Continuous (Machine) Learning - A Comprehensive Review

by Anaïs Lavorel

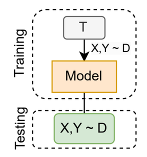
La Région Auvergne-Rhône-Alpes

Continuous (Machine) Learning for Artificial Neural Network

1. Knowledge Acquisition

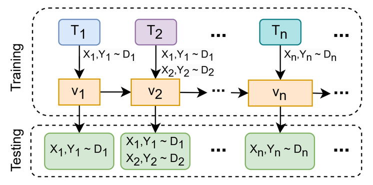
1.A. Data Availability

1.A.I. Offline Learning



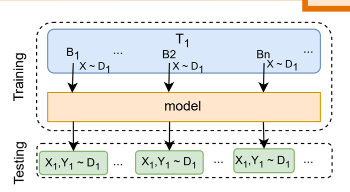
Data were fully collected before training occurred.

1.A.II. Continual Learning (CL)



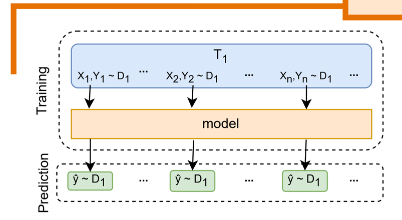
Datasets are acquired over time to encompass new tasks, classes, and domains, prompting model expansion for their support. **Continual learning enables the acquisition of new knowledge while mitigating critical forgetting of prior tasks.**

1.A.III. Online Learning (OL)



Data are receive incrementally, enabling the **models to adjust swiftly and effectively manage data drift.**

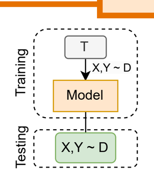
1.A.IV. Stream Learning



Data is **ingested and analyzed sequentially, without any storage**. Models rely on unsupervised learning, generally employing simplified techniques to support analysis.

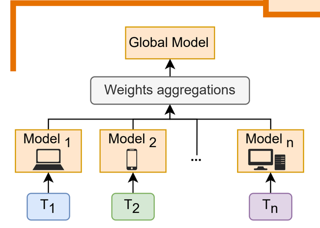
1.B. Ressources Accessibility

1.B.I. Centralized Learning



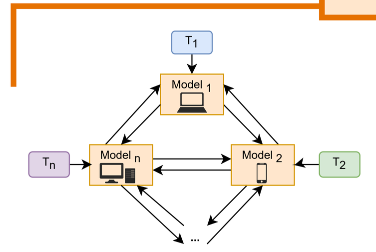
Once all data has been aggregated in a central location, **training is conducted on an individual device.**

1.B.II. Federated Learning



A replica of the model is 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**.

1.B.II. Federated 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.

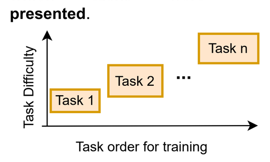
1.C. Learning Mechanism

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.C.II. Curriculum Learning

The efficacy of learning was augmented by **refining the sequence in which tasks were presented**.

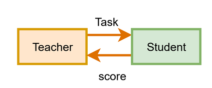


1.C.I.a. Hand-crafted Curriculum

An expert chose the order in which the task will be encountered

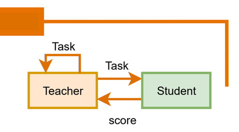
1.C.I.b. Teacher-Student based

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



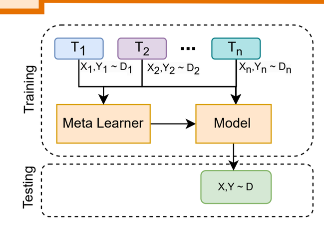
1.C.I.c. Self-Play

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.



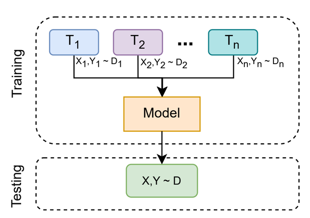
1.C.III. Meta Learning

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



1.C.IV. Multi-Task Learning

Models trained concurrently on a variety of discrete tasks.

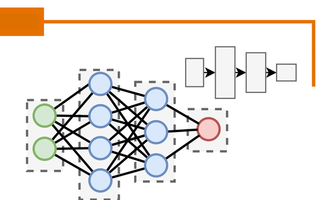


1.C.V. Modular Learning

Modularly constructed models

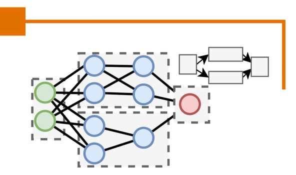
1.C.V.a. Layer Modular Network

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



1.C.V.b. Path Modular Network

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



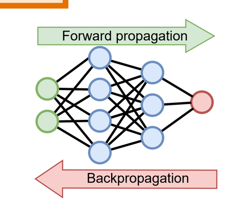
1.C.VI. NeuroEvolution

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

1.D. Learning Assessment

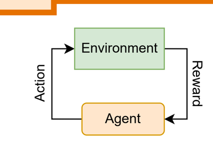
1.D.I. Backpropagation

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.**



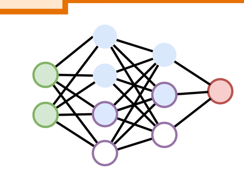
1.D.II. Reinforcement Learning

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



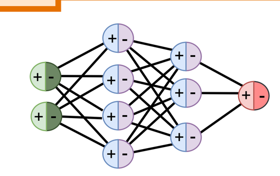
1.D.III. Hebbian Learning

Review the weight using the principle that neurons **that work together fire together**.∆Wij= n× xi× yi, where xiis the activation of the presynaptic neurone,yiis one of the post-synaptic neurones, and n is the learning rate.



1.D.IV. Predictive Learning

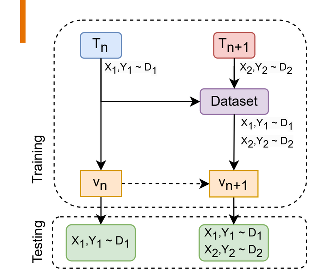
To emulate the brain's unsupervised learning, the network consistently predicts subsequent inputs. This is accomplished by **employing positive and negative neurons to simulate simultaneous forward and backward activations.**



2. Knowledge Retention

2.A. Replay based

Expand the training dataset to maintain prior knowledge retentions



3.A.I. Real Data Replay

Store key data to reinject them in the training dataset.

3.A.II. Real Feature Replay

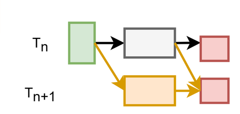
Rely on an feature extracting model to store feature of previous task that can be replay. Similar to Data replay

3.A.III. Generative Replay

A model trained to generate data is used to **generate on demande data to trained a model**. Thus expanding training dataset

2.B. Lock based

3.B.I. Lock & Expand

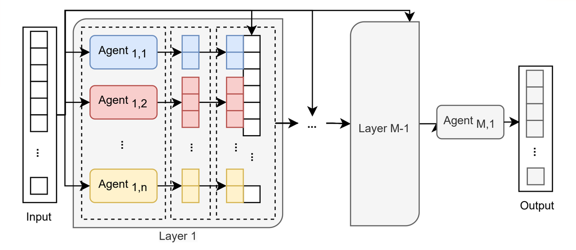


To support a new task T_{n+1} the model add a new path+poutput and only allow the new path to access previous learn information. Thus exploiting the Path Modular Network behavior

3.B.II. Multi-Agent based

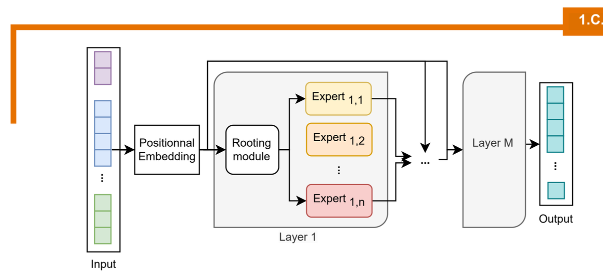
Rely on multiple agent handling each specific tasks

3.B.II.A. Mixture of Agent (MOA)



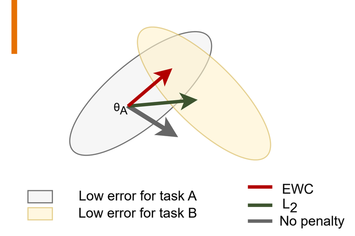
A Models that contains a set of **agent that collaborate to performed each task**

1.C.v.c. Mixture of Expert (MOE)



A model is composed of module which correspond to expert model. **Relying on a rooting system, different expert are used to performed a prediction.**

2.C. Regularisation based



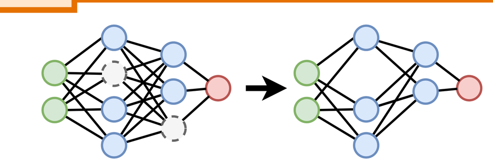
Inspired from the Neuroplasticity and Neuromodulation, the **pertinence for each synapse/neurone tuple are compute** for the prior task, **allowing to freeze them or restrain their plasticity** to ensure knowledge retention

3. Knowledge Transfert

3.A. Model Compression

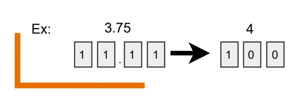
3.A.I. Prunning

Delete part of the structure(neurons/synapses) that do not play a major role for any task. Those neurones/synapse have weights close to 0



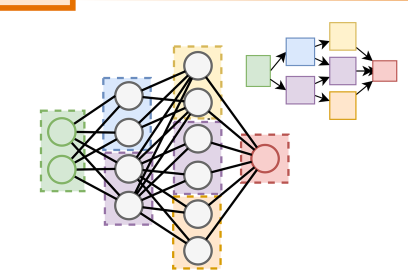
3.A.II. Quantisation

Reduce the precision of the model byrounding vlaues (ex: 2.75 -> 3)



3.A.III. Modular Network based

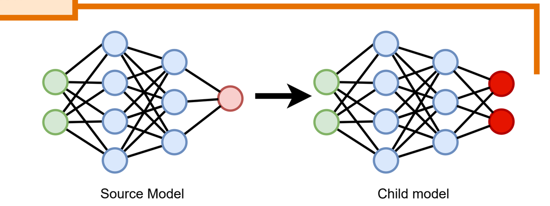
Rely on Neural Modular Network principal and Principal Component Analysis method to merge redundant component of the structure toggether



3.B. Transfert Learning

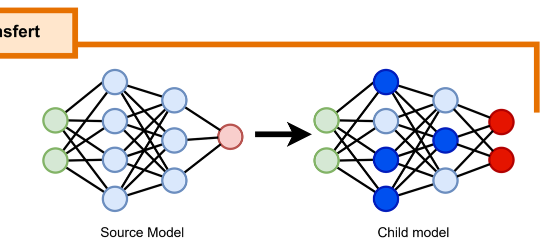
3.B.I. Full Feature Extraction

The new model reuse the full input/hidden layer strructure fro the previous trained solution. Mostly used when similar task / similar learning is required



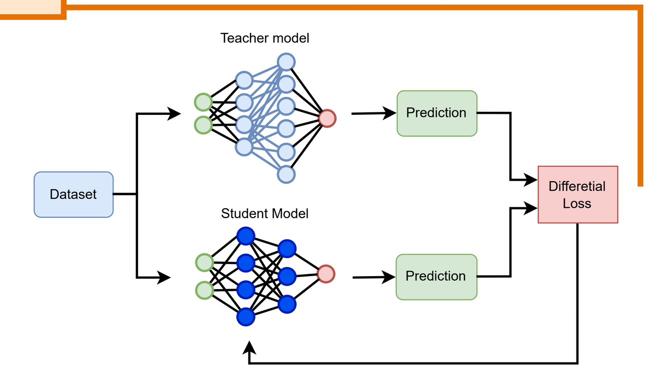
3.B.II. Partial Feature Transfert

Reuse only fragment of previous trained model by reusing the same weights



3.B.III. Knowledge Distillation

The trained model is used as a teacher to teach a new model



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. Gracefull Forgetting
  - ii. Knowledge Generalisation
  - iii. Knowledge Specialization

Contact secrétariat : 5 avenue Pierre Mendès-France - 69676 Bron Cedex
Téléphone : +33 (0)4 78 77 31 54 - Télécopie : +33 (0)4 78 77 23 75
Courriel : secretariat@eric.univ-lyon2.fr
http://eric.univ-lyon2.fr