REDUCR: Robust Data Downsampling using Class Priority Reweighting

William Bankes, George Hughes, Ilija Bogunovic* and Zi Wang*

*Co-senior authors

Overview

Goal: Train a model on a large streamed dataset.

Problem: Downsample the data to a manageable size whilst maintaining model performance.

Challenge: Downsampling datasets with class imbalance can lead to poor worst-class generalisation. Excerbated by distributional shift.

Solution: REDUCR downsamples data in a robust manner to improve model performance under class imbalance and distributional shift.

Formal Problem Setting

- Vanilla data downsampling:

$$D_T = \arg\max_{D \subset \mathcal{D}} \underbrace{\log p(y_{ho} | x_{ho}, D)}_{\text{Likelihood of the holdout dataset}}$$

- Robust data downsampling:

$$D_T = \arg\max_{D \subset \mathcal{D}} \underbrace{\min_{c \in C} \log p(y_{ho}^{(c)} | x_{ho}^{(c)}, D)}_{\text{Likelihood of the worst-class in the holdout dataset}}.$$

- This problem is NP hard

Our Solution: REDUCR

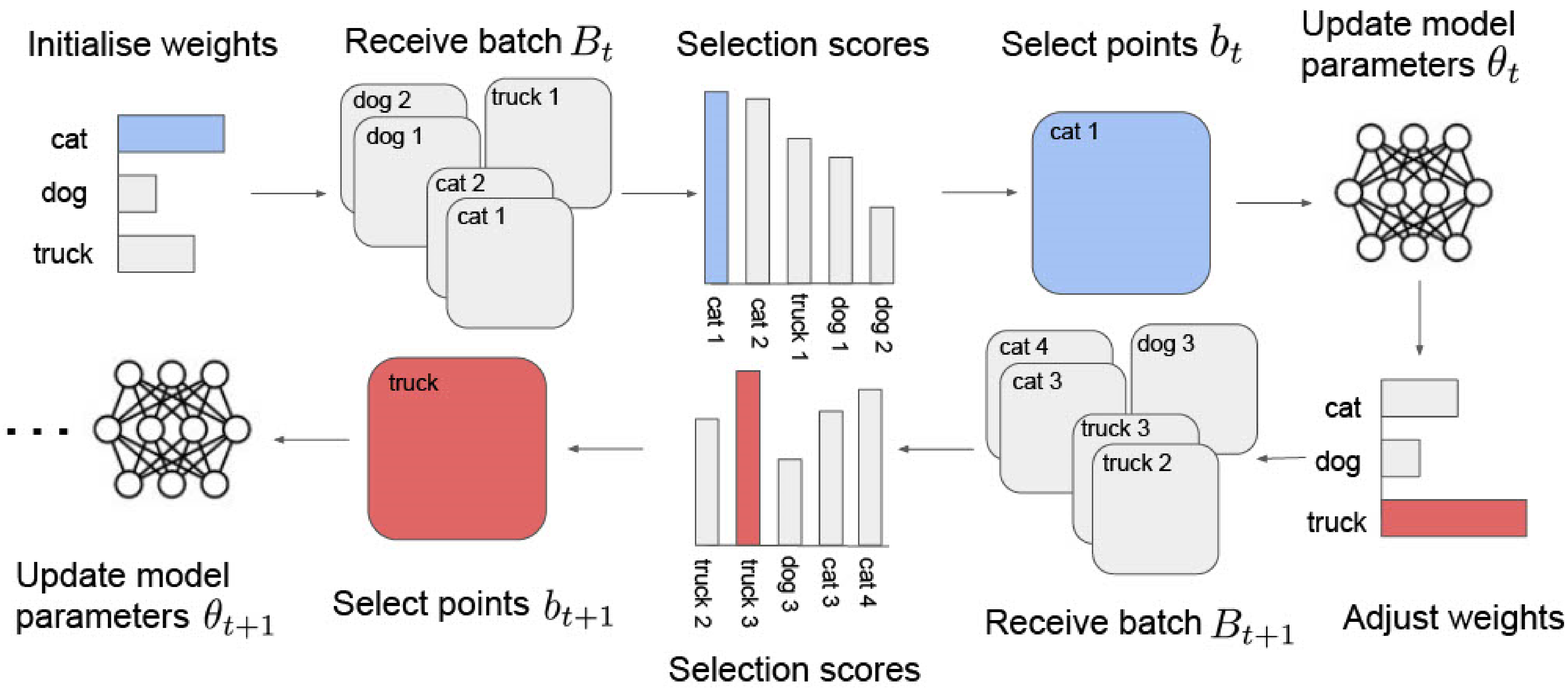
1. Robust online batch selection – suitable for streamed data

$$b_t = \arg\max_{b \subset B_T} \min_{c \in C} \log p(y_{ho}^{(c)} | x_{ho}^{(c)}, D_t \cup b) .$$

2. Solve the max min problem using a multiplicative weights algorithm

$$b_t = \arg\max_{b \subset B_T} \sum_{c=1}^{C} w_c \underbrace{\log p(y_{ho}^{(c)} | x_{ho}^{(c)}, D \cup b)}_{\text{Class specific likelihood given new data}} .$$

3. We design a novel selection scoring function which approximates the class specific likelihood given new data



Initialise weights
cat
dog
truck
Receive batch B_t
dog 2
truck 1
dog 1
cat 2
cat 1
Selection scores
cat 1
cat 2
truck 1
dog 1
dog 2
Select points b_t
cat 1
Update model parameters θ_t
cat
dog
truck
Adjust weights
cat 4
dog 3
cat 3
truck 3
truck 2
Receive batch B_{t+1}
truck 2
truck 3
dog 3
cat 3
cat 4
Selection scores
truck
Select points b_{t+1}
. . .
Update model parameters θ_{t+1}

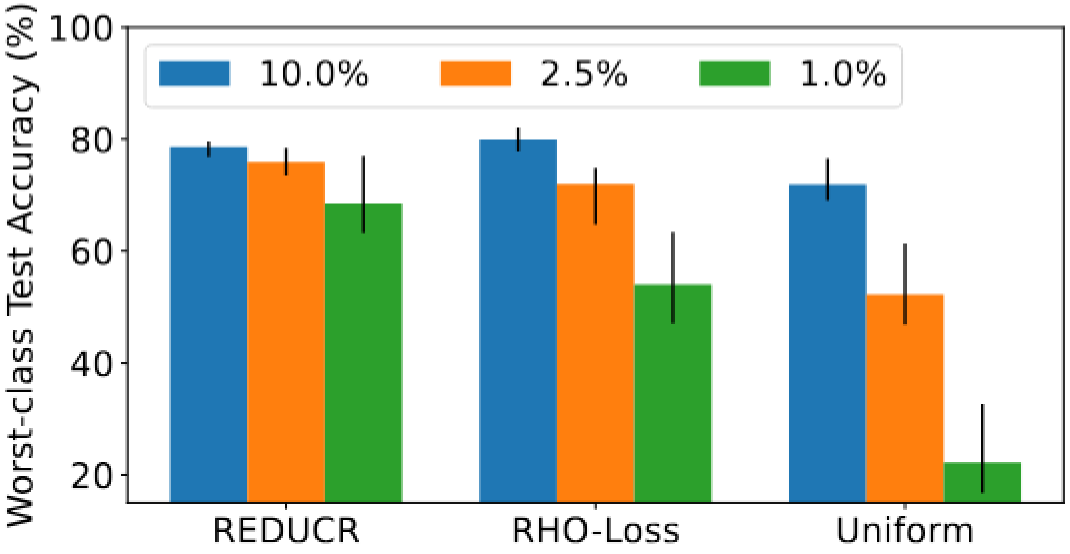
Results

- Experiments are run on a variety of NLP and vision classification tasks.
- REDUCR improves the worst-class accuracy when compared with relevant online batch selection baselines.
- REDUCR maintains average test accuracy despite prioritising the worst-class during training.

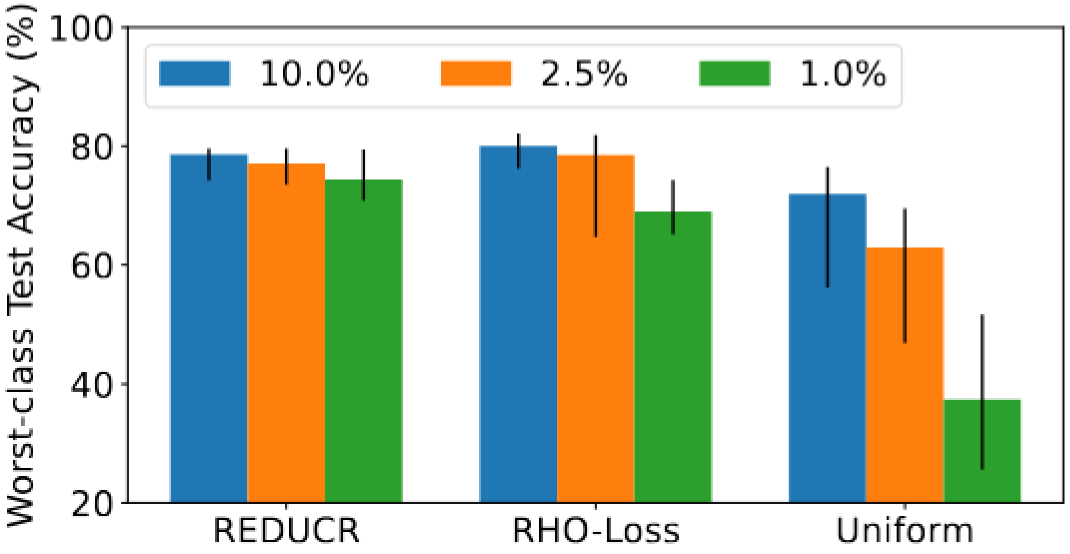
Dataset	Worst-Class Test Accuracy		Average Test Accuracy	
	Best Baseline	REDUCR	Best Baseline	REDUCR
CIFAR10 (10 runs)	78.80 ± 2.09	$\mathbf{83.29 \pm 0.84}$	$\mathbf{90.00 \pm 0.33}$	$\mathbf{90.02 \pm 0.44}$
CINIC10 (10 runs)	69.39 ± 3.56	$\mathbf{75.30 \pm 0.85}$	$\mathbf{82.09 \pm 0.30}$	$\mathbf{81.68 \pm 0.47}$
CIFAR100 (5 runs)	17.59 ± 5.17	$\mathbf{26.00 \pm 2.65}$	60.95 ± 0.64	$\mathbf{62.21 \pm 0.62}$
Clothing1M (5 runs)	40.37 ± 3.58	$\mathbf{53.91 \pm 2.42}$	71.07 ± 0.46	$\mathbf{72.69 \pm 0.42}$
MNLI (5 runs)	76.74 ± 0.93	$\mathbf{79.45 \pm 0.39}$	$\mathbf{80.89 \pm 0.31}$	$\mathbf{80.28 \pm 0.33}$
QQP (5 runs)	79.96 ± 2.34	$\mathbf{86.61 \pm 0.49}$	$\mathbf{86.88 \pm 0.31}$	$\mathbf{86.99 \pm 0.49}$

Further Results

REDUCR's performance remains consistent as specific classes are under-sampled whilst other online batch selection approaches deteriorates. Results shown on the CIFAR10 dataset.



(a) Under-sampling on class 3



(b) Under-sampling on class 5

Conclusion

- REDUCR – Smart, robust data downsampling algorithm
- Empirically observe that REDUCR improves the worst-class accuracy without sacrificing average accuracy

Contact E-mail: william.bankes.21@ucl.ac.uk