HAL open science

Bias mitigation for collaborative and ethical learning on dynamic data

Lynda Ferraguig, Alexandre Benoit, Faiza Loukil, Mickaël Bettinelli

To cite this version:

Lynda Ferraguig, Alexandre Benoit, Faiza Loukil, Mickaël Bettinelli. Bias mitigation for collaborative and ethical learning on dynamic data. Symposium on Intelligent Data Analysis (IDA 2023), Apr 2023, Louvain-la-Neuve,, Belgium. hal-04106967

HAL Id: hal-04106967

https://hal.science/hal-04106967v1

Submitted on 26 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bias mitigation for collaborative & ethical learning on dynamic data

Ferraguig Lynda, Benoit Alexandre, Loukil Faiza, Mickael Bettinelli

{lynda.ferraguig,alexandre.benoit,faiza.loukil,mickael.bettinelli}@univ-savoie.fr

UNIVERSITÉ SAVOIE MONT BLANC

LISTIC

Problem Statement

- Data growth drives the proliferation of machine learning and distributed machine learning.
- Federated learning (FL) addresses data privacy concerns but worsens bias in models.
- Bias refers to unfairness in model decisions, and privacy constraints in FL hinder traditional bias mitigation approaches.
- FL exacerbates bias, and classical mitigation techniques are not suitable due to privacy constraints.

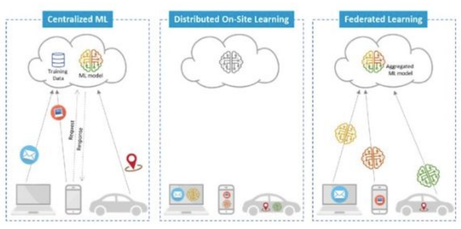
Goal

- Detect similar populations in federated learning (FL).
- Propose population-specific bias detection and correction solutions.
- Build personalized models for each population to improve generalization.
- Account for FL constraints: decentralized nature, non-IID data, dynamic and privacy.

DML

Evolution of DML architectures

- Classic DML involves grouping data together and training a distributed ML model, but may not consider data confidentiality or new regulations.
- On-site DML trains pre-trained models on data kept at the client's location, but performance may suffer due to clients not benefiting from each other's learned models.
- Federated Learning enables multiple clients to train the same ML model without sharing data, ensuring privacy and security while improving performance.

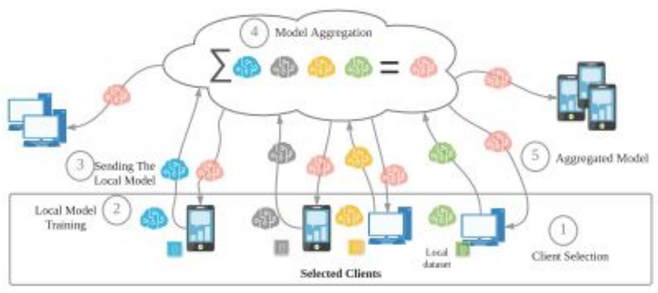


Centralized vs Distributed On-Site vs Federated Learning Architectures [1]

Federated Learning

FL an overview

- FL was proposed by Google in 2016 to address the challenge of training models on distributed data without compromising data privacy. [2]
- FL is a type of Distributed Machine Learning (DML) approach that emphasizes data confidentiality.
- FL allows clients to participate in model training without sharing data by training the model locally and sending only updates to a central server.



Federated Learning steps [3]

Fairness in Distributed Machine Learning

Why care about bias in distributed machine learning?

Definitions

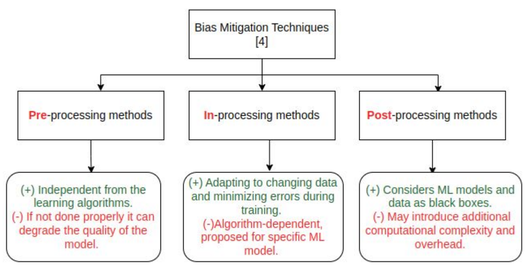
- Fairness: Algorithm output should not disadvantage protected groups.[4]
- Bias: Systematic error leading to incorrect or unfair outcomes. [5]
- Sensitive attribute: Protected characteristic used to define subgroups for fairness analysis.[5]

Impact of bias on society [4]
Biased ML is sexist, racist and sends people wrongly to jail.

Fairness in DML vs Fairness in FL [6]

ML paradigm Comparison criteria	Classical Machine Learning	Federated Learning
Data	-Full Access to data -Homogeneous dataset -Centralized in one location	-Data remains local -Non-IDD data -Distributed among clients
Source of Bias	- Historical Bias - Human Bias - Measurement Bias - Representation Bias	Classical one + : - Data heterogeneity -Aggregation -Client selection
Applicable Bias mitigation techniques	All Techniques are directly applicable	Need to be adjusted to fit FL constraints

Mitigating Bias in DML systems



Bias correction techniques

Ongoing Work

- Clustering in federated learning to detect similar subpopulations
- Propose novel bias mitigation techniques that takes into account FL constraints
- Evaluate the proposed approaches on real world datasets

References

[1] Abdulrahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C. & Guizani, M. (2020).A Survey on Federated Learning: The Journey from Centralized to Distributed On-Site Learning and Beyond. IEEE Internet of Things Journal

[2] McMahan, H. B. et al. (2021). Advances and Open Problems in Federated Learning. Foundations and Trends® in Machine Learning, 14(1).

[3] Ferraguig, L., Djebrouni, Y., Bouchenak, S., & Marangozova, V. (2021, July). Survey of Bias Mitigation in Federated Learning. In *Conférence francophone d'informatique en Parallélisme, Architecture et Système*.

[4]Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1-35.

[5] Abay (A.), Zhou (Y.), Baracaldo (N.), Rajamoni (S.), Chuba (E.) et Ludwig (H.). – Mitigating bias in federated learning.

[6] Abay (A.), Chuba (E.), Zhou (Y.), Baracaldo (N.) et Ludwig (H.). – Addressing unique fairness obstacles within federated learning. 2021