Confirm Smart Manufacturing

Imbal-OL: Online Machine Learning from Imbalanced Data Streams in Real-world IoT

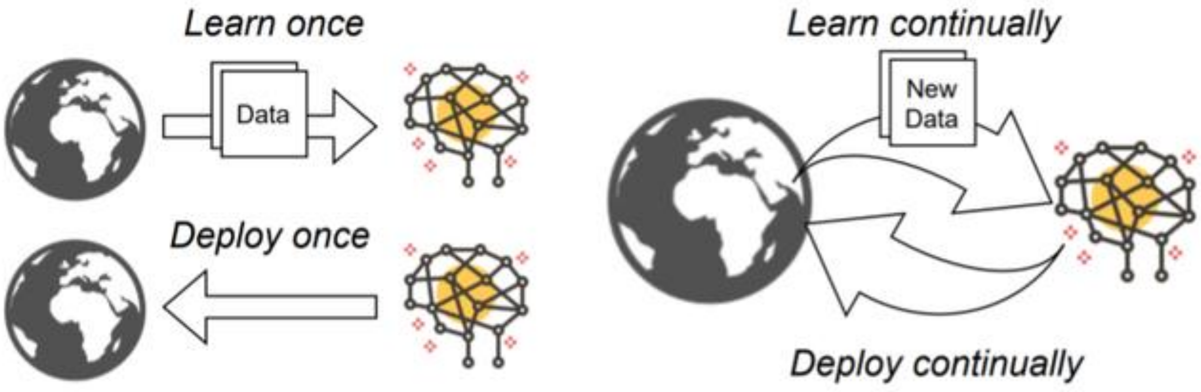
Bharath Sudharsan, John G. Breslin, and Muhammad Intizar Ali

A World Leading SFI Research Centre

Science Foundation Ireland For what's next

Online Machine Learning

Confirm Smart Manufacturing

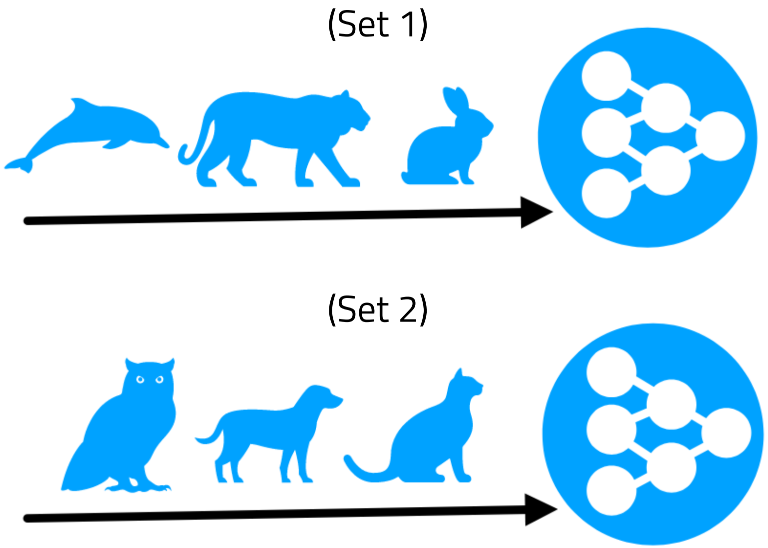


- Typically a Neural Networks (NN) is trained on data centers using historic datasets, then a C source file (model as a char array) of the trained model is generated and flashed on IoT devices

- This standard process impedes the flexibility of billions of deployed ML-powered devices as they cannot learn unseen data patterns (static intelligence) and are impossible to adapt to dynamic scenarios

- Online Machine Learning (OL) provide devices the ability to locally re-train themselves - continuously updating the last few NN layers using unseen data patterns encountered after deployment

Catastrophic Forgetting

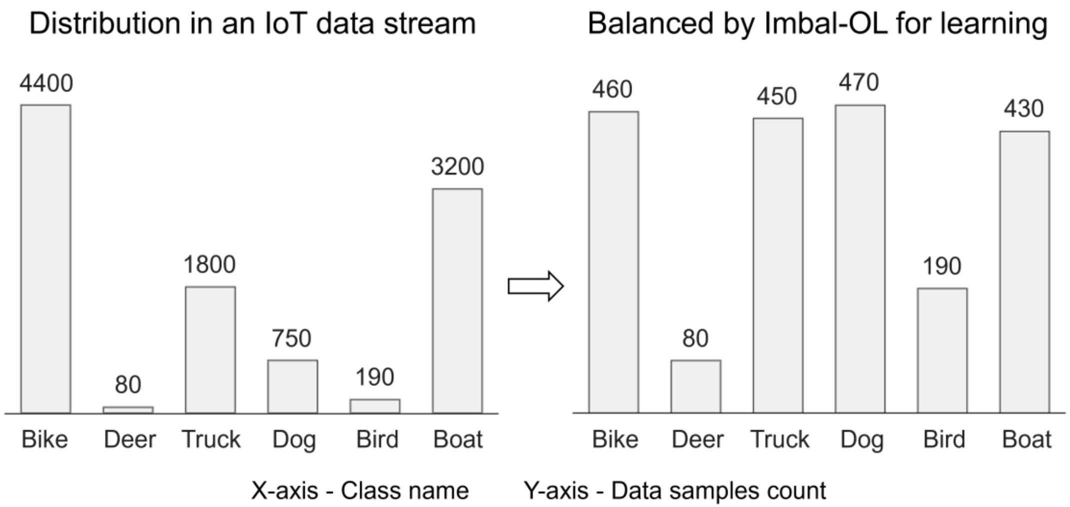


- NNs are known to catastrophically forget the information that are not frequently or recently seen. For example
  - We have a NN trained to recognize dolphins, rabbits and lions (Set 1)
  - We then train it with other classes such as birds, cats and dogs (Set 2)
  - After some time, the NN can start getting poor results in recognizing any class in Set 1
  - **Why?** Because the weights are biased to recognize the classes in Set 2

OL Meets IoT Data Streams

- In OL, catastrophic forgetting is common when NNs are trained using **non-stationary data distribution**. The majority of recent work in the OL domain embraces the implicit assumption that the distribution of local training data is balanced
- But the sensor data streams in real-world IoT are **severely imbalanced and temporally correlated**
- Need a lightweight technique to provide devices the ability to deal with imperfect data streams and manage to produce high-quality models even under **challenging learning settings**

Imbal-OL



- Imbal-OL is a resource-friendly technique that can be used as an **OL plugin** to balance the size of classes in a range of data streams. When Imbal-OL processed stream is used for OL, the models can adapt faster to changes in the stream while parallelly preventing catastrophic forgetting

- As shown, **without discarding** any data samples from the significantly underrepresented deer and bird classes, Imbal-OL **balances** the remaining classes

Imbal-OL Characteristics

Confirm Smart Manufacturing

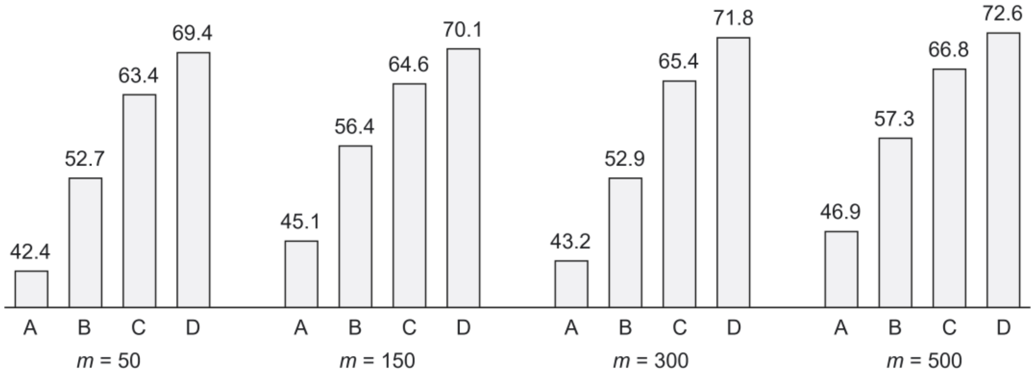
- **Imbal-OL is designed to show the following characteristics:**
  - **Underrepresented Classes.** Imbal-OL identifies and stores all data representing the minority classes and uses it for learning
  - **Sampling.** A subset of data samples from each class stored in device memory for learning is iid concerning the data samples from the stream
  - **Memory Utilization.** To avoid inefficient utilization of resources, Imbal-OL fills the entire memory, then balances the classes
  - **Weighted Replay.** Imbal-OL replays samples considering where they belong to rather than uniform replay where data samples from all classes have a similar probability for replay

Imbal-OL Evaluation

- **Datasets.** Standard CIFAR-10 and CIFAR-100 for image classification task
- **Stream.** Each learning task is assigned with classes from the dataset in an incremental fashion – to create distinct talks, inducing complexity during OL from the stream
- **IoT Boards.** Heterogenous setup - Google Coral Dev board, Intel Movidius NCS accelerated Raspberry Pi 4, and NVIDIA Jetson Nano
- **Models.** ResNet18 pre-trained on ImageNet used for OL over data stream
- **Imbalances Simulation.** A vector *r* is defined containing *k* retention factors - selection of *r* decides the level of imbalances. In this setup, the imbalance ratio between classes in the stream can range from 1:3 to 1:60

Results

Confirm
Smart Manufacturing



X-axis - Scheme, memory unit size Y-axis - Accuracy (%)

A - No-replay B - Random-replace C - GSS-greedy D - Imbal-OL

- **Memory unit size vs accuracy**. Models produced by training using Imbal-OL processed streams have
  - Higher performance than the baselines
  - Consistent performance across different *m* values
  - Higher improvements for smaller *m* - even devices with low memory can produce better quality models

Results

Confirm
Smart Manufacturing

TABLE I
ACCURACY (%) COMPARISON OF MODELS AFTER OL USING CIFAR-10.

Scheme	$m = 100$		$m = 500$	
	Uniform	Weighted	Uniform	Weighted
GSS-greedy	63.6	65.2	62.2	66.8
Imbal-OL	69.4	70.7	70.2	72.6

- **Uniform vs Weighted Replay**. Investigate the benefits of weighted-replay Imbal-OL characteristic
  - Performance gap between Imbal-OL and GSS-greedy is higher under uniform replay. This is because, for GSS-greedy, the on-device learned model is highly **influenced** by the imbalances in the stream, causing the model to show top classification performance only for a few classes. So, it resulted in reduced accuracy when evaluating the learned model using the balanced test set
  - Using weighted replay has higher benefits than uniform replay because it partially **masks** the effect of imbalanced streams by oversampling the minority classes

Results

Confirm Smart Manufacturing

TABLE II
TIME, MEMORY CONSUMED BY SCHEMES TO MAKE STREAMS OL READY.

Scheme	Time (sec)		Memory (%)	
	CIFAR-10	CIFAR-100	CIFAR-10	CIFAR-100
Random-replace	4.17	4.9	3.7	3.4
GSS-greedy	24.3	23.6	31.8	32.6
Imbal-OL	8.18	9.06	4.6	5.1

- **Time and Memory Consumption**. Report time elapsed and space used for processing streams
  - Random-replace and Imbal-OL consume roughly the same time. But GSS-greedy consumes a considerably higher amount as it performs additional computation to calculate a similarity score for incoming data samples, making **GSS-greedy non-suitable** for processing streams on tiny devices
  - Random-replace requires the least space, and GSS-greedy consumes the highest as it uses memory to store gradient vectors in 32-bit floating-point numbers. To speed up the storage process after sampling data from the stream, Imbal-OL consumes space to store the **memory address** of the classes - this results in using more memory than Random-replace

Conclusion

Confirm Smart Manufacturing

- This work investigated the catastrophic forgetting issue when practicing online machine learning on tiny devices using **severely imbalanced** data streams

- Imbal-OL was proposed as an **OL plugin** to process real-world IoT streams before feeding it to the learner that updates the local on-device model

- Future work plans to extend Imbal-OL to be applicable in **federated learning** settings where OL needs to be performed using imbalanced and also **incomplete** data streams

Confirm
Smart Manufacturing

Contact: Bharath Sudharsan
Email: bharath.sudharsan@insight-centre.org

www.confirm.ie

Confirm
Smart Manufacturing

Science Foundation Ireland
For what's next