Machine Unlearning

Lucas Bourtoule, Varun Chandrasekaran, Christopher A. Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, Nicolas Papernot

Presented by Peter Li

Problem Statement

People’s privacy rights

You are a data controller and/or a data processor. But as a person who uses the Internet, you’re also a data subject. The GDPR recognizes a litany of new privacy rights for data subjects, which aim to give individuals more control over the data they loan to organizations. As an organization, it’s important to understand these rights to ensure you are GDPR compliant.

Below is a rundown of data subjects’ privacy rights:

1. The right to be informed
2. The right of access
3. The right to rectification
4. The right to erasure
5. The right to restrict processing
6. The right to data portability
7. The right to object
8. Rights in relation to automated decision making and profiling.

PIPEDA Fair Information Principle 5 – Limiting Use, Disclosure, and Retention

Reviewed: August 2020

Your responsibilities

- Unless someone consents otherwise—or unless doing so is required by law—your organization may use or disclose personal information only for the identified purposes for which it was collected. Keep personal information only as long as it is needed to serve those purposes.
- Know what personal information you have, where it is, and what you are doing with it.
- Obtain fresh consent if you intend to use or disclose personal information for a new purpose.
- Collect, use or disclose personal information only for purposes that a reasonable person would consider appropriate in the circumstances.
- Put guidelines and procedures in place for retaining and destroying personal information.

California Consumer Privacy Act (CCPA)

Home / Privacy / *California Consumer Privacy Act (CCPA)*

Updated on May 10, 2023

The California Consumer Privacy Act of 2018 (CCPA) gives consumers more control over the personal information that businesses collect about them and the CCPA regulations provide guidance on how to implement the law. This landmark law secures new privacy rights for California consumers, including:

- The right to know about the personal information a business collects about them and how it is used and shared;
- The right to delete personal information collected from them (with some exceptions);
- The right to opt-out of the sale or sharing of their personal information; and
- The right to non-discrimination for exercising their CCPA rights.

1. https://gdpr.eu/what-is-gdpr/
2. https://oag.ca.gov/privacy/ccpa
3. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/p_principle/principles/p_use/

Goals and Formalisation

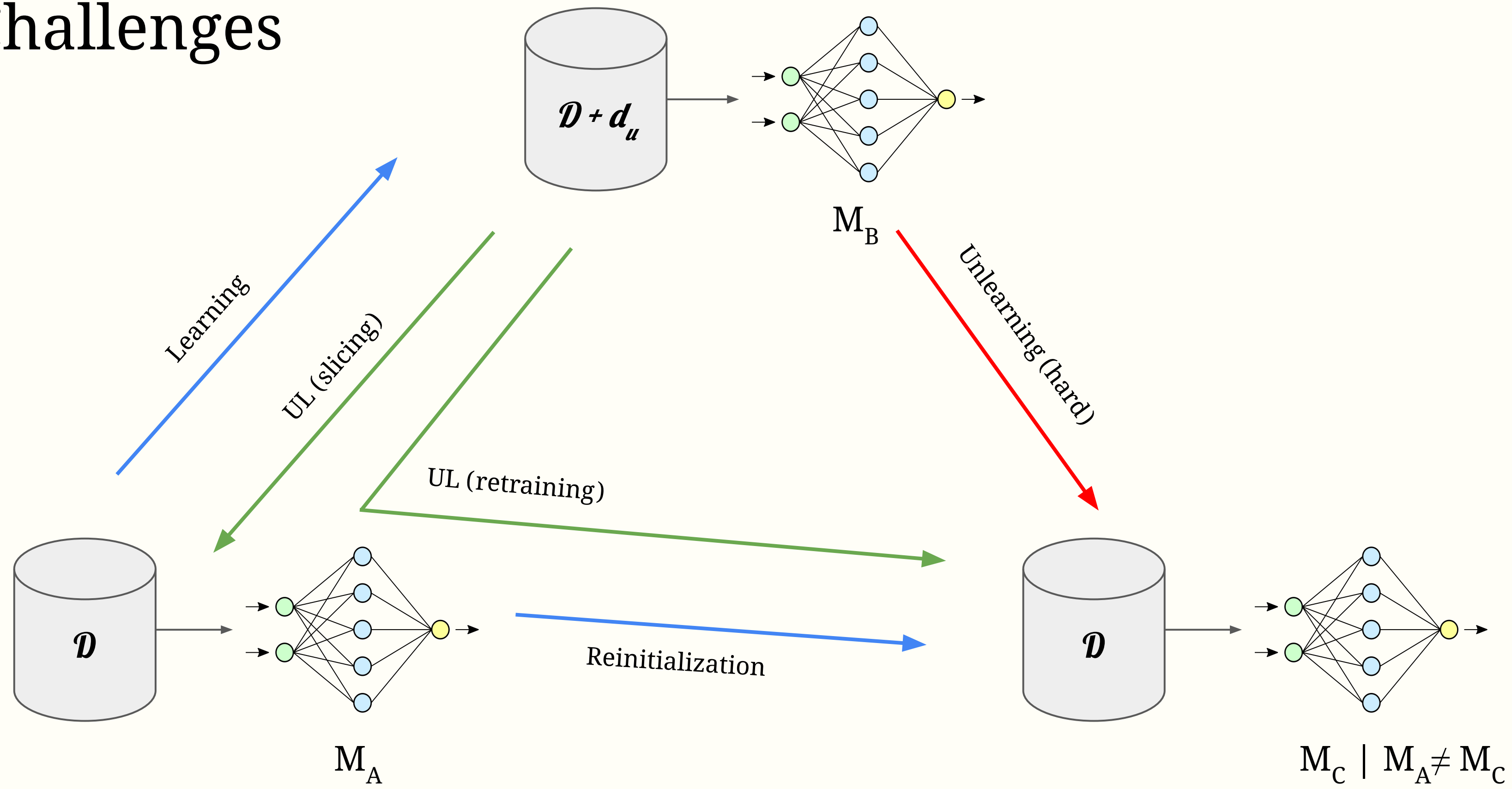
Definition III.1. Let $D = \{d_i : i \in U\}$ denote the training set collected from population U. Let $D' = D \cup d_u$. Let D_M denote the distribution of models learned using mechanism M on D' and then unlearning du. Let D_{real} be the distribution of models learned using M on D. The mechanism M facilitates unlearning when these two distributions are identical.

<u>Goals</u>

G1: Intelligibility
G2: Comparable Accuracy
G3: Reduced Unlearning Time
G4: Provable Guarantees
G5: Model Agnostic
G6: Limited Overhead

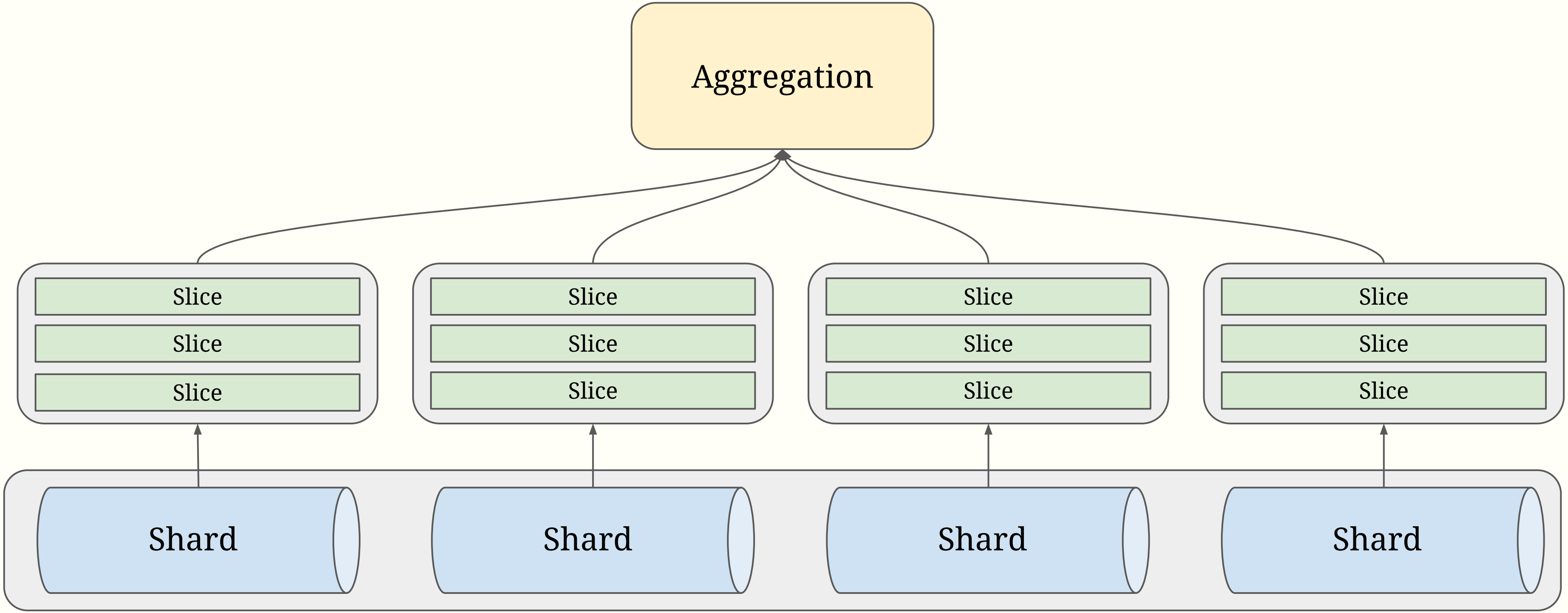
1. Bourtole et al. (pg. 4-5)

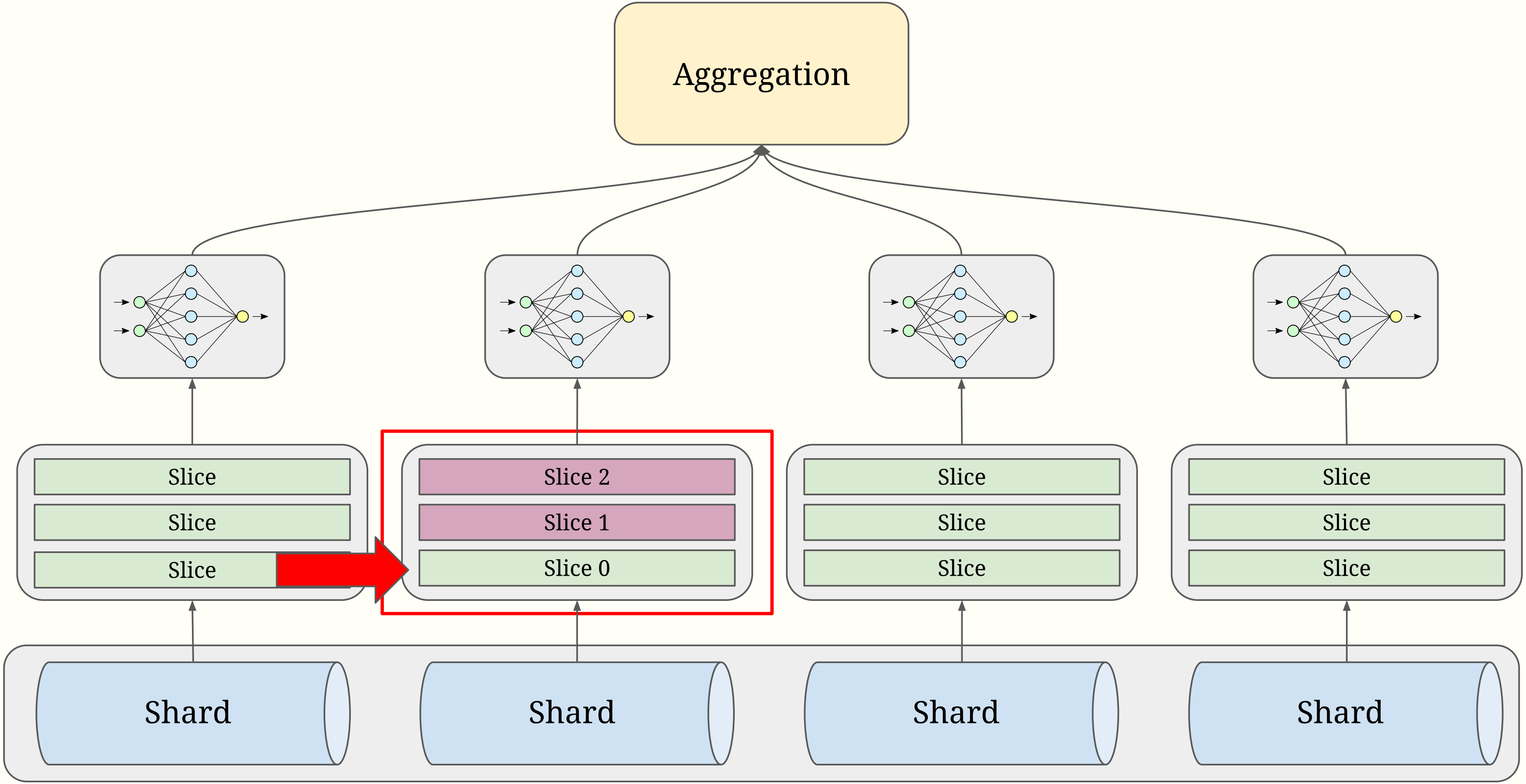
Challenges



1. Bourtole et al. (pg. 4)

Shared Isolated Sliced Aggregated





Neural Network Image Source: https://www.methodpark.de/blog/introduction-to-neural-networks/

Trade Offs

Strategy Trade Offs			
	Retraining Speed	Storage Cost	Accuracy
Sharding	↑		↓
Slicing	↑	↑	
Aggregation Model			↑

1. https://www.youtube.com/watch?v=xUnMkCB0Gns&themeRefresh=1

Measuring Time (Sharding)

$$\binom{i-1}{j}\left(\frac{1}{S}\right)^{j}\left(1-\frac{1}{S}\right)^{i-j-1}$$

By first summing over all possible combinations of points that are unlearned in a shard at a specific step, and then summing over all requests (K in total), we are able to obtain the expected number of points to be retrained ($\mathbb{E}(C)$) as:

$$\sum_{i=1}^{K}\sum_{j=0}^{i-1}\binom{i-1}{j}\left(\frac{1}{S}\right)^{j}\left(1-\frac{1}{S}\right)^{i-j-1}\left(\frac{N}{S}-1-j\right)$$

This expression can be simplified using the binomial theorem, as described in Appendix D to obtain:

$$\mathbb{E}[C]=\left(\frac{N}{S}+\frac{1}{2S}-1\right)K-\frac{K^2}{2S} \tag{2}$$

1. Bourtoule et al. (pg. 8)

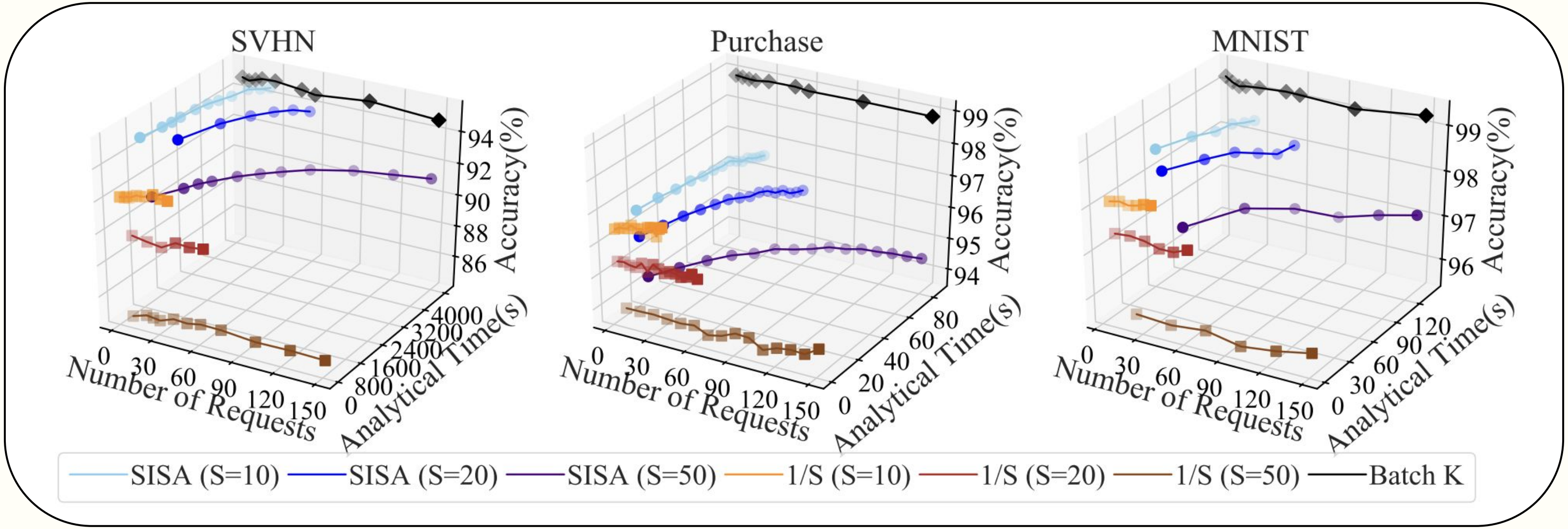
Model Architecture

Dataset	Model Architecture
MNIST [43]	2 conv. layers followed by 2 FC layers
Purchase [49]	2 FC layers
SVHN [50]	Wide ResNet-1-1
CIFAR-100 [51]	ResNet-50
Imagenet [44]	ResNet-50
Mini-Imagenet [48]	ResNet-50

TABLE II: Salient features of DNN models used.

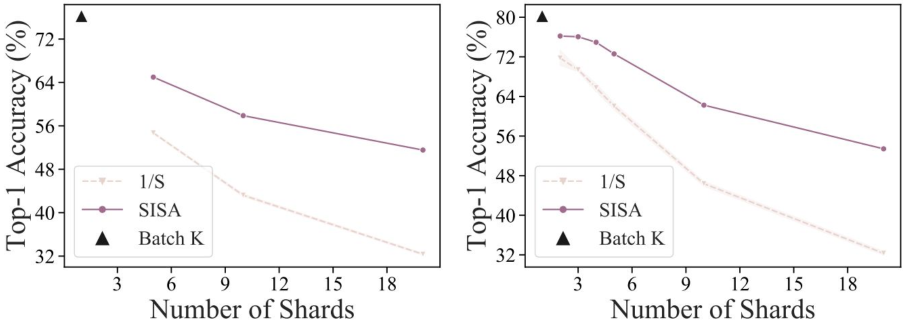
1. Bourtoule et al. (pg. 10)

Experimental Results



1. Bourtoule et al. (pg. 10)

Experimental Results (Complex Tasks Are Difficult for SISA)



(a) Imagenet dataset (b) Mini-Imagenet dataset

Fig. 6: For complex learning tasks such as those involving Imagenet and Mini-Imagenet, **SISA** training introduces a larger accuracy gap in comparison to the batch K baseline. However, it is still more performant than the $\frac{1}{S}$ fraction baseline. Each constituent (and baseline) utilized the prediction vector aggregation strategy.

1. Bourtoule et al. (pg. 10)

Strengths

- Simple & Intelligible Strategy
- Experimentally and analytically proven benefits
- Not model restrictive

Limitations

- Accuracy suffers on more complex tasks
- Storage costs