New Metrics and Experimental Paradigms for Continual Learning

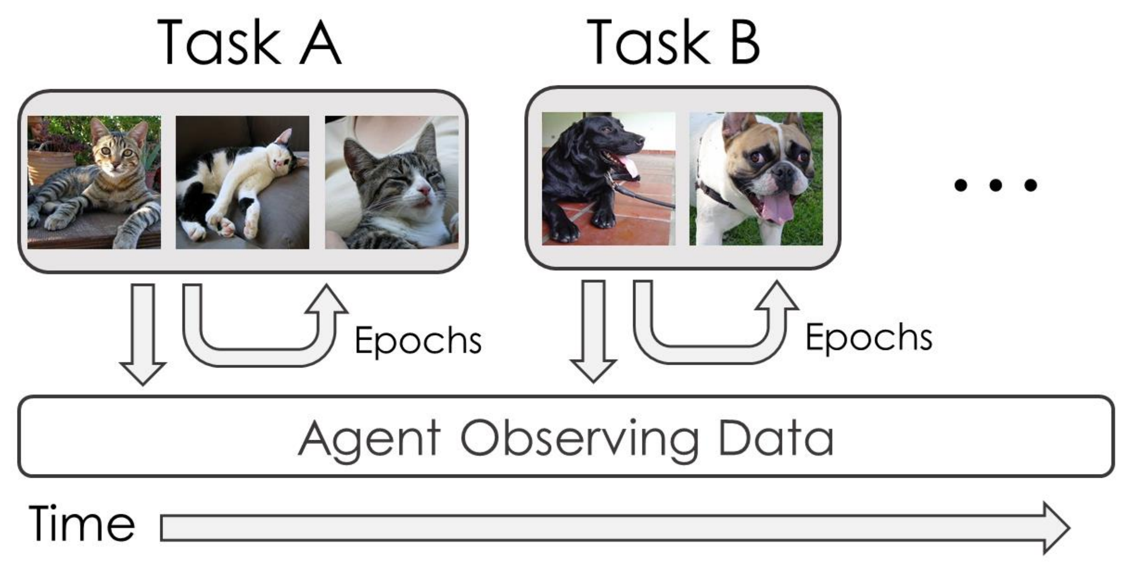
Tyler L. Hayes, Ronald Kemker, Nathan D. Cahill, and Christopher Kanan
Rochester Institute of Technology, Rochester NY
{tlh6792, rmk6217, ndcsma, kanan}@rit.edu

Overview

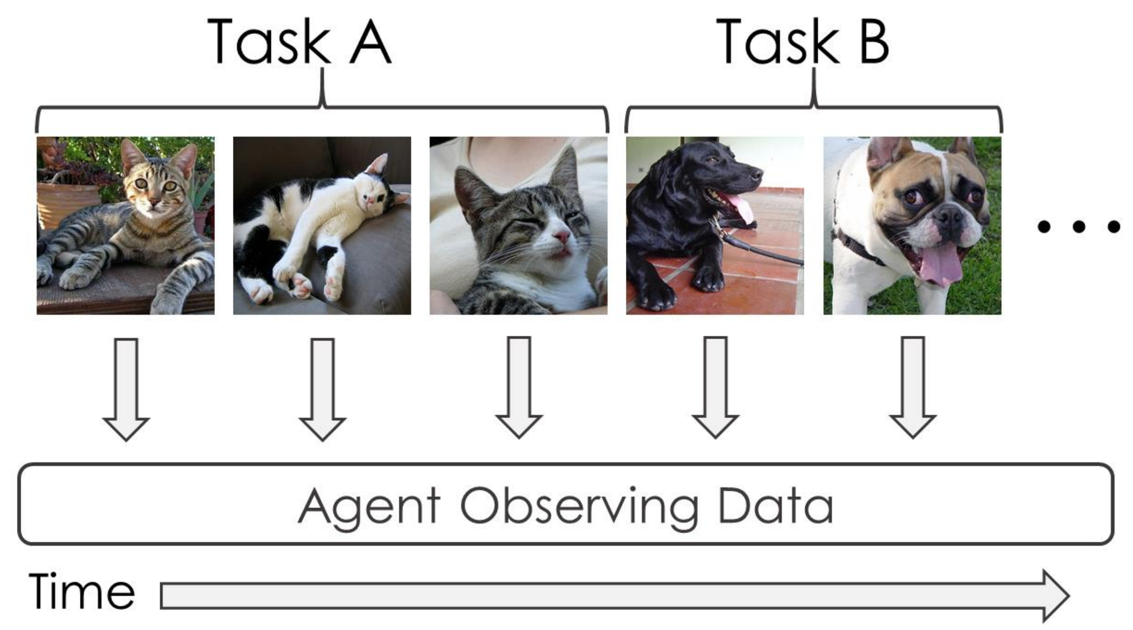
- In real-world environments, agents must quickly alter their behavior to **learn and adapt in real-time**.
- Deep Neural Networks are the dominant approach for machine perception, but they cannot learn new instances immediately and learning requires multiple loops over a dataset. They are also susceptible to **catastrophic forgetting** when streams of instances are not independent and identically distributed (iid).
- In **continual learning**, an algorithm must be able to immediately make inferences from new examples and must have the ability to learn from non-iid data streams.

Problem Formulation

Incremental Batch Learning



Continual Learning



Motivation

- Assess the influence of data ordering
- Easily compare methods **quantitatively**
- Flexible metrics for non-uniform testing events

Experimental Paradigms and Metrics

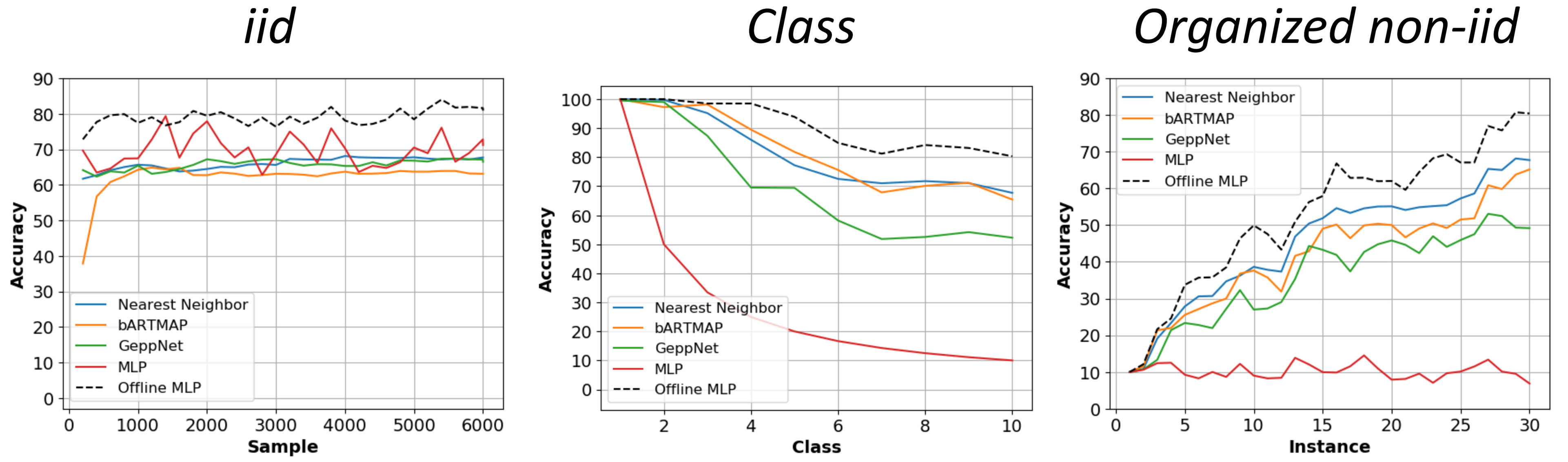
Data	iid	Class	Organized non-iid
Data Stream Organization	Completely unordered	Ordered by class	Temporally ordered by instances
Tests Learner's Ability	To learn quickly	To learn new classes incrementally	To learn classes/objects and then revisit them much later
Accuracy Computed at Regular Intervals on	All test data	Test data belonging to all previously observed classes	All test data
Notes	Easiest for continual learner to rival offline learner	Popular in incremental batch learning literature [1]	Closely matches how a robot would experience stimuli

- Evaluating a continual learner means evaluating its ability to learn quickly from non-iid data streams and measuring the learner's memory usage.
- Applying the metric from [2] , overall performance of a continual learner is given by

$$\Omega_{\text{all}} = \frac{1}{T}\sum_{t=1}^{T} \frac{\alpha_{\text{all},t}}{\alpha_{\text{offline},t}},$$

where $\alpha_{\text{all},t}$ is the accuracy on all test data seen at time t, $\alpha_{\text{offline},t}$ is the accuracy of an optimized offline model on all test data seen at time t, and T is the number of testing events.

Baseline Experiments



- No algorithm reaches the performance of the ideal learner, even on this easy dataset, demonstrating the difficulty the continual learning problem poses for existing models.

Method	iid	Class	Organized
Online MLP	**0.881**	0.308	0.255
1NN	0.836	0.894	**0.863**
bARTMAP [6]	0.787	**0.898**	0.800
GeppNet [7]	0.832	0.757	0.694
Offline (Ideal)	1.000	1.000	1.000

Table: Ω_{all} metrics computed on iCub World-1.

Conclusions and Open Questions

- To make continual learning agents more robust, we must create **larger datasets** with **more diversity**, e.g., face, scene, activity recognition.
- An interesting idea would be to create a dataset with classes in the test set that are not in the training set and require a model to account for this.
- Develop metrics that account for both performance and **memory usage**.
- Develop agents capable of continual learning.
- Overcoming the constraints of continual learning would allow agents to learn from non-iid, temporally organized data streams, adapt to changes over time, and have improved computational and memory efficiency.

Acknowledgements:

Thanks to the Naval Research Laboratory for financial support for this research.