PICSR: Prototype-Informed Cross-Silo Router for Federated Learning (Student Abstract)

Eric Enouen[1,3], **Sebastian Caldas**[2,3], **Mononito Goswami**[3], **Artur Dubrawski**[3]

[1]The Ohio State University
[2]Princeton University
[3]Auton Lab, School of Computer Science, Carnegie Mellon University
enouen.9@osu.edu, scaldas@princeton.edu, {mgoswami, awd}@cs.cmu.edu

Abstract

Federated Learning is an effective approach for learning from data distributed across multiple institutions. While most existing studies are aimed at improving predictive accuracy of models, little work has been done to explain knowledge differences between institutions and the benefits of collaboration. Understanding these differences is critical in cross-silo federated learning domains, *e.g.*, in healthcare or banking, where each institution or silo has a different underlying distribution and stakeholders want to understand how their institution compares to their partners. We introduce *Prototype-Informed Cross-Silo Router* (PICSR) which utilizes a mixture of experts approach to combine local models derived from multiple silos. Furthermore, by computing data similarity to prototypical samples from each silo, we are able to ground the router's predictions in the underlying dataset distributions. Experiments on a real-world heart disease prediction dataset show that PICSR retains high performance while enabling further explanations on the differences among institutions compared to a single black-box model.

Introduction

Data heterogeneity is a limiting factor for federated learning in practice, especially in cross-silo settings, where institutions often have large datasets with different distributions, representing distinct yet complementary knowledge bases. To learn effective local and global models, federated learning methods must be able to leverage the intrinsic heterogeneity in such distributed datasets.

Heterogeneous federated learning approaches can learn and utilize the differences amongst institutions, training a model that performs better despite the heterogeneity. However, the vast majority of these methods focus on performance, resulting in a final black-box model that must be trusted blindly. This is insufficient in many real-world application domains, where stakeholders need to understand how institutions differ from their own, what benefits they are receiving from participation in model sharing, and the limitations of their own models (Caldas et al. 2021).

We aim to answer the following question: **how do other institutions differ from my own?** To accomplish this we propose a novel Prototype-Informed Cross-Silo Router (PICSR) to learn a mapping from samples to silos. It utilizes a mixture of experts framework to enable each expert to specialize to its own institution's data while the router decides how to ensemble their predictions. This design allows for not only high performance in the cross-silo setting, but the model framework allows for stakeholders to analyze the router directly by inspecting the weights it assigns to the different local experts.

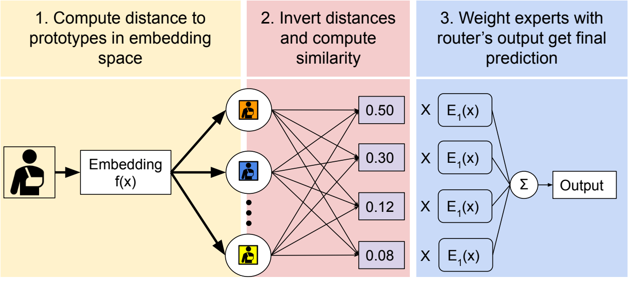
1. Compute distance to prototypes in embedding space
2. Invert distances and compute similarity
3. Weight experts with router's output get final prediction
Embedding f(x)
0.50
0.30
0.12
0.08
X
E1(x)
X
E1(x)
X
E1(x)
X
E1(x)
Σ
Output

Figure 1: PICSR ensembles predictions from silo-specific expert models (E_i) using a router grounded in prototypical samples. The output weights of the router can help stakeholders understand how their institution differs from others.

Methodology

Model Architecture

Our framework is depicted in Figure 1. We utilize a prototype-informed architecture to train a routing model to ensemble a group of expert models together. There are two core components to our method: our federated mixture of experts approach to learning an appropriate weighted ensemble to use for each sample, and the prototype embedding in the router to ensure the decisions are grounded in the silo's data distributions.

Federated Mixture of Experts We train a router with parameters Θ, which weighs predictions from K experts with parameters $\mathbf{w}_i$ denoted by $E_i(\mathbf{x}, \mathbf{w}_i), i \in [K]$ on a given data point $\mathbf{x}$. The router assigns each expert a real-valued normalized scalar weight denoted by $h(\mathbf{x}; \Theta)_i \in \mathbb{R}$ and per-

	Local	Ensemble	FedAvg	FedAvg + FT	FedProx	PICSR
Cleveland	77.31±0.77	**79.81±1.92**	75.77±1.96	77.12±0.38	76.92±0.61	77.88±2.11
Hungary	77.75±2.18	78.88±2.29	79.78±1.00	78.65±0.00	78.88±0.45	**80.67±1.31**
Switzerland	90.00±5.00	81.25±3.95	90.00±5.00	**93.75±0.00**	88.75±4.68	90.00±3.06
VA	76.89±2.27	77.78±3.14	80.44±1.66	81.78±0.89	81.33±1.09	**82.67±1.66**

Table 1: PICSR Improves Performance. Mean accuracy over 5 runs, best models in bold, second best underlined.

forms soft-routing as follows:

$$\hat{y} = \sum_{i=1}^{K} h(\mathbf{x};\Theta)_i \times E_i(\mathbf{x},\mathbf{w}_i), \ \sum_{i=1}^{K} h(\mathbf{x};\Theta)_i = 1 \tag{1}$$

This allows each expert to learn to specialize to the distribution that they are most equipped to handle, and allows the router to divide up the problem space.

Prototype-Informed Router By allowing the router to learn how to divide the problem space, stakeholders can analyze the outputs of the router to understand how their local model differs from other participants. We choose to ground the output of the router in the similarity scores from each institution by utilizing prototypes.

We define a prototype for a given silo as the mean sample in the silo's training dataset. In the tabular setting, each silo computes this mean sample as the average of each feature.

As shown in Figure 1, we first use the embedding function f to map a sample x to a latent space. We compare the embedded samples to the embedded prototypes using the L_2 distance, $dist(x, p_i) = ||f(x) - f(p_i)||_2$, where i indexes each prototype. These distance scores are then inverted and fed into a final linear prediction layer to compute the importance of each prototype for the final prediction. The output of the routing model is then used to weight the expert models and compute the final prediction.

Case Study: Heart Disease

We first evaluate our approach on the Fed-Heart-Disease (FHD) dataset, following the implementation of (du Terrail et al. 2022). The dataset has 740 samples naturally split up into four silos based on the hospital: Cleveland ($n = 303$), Hungarian ($n = 261$), Switzerland ($n = 46$), and VA Long Beach ($n = 130$). Each sample has 13 features, and the task is to predict whether a patient has heart disease or not.

We compare our method to five baselines. "Local" reflects local models trained only on their local datasets. "Ensemble" averages the local model outputs for the final prediction. We also test against three common federated learning approaches: "FedAvg", "FedAvg + Fine-tuning", "FedProx" (see Appendix).

PICSR Improves Performance As shown in Table 1, PICSR outperforms an ensemble on most hospitals, demonstrating the value of collaboration and intelligently weighing the predictions of local models. PICSR also performed better than most federated approaches. Hence, we can conclude that PICSR improves predictive performance over frequently-used baselines on the FHD dataset.

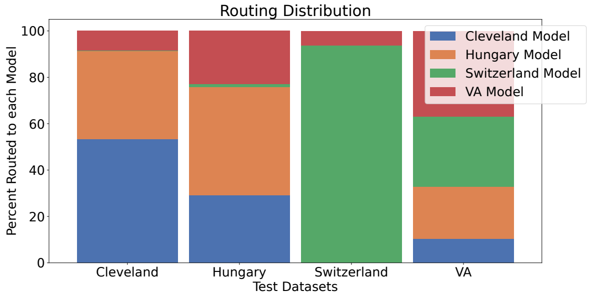

Figure 2: Routing Distribution for the test datasets of all four silos in the FHD dataset. Such plots can enable stakeholders to better understand differences among institutions.

PICSR Benefits Stakeholders By analyzing the weights produced by the router (Figure 2), stakeholders can better understand the differences amongst institutions as well as the benefits of collaboration. For example, Switzerland does not benefit much from collaboration with other hospitals, and that is because its patient population differs a lot from the other institutions. The router reflects this by routing the majority of samples in the Switzerland test dataset back to the Switzerland model. However, the patient distributions of Cleveland and Hungary are similar, and the hospitals can benefit from collaborating together as reflected in the routing distribution.

Acknowledgements

This work was partially supported by the Research Experiences for Undergraduates program under National Science Foundation grant 1730147, and by the U.S. Army Research Office and the U.S. Army Futures Command under contract W911NF-20-F-0020.

References

Caldas, S.; Yoon, J. H.; Pinsky, M. R.; Clermont, G.; and Dubrawski, A. 2021. Understanding clinical collaborations through federated classifier selection. In *Machine Learning for Healthcare Conference*, 126–145. PMLR.

du Terrail, O.; et al. 2022. FLamby: Datasets and Benchmarks for Cross-Silo Federated Learning in Realistic Healthcare Settings. In *Advances in Neural Information Processing Systems*, volume 35, 5315–5334. Curran Associates, Inc.