International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.052 ISSN: 2581-4621

Federated Learning: Collaborative Machine Learning without Centralized Training Data

Abhishek V A[1], Binny S[2], Johan T R[3], Nithin Raj[4], Vishal Thomas[5]
[1] *PG – Master of Computer Application, Kristu Jyoti College of Management and Technology, Changanassery, Kerala*
[2] *Associate Professor - Master of Computer Application, Kristu Jyoti College of Management and Technology, Changanassery, Kerala*
[3] *PG Master of Computer Application, Kristu Jyoti College of Management and Technology, Changanassery, Kerala*
[4] *PG - Master of Computer Application, Kristu Jyoti College of Management and Technology, Changanassery, Kerala*
[5] *PG - Master of Computer Application, Kristu Jyoti College of Management and Technology, Changanassery, Kerala*

ABSTRACT
Federated learning (also known as collaborative learning) is a machine learning technique that trains an algorithm without transferring data samples across numerous decentralized edge devices or servers. This strategy differs from standard centralized machine learning techniques in which all local datasets are uploaded to a single server, as well as more traditional decentralized alternatives, which frequently presume that local data samples are uniformly distributed.
Federated learning allows several actors to collaborate on the development of a single, robust machine learning model without sharing data, allowing crucial issues such as data privacy, data security, data access rights, and access to heterogeneous data to be addressed. Defence, telecommunications, internet of things, and pharmaceutical industries are just a few of the sectors where it has applications.

1. Introduction

Standard machine learning approaches require centralizing the training data on one machine or in a datacenter. And Google has built one of the most secure and robust cloud infrastructures for processing this data to make our services better. Now for models trained from user interaction with mobile devices, we're introducing an additional approach: Federated Learning.
Federated Learning enables mobile phones to collaboratively learn a shared prediction model while keeping all the training data on device, decoupling the ability to do machine learning from the need to store the data in the cloud. This goes beyond the use of local models that make predictions on mobile devices (like the Mobile Vision API and On-Device Smart Reply) by bringing model training to the device as well. Federated Learning allows for smarter models, lower latency, and less power consumption, all while ensuring privacy. And this approach has another immediate benefit: in addition to providing an update to the shared model, the improved model on your phone can also be used immediately, powering experiences personalized by the way you use your phone
Federated learning was studied by Google in a research paper published in 2016 on arxiv. Since then, it has been an area of active research in the AI community as evidenced by the fast-growing volume of preprints appearing on arxiv. Recent research work on federated learning are mainly focused on improving security and statistical challenges

2. Methodology

Your phone participates in Federated Learning only when it won't negatively impact your experience.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.052 ISSN: 2581-4621

The process goes as follows: your device gets the most recent model, refines it using data from your phone, and then compiles the modifications into a brief, targeted update.

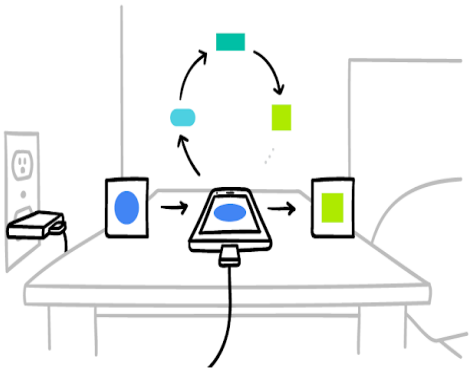


Fig 1. Your phone personalizes the model locally, based on your usage (A). Many users' updates are aggregated (B) to form a consensus change (C) to the shared model, after which the procedure is repeated.

Only this model update is transmitted via encrypted communication to the cloud, where it is quickly averaged with updates from other users to enhance the shared model.
No specific updates are saved in the cloud; all training data is kept on your device.
We can use Google Keyboard, G-board with Federated Learning. Your phone keeps information about the current context and whether you clicked a suggested query when G-board displays one locally. Federated Learning analyses that history on-device to make suggestions for enhancements to the query suggestion model in G-board.

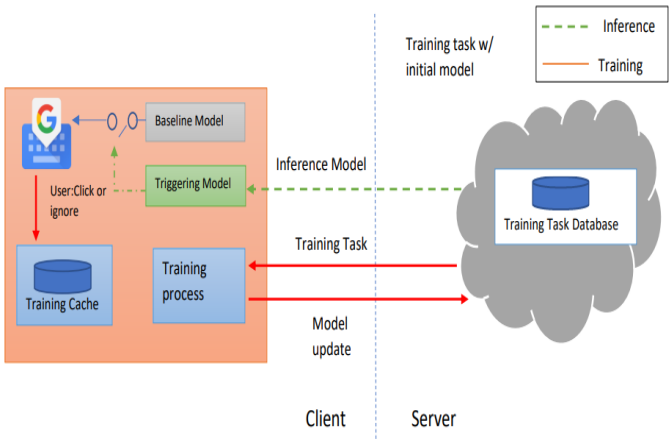


Figure 2: overview of the architecture. On-device inference and training are used; during training rounds, model updates are communicated to the server, and trained models are manually deployed to clients.
This section walks through the process of completing training, evaluation, and inference of the question suggestion triggering model to provide a concise technical overview of the client and server-side runtime that supports FL in G-board.

- **2.1 Working**

A) The devices download a shared prediction model that has been initially trained on proxy data.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.052 ISSN: 2581-4621

B) The edge devices use the local data they have access to improve the downloaded model(mobile phones)
C) Creates a brief update that summarizes the changes, and just this update is sent to the cloud, where it is quickly averaged with updates from other users, improving the shared prediction model.
(This iterative procedure is carried out up until the development of a high-quality model.)
D) As a result, privacy is also maintained because all training data is kept on the device and individual updates are not saved in the cloud.

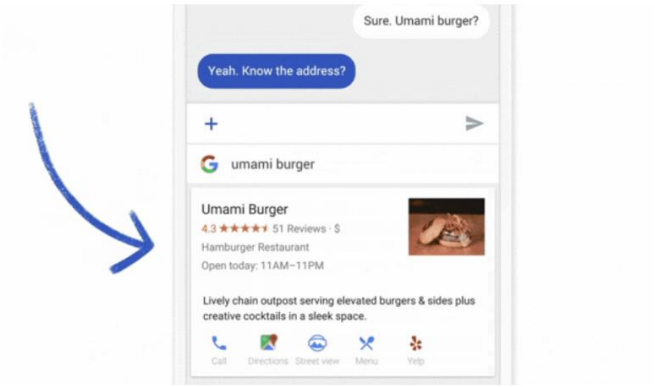


Figure 3

Whether we clicked on the recommendation query or not, when we receive a recommendation on G-board, our phone remembers our activities locally, and that on-device history of data proposes improvements to the shared prediction model.

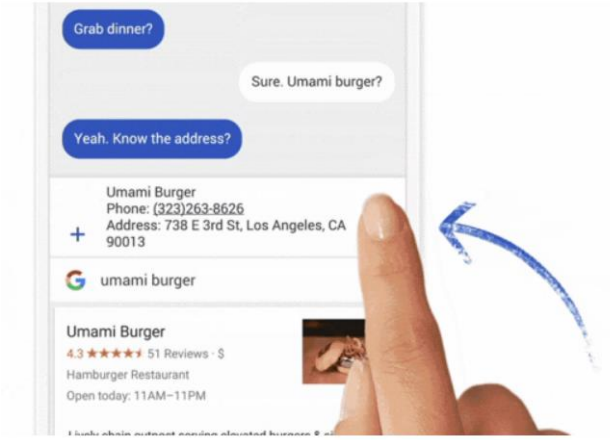


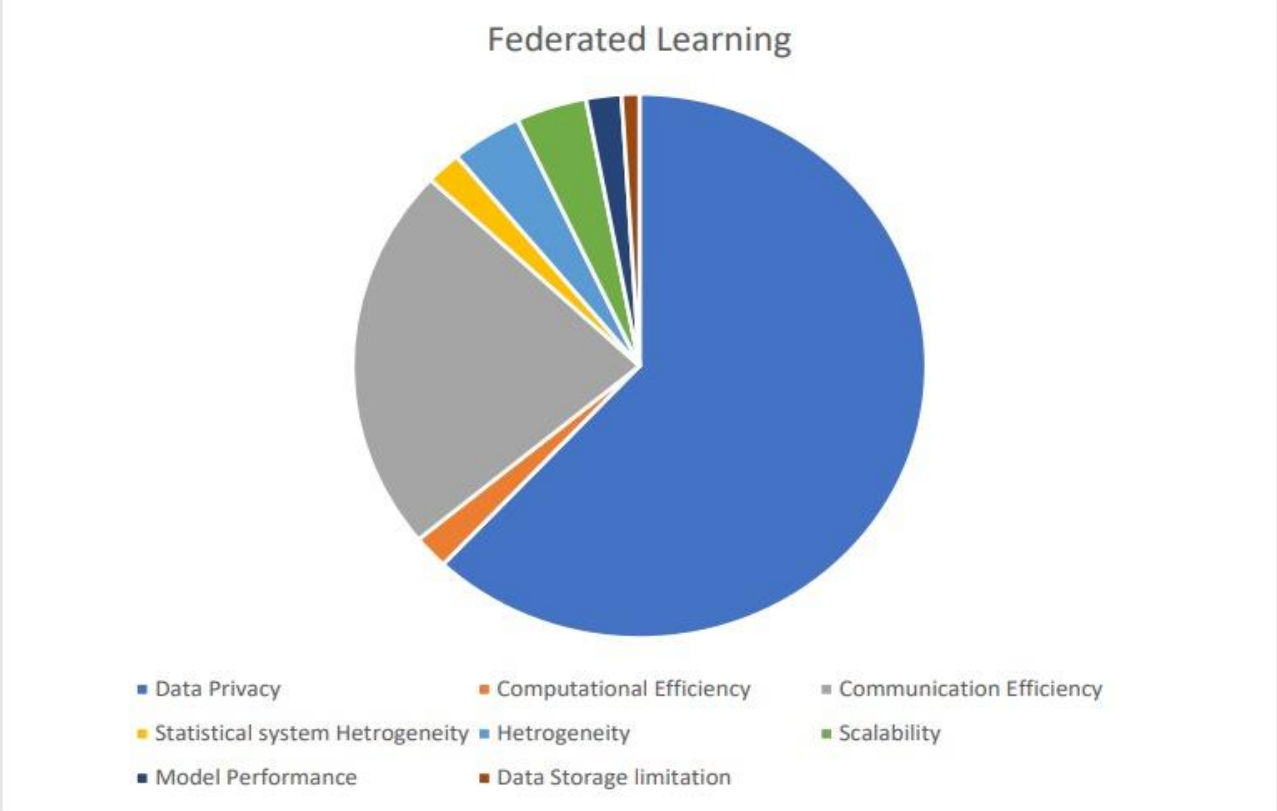
Figure 4

As was previously indicated, on-device training using a scaled-down version of TensorFlow makes it possible to deploy this technology to millions of heterogeneous phones running G-board.
In order to prevent any effects on the phone's performance, the training only takes place when the device is unplugged and plugged into a wi-fi connection.

3. Results and Discussion

- Based on what you actually type on your phone, improve the language models that drive your keyboard.
- Enhanced precision
- Predict the next word based on the text you've written so far
- Powers the predictions strip

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.052 ISSN: 2581-4621



The merits include Privacy – The data remains on the user's device. Low latency – The model trained is available locally and can be utilised to produce predictions. Accurate predictions – Because the model is trained on local data rather than proxy data, the outcomes are more accurate and tailored. Low power usage – The local model training aids in power consumption optimization. Running on low-end hardware, such as mobile phones, minimal hardware reduces the amount of hardware infrastructure required.
It also has challenges Because federated networks are often made up of many devices, communication in that network can be slower than local computing, so effective and economical communication is required!
Data heterogeneity – Different local datasets may introduce bias, and dataset sizes will differ!
Temporal heterogeneity – The distribution of each dataset may change over time!
Update loss – A partial or total loss of model updates owing to device failure may have an impact on the shared prediction model.

CONCLUSION

Our work has merely touched the surface of what is conceivable. Federated Learning can't solve all machine learning issues (for example, learning to recognise different dog breeds by training on carefully labelled examples), and for many other models the necessary training data is already available in the cloud (like training spam filters for Gmail) (like training spam filters for Gmail).

The state-of-the-art for cloud-based ML will thus continue to be advanced by Google, but we are also dedicated to continuous research to broaden the spectrum of issues that Federated Learning can address. In addition to G-board inquiry recommendations, for instance, we intend to enhance the language models that power your keyboard depending on what you actually type on your phone (which can have a style all its own).

DOI:10.46647/ijetms.2022.v06i05.052 ISSN: 2581-4621

Model building, training, and evaluation without direct access to or labelling of raw data, with communication cost as a limiting factor, necessitate machine learning practitioners to embrace new tools and a new way of thinking in order to apply federated learning. The user benefits of Federated Learning, in our opinion, outweigh the technical difficulties, and we are sharing our work in the hope that it will spark a broad discussion among machine learning experts.

References

1. McMahan, B. (n.d.). Federated Learning, from Research to Practice. Parallel Data Lab. Retrieved July20, 2022, from https://www.pdl.cmu.edu/SDI/2019/slides/2019-09-05Federated20Learning.pdf
2. McMahan, B., & Ramage, D. (2017, April 6). *Federated learning: Collaborative machine learning without centralized training data*. Google AI Blog. Retrieved July 19, 2022, from https://ai.googleblog.com/2017/04/federated-learning-collaborative.html
3. Yu, H., Yang, Q., Kang, Y., Cheng, Y., Liu, Y., & Chen, T. (2019). *Federated learning*. Morgan & Claypool Publisher.
4. Jakub Konecny, H. Brendan McMahan, Felix X. Yu, Peter Richtarik, Ananda Theertha Suresh, and Dave Bacon, "Federated learning: Strategies for improving communication efficiency", in NIPS Workshop on Private Multi-Party Machine Learning, 2016.
5. Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar, "Federated multi-task learning," in Advances in Neural Information Processing Systems 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., pp. 4424–4434. Curran Associates, Inc., 2017.
6. Blake E Woodworth, Jialei Wang, Adam Smith, Brendan McMahan, and Nati Srebro, "Graph oracle models, lower bounds, and gaps for parallel stochastic optimization," in Advances in Neural Information Processing Systems 2018.
7. Naman Agarwal, Ananda Theertha Suresh, Felix Xinnan X Yu, Sanjiv Kumar, and Brendan McMahan, "cpsgd: Communication-efficient and differentially-private distributed sgd," in Advances in Neural Information Processing Systems(2018)