SCUOLA NORMALE SUPERIORE

UNIVERSITÀ DI PISA

Continual Learning with Echo State Networks

are random recurrent networks suitable for continual learning?

Andrea Cossu, Davide Bacciu, Antonio Carta, Claudio Gallicchio, Vincenzo Lomonaco
Scuola Normale Superiore, University of Pisa

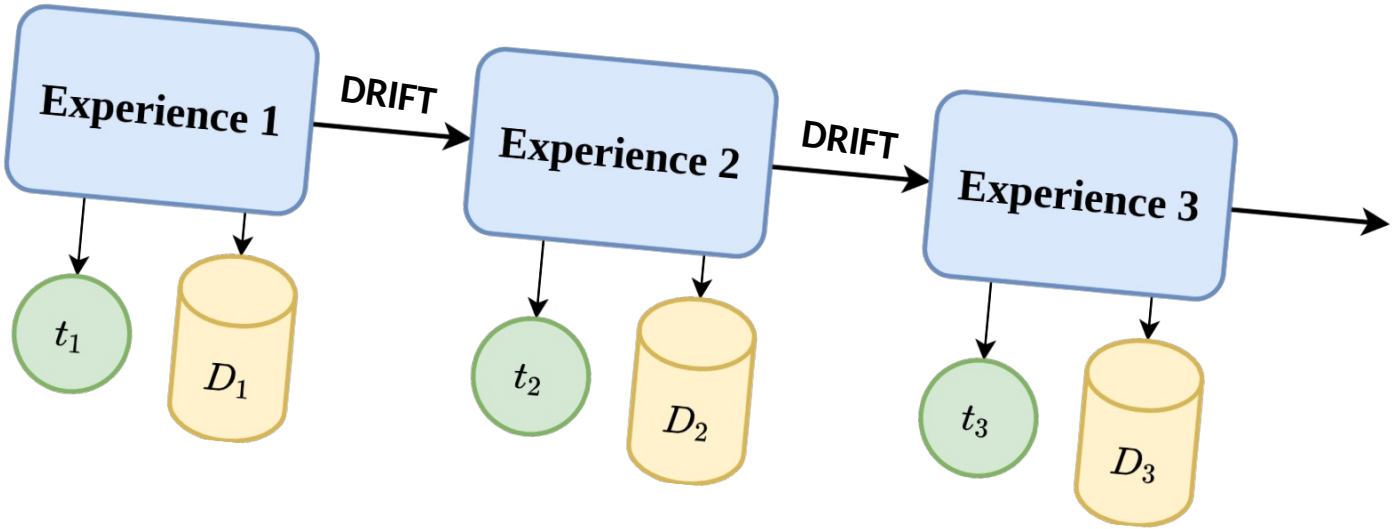
ESANN 2021

SCUOLA NORMALE SUPERIORE

UNIVERSITÀ DI PISA

Learning continuously without forgetting

an open challenge



Quickly learn new experiences

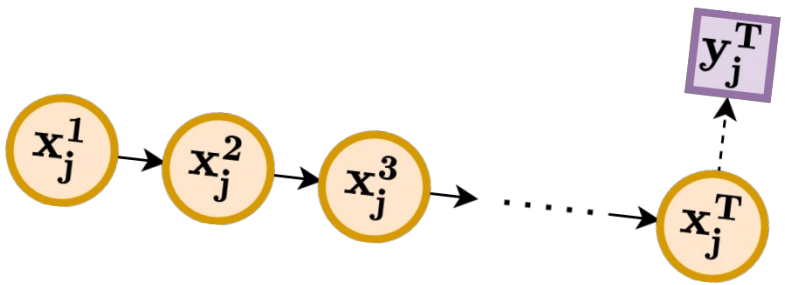
Exploit existing knowledge

Mitigate catastrophic forgetting

…

Recurrent models in continual learning

- Everything is perceived as a sequence… more or less!



- *What is the impact of existing CL strategies on recurrent models?*

A. Cossu, A. Carta, V. Lomonaco, D. Bacciu, *Continual Learning for Recurrent Neural Networks: An Empirical Evaluation,* Neural Networks, 2021.

Random recurrent networks for CL

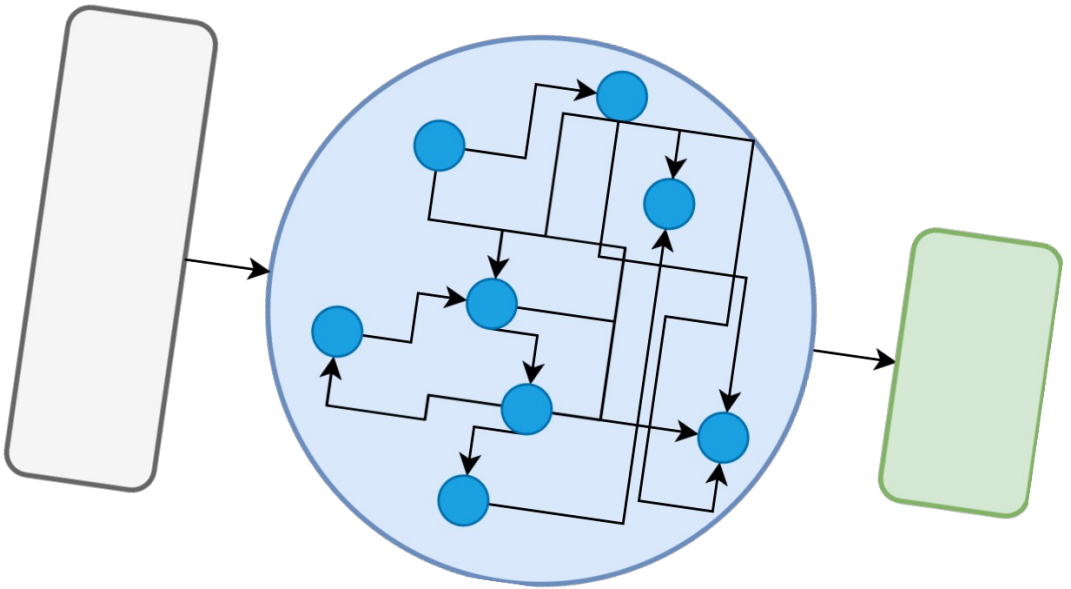

You can’t forget if you are not changing


No recurrent parameters to learn


Apply CL strategies on trainable parameters only


Treat the untrained components as pre-trained network

Echo State Network



$$\mathbf{x_{t+1}} = \sigma(\mathbf{W}\mathbf{x_t} + \mathbf{W}^{\text{in}}\mathbf{u_{t+1}})$$

$$\mathbf{y_{t+1}} = \mathbf{W}^{\text{out}}\mathbf{x_{t+1}}$$

SCUOLA NORMALE SUPERIORE

UNIVERSITÀ DI PISA

Experiments setting

- Avalanche library
- Split row-MNIST
  - 5 experiences, 2 classes each
- Synthetic Speech Commands
  - 10 experiences, 2 classes each
- EWC, LwF, Replay, Streaming Deep LDA, Naive and Joint Training
- ESN and LSTM

https://github.com/ContinualAI/avalanche

Results

SMNIST	LSTM†	ESN
EWC	$0.21_{\pm 0.02}$	$0.20_{\pm 0.00}$
LWF	$0.31_{\pm 0.07}$	$0.47_{\pm 0.07}$
REPLAY	$\mathbf{0.85_{\pm 0.03}}$	$0.74_{\pm 0.03}$
SLDA	—	$\mathbf{0.88_{\pm 0.01}}$
NAIVE	$0.20_{\pm 0.00}$	$0.20_{\pm 0.00}$
JOINT	$0.97_{\pm 0.00}$	$0.97_{\pm 0.01}$

SSC	LSTM†	ESN
EWC	$0.10_{\pm 0.00}$	$0.09_{\pm 0.02}$
LWF	$0.12_{\pm 0.01}$	$0.12_{\pm 0.02}$
REPLAY	$\mathbf{0.74_{\pm 0.07}}$	$0.36_{\pm 0.07}$
SLDA	—	$\mathbf{0.57_{\pm 0.03}}$
NAIVE	$0.10_{\pm 0.00}$	$0.10_{\pm 0.00}$
JOINT	$0.89_{\pm 0.02}$	$0.91_{\pm 0.02}$

A promising future

- Expand the analysis
  - reservoir topologies
  - unsupervised finetuning
- Continual learning on low-powered devices
  - neuromorphic hardware
- Ad-hoc strategies
  - exploit readout linearity

SCUOLA NORMALE SUPERIORE

UNIVERSITÀ DI PISA

Questions?