Auditions CNRS 2023

Concours 06/02 (CRCN)

Batiste Le Bars

Inria Lille

Wednesday, March 22nd, 2023

About me

2011 - 2016 Education in applied mathematics

- Master M1 MAEF (Université Paris 1)
- Master M2 MVA (ENS Paris-Saclay)

2017 - 2021 PhD in machine learning

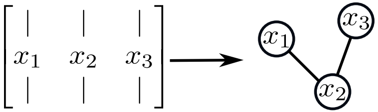
- Centre Borelli (UMR 9010, ENS Paris-Saclay), Sigfox (CIFRE PhD)
- Advisors: Nicolas Vayatis, Argyris Kalogeratos

2021 - now Post-doc

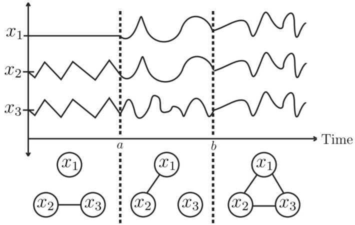
- Inria Lille, CRIStAL (UMR 9189)
- Working with: Marc Tommasi, Aurélien Bellet, Anne-Marie Kermarrec (EPFL)
- Inria-EPFL postdoc fellowship

Statistical Learning with graph-structured data

Static graph learning



Time-varying learning



- **Objective**: Infer similarity/dependency structure
- **Motivation:** Anomaly detection, Change-point detection, Application to Sigfox network
- **Tools**: Signal processing, Statistical inference, Optimization
- 4 publications (INFOCOM, ICASSP, ICML, JMLR)

Trustworthy Machine Learning

- Ethical concerns, new regulations
- Fairness, Privacy, Robustness

Contributions:

- Outlier-robust density estimation (1 paper at ICML 2022)
- Decentralized learning (1 paper at AISTATS 2023)

Trustworthy Machine Learning

- Ethical concerns, new regulations
- Fairness, Privacy, Robustness

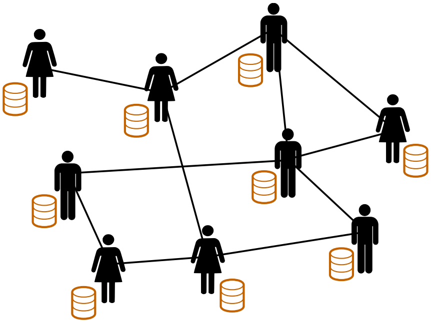
Contributions:

- Outlier-robust density estimation (1 paper at ICML 2022)
- **Decentralized learning** (1 paper at AISTATS 2023)
  - Federated learning
  - Privacy by decentralization

Federated Learning

- Decentralized Learning with decentralized data
- Centralization can be costly and implies a risk to privacy
- Collaboration is necessary (local datasets can be small or biased)

Fully decentralized FL



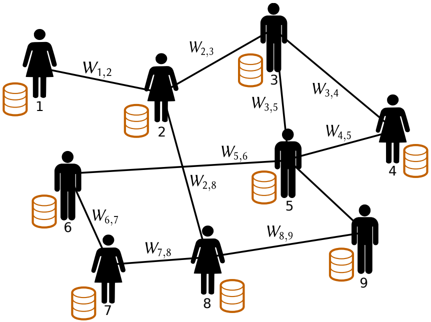
Objective: $\min_\theta \left[f(\theta) \triangleq \frac{1}{n} \sum_{i=1}^n f_i(\theta) \right]$ with f_i local loss of agent i

Algorithm: Decentralized SGD with weighted graph W

Federated Learning

- Decentralized Learning with decentralized data
- Centralization can be costly and implies a risk to privacy
- Collaboration is necessary (local datasets can be small or biased)

Fully decentralized FL



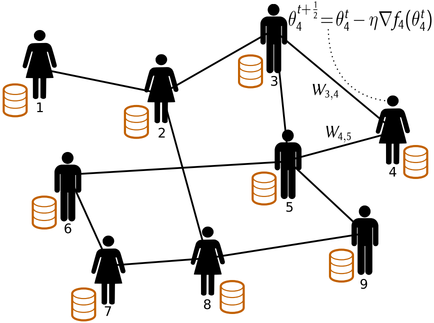
Objective: $\min_\theta \left[f(\theta) \triangleq \frac{1}{n} \sum_{i=1}^{n} f_i(\theta) \right]$ with f_i local loss of agent i

Algorithm: Decentralized SGD with weighted graph W

Federated Learning

- Decentralized Learning with decentralized data
- Centralization can be costly and implies a risk to privacy
- Collaboration is necessary (local datasets can be small or biased)

Fully decentralized FL



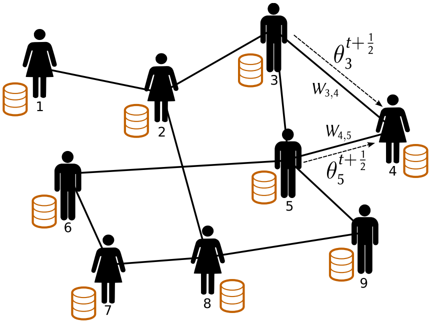
Objective: $\min_\theta \left[f(\theta) \triangleq \frac{1}{n} \sum_{i=1}^{n} f_i(\theta) \right]$ with f_i local loss of agent i

Algorithm: Decentralized SGD with weighted graph W

Federated Learning

- Decentralized Learning with decentralized data
- Centralization can be costly and implies a risk to privacy
- Collaboration is necessary (local datasets can be small or biased)

Fully decentralized FL



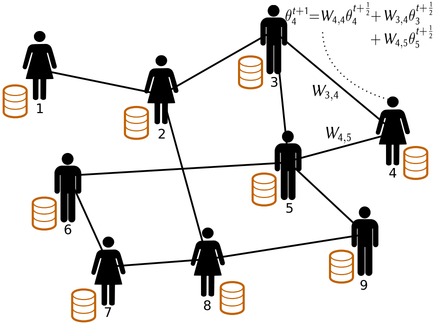
Objective: $\min_\theta \left[f(\theta) \triangleq \frac{1}{n}\sum_{i=1}^n f_i(\theta)\right]$ with f_i local loss of agent i

Algorithm: Decentralized SGD with weighted graph W

Federated Learning

- Decentralized Learning with decentralized data
- Centralization can be costly and implies a risk to privacy
- Collaboration is necessary (local datasets can be small or biased)

Fully decentralized FL



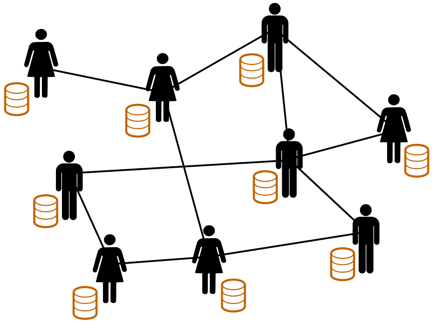
Objective: $\min_\theta \left[f(\theta) \triangleq \frac{1}{n} \sum_{i=1}^n f_i(\theta) \right]$ with f_i local loss of agent i

Algorithm: Decentralized SGD with weighted graph W

Federated Learning

- Decentralized Learning with decentralized data
- Centralization can be costly and implies a risk to privacy
- Collaboration is necessary (local datasets can be small or biased)

Fully decentralized FL



Objective: $\min_\theta \left[f(\theta) \triangleq \frac{1}{n}\sum_{i=1}^{n} f_i(\theta)\right]$ with f_i local loss of agent i

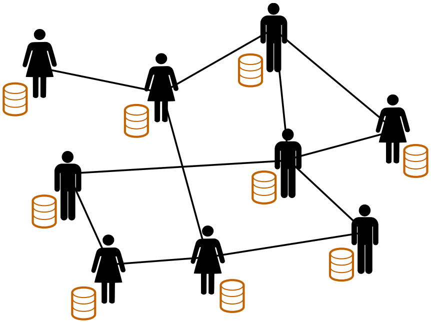
Algorithm: Decentralized SGD with weighted graph W

Challenges: Data heterogeneity, privacy, robustness, communication cost

Federated Learning

- Decentralized Learning with decentralized data
- Centralization can be costly and implies a risk to privacy
- Collaboration is necessary (local datasets can be small or biased)

Fully decentralized FL



Objective: $\min_\theta \left[f(\theta) \triangleq \frac{1}{n}\sum_{i=1}^{n} f_i(\theta)\right]$
with f_i local loss of agent i

Algorithm: Decentralized SGD with weighted graph W

Challenges: Data heterogeneity, privacy, robustness, communication cost

$\rightarrow$ How to chose the communication graph?

Impact of the communication graph - Overview

Known results

- Convergence is *strongly* impacted by **data heterogeneity**
- W well-connected $\Rightarrow$ $\searrow$ convergence time $\nearrow$ communication

Impact of the communication graph - Overview

Known results

- Convergence is *strongly* impacted by **data heterogeneity**
- W well-connected $\Rightarrow$ $\searrow$ convergence time $\nearrow$ communication

Questions

- Can the choice of graph mitigate the impact of data heterogeneity?

Impact of the communication graph - Overview

Known results

- Convergence is *strongly* impacted by **data heterogeneity**
- W well-connected $\Rightarrow$ $\searrow$ convergence time $\nearrow$ communication

Questions

- Can the choice of graph mitigate the impact of data heterogeneity?

Contribution

- First work to show that a sparse **graph can compensate the heterogeneity**
- Algorithm that learns a **sparse and data-dependent graph**

Impact of the communication graph - Overview

Known results

- Convergence is *strongly* impacted by **data heterogeneity**
- W well-connected $\Rightarrow$ $\searrow$ convergence time $\nearrow$ communication

Questions

- Can the choice of graph mitigate the impact of data heterogeneity?

Contribution

- First work to show that a sparse **graph can compensate the heterogeneity**
- Algorithm that learns a **sparse and data-dependent graph**

$\rightarrow$ A work between *decentralized optimization*, *statistical modeling* and *graph learning*

A bit of technical details

- Local heterogeneity: $\frac{1}{n}\sum_i \|\nabla f_i(\theta) - \nabla f(\theta)\|^2 \leq \zeta^2$ (previous work)
- **Neighborhood heterogeneity**: $\frac{1}{n}\sum_i \|\sum_j W_{ij}\nabla f_j(\theta) - \nabla f(\theta)\|^2 \leq \bar{\tau}^2$

  $\longrightarrow$ impact of the graph *with* the data-heterogeneity

A bit of technical details

- Local heterogeneity: $\frac{1}{n}\sum_i \|\nabla f_i(\theta) - \nabla f(\theta)\|^2 \leq \zeta^2$ (previous work)
- **Neighborhood heterogeneity**: $\frac{1}{n}\sum_i \|\sum_j W_{ij}\nabla f_j(\theta) - \nabla f(\theta)\|^2 \leq \bar{\tau}^2$

$\longrightarrow$ impact of the graph *with* the data-heterogeneity

Theorem (Informal)

The decentralization error reaches a value ε after T iterations with

$$T = \mathcal{O}\left(\frac{\bar{\tau}}{p\varepsilon^{3/2}}\right)$$

and where p is the spectral gap of W.

A bit of technical details

- Local heterogeneity: $\frac{1}{n}\sum_i \|\nabla f_i(\theta) - \nabla f(\theta)\|^2 \leq \zeta^2$ (previous work)
- **Neighborhood heterogeneity**: $\frac{1}{n}\sum_i \|\sum_j W_{ij}\nabla f_j(\theta) - \nabla f(\theta)\|^2 \leq \bar{\tau}^2$

  $\longrightarrow$ impact of the graph *with* the data-heterogeneity

Theorem (Informal)

The decentralization error reaches a value ε after T iterations with

$$T = \mathcal{O}\left(\frac{\bar{\tau}}{p\varepsilon^{3/2}}\right)$$

and where p is the spectral gap of W.

- W impacts the rate through p AND $\bar{\tau}$
- Sparse W can still make $\bar{\tau}$ small $\Rightarrow$ **Learn W by minimizing $\bar{\tau}$**

Federated learning: beyond optimization

Objectives

- Current FL techniques focus on the optimization of training errors
- In general optimizing the training performance is not enough
  $\longrightarrow$ models must **generalize** to unseen data!

Objectives

- Current FL techniques focus on the optimization of training errors
- In general optimizing the training performance is not enough
  $\longrightarrow$ models must **generalize** to unseen data!
- Optimization is only a step of the learning pipeline:
  - Anomaly detection, missing data imputation
  - Model selection, cross-validation
  - Uncertainty quantification
  - And many more
- **FL should consider these questions for real-world deployments**

Research Axes

Axis 1. Generalization in Federated Learning

Axis 2. Uncertainty Quantification in Federated Learning

$\rightarrow$ Project at the interface of *statistical learning*, *trustworthy machine learning* and *decentralized optimization*

Axis 1. Generalization in Federated Learning

- $R(\theta) = \mathbb{E}_{Z\sim\mathcal{D}}[\ell(\theta, Z)]$ (population risk)

Axis 1. Generalization in Federated Learning

- $R(\theta) = \mathbb{E}_{Z\sim\mathcal{D}}[\ell(\theta, Z)]$ (population risk)
- $R_S(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, Z_i)$ (empirical risk)

Axis 1. Generalization in Federated Learning

- $R(\theta) = \mathbb{E}_{Z\sim\mathcal{D}}[\ell(\theta, Z)]$ (population risk)
- $R_S(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, Z_i)$ (empirical risk)
- $\hat{\theta}_S = \arg\min R_S(\theta)$ (ERM)

Axis 1. Generalization in Federated Learning

- $R(\theta) = \mathbb{E}_{Z\sim\mathcal{D}}[\ell(\theta, Z)]$ (population risk)
- $R_S(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ell(\theta, Z_i)$ (empirical risk)
- $\hat{\theta}_S = \arg\min R_S(\theta)$ (ERM)
- $A(S)$, $S = \{Z_i\}_{i=1}^{n}$ (Iterative algorithm)

Axis 1. Generalization in Federated Learning

- $R(\theta) = \mathbb{E}_{Z \sim \mathcal{D}}[\ell(\theta, Z)]$ (population risk)
- $R_S(\theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(\theta, Z_i)$ (empirical risk)
- $\hat{\theta}_S = \arg\min R_S(\theta)$ (ERM)
- $A(S)$, $S = \{Z_i\}_{i=1}^{n}$ (Iterative algorithm)

$$R(A(S)) - R(\theta^*) \leq \underbrace{R(A(S)) - R_S(A(S))}_{\textit{Generalization}} + \underbrace{R_S(A(S)) - R_S(\hat{\theta}_S)}_{\textit{Optimization}}$$

Axis 1. Generalization in Federated Learning

Short/mid-term objectives (1-3 years)

- Reveal the **impact of decentralization on generalization**: communication graph, data heterogeneity, asynchronous communication

  $\rightarrow$ using stability analysis, Information-Theoretic generalization bounds

- **Algorithmic developments**: improve generalization performance

Mid-long-term objectives (3-5 years)

- Better generalization with **personalized models**
- Propose **unified framework** for consensus vs personalized
- **Contribution to generalization analysis** for ML in general

Axis 1. Generalization in Federated Learning

Axis 2. Uncertainty Quantification in Federated Learning

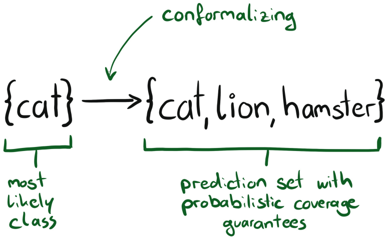
Axis 2. Uncertainty Quantification in Federated Learning

Measuring data-heterogeneity

- Heterogeneity has a strong impact on optimization; and generalization?
- Motivation: data-analysis, **model selection**, hyperparameter tuning

Uncertainty in the prediction

- Strong variance in the prediction
- Scalar prediction are not sufficiently conservative
  $\to$ predict intervals
- **Conformal prediction** in FL



Integration project

UMR 7243 Laboratoire d'analyse et modélisation de systèmes pour l'aide à la décision (LAMSADE)

- MILES team (head: Jamal Atif)
- Trustworthy ML (Privacy and robustness)
- Optimization, high-dimensional learning

UMR 9189 Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL)

- MAGNET team (head: Marc Tommasi)
- Trustworthy ML (Fairness, Privacy, Federated Learning)

List of publications

- **B. Le Bars**, A. Bellet, M. Tommasi, E. Lavoie, A-M. Kermarrec. *Refined convergence and topology learning for decentralized sgd with heterogeneous data.* AISTATS, 2023.
- P. Humbert*, **B. Le Bars***, L. Minvielle. *Robust kernel density estimation with median-of-means principle.* ICML,2022.
- P. Humbert*, **B. Le Bars***, L. Oudre, A. Kalogeratos, N. Vayatis. *Learning laplacian matrix from graph signals with sparse spectral representation.* JMLR, 2021.
- **B. Le Bars**, P. Humbert, A. Kalogeratos, N. Vayatis. *Learning the piece-wise constant graph structure of a varying ising model.* ICML 2020.
- **B. Le Bars***, P. Humbert*, L. Oudre, A. Kalogeratos. *Learning laplacian matrix from bandlimited graph signals.* ICASSP 2019.
- **B. Le Bars**, A. Kalogeratos. *A probabilistic framework to node-level anomaly detection in communication networks.* INFOCOM 2019.

STL-FW - Objective

Proposition

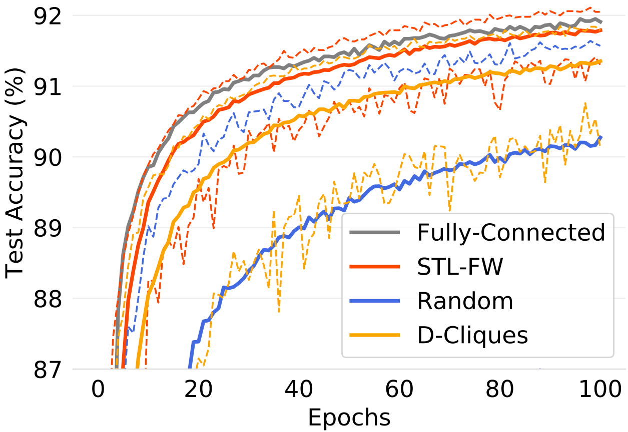
$\exists\, \lambda > 0$ s.t. neighborhood heterogeneity H is upper bounded by

$$H \leq g(W) \triangleq \frac{1}{n}\left\| W\Pi - \frac{\mathbf{1}\mathbf{1}^\top}{n}\Pi \right\|_F^2 + \frac{\lambda}{n}\left\| W - \frac{\mathbf{1}\mathbf{1}^\top}{n} \right\|_F^2$$

Objective: Minimize $g(W)$ s.t. W **doubly stochastic**

- Avoid trivial (dense) solution $W = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$
- Find W sparse instead: using Frank-Wolfe!

STL-FW - Results



$$d_{\text{max}} = 5$$