Report date 10/02/2025

Paper report:

**Overview:**

The paper builds on top of another unlearning paper (https://arxiv.org/abs/1912.03817) that works on ensemble learning and unlearns by simply identifying the learner that dealt with the data to be removed and retrains that learner only with the updated shard.

This paper also adds a data preparation step before unlearning.

The whole point of the paper is to show that adding this data preparation step allows more efficient unlearning and increases the unlearnt model performances.

The model grants a perfect unlearning due to the fact that the underlying model(s) are fully retrained when the data changes, this paper is not much about ML but more about data management/data preparation.

The weak learners (the multiple learners from which the final model aggregates the results from) are regression models and the results are aggregated by averaging the results.

The relevant thing here is the linear encoder which shuffles and takes the data to be place in each shard. The shards are not partitions of the starting training data but they do overlap. The idea here is to have some overlap so that more shard pieces of the uncoded algorithm (without the linear encoder), are present in a single shard of the coded one.This redundancy is defined by a parameter of the linear encoder which defines how many uncoded shards end up in a single coded shard.

The unlearning algorithm then only finds out which weak learner has learnt the data to be removed, removes that data from the weak learner shard and retrains the weak learners affected.

An interesting observation is that with coding, influential training samples may end up in multiple shards leading to better results in more learners.

To prove this point another experiment is run where outliers or inliers are removed from the training sample, the performances are severely degraded when outliers are removed, the author then concludes that influential data in this case is the outliers and removing those data we worsen the model performances.

**Experimental metrics and scenarios:**

Three dataset are used:

* Physicochemical Properties of Protein Tertiary Structure dataset
* Computer Activity dataset
* Combined Cycle Power Plant dataset

The evaluated metrics are:

* MSE
* Unlearning time

**Benchmarked against:**

The implementation is provided by the paper Machine Unlearning (<https://arxiv.org/abs/1912.03817>) which does not code the training set (no linear encoder).

**Limitations:**

**Repo:**

**Authors:**

Nasser Aldaghri, Hessam Mahdavifar, Ahmad Beirami

**Year:**

2023