GAL:
Gradient Assisted Learning for Decentralized Multi-Organization Collaborations

Presenter: Enmao Diao

Paper

Code

Enmao Diao[1] Jie Ding[2] Vahid Tarokh[1]

[1]Duke University [2]University of Minnesota

Overview

- **Motivation**
- **Gradient Assisted Learning**
  - Vertically-distributed data
  - Objectives
  - The GAL algorithm
- **Experiments**
  - Model Autonomy
  - Comparison with Assisted Learning (AL)
  - Case Studies
- **Conclusion**

Motivation

- Fusion of knowledge from numerous decentralized organizations
- Unable to centralize their data and fully collaborate to learn a shared model

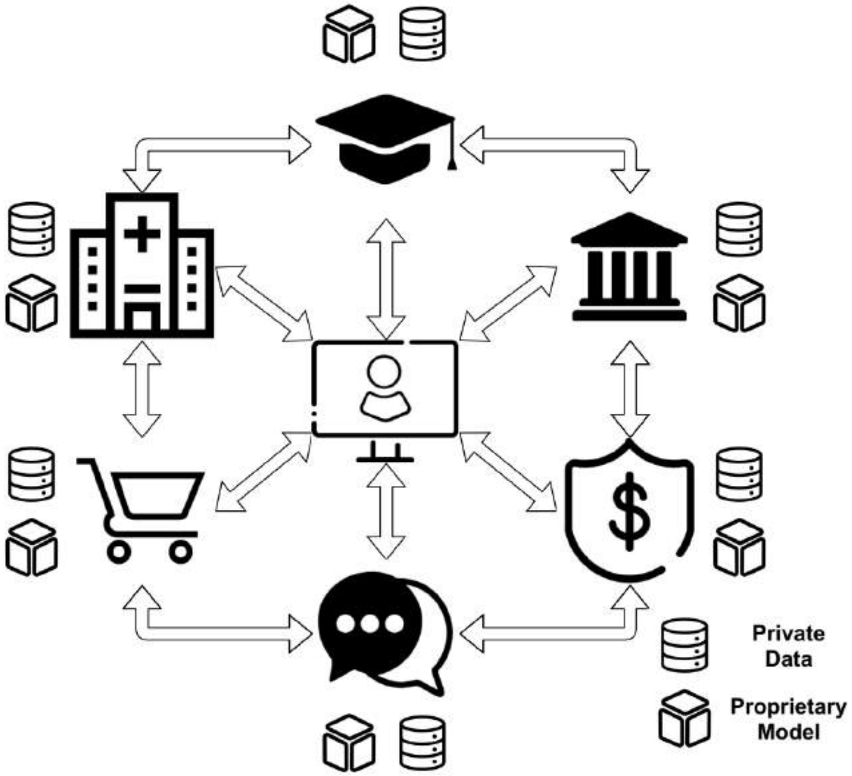

Figure 1. Decentralized organizations form a community of shared interest to provide better Machine-Learning-as-a-Service.

Gradient Assisted Learning (GAL)

- **Vertically-distributed data**
  - Suppose there are M organizations. Each organization m only holds X_m, a sub-vector of the joint data X
  - Sponsor (Alice) : X_1, y_1
  - Assistors: $X_2, \dots, X_M$

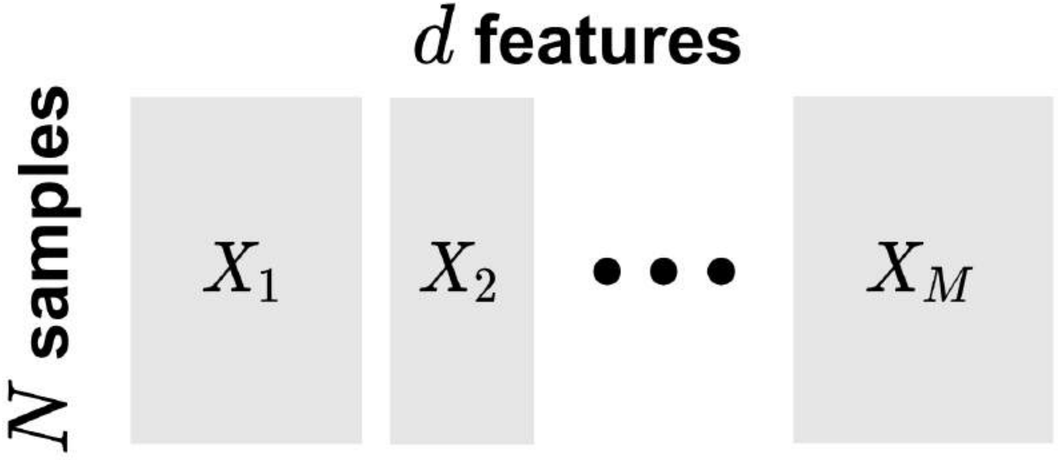

Figure 2. An illustration of organizations' vertically distributed data.

Gradient Assisted Learning (GAL)

- **Objectives**
  - Ideal case

$$F_{\text{Joint}} = \underset{F \in \mathcal{F}}{\operatorname{argmin}} \, \mathbb{E}_N L_1(y_1, F(x))$$

  - Without assistance

$$F_{\text{Alone}} = \underset{F_1 \in \mathcal{F}_1}{\operatorname{argmin}} \, \mathbb{E}_N L_1(y_1, F_1(x_1))$$

  - Gradient assistance
    - Our objective

$$f_m = \underset{f \in \mathcal{F}_m}{\operatorname{argmin}} \, \mathbb{E}_N \ell_m(r_1, f(x_m)) = \underset{f \in \mathcal{F}_m}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \ell_m\left(r_{i,1}, f(x_{i,m})\right)$$

    - Pseudo-residuals

$$r_1 = -\left[\frac{\partial L_1(y_1, F(x))}{\partial F(x)}\right]$$

Gradient Assisted Learning (GAL)

- **The GAL algorithm**
  - Asymptotic convergence analysis
  - Reduce to the standard gradient boosting algorithm when there is only one organization

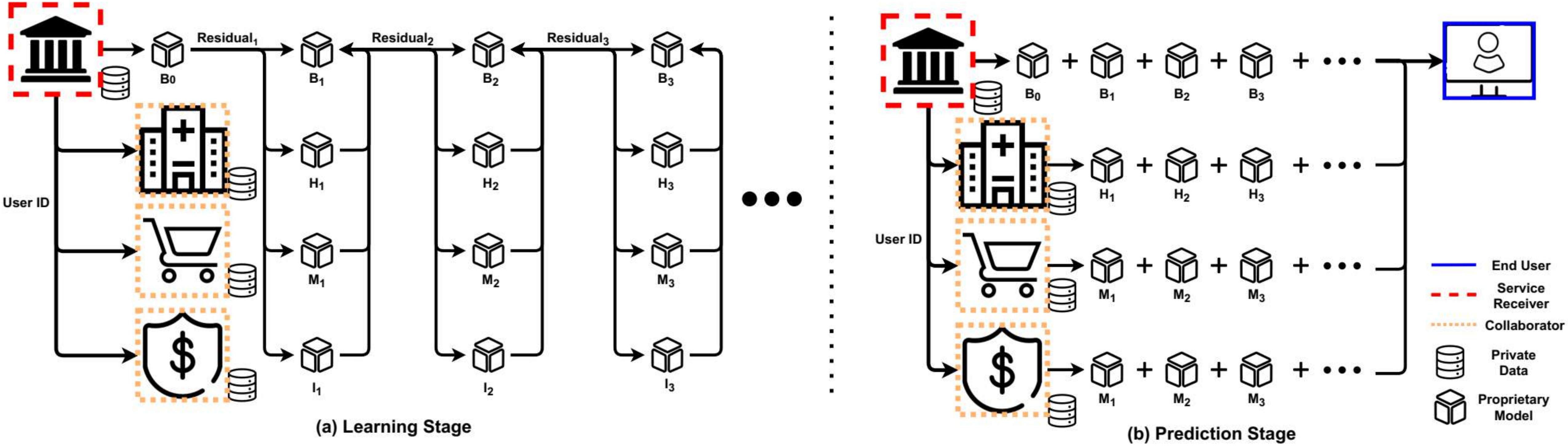

Figure 3. Learning and Prediction Stages for Gradient Assisted Learning (GAL).

Experiments

- **Model Autonomy**
  - An organization with little informative data and free choice of its local model (model autonomy) can leverage others' local data and models and even achieve near-oracle performance.

Table 1: Results of the UCI datasets ($M = 8$) with Linear, GB, SVM and GB-SVM models. The Diabetes and Boston Housing (regression) are evaluated with Mean Absolute Deviation (MAD), and the rest (classification) are evaluated with Accuracy.

Dataset	Model	Diabetes(↓)	BostonHousing(↓)	Blob(↑)	Wine(↑)	BreastCancer(↑)	QSAR(↑)
Late	Linear	136.2(0.1)	8.0(0.0)	100.0(0.0)	100.0(0.0)	96.9(0.4)	76.9(0.8)
Joint	Linear	43.4(0.3)	3.0(0.0)	100.0(0.0)	100.0(0.0)	98.9(0.4)	84.0(0.2)
Alone	Linear	59.7(9.2)	5.8(0.9)	41.3(10.8)	63.9(15.6)	92.5(3.4)	68.8(3.4)
AL	Linear	51.5(4.6)	4.7(0.6)	97.5(2.5)	95.1(3.6)	97.7(1.1)	70.6(5.2)
GAL	Linear	42.7(0.6)	3.2(0.2)	100.0(0.0)	96.5(3.0)	98.5(0.7)	82.5(0.8)
GAL	GB	56.5(2.8)	3.8(0.5)	96.3(2.2)	95.8(1.4)	96.1(1.0)	84.8(0.9)
GAL	SVM	46.6(1.4)	2.9(0.2)	96.3(4.1)	96.5(1.2)	99.1(1.1)	85.5(0.7)
GAL	GB-SVM	49.8(2.6)	3.4(0.8)	70.0(7.9)	95.8(1.4)	93.2(1.6)	82.9(1.5)

Experiments

- **Comparison with Assisted Learning (AL)**

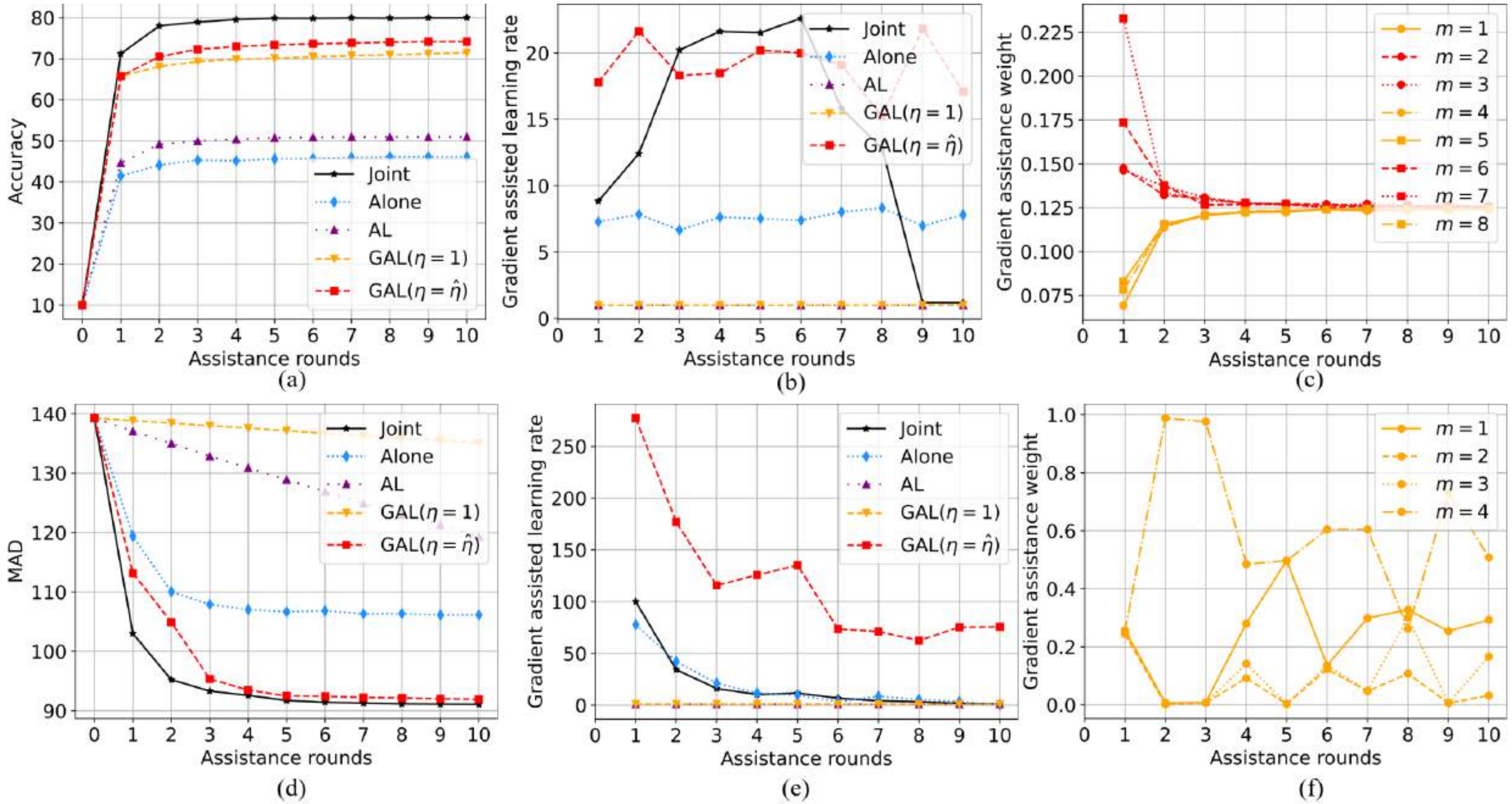

Figure 4. Results of the CIFAR10 (a-c) ($M = 8$) and MIMICL (d-f) ($M = 4$) datasets. GAL significantly outperforms 'Alone' and 'AL' and performs close to the centralized baselines.

Experiments

- **Case Studies**
  - Deep Model Sharing
  - Three-dimensional object recognition
  - Medical time series forecasting

Table 2: Results of case studies of 3D object recognition and medical time series forecasting.

Dataset	ModelNet40(↑)	ShapeNet55(↑)	MIMICL(↓)	MIMICM(↑)
Interm	75.3(18.2)	88.6(0.1)	64.6(0.9)	0.90(0.0)
Late	86.6(0.2)	88.4(0.1)	71.4(0.2)	0.91(0.0)
Joint	46.3(1.4)	16.3(0.0)	91.1(0.7)	0.82(0.0)
Alone	76.4(1.1)	81.3(0.6)	106.1(0.3)	0.78(0.0)
AL	77.3(2.8)	83.8(0.0)	119.3(0.3)	0.86(0.0)
GAL	83.0(0.2)	84.1(0.6)	**91.9(2.3)**	**0.88(0.0)**
GAL_{DMS}	**83.2(0.3)**	**85.3(0.2)**	97.7(2.9)	0.81(0.0)

Conclusion

- We propose a Gradient Assisted Learning (GAL) algorithm that is suitable for large-scale autonomous decentralized learning.

- We provide asymptotic convergence analysis and practical case studies of GAL. For the case of vertically distributed data, GAL generalizes the classical Gradient Boosting algorithm.

- Our proposed method can significantly outperform learning baselines and achieve near-oracle performance on various benchmark datasets.

- Future works can study GAL with Adversarial Learning, Fairness, and Automated machine learning

Thank you!