Federated Learning

Author

FELIX (1946566) &
MICHELLE (6247808)

Date

13.12.2022

Wie können wir Machine Learning mit großen Datensätzen, Datenschutz und Realdaten vereinen?

LEITFRAGE

Agenda

- Konzepte und Begriffe
- Funktionsweise
- Anwendungsbeispiele
- Zusammenfassung
  - Vorteile
  - Nachteile
  - Herausforderungen

KONZEPTE UND BEGRIFFE

Client / Server

Vernetzung mehrerer Clients mit einem oder mehrerer Server (Distributed Training)

Datenvertraulichkeit

Datensätze sind nur bei einem Client gespeichert. Parameter / Model wird anschließend vom Server aggregiert.

Heterogenität / Homogenität

Verschiedenartigkeit oder Gleichheit der Daten kann Konvergenz beeinträchtigen. Daten von unterschiedlichen Clients können verschiedene Bias haben.

Konvergenz

Abhänig von Datenqualität und Bias kann das Model konvergieren oder divergieren.

[3, 5, 6]

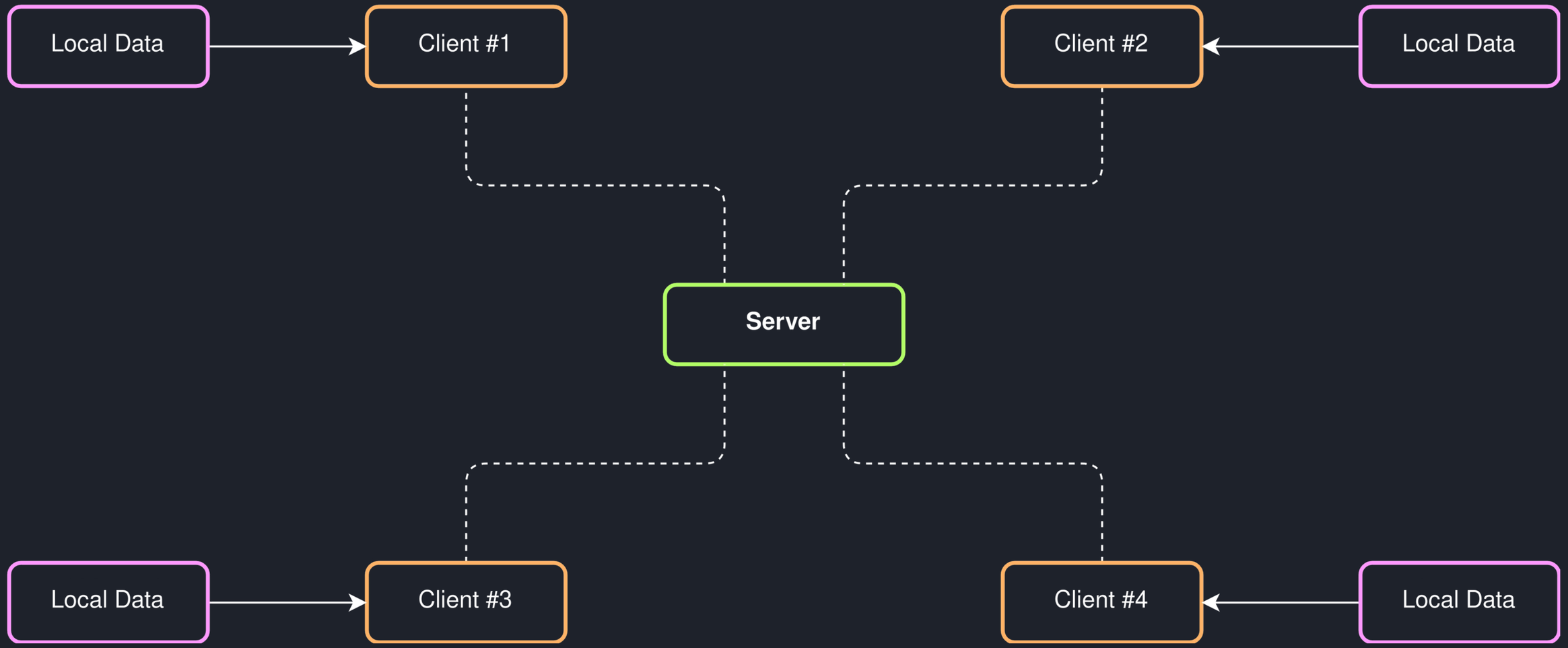
</>

Funktionsweise

INFRASTRUKTUR UND ABLAUF VON CENTRALIZED FEDERATED LEARNING

[1, 5]

Server verteilt Machine Learning Modell- mit Trainingsparameter an die Clients



Clients trainieren Modell mit lokalen Daten

Clients synchronisieren Modell- und Trainingsparameter mit dem Server
(Es werden keine Trainingsdaten weitergegeben)

ANWENDUNGSBEISPIELE

Digital Health

Medizinische Daten wie Krankheitsverläufe, oder Diagnosen sind personenbezogene Daten, welche hochsensibel sind und daher geschützt werden müssen.

Smartphone Tastaturen (GBoard)

Kleine lokale Datensätze, welche mittels des Handys vorverarbeitet werden können. Außerdem verlassen die privaten Daten das Gerät nicht.

Selbstfahrende Autos

Anpassen des globalen Models an lokale Gegebenheiten, die nicht im Hauptdatensatz existieren.

[3, 4]

Zusammenfassung

[2, 4, 5]

VORTEILE

- Unterstützung zum Einhalten von Datenschutzbestimmungen
- Steigerung der Trainingsgeschwindigkeit mittels paralleles Training (Distributed Training)
- Reduzierung von Netzwerkauslastung (Datentransfer in die Cloud)

NACHTEILE & HERAUSFORDERUNGEN

- Datenschutz nicht zwangsläufig gewährleistet
- Statistische Uneinheitlichkeit der Trainingsdaten
- Rechenleistung von Clients kann stark abweichen (Clients werden potenziell gezwungen zu warten)
- Sicherstellung von Datenqualität und Vermeidung von Bias

QUELLEN

ID	Name	URL/DOI
[1]	Federated Learning With Cooperating Devices	https://ieeexplore.ieee.org/document/8950073
[2]	Föderales maschinelles Lernen	https://publikationen.bibliothek.kit.edu/1000150233
[3]	A Systematic Review of Federated Learning in the Healthcare Area	https://www.mdpi.com/2076-3417/11/23/11191
[4]	Collaborative Machine Learning without Centralized Training Data	https://ai.googleblog.com/2017/04/federated-learning-collaborative.html
[5]	Federated Learning through Revolutionary Technology	https://www.intel.de/content/www/de/de/financial-services-it/federated-learning-solution.html
[6]	HeteroFL	https://arxiv.org/abs/2010.01264