[Online Coreset Selection for Rehearsal-based Continual Learning]

- Our proposed method maximizes the model's adaptation to a target dataset while selecting high-affinity samples to past tasks, which directly inhibits catastrophic forgetting
- Continual learning: model continuously learns over a sequence of tasks, does not lose the fidelity for past tasks after adapting the previously learned knowledge to future tasks

Online Coreset Selelction (OCS)

- **Minibatch similarity**: selects samples that are representative to the target task T
- **Cross-batch diversity:** encourages minimal redundancy among the samples of target task T
- **Coreset affinity:** promotes minimum interference between selected samples and knowledge of their previous tasks

1. **Minibatch similarity**: Select subset with the **smallest** possible distance between subset gradients and target dataset gradients → The gradient of any arbitrary subset approximates the gradient of the whole subset
2. **Cross-batch diversity:** make sure subset has **largest** possible negative averaged similarity with other peer instances

Lit review of coreset selection

- **Importance sampling:** strengthens loss/gradients of important samples based on influence functions
- **Stochastic Gumbel-top-k trick and beam search**: hierarchically sample sequences without replacement
- **Herding based strategy**
- **Online variational inference**
- **Select replay buffer:** maximise variance in gradient-space
- **Bilevel optimisation framework**: with cardinality constraints

Coreset selection. There exist various directions to obtain a coreset from a large dataset. Importance sampling [13, 14, 38] strengthens the loss/gradients of important samples based on influence functions. Kool et al. [16] connect stochastic Gumbel-top-k trick and beam search to hierarchically sample sequences without replacement. Rebuffi et al. [32] propose a herding based strategy for coreset selection. Nguyen et al. [31] formulate the coreset summarization in continual learning using online variational inference [35, 6]. Aljundi et al. [2] select the replay buffer to maximize the variance in the gradient-space. Contrary to these methods, OCS considers the diversity, task informativity and relevancy to the past tasks. Recently, Borsos et al. [5] propose a bilevel optimization framework with cardinality constraints for coreset selection. However, their method is extremely limited in practice and inapplicable in large-scale settings due to the excessive computational cost incurred during training. In contrast, our method is simple, and scalable since it can construct the coreset in the online streaming data continuum without additional optimization constraints.

Coresets for Data-efficient Training of Machine Learning Models

- Our key idea is to select a weighted subset S of training data V **that best approximates the full gradient of V** . We prove that the subset S that minimizes an upper-bound on the error of estimating the full gradient maximizes a submodular facility location function. Hence, S can be efficiently found using a fast greedy algorithm.

Lit review of coreset selection

- Importance sampling with respect to sensitivity score to provide high-probability solutions for specific problems
- our goal in CRAIG is to find the smallest subset $S \subseteq V$ and corresponding per-element stepsizes $\gamma_j > 0$ that approximate the full gradient with an error at most $\epsilon > 0$ for all the possible values of the optimization parameters $w \in W$.