**Federated Learning**: where machine learning and data privacy can coexist

Mattia Dutto s287598 *(mattia.dutto@studenti.polito.it)*

Simone Alberto Peirone s286886 *(simonealberto.peirone@studenti.polito.it)*

Nolan Zerigue s293622 *(nolan.zerigue@studenti.polito.it)*

**TA** : Debora Caldarola *(debora.caldarola@polito.it)*

INTRODUCTION

Mobile phones, healthcare devices, bank transactions, and self-driving cars are just a few examples of modern distributed networks that generate vast amounts of data. Because of machine’s rising computational capacity, as well as concerns about transferring private information, it is becoming more appealing to store data locally in order to avoid data breaches.

That’s why Google introduced in 2016 the term of *Federated Learning* which allows machine learning models to gain experience from many datasets located at separate locations without exchanging training data. This enables personal data to stay on local devices, lowering data leak risk.

In this work, we studied federated scenarios, its architecture and the main challenges involved. Then moved on the implementation of our own baseline, trying to find the best parameters configuration for our model and understand the issues occurred over different local data distributions. Finally, recognized the effect of normalization layers and pushed some proposed variations to improve our model.

For all those previous incremental steps, we compared our results in both centralized and decentralized (federated) manner.

RELATED WORKS

As the first Federated Learning papers were published by Google, we firstly studied these two following articles [1] [2] that provides us an introduction about the standard FederatedAveraging (FedAvg) algorithm use on our project.

Moreover, we read the first application paper on Android keyboard [7] to see how it works on concrete situation. It provides experiments on federated scenarios using FedAvg algorithm for next word predictions.

Then, to get more advanced knowledge on the Federated Learning current state and challenges, we studied these following surveys:

* Federated Learning: Challenges, Methods, and Future Directions [4] where it describes the core challenges which are expensive communication, system heterogeneity, statistical heterogeneity and privacy concerns.
* Advances and Open Problems in Federated Learning [5] that shows current improvements for federated learning models and protections against data privacy attacks.
* To learn more about federated learning in mobile edge computing, Federated Learning in Mobile Edge Networks: A Comprehensive Survey [6] provides current details.

Finally, we focus our study on the Federated Visual Classification with Real-World Data Distribution article proposed by Hsu TM.H. et al [3] because they show several possible divisions of our studied dataset. That allowed us to understand how the data distribution influence the model.

METHODOLOGY

FedAvg

SGD

LeNet5

EXPERIMENTS AND DISCUSSIONS

1. **Our baseline**
2. **Ablation studies**
3. **Normalization layers**
4. **Variations**

CONCLUSION

REFERENCES

[1] Google AI Blog, [Federated Learning: Collaborative Machine Learning without Centralized Training Data](https://ai.googleblog.com/2017/04/federated-learning-collaborative.html)

[2] McMahan, Brendan et al. “[Communication-Efficient Learning of Deep Networks from Decentralized Data](https://arxiv.org/abs/1602.05629)” *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR* 54:1273-1282, (2017).

[3] Hsu TM.H. et al. “[Federated Visual Classification with Real-World Data Distribution](https://arxiv.org/abs/2003.08082)” *European Conference on Computer Vision . ECCV 2020*. Lecture Notes in Computer Science, vol 12355. Springer, Cham.

[4] Li, Tian, et al. “[Federated Learning: Challenges, Methods, and Future Directions](https://arxiv.org/abs/1908.07873)” *IEEE Signal Processing Magazine* 37.3 (2020): 50-60.

[5] Kairouz, Peter, et al. “[Advances and Open Problems in Federated Learning](https://arxiv.org/abs/1912.04977)” *arXiv preprint arXiv:1912.04977* (2019).

[6] Lim, Wei Yang Bryan, et al. “[Federated Learning in Mobile Edge Networks: A Comprehensive Survey](https://ieeexplore.ieee.org/document/9060868)” *IEEE Communications Surveys & Tutorials* 22.3 (2020): 2031-2063.

[7] Hard, Andrew, et al. “[Federated Learning for Mobile Keyboard Prediction](https://arxiv.org/abs/1811.03604)” *arXiv preprint arXiv:1811.03604* (2018).