UE21CS343BB2

Topics in Deep Learning

Introduction to Federated Learning

Dr. Shylaja S S
Director of Cloud Computing & Big Data (CCBD), Centre for Data Sciences & Applied Machine Learning (CDSAML)
Department of Computer Science and Engineering
shylaja.sharath@pes.edu

Ack: Divya K,
Teaching Assistant

Topics in Deep Learning

Overview

PES UNIVERSITY

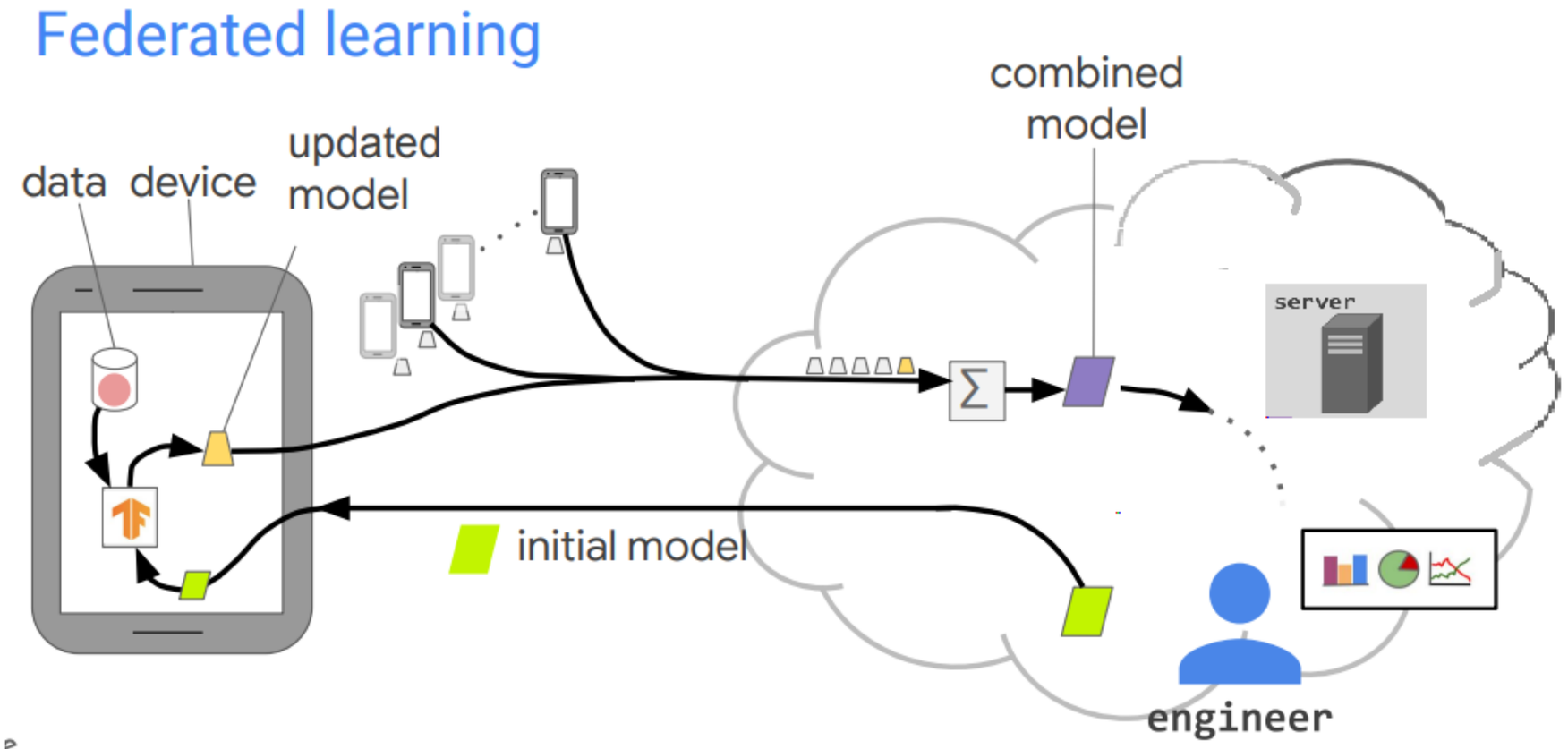
- Why Federated Learning?
- What is Federated Learning?
- Benefits of Federated Learning
- How does Federated Learning work?
- Key Components of Federated Learning
- Challenges in Federated Learning
- Applications of Federated Learning
- Conclusion

Why Federated Learning?

- A conventional approach to train machine learning models is to gather all data at a central server and use it to train the model. But this method, while easy, has raised concerns about data privacy, leaving a lot of valuable but sensitive data inaccessible.
- Centralized ML may also face other challenges such as scaling due to centralized processing limitations and security vulnerabilities.
- To address this issue, AI models started to shift to a decentralized approach, and a new concept called "federated learning" has emerged.
- Federated Learning serves as a great solution for encouraging collaboration while respecting data privacy.

What is Federated Learning?

- Federated Learning (often referred to as collaborative learning) is a decentralized approach to training machine models.
- It doesn't require an exchange of data from client devices to global servers. Instead, the raw data on edge devices is used to train the model locally, increasing data privacy. The final model is formed in a shared manner by aggregating the local updates.



Benefits of Federated Learning

- **Privacy:** In contrast to traditional methods where data is sent to a central server for training, federated learning allows for training to occur locally on the edge device, preventing potential data breaches.

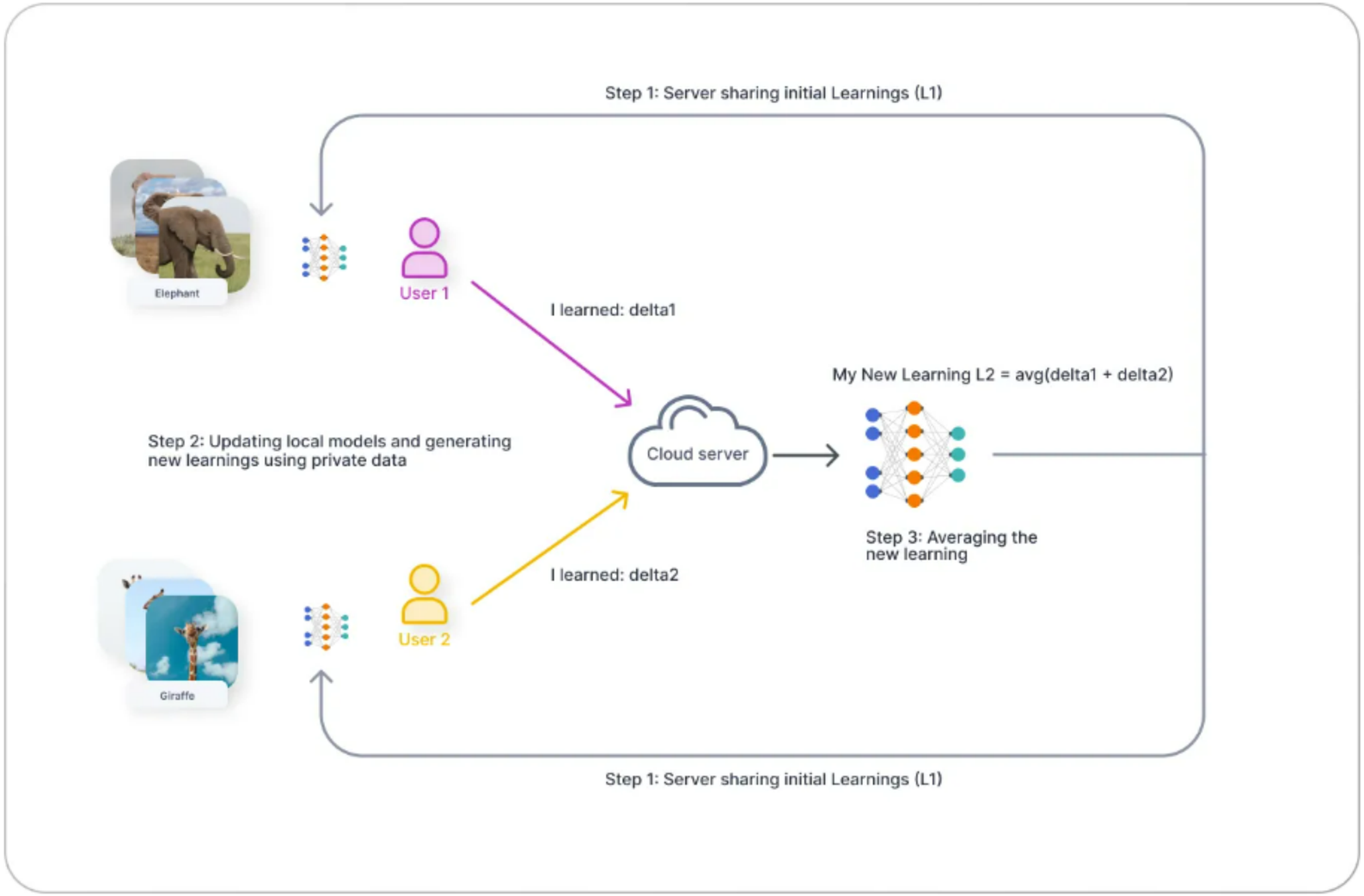
- **Data security:** Only the encrypted model updates are shared with the central server, assuring data security. Additionally, secure aggregation techniques such as Secure Aggregation Principle allow the decryption of only aggregated results.

- **Access to heterogeneous data:** Federated learning guarantees access to data spread across multiple devices, locations, and organizations. It makes it possible to train models on sensitive data, such as financial or healthcare data while maintaining security and privacy. And thanks to greater data diversity, models can be made more generalizable.

Topics in Deep Learning

How does Federated learning work?

- A baseline model is initially stored on a central server and copies are shared with client devices. These devices train their own personalized models based on local data, enhancing user experience.
- Updates (model parameters) from these local models are shared with the main model located at the central server using secure aggregation techniques. This model combines and averages different inputs to generate new learnings, leveraging diverse data for generalizability.
- This retrained central model is then redistributed to client devices, iterating the process without compromising privacy.



Key Components of Federated Learning

- Central Server: the entity responsible for managing the connections between the entities in the FL environment and for aggregating the knowledge acquired by the FL clients;
- Clients: all computing devices with data that can be used for training the global model, including but not limited to: personal computers, servers, smartphones, smartwatches, computerized sensor devices, and many more;
- Communication Framework: consists of the tools and devices used to connect servers and parties and can vary between an internal network, an intranet, or even the Internet;
- Aggregation Algorithm: the entity responsible for aggregating the knowledge obtained by the parties after training with their local data and using the aggregated knowledge to update the global model

Challenges in Federated Learning

➢ Data Heterogeneity
- Data in various formats and structures pose challenges for model training.
- Diverse data sources require complex preprocessing to ensure compatibility.

➢ Communication Efficiency
- High communication overhead between devices and the central server impacts performance.
- Efficient protocols are crucial to minimize data transmission and latency issues.

Applications of Federated Learning

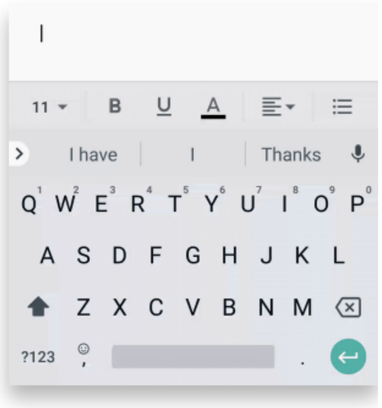
➢Smartphones: Smartphones are one of the most common ways to witness federated learning in action. Word prediction, face recognition for logging, or voice recognition while using Siri or Google Assistant are all examples of federated-learning-based solutions.

Gboard: language modeling

- Predict the next word based on typed text so far
- Powers the predictions strip

When should you consider federated learning?

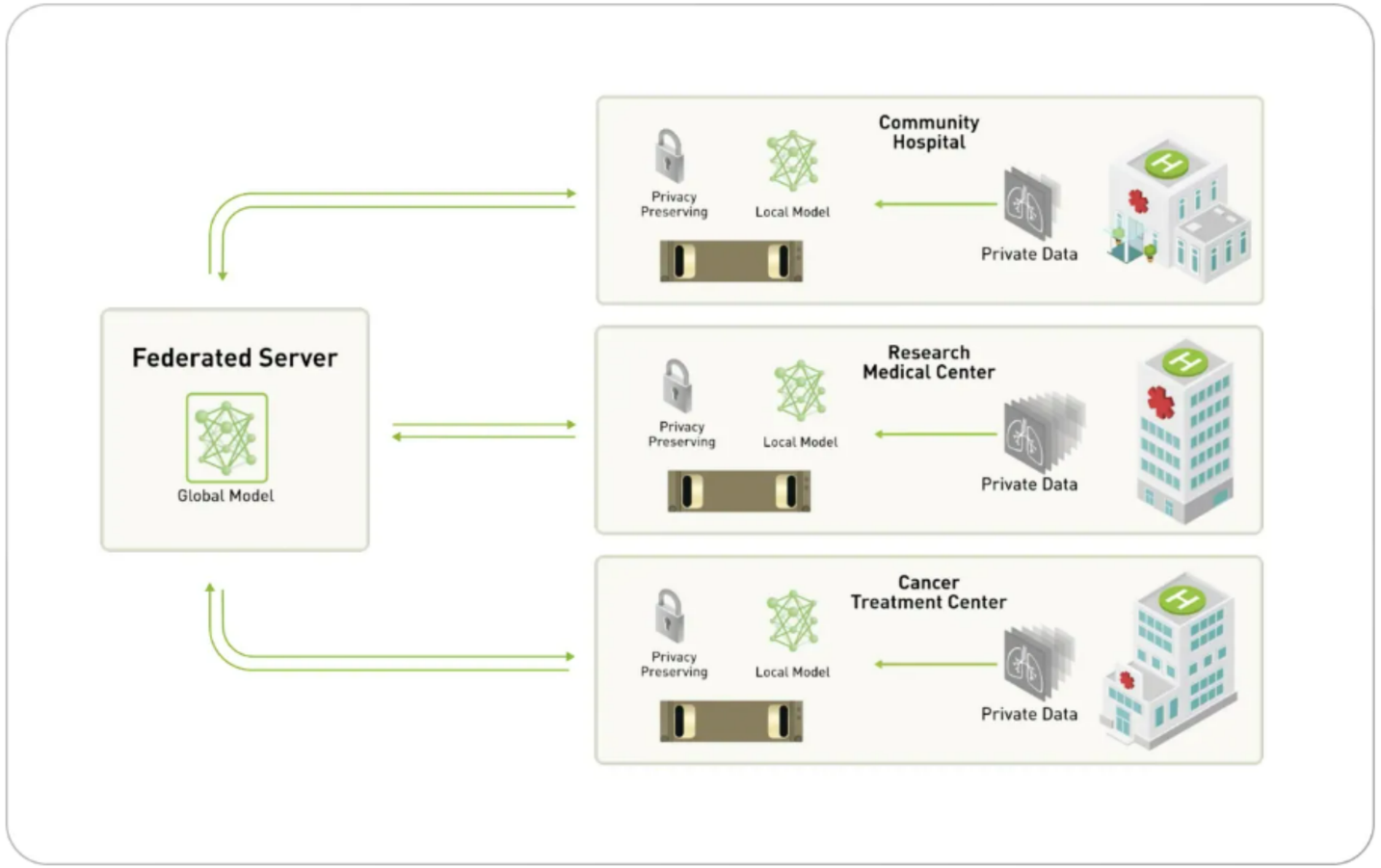
- On-device data is more relevant than server-side proxy data
- On-device data is privacy sensitive or large
- Labels can be inferred naturally from user interaction



Google

Applications of Federated Learning

➢ Healthcare: The sensitive nature of healthcare data and its restricted access due to privacy issues make it difficult to scale machine learning systems in this industry globally. With federated learning, models can be trained through secure access to data from patients and medical institutions while the data remains at its original premises.



Conclusion

In conclusion, we have explored the concept of federated learning and its paradigm shift towards moving computations to data, ensuring robust data privacy.

In our next lecture, we will focus on the categories of federated learning which include Horizontal Federated Learning, Vertical Federated Learning and Federated Transfer Learning (FTL).

Acknowledgements

https://www.v7labs.com/blog/federated-learning-guide

https://towardsdatascience.com/introduction-to-federated-learning-and-challenges-ea7e02f260ca

https://www.pdl.cmu.edu/SDI/2019/slides/2019-09-05Federated%20Learning.pdf

THANK YOU

Dr. Shylaja S S
Director of Cloud Computing & Big Data (CCBD), Centre for Data Sciences & Applied Machine Learning (CDSAML)
Department of Computer Science and Engineering
shylaja.sharath@pes.edu

Ack: Divya K,
Teaching Assistant