Discussion on Bonus 2

1. The influence on FL model performance that caused by data splitting.

   In early 2016, McMahan [i] et al. proposed Federated Averaging, which was a pioneering work in federated learning. The authors emphasize that the performance of federated learning may suffer when client data is presented in a non-IID distribution. The experimental results show that compared with IID data, the convergence speed of the model will be reduced under non-IID data, and the final performance of the model may also be affected.

   To be specific, McMahan adopts different data-splitting strategies (e.g., some clients have only a subset of number categories) on the MNIST dataset to simulate non-IID data distribution.

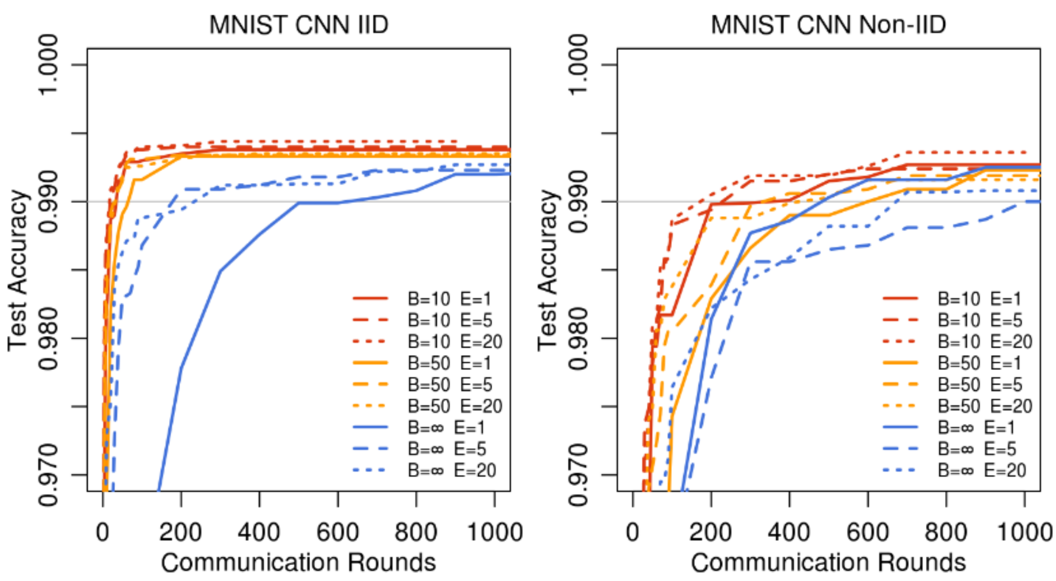

Figure 1 Comparison of two types of data

   As we can see in Figure 1, though the use of the FedAvg algorithm can largely converge the non-IID data finally, the speed is much slower, and the final accuracy can be slightly lower than the IID data. How to improve performance in these two aspects could be an interesting and challenging topic.

2. A potential method to reduce the influence of non-IID data.

   Meta-learning [ii] is a machine learning method whose intention is to design models that can quickly learn new skills or adapt to new environments through a small number of training examples. Finn et al. [iii] proposed a Model-Agnostic Meta-Learning(MAML) method to enable it to quickly adapt to new tasks with a small number of gradient updates.

   In order to improve the performance of the federated learning model, Chen [iv] et al. proposed a combining method between meta-learning and federated learning.

Instead of using a global model, the author uses meta-learners to share the parameters, reducing communication costs, bringing the convergence quicker, and providing much higher accuracy. Specifically, Chen applies MAML to initialize a meta-learning model as a framework and distribute it to the clients. Then the clients train their model locally with a respective dataset. To reduce communication costs and make the model convergent quicker, it adopts Gradient compression and sparsity techs. About the evaluation, Chen selects 4 datasets to present their model preference, FEMNIST, Shakespeare, Sent140, and Production Dataset.

Table 1 Accuracy results on LEAF Datasets

		20% Support	50% Support	90% Support
FEMNIST	FedAvg	76.79% ± 0.45%	75.44% ± 0.73%	77.05% ± 1.43%
	FedAvg(Meta)	83.58% ± 0.13%	87.84% ± 0.11%	88.76% ± 0.78%
	FedMeta(MAML)	88.46% ± 0.25%	89.77% ± 0.08%	89.31% ± 0.15%
	FedMeta(Meta-SGD)	**89.26%** ± 0.12%	**90.28%** ± 0.02%	**89.31%** ± 0.09%
Shakespeare	FedAvg	40.76% ± 0.62%	42.01% ± 0.43%	40.58% ± 0.55%
	FedAvg(Meta)	38.71% ± 0.51%	42.97% ± 0.97%	43.48% ± 0.64%
	FedMeta(MAML)	**46.06%** ± 0.85%	**46.29%** ± 0.84%	**46.49%** ± 0.77%
	FedMeta(Meta-SGD)	44.72% ± 0.72%	45.24% ± 0.53%	46.25% ± 0.63%
Sent140	FedAvg	71.53% ± 0.18%	72.29% ± 0.49%	73.38% ± 0.38%
	FedAvg(Meta)	70.10% ± 0.66%	73.88% ± 0.06%	75.86% ± 0.46%
	FedMeta(MAML)	76.37% ± 0.06%	78.63% ± 0.19%	79.53% ± 0.25%
	FedMeta(Meta-SGD)	**77.24%** ± 0.32%	**79.38%** ± 0.09%	**80.94%** ± 0.29%

As Table 1 shown, FedMeta frameworks have the best performance during these three datasets.

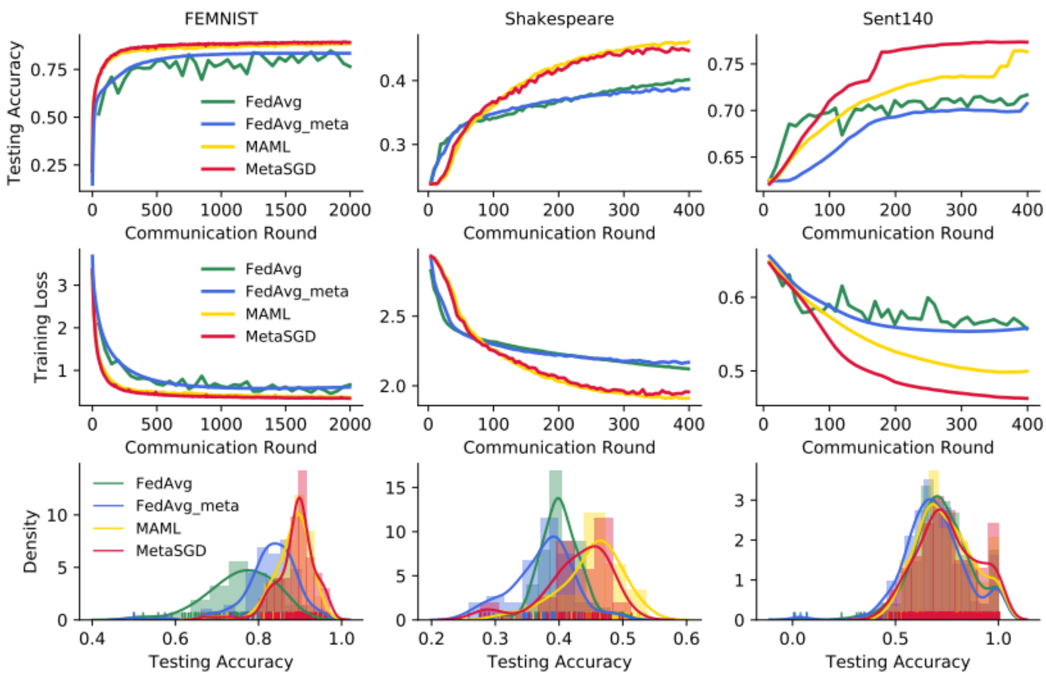

Figure 2 Performance on LEAF datasets for FedAvg and three running examples of FedMeta.

As Figure 2 shown, compared with the original FedAvg, all the other examples within FedMeta frameworks provide faster convergence and higher accuracy.

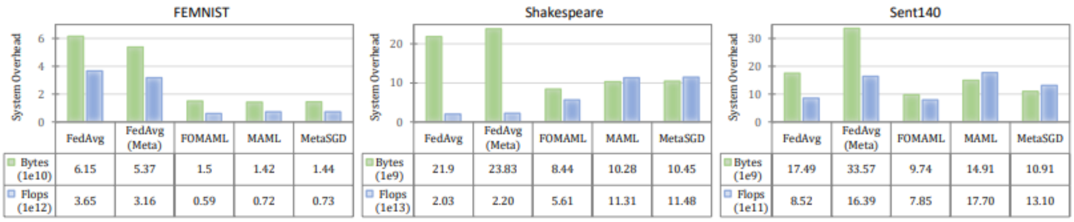

Figure 3 Communication cost

As Figure 3 shown, FedMeta nearly achieves a reduction in required communication cost by several times in all cases.

Overall, Although Chen doesn't consider non-IID data specifically, the FedMeta frameworks could be an effective way to reduce the non-IID data distribution influence, since it combines meta-learning in a federated learning model and, indeed, performs better than a pure federal model. Furthermore, meta-learning can adapt to new tasks quickly, therefore, it has the probability that adapts to non-IID data.

Reference

[i] McMahan, H.B., Moore, E., Ramage, D., Hampson, S., & Arcas, B.A. (2016). Communication-Efficient Learning of Deep Networks from Decentralized Data. International Conference on Artificial Intelligence and Statistics.

[ii] https://www.jianshu.com/p/0673fc7d7760

[iii] Finn, C., Abbeel, P., & Levine, S. (2017). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. ArXiv, abs/1703.03400.

[iv] Chen, F., Luo, M., Dong, Z., Li, Z., & He, X. (2018). Federated Meta-Learning with Fast Convergence and Efficient Communication. arXiv: Learning.