When The Model Goes to the Data: a journey through Federated Learning

Scuola Ortogonale

Fondazione ELICSIR

Nicola Figus

Anno Accademico 2025/2026

Mentore: Prof.ssa Valeria Cardellini

elicsir

Elevating Italian Computer Science Innovation and Research

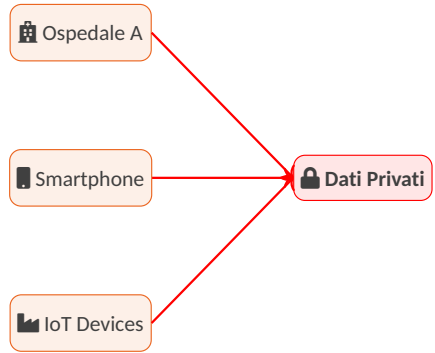
Il Dilemma dei Dati

Il paradosso del Deep Learning moderno:

- I modelli richiedono **enormi quantità di dati**
- I dati più preziosi sono **distribuiti** e **sensibili**
- Regolamenti (GDPR, HIPAA) **impediscono la centralizzazione**

La domanda chiave

Come addestrare modelli **accurati** senza mai spostare i dati dal dispositivo che li ha generati?



Dati Decentralizzati: Lo Scenario Reale

Approccio Tradizionale (Centralizzato):

1. Raccogliere tutti i dati in un server
2. Addestrare il modello centralmente
3. Distribuire il modello addestrato ai client
4. **Problemi:**
   - Regolamenti sulla privacy
   - Costi di trasferimento

Approccio Federato:

1. I dati **restano sui dispositivi**
2. Ogni client addestra **localmente**
3. Si condividono solo i **parametri del modello**
4. **Risultato:**
   - Privacy preservata
   - Modello globale accurato

"Il modello va ai dati, non i dati al modello." Essenza del FL

Formulazione del Problema

Funzione Obiettivo Globale — K client, ciascuno con dataset locale $\mathcal{D}_k$ di dimensione n_k:

$$\min_{w \in \mathbb{R}^d} f(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w)$$

dove $n = \sum_k n_k$ e la **funzione obiettivo locale** di ciascun client k è:

$$F_k(w) = \frac{1}{n_k} \sum_{i \in \mathcal{D}_k} \ell(w; x_i, y_i)$$

Vincoli

- Nessun accesso diretto a $\mathcal{D}_k$
- Comunicazione limitata
- Client eterogenei

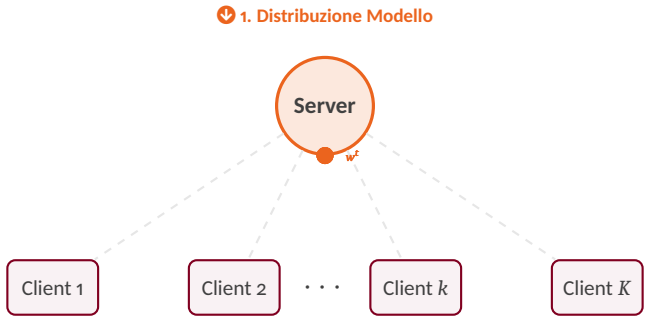
Sfide

- $\mathcal{D}_k$ possono essere **non-IID**
- Client **eterogenei**
- Client possono **disconnettersi**

Indice

FedAvg: L'Idea Chiave

1 FedAvg: Federated Averaging



Vantaggio chiave: ogni client esegue **più passi locali** di Stochastic Gradient Descent (SGD) prima della comunicazione $\Rightarrow$ **riduzione drastica** dei round di comunicazione

FedAvg: L'Algoritmo

Federated Averaging (McMahan et al. 2017)

Input: K client, learning rate η, epoche locali E, frazione di client C, dimensione batch B

ServerComputation:

Inizializza w^0

for round $t = 0, 1, 2, \ldots$ **do**

$S_t \leftarrow$ campiona $\max(C \cdot K, 1)$ client

for ogni client $k \in S_t$ **in parallel do**

$w_k^{t+1} \leftarrow \text{ClientUpdate}(k, w^t)$

$$w^{t+1} \leftarrow \sum_{k \in S_t} \frac{n_k}{\sum_{j \in S_t} n_j} w_k^{t+1}$$

ClientUpdate(k, w):

$\mathcal{B} \leftarrow$ split $\mathcal{D}_k$ in mini-batch di B

for epoca $e = 1, \ldots, E$ **do**

for batch $b \in \mathcal{B}$ **do**

$w \leftarrow w - \eta \nabla \ell(w; b)$

return w

FedAvg: Iperparametri Chiave

C — Frazione di client

- $C = 1$: tutti i client partecipano
- $C \ll 1$: solo un sottoinsieme
- Trade-off **velocità vs. varianza**
- McMahan: $C = 0.1$ (default)

B — Dimensione batch

- $B = \infty$: Full-batch (`FedSGD`)
- B piccolo: più passi locali

E — Epoche locali

- $E = 1$: un'epoca per round
- E alto $\Rightarrow$ meno round
- **Attenzione:** E troppo grande $\Rightarrow$ *client drift*

η — Learning rate

- Learning rate locale (client)
- Può richiedere *decay*
- Interagisce con E e B

FedAvg: Risultati Chiave

1 FedAvg: Federated Averaging

Configurazione	Round per target accuracy	Speedup vs FedSGD
FedSGD ($E=1,\ B=\infty$)	$\sim$1000	1$\times$
FedAvg ($E=1,\ B=10$)	$\sim$300	3.3$\times$
FedAvg ($E=5,\ B=10$)	$\sim$50	**20**$\times$

Risultati indicativi su MNIST (2NN, IID). Adattato da McMahan et al. (2017).

✓ Punti di forza

- Riduzione $10-100\times$ della comunicazione
- Semplice da implementare
- Funziona con varie architetture

✗ Limitazioni

- Convergenza **non garantita** con dati non-IID
- Sensibile a **eterogeneità di sistema**
- Nessun meccanismo per **straggler**

Indice

2 Sfide: Eterogeneità Statistica e di Sistema

Eterogeneità Statistica: Dati Non-IID

2 Sfide: Eterogeneità Statistica e di Sistema

- **Label distribution skew:** ogni client ha solo alcune classi
- **Feature distribution skew:** stessa label, feature diverse (es. foto con diversi dispositivi)
- **Quantity skew:** client con dataset di dimensioni molto diverse

Impatto: con dati fortemente non-IID, FedAvg può **divergere** o convergere a soluzioni **sub-ottimali**

Eterogeneità di Sistema

2 Sfide: Eterogeneità Statistica e di Sistema

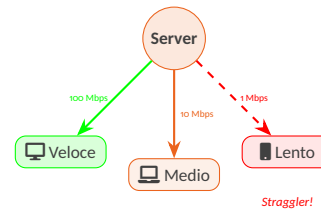
I client non sono tutti uguali!

Risorse Computazionali

- CPU/GPU diverse
- Memoria RAM e Storage
- Alimentazione (batteria)

Connettività

- Banda di rete variabile
- Latenza diversa
- Connessione intermittente
- WiFi vs. 4G/5G



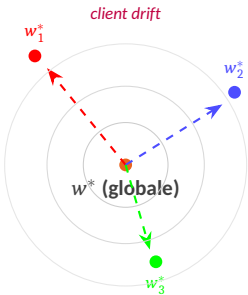
Problema degli *straggler*:
Il server deve aspettare il client **più lento**. Se un client non risponde, il round è incompleto.

Il *Client Drift*: Quando i Modelli Divergono

2 Sfide: Eterogeneità Statistica e di Sistema

Con dati Non-IID e molte epoche locali (E grande):

- Ogni client ottimizza verso il **suo ottimo locale** w_k^*
- Gli ottimi locali possono essere **molto diversi** dall'ottimo globale w^*
- Più epoche locali $\Rightarrow$ più *drift*
- L'aggregazione combina modelli che convergono verso **ottimi divergenti**



Conseguenza: il modello globale **converge** a una soluzione sub-ottimale. In casi estremi: **divergenza**.

Indice

3 FedProx: Ottimizzazione Federata Eterogenea

FedProx: L'Intuizione

Li et al. (2020) — *Federated Optimization in Heterogeneous Networks*

Due problemi chiave di FedAvg:

1. **Eterogeneità Statistica:**
   dati non-IID $\Rightarrow$ client drift
2. **Eterogeneità dei Sistemi:**
   straggler $\Rightarrow$ round incompleti

Soluzione di FedProx:

- Aggiungere un **termine prossimale** all'obiettivo locale
- Permettere **soluzioni inesatte** (lavoro parziale)

FedAvg: ogni client minimizza la propria $F_k(w)$

$\Downarrow$

FedProx: ogni client minimizza $F_k(w)$ **restando vicino** al modello globale w^t

FedAvg è un caso speciale di FedProx

Il Termine Prossimale

3 FedProx: Ottimizzazione Federata Eterogenea

Obiettivo locale modificato per il client k al round t:

$$h_k(w; w^t) = F_k(w) + \underbrace{\frac{\mu}{2}\|w - w^t\|^2}_{\text{termine prossimale}}$$

Cosa fa il termine $\frac{\mu}{2}\|w - w^t\|^2$?

- **Penalizza** soluzioni locali lontane da w^t
- **Limita** il client drift
- Equivale a **regolarizzazione L2** centrata su w^t

Il ruolo di μ

- $\mu = 0$: si riduce a **FedAvg**
- μ piccolo: leggera regolarizzazione
- μ grande: locale resta **vicino** a w^t
- Scelta ottimale non banale

Soluzioni Inesatte e Partecipazione Parziale

3 FedProx: Ottimizzazione Federata Eterogenea

Innovazione chiave: FedProx **non richiede** che ogni client completi E epoche!

γ-inexact solution

Il client k restituisce w_k tale che:

$$\|\nabla h_k(w_k; w^t)\| \leq \gamma_k \|\nabla h_k(w^t; w^t)\|, \quad \gamma_k \in [0, 1)$$

Vantaggi per l'eterogeneità di sistema:

- Client veloci: più iterazioni
- Client lenti: meno iterazioni
- **Nessun client viene scartato**
- Convergenza **garantita**

Confronto con FedAvg:

- FedAvg: **ignora** chi non completa
- FedProx: **include** contributi parziali
- Migliore utilizzo di **tutti** i client
- Robusto rispetto agli **straggler**

FedAvg vs FedProx: Confronto

3 FedProx: Ottimizzazione Federata Eterogenea

Proprietà	FedAvg	FedProx
Obiettivo locale	$F_k(w)$	$F_k(w) + \frac{\mu}{2}\|w - w^t\|^2$
Dati IID	✓ Buono	✓ Buono
Dati Non-IID	✗ Instabile	✓ Stabile
Straggler	✗ Scartati	✓ Inclusi
Soluzioni inesatte	✗ No	✓ γ-inexact
Convergenza	Empirica	✓ Garanzie formali
Iperparametro extra	—	μ (prossimale)

FedProx aggiunge **minima complessità** (μ) ottenendo **convergenza robusta** in scenari eterogenei

FedProx: Risultati Sperimentali (Li et al. 2020)

Dataset utilizzati

- **MNIST** (Non-IID partition)
- **FEMNIST** (Federated EMNIST)
- **Shakespeare** (next-character)
- **Sent140** (sentiment analysis)

Risultati principali

- Fino a **22%** miglioramento di accuratezza vs FedAvg su dati non-IID
- Convergenza **più stabile**
- Robusto a **straggler attivi**

Impatto degli Straggler

% Straggler	FedAvg	FedProx
0%	95.2%	95.4%
50%	91.1%	**94.8%**
90%	82.3%	**93.5%**

Accuratezze indicative su FEMNIST

Indice

4 Oltre FedProx: Efficienza e Personalizzazione

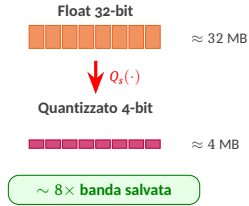
Comunicazione Efficiente: FedPAQ

4 Oltre FedProx: Efficienza e Personalizzazione

Reisizadeh et al. (2020) — *Federated Learning with Periodic Averaging and Quantization*

Il problema:

- Ogni round: trasferimento dell'**intero modello**
- Reti moderne: $10^7 - 10^9$ parametri
- Client edge: banda **costosa** e **intermittente**



L'Idea

Quantizzare i messaggi: $w \to Q_s(w)$ + **periodic averaging** (τ passi locali)

Personalizzazione: Un Modello per Ciascuno?

4 Oltre FedProx: Efficienza e Personalizzazione

Limite implicito di FedAvg/FedProx: producono **un solo modello globale** per client profondamente eterogenei

Fine-Tuning Locale

- Train globale → adattamento locale
- Globale come *starting point*
- **Per-FedAvg, pFedMe**: meta-learning

Clustered FL

- Gruppi di client con dati **simili**
- Un modello **per cluster**
- **FedBN, IFCA, CFL**

Nel campo medicale: la personalizzazione è l'essenza della **medicina di precisione**

Indice

5 Caso di Studio: Analisi di Immagini Mediche

Caso di Studio: Diagnosi di PTT da Immagini di Sangue

5 Caso di Studio: Analisi di Immagini Mediche

Porpora Trombotica Trombocitopenica (PTT):

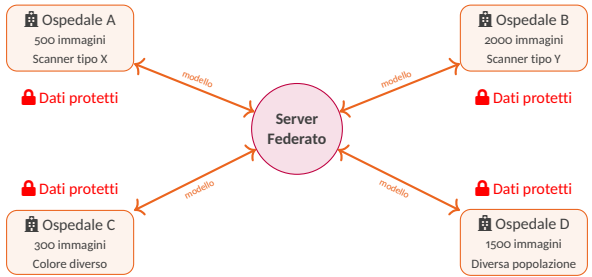
- Malattia rara e potenzialmente letale
- Diagnosi tramite **analisi di strisci di sangue periferico**
- Marcatore diagnostico: **schistociti**
- Diagnosi manuale: lenta, soggettiva, variabile

Soluzione: Classificazione automatica con **Deep Learning**

Problema: **Come gestire i dataset sensibili? ⇒ Federated Learning!**

Perché il Federated Learning per la Diagnostica?

5 Caso di Studio: Analisi di Immagini Mediche



Non-IID
Diverse popolazioni di pazienti, diversi protocolli

Quantity skew
Ospedali grandi vs. piccoli (300 vs 2000 img)

Feature skew
Diversi scanner, colorazioni, risoluzioni

FL per Imaging Medico: Opportunità e Sfide

5 Caso di Studio: Analisi di Immagini Mediche

✓ Opportunità

- **Privacy:** dati mai centralizzati (compliance GDPR/HIPAA)
- **Scale:** accesso a dataset distribuiti molto più grandi
- **Generalizzazione:** modello robusto a diverse condizioni di acquisizione
- **Collaborazione:** ospedali possono cooperare senza condividere dati

✗ Sfide aperte

- **Domain shift:** forte eterogeneità di feature
- **Annotazioni:** qualità variabile tra ospedali
- **Class imbalance:** malattie rare $=$ pochi positivi
- **Comunicazione:** immagini mediche $\Rightarrow$ modelli pesanti

FedProx è particolarmente adatto al contesto medico: gestisce l'eterogeneità statistica tra ospedali e la partecipazione parziale di strutture con risorse limitate

Indice

6 Conclusioni e Direzioni Future

Direzioni Future e Conclusioni

6 Conclusioni e Direzioni Future

In Sintesi

1. **FedAvg**: paradigma fondante
2. **FedProx**: robustezza all'eterogeneità
3. **FedPAQ**: efficienza di comunicazione
4. **Personalizzazione**: fine-tuning oltre il modello unico

Sistemi Reali

Flower, NVIDIA FLARE, OpenFL

Framework open source

Frontiere ancora aperte

- **Privacy:** *differential privacy*, *secure aggregation*
- **Robustezza:** difesa da client *bizantini*
- **Fairness:** equità tra client eterogenei
- **Asincronia:** FL senza sincronizzazione

Riferimenti Bibliografici

6 Conclusioni e Direzioni Future

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas,
"Communication-Efficient Learning of Deep Networks from Decentralized Data,"
Proc. AISTATS, PMLR 54, pp. 1273–1282, 2017.

[2] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, V. Smith,
"Federated Optimization in Heterogeneous Networks,"
Proc. MLSys, 2020.

[3] A. Reisizadeh, A. Mokhtari, H. Hassani, A. Jadbabaie, R. Pedarsani,
"FedPAQ: A Communication-Efficient Federated Learning Method with Periodic Averaging and Quantization,"
Proc. AISTATS, PMLR 108, pp. 2021–2031, 2020.

Riferimenti Bibliografici

6 Conclusioni e Direzioni Future

[4] A. Fallah, A. Mokhtari, A. Ozdaglar,

"Personalized Federated Learning with Theoretical Guarantees: A Model-Agnostic Meta-Learning Approach,"

Proc. NeurIPS, 2020.

[5] C. T. Dinh, N. H. Tran, T. D. Nguyen,

"Personalized Federated Learning with Moreau Envelopes,"

Proc. NeurIPS, 2020.

[6] D. J. Beutel et al.,

"Flower: A Friendly Federated Learning Research Framework,"

arXiv:2007.14390, 2020.

Riferimenti Bibliografici

6 Conclusioni e Direzioni Future

[7] T. Li, A. K. Sahu, A. Talwalkar, V. Smith,
"Federated Learning: Challenges, Methods, and Future Directions,"
IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, 2020.

[8] P. Kairouz et al.,
"Advances and Open Problems in Federated Learning,"
Foundations and Trends in Machine Learning, vol. 14, no. 1–2, 2021.

[9] N. Rieke et al.,
"The Future of Digital Health with Federated Learning,"
npj Digital Medicine, vol. 3, no. 119, 2020.

Grazie per l'attenzione!