Tao SHEN, Jie ZHANG, Xinkang JIA, Fengda ZHANG, Zheqi LV, Kun KUANG, Chao WU, Fei WU, 2023. Federated mutual learning: a collaborative machine learning method for heterogeneous data, models, and objectives. *Frontiers of Information Technology & Electronic Engineering*, 24(10):1390-1402.
https://doi.org/10.1631/FITEE.2300098

Federated mutual learning: a collaborative machine learning method for heterogeneous data, models, and objectives

Key words: Federated learning; Knowledge distillation; Privacy preserving; Heterogeneous environment

Corresponding authors: Chao WU, Fei WU
E-mail: chao.wu@zju.edu.cn, wufei@zju.edu.cn
ORCID: https://orcid.org/0000-0003-0885-6869; https://orcid.org/0000-0003-2139-8807

Motivation

Traditional federated learning (FL) has heterogeneity problems, including:

1. Data heterogeneity (DH). In FL, the data collected from multiple clients are non-independent and identically distributed (non-IID) as opposed to centralized deep learning, where data are independent and identically distributed (IID).
2. Model heterogeneity (MH). In FL, the global model obtained through FedAvg by aggregating the weights of local models cannot be customized for various scenarios and tasks.
3. Objective heterogeneity (OH). In FL, OH has two aspects, referring to the existence of different objectives between the global model and local models in FL, as well as across different clients.

Method

The heterogeneity problems can be depicted as follows:

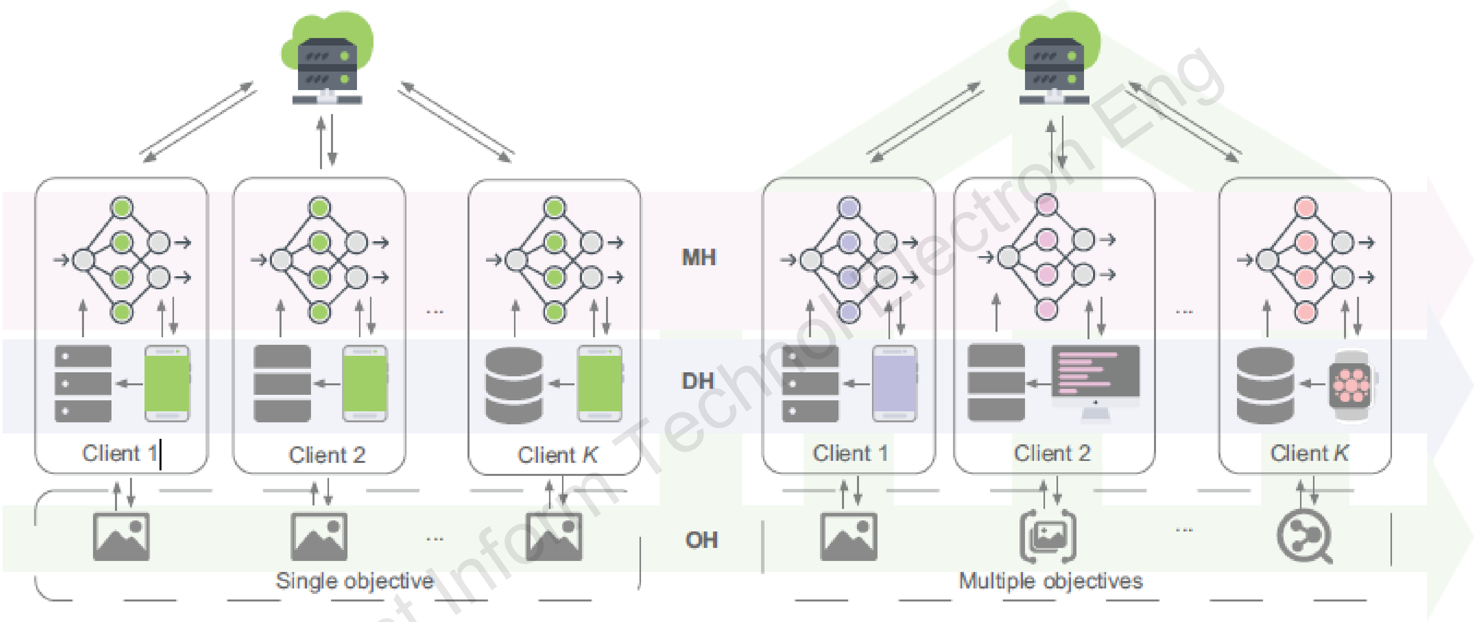


Fig. 1 The heterogeneities in FL can be divided into three types: DH, MH, and OH. DH refers to the fact that data generated by different clients are non-independent and identically distributed like in centralized deep learning. This statistical heterogeneity of data can result in significant accuracy reduction when model weights are averaged due to weight divergence. MH refers to the fact that clients may have different hardware capabilities, different representations of local data, or different tasks, and they need to design their own models. However, FedAvg cannot provide customized models for various scenarios and tasks, because it needs to aggregate the weights of local models with the same architecture. OH arises from the inconsistent objectives of the server and clients in FL. The server aims to construct a single generalized model from data contributed by all clients, while clients aim to train a personalized model for themselves. As a result, this trade-off between these two objectives can lead to the loss of both generalization and personalization. Additionally, clients may have data of similar features but different tasks, thereby complicating the model aggregation process. DH: data heterogeneity; FL: federated learning; MH: model heterogeneity; OH: objective heterogeneity

Method

FML: We incorporate knowledge distillation into the FL process during the local update stage.

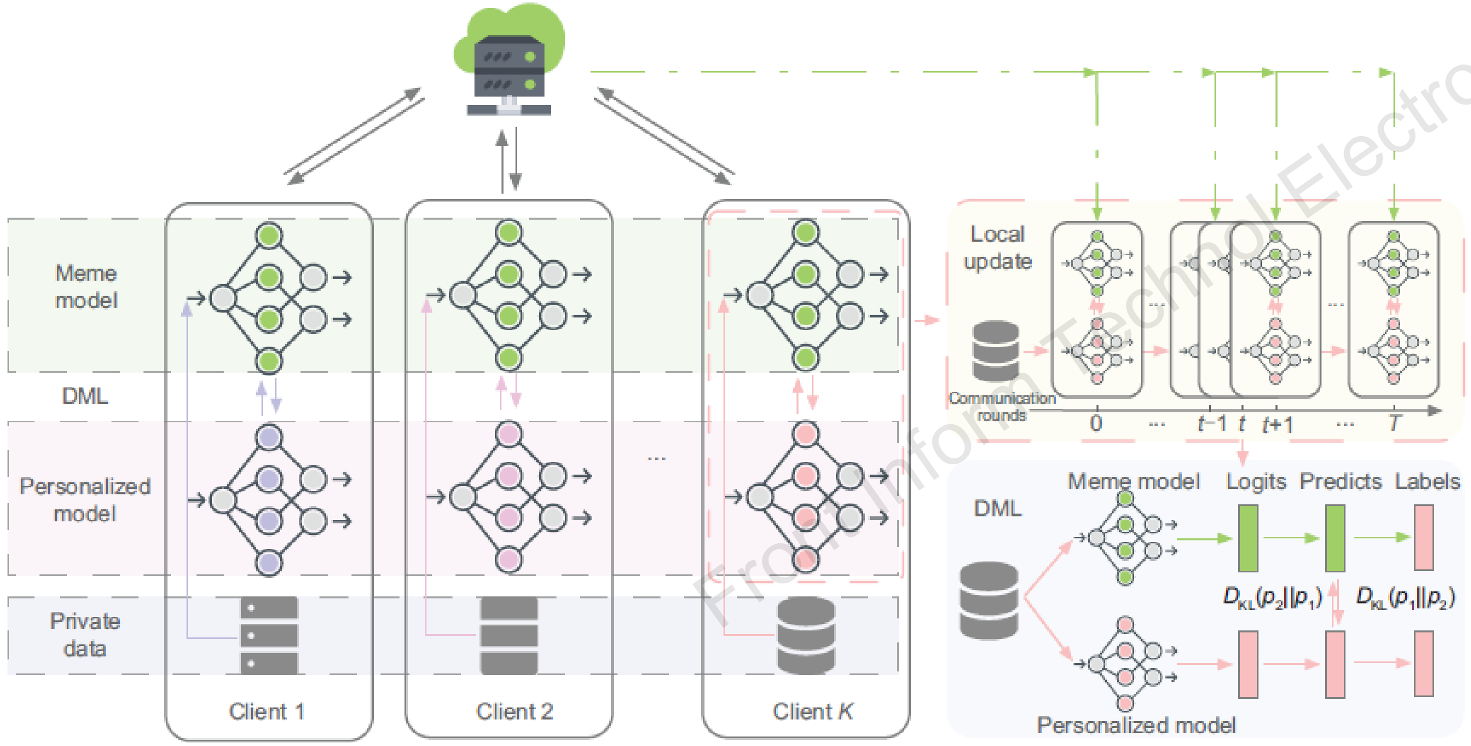


Fig. 2 In the FML method, each client trains two models over its private data during local update: the meme model and the personalized model. At each communication round, the clients fork the new generation of the global model as their meme model, while the personalized model is trained privately and continuously. During each local update, the two models in the clients engage in DML for several epochs, learning mutually. FML: federated mutual learning; DML: deep mutual learning

Algorithm 1 Federated mutual learning

```
Server execution:
for each round t = 1, 2, ..., T do
    for each client k in parallel do
        meme^k_{t+1} ← ClientUpdate(meme^k_t)
    end for
    Merge: global_{t+1} ← (1/K) Σ_{k=1}^{K} meme^k_{t+1}
end for
ClientUpdate:
for each client k do
    Fork: meme^k_0 ← global_0
    for each epoch e = 1, 2, ..., E do
        Conduct DML between meme^k_t and local^k_t over private data (X_k, Y_k)
    end for
end for
```

Method

FML performance in traditional FL settings (global model)

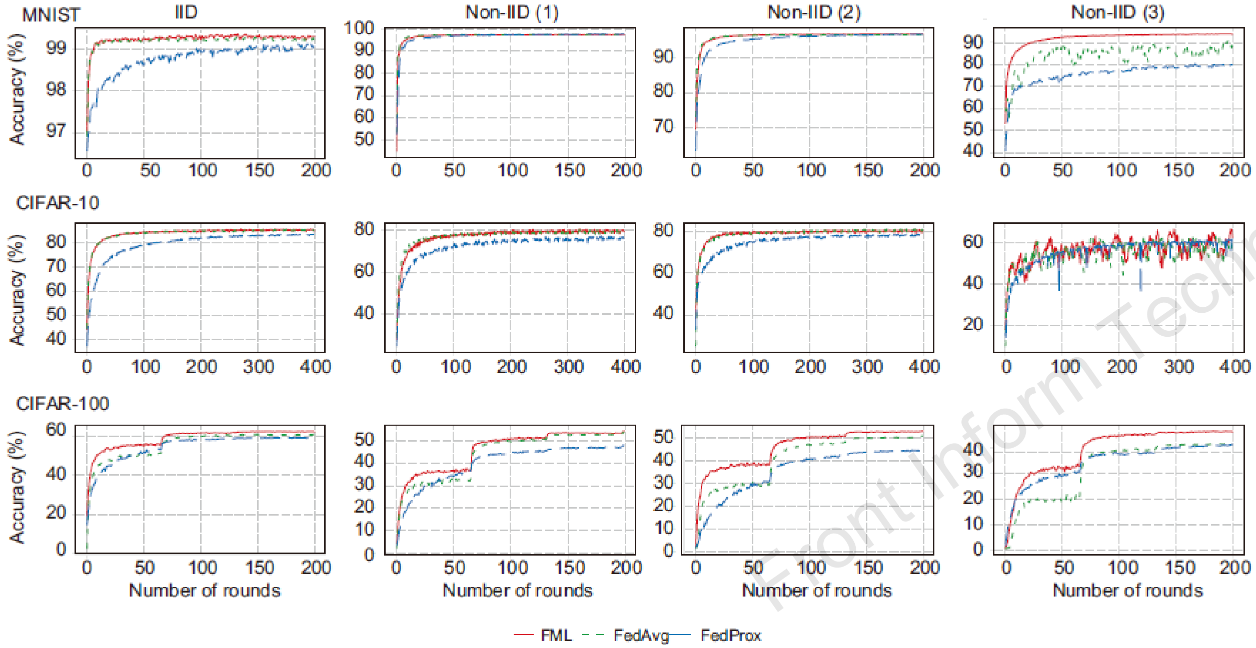


Fig. 3 Our findings indicate that FML offers better performance improvements than the FedAvg and FedProx approaches across four different data settings. We simulate various levels of DH, where the difficulty level gradually increases from left to right. With the aid of DML, the D_{KL} loss component acts as a powerful regularizer during the training process. In the non-IID setting, we observe that FML performs better with a stable trajectory compared to FedProx and FedAvg, which exhibit severe oscillations. As noted in Zhang Y et al. (2018), FML can identify a more stable and robust minimum. FML: federated mutual learning; DH: data heterogeneity; DML: deep mutual learning; non-IID: non-independent and identically distributed

In non-IID setting, the dataset is divided into *Kp* shards of size *n*/(*Kp*) (where *K*=5 and *n*=50 000), and *p* shards are assigned to each client.

Table 1 Top-1 accuracies of global models in typical FL settings

Setting	Method	Accuracy (%)					
		MNIST		CIFAR-10		CIFAR-100	
		MLP	LeNet5	CNN1	CNN2	CNN1	CNN2
IID	FedAvg	98.44	99.29	85.90	**87.49**	56.11	60.88
	FedProx	98.14	99.13	83.91	86.15	32.41	59.23
	FML (ours)	**98.49**	**99.37**	**85.93**	87.41	**57.11**	**62.50**
Non-IID (1)	FedAvg	97.40	98.92	80.41	82.64	53.77	57.76
	FedProx	97.35	98.75	77.46	80.88	47.83	55.60
	FML (ours)	**97.70**	**99.07**	**80.86**	**82.69**	**54.21**	**59.77**
Non-IID (2)	FedAvg	96.84	98.67	**78.85**	**81.17**	50.86	**56.82**
	FedProx	96.98	98.50	76.53	78.87	45.46	55.34
	FML (ours)	**97.00**	**98.71**	78.64	80.85	**52.92**	55.93
Non-IID (3)	FedAvg	90.46	96.45	**63.22**	64.12	41.48	50.36
	FedProx	80.03	87.55	58.07	62.01	41.29	49.51
	FML (ours)	**93.77**	**96.70**	62.42	**66.75**	**46.30**	**51.86**

Best results are in bold. FL: federated learning; FML: federated mutual learning

Method

FML in DMO settings

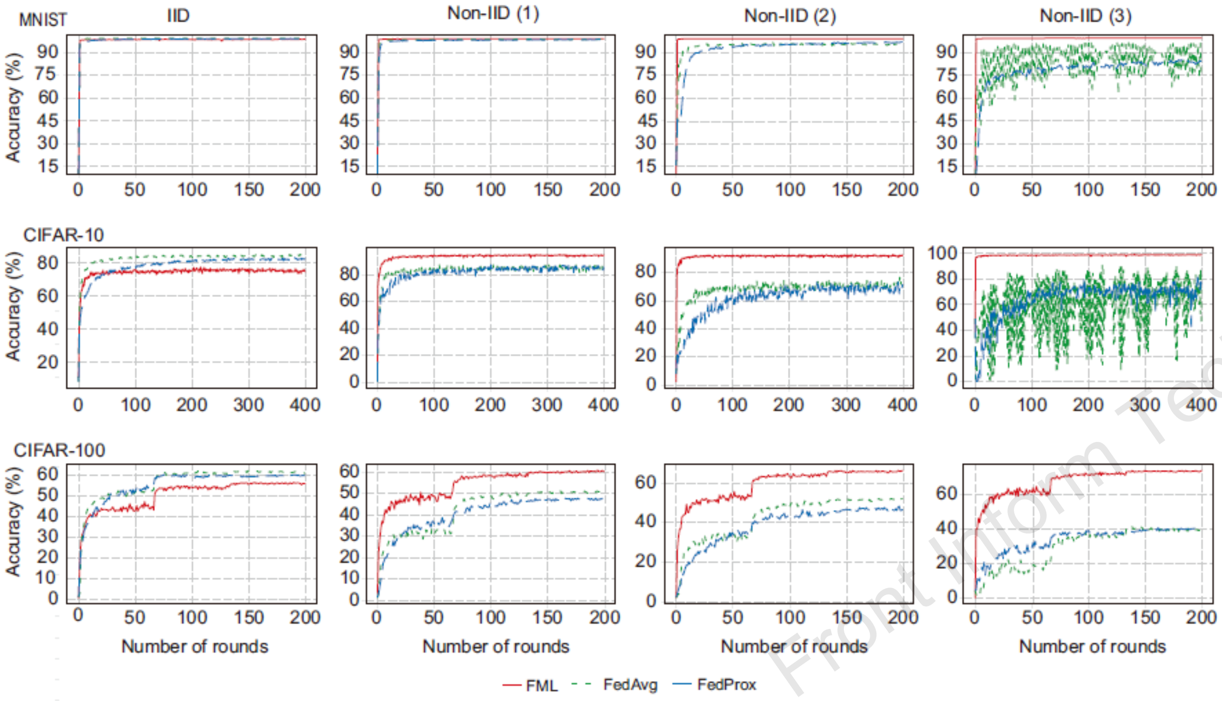


Fig. 4 The global model trained in an FL system functions as a generalized model, which exhibits poor performance on private data in non-IID settings. Through our analysis of the three curves, we observe that FedAvg exhibits more severe oscillations as the level of DH increases, particularly in non-IID (3). Although FedProx adds a proximal term to alleviate the oscillation, it fails to achieve a high accuracy. In contrast, FML rapidly improves and stabilizes at a high level, demonstrating superior performance in terms of both stability and accuracy. FL: federated learning; non-IID: non-independent and identically distributed; DH: data heterogeneity; FML: federated mutual learning

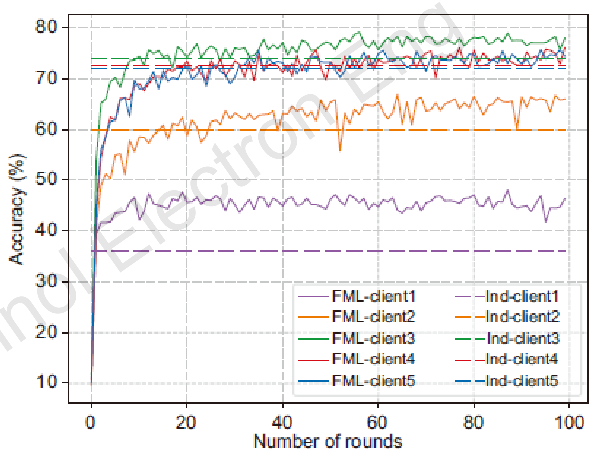


Fig. 5 We evaluate the performance of personalized models trained using FML (represented by the solid curves) and compare it to the highest accuracy achieved by the personalized models through independent training (represented by the dashed lines), using the private validation set. We show the first 100 rounds of the training process. Our results indicate that the use of a shared model through FML leads to improved accuracy for personalized models across all clients, regardless of the specific model architecture employed. FML: federated mutual learning

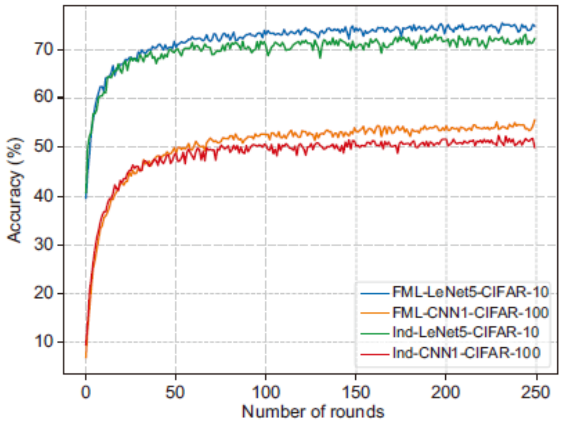
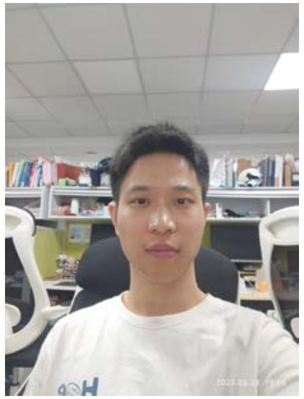


Fig. 6 We illustrate the performance of LeNet5 and CNN1 models, which were trained independently on CIFAR-10 and CIFAR-100 datasets, using the green and red curves, respectively. The blue and orange curves represent the performances of the two models trained using FML. We show the first 250 rounds of the training process. Our results demonstrate that the use of a shared representation through FML can effectively improve the accuracy of all models, despite the presence of different tasks assigned to each client. References to color refer to the online version of this figure

Conclusions

1. We explored the challenges of data, model, and objective (DMO) heterogeneities in federated learning.
2. We propose a novel federated mutual learning (FML) framework that effectively addresses the proposed heterogeneity challenges.
3. FML outperforms alternatives in different FL scenarios, thus establishing its effectiveness in dealing with DMO challenges.



Tao SHEN received his BS degree in control science and technology from China University of Petroleum, Qingdao, China, in 2015. He is currently pursuing his PHD degree in computer science and technology at Zhejiang University. His research interests focus on federated learning.

Fei WU received his BS degree from Lanzhou University, Lanzhou, Gansu, China, his MS degree from Macao University, Taipa, Macau, China, and his PhD degree from Zhejiang University, Hangzhou, China, in 1996, 1999, and 2002, respectively. He was a visiting scholar with Prof B Yu's Group, University of California at Berkeley, Berkeley, CA, USA, from 2009 to 2010. He is currently a full professor with the College of Computer Science and Technology, Zhejiang University. His current research interests include machine learning, sparse representation, and multimedia retrieval.