Contribution ID: **68**

Type: **not specified**

Model Training and Real World Analysis Using Health Data with Federated Learning

Friday, December 27, 2024 2:30 PM (15 minutes)

Machine learning has emerged in many fields thanks to its ability to extract meaningful information from large data sets, solve complex problems, and make predictions about the future. In healthcare, machine learning is used in various applications such as disease diagnosis, treatment planning, patient monitoring, and personalized healthcare. However, for these applications to be successful, large size of data are needed. Concerns about the privacy and confidentiality of health data make it difficult to collect this data on centralized servers and limit the effectiveness of data-driven models. Because of the need to protect privacy, collecting and analyzing individual patient data in a centralized system becomes a major challenge. Federated learning is a machine learning approach that allows data to be processed locally without being collected on a central server, and only model parameters are sent to the central server. This method provides secure model training by reducing network traffic while maintaining the confidentiality of the data. In this paper, we focus on diabetes prediction using the duCBA method developed for federated learning architecture. For model training, a dataset called Diabetes Health Indicators on the Kaggle platform was used. To manage the data flow and perform model training securely, three different environments were used: a central server, a Flask-based API server, and clients. Clients provide secure data privacy by performing distributed model training on their local data without sharing the data. The API and the central server are run in the Google Cloud Platform environment. The API server collects local model updates from clients and sends them to the central server using WebSocket when it has enough models. After the central server performs the model merge, it sends the updated global model to the API server. The API server then distributes this model to the clients. With this structure, clients participate in a centralized global model training process through the API without directly accessing the central server, and communication security is ensured through the WebSocket protocol. As a result of the tests conducted in the prototype application, the accuracy value of the model was calculated as 70%. These results support the wide applicability of federated learning in the healthcare domain.

Paper Language

English

Contribution Type

Virtual

Primary authors: DEMIR, Abdülbaki; KULAKSIZ, Ahmet Yasir; BÜYÜKTANIR, Büşra; KARATAS BAYDOGMUS, Gozde; YILDIZ, Kazim (Marmara University)

Presenters: DEMIR, Abdülbaki; KULAKSIZ, Ahmet Yasir

Session Classification: Session: Virtual (English Language)

Track Classification: General Track