through the rounding of its numerical values** [11].

Ex: 3.75 → 4

3.A.III. Modular Network based

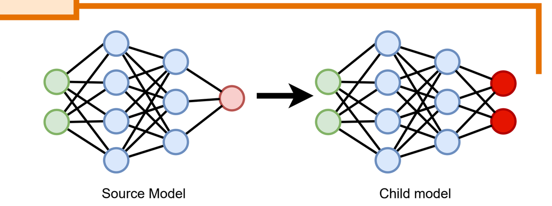
Rely on Neural Modular Network and Principal Component Analysis to **merge the redundant components together** [6].



3.B. Transfert Learning

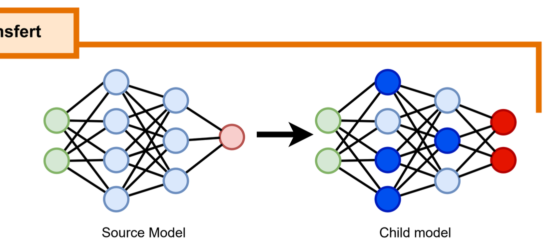
3.B.I. Full Feature Extraction

The revised model **retains the complete input/hidden layer configuration from the prior trained solution**, primarily serving scenarios requiring comparable tasks or learning [11].



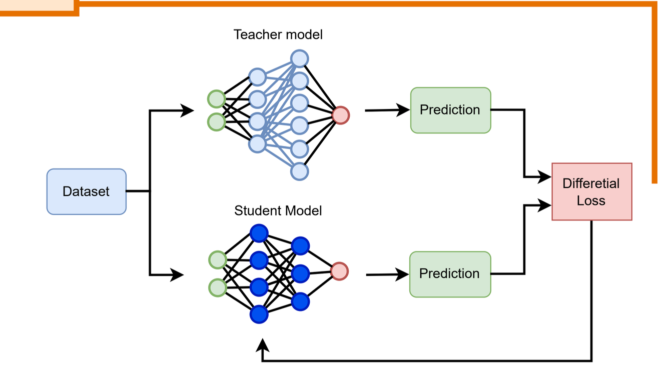
3.B.II. Partial Feature Transfert

Resuing fragments of the previously trained model by maintaining the same weights [11].



3.B.III. Knowledge Distillation

The trained model is used as a teacher to teach a new model [11].



Summary:

Continuous Machine Learning can be viewed as a sub-domain of Machine Learning focused on enabling continuous learning behaviors, drawing inspiration from biological ontogenesis and phylogenesis. Ontogenesis Learning encompasses the learning capabilities exhibited during the lifetime of a single individual, as observed for instance in Continual Machine Learning [1-4,7,9], Online Learning [1,3], and Curriculum Learning [5]. In contrast, Phylogenesis Learning, specifically cover learning related to a group of individuals, it incorporates legacies from ancestors and interactions between individuals/agents. This is emulated by Evolutionary Algorithms, such as NeuroEvolution [7], and can be view to some extend as an inspiration for multi-agent systems [10].

While numerous reviews delineate specific topics, they often lack a comprehensive overview regarding the co-existence, interconnections, and distinct characteristics of these domains. In this overview, a broad definition of Continuous Machine Learning is established based on an the literature review, as illustrated on the right. This framework facilitates the distinction of three primary axes: Knowledge Acquisition [1-8], Knowledge Retention [1,9,10], and Knowledge Transfert [6,11]. These axes collectively form the aforementioned taxonomic representation, which serves to elucidate their key features and enhance the understanding of how Continuous Machine Learning operates:

- Knowledge Acquisition encompasses the mechanisms by which a model acquires knowledge, considering factors such as data availability, resource accessibility, learning mechanisms, and knowledge assessment methodologies. This axis delineates the distinctions among Continual Learning [1-4,7,9], Online Learning [1,3], and Curriculum Learning [5], while also highlighting the reliance of many continuous models on established methods such as Federated Learning [4], Distributed Learning [4], and Incremental Learning [2].
- Knowledge Retention, however, is predominantly addressed within Continual Learning [1-4,7,9], enabling a model to expand its capabilities and support new tasks, classes, and domains over time, while simultaneously mitigating catastrophic forgetting [1]. Nevertheless, recent approaches leveraging multi-agent systems are beginning to emerge, to decrease update times and reduce hallucination risks, thereby supporting more adaptive continuous solutions that permit on-demand updates of specific components [10].
- Knowledge Transmission, lastly, relies on Compression and Transfer Learning techniques employed in domains such as TinyML and Continual Learning to diminish model size and complexity [6,11], thereby mitigating model growth over time before subsequent expansion to accommodate new tasks, classes, or domains.

Continuous (Machine) Learning

1. Inspirations
- A. Ontogenesis
- B. Phylogenesis

2. Caracteristics
- A) Knowledge Acquisition
  - i. Task Learning
  - ii. Task Agnostic Learning
  - iii. Exploit Task Similarity
- B) Knowledge Retention
  - i. Overcoming Catastrophic Forgetting
- C) Knowledge Transfert
  - i. Graceful Forgetting
  - ii. Knowledge Generalisation
  - iii. Knowledge Specialization

Sources :

[1] Y. Li *et al.*, "Unleashing the Power of Continual Learning on Non-Centralized Devices: A Survey," *IEEE Communications Surveys & Tutorials*, 2025, pp. 1-1, doi: 10.1109/COMST.2025.3568279 (Li et al., 2025).

[2] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, and T. Tuytelaars, "A Continual Learning Survey: Defying Forgetting in Classification Tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3366-3385, 2022, doi: 10.1109/TPAMI.2021.3057446 (Delange et al., 2021).

[3] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias, "Three types of incremental learning," *Nature Machine Intelligence*, vol. 4, no. 12, pp. 1185-1197, 2022, doi: 10.1038/s42256-022-00568-3 (Ven et al., 2022).

[4] A. Cossu *et al.*, "Don't Drift Away: Advances and Applications of Streaming and Continual Learning," in *ESANN*, 2025, pp. 23-32.

[5] S. Narvekar, B. Peng, M. Leonetti, J. Sinapov, M. E. Taylor, and P. Stone, "Curriculum Learning for Reinforcement Learning Domains: A Framework and Survey," *Journal of Machine Learning Research*, vol. 21, no. 181, pp. 1-50, 2020.

[6] M. Amer and T. Maul, "A review of modularization techniques in artificial neural networks," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 527-561, 2019, doi: 10.1007/s10462-019-09706-7 (Amer & Maul, 2019).

[7] S. Schmidgall *et al.*, "Brain-inspired learning in artificial neural networks: A review," *APL Machine Learning*, vol. 2, no. 2, p. 021501, 2024, doi: 10.1063/5.0186054 (Schmidgall et al., 2024).

[8] B. Millidge *et al.*, "Predictive Coding: Towards a Future of Deep Learning beyond Backpropagation?," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, 2022, pp. 5538-5545, doi: 10.24963/ijcai.2022/774 (Millidge et al., 2022).

[9] L. Wang *et al.*, "A Comprehensive Survey of Continual Learning: Theory, Method and Application," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 8, pp. 5362-5383, 2024, doi: 10.1109/TPAMI.2024.3367329 (Wang et al., 2024).

[10] W. Jin *et al.*, "A Review of AI-Driven Automation Technologies: Latest Taxonomies, Existing Challenges, and Future Prospects," *Computers, Materials & Continua*, vol. 84, no. 3, pp. 3961-4018, 2025, doi: 10.32604/cmc.2025.067857.

[11] P. V. Dantas *et al.*, "A comprehensive review of model compression techniques in machine learning," *Applied Intelligence*, vol. 54, no. 22, pp. 11804-11844, 2024, doi: 10.1007/s10489-024-05747-w (Dantas et al., 2024).

Contact : anais.lavorel@etu.univ-lyon1.fr