Towards Practical Few-shot Federated NLP

Dongqi Cai, Yaozong Wu, Haitao Yuan, Shangguang Wang, Felix Xiaozhu Lin, Mengwei Xu

Beiyou Shenzhen Institude

University of Virginia

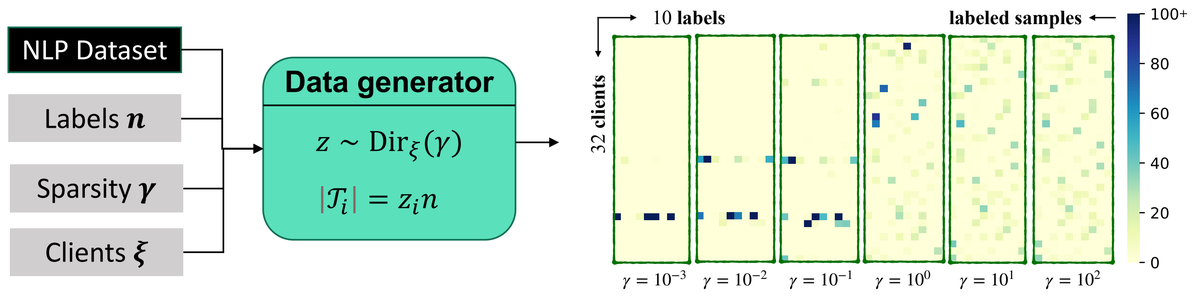
Abstract

Transformer-based pre-trained models have emerged as the predominant solution for natural language processing (NLP). Fine-tuning such pre-trained models for downstream tasks often requires a considerable amount of labeled private data. In practice, private data is often distributed across heterogeneous mobile devices and may be prohibited from being uploaded. Moreover, well-curated labeled data is often scarce, presenting an additional challenge.

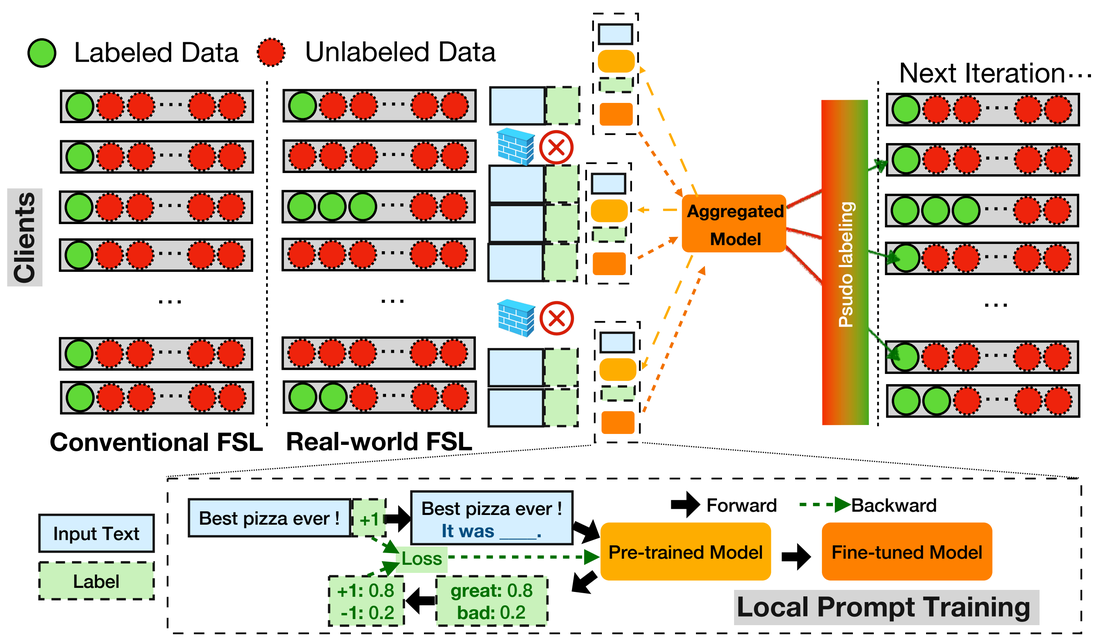
To address these challenges, we first introduce a data generator for federated few-shot learning tasks, which encompasses the quantity and skewness of scarce labeled data in a realistic setting. Subsequently, we propose AUG-FedPrompt, a prompt-based federated learning system that exploits abundant unlabeled data for data augmentation. Our experiments indicate that AUG-FedPrompt can perform on par with full-set fine-tuning with a limited amount of labeled data. However, such competitive performance comes at a significant system cost.

Data generator

We propose a data generator to simulate federated few-shot dataset.



Our System: AUG-FedPrompt



Key building blocks:

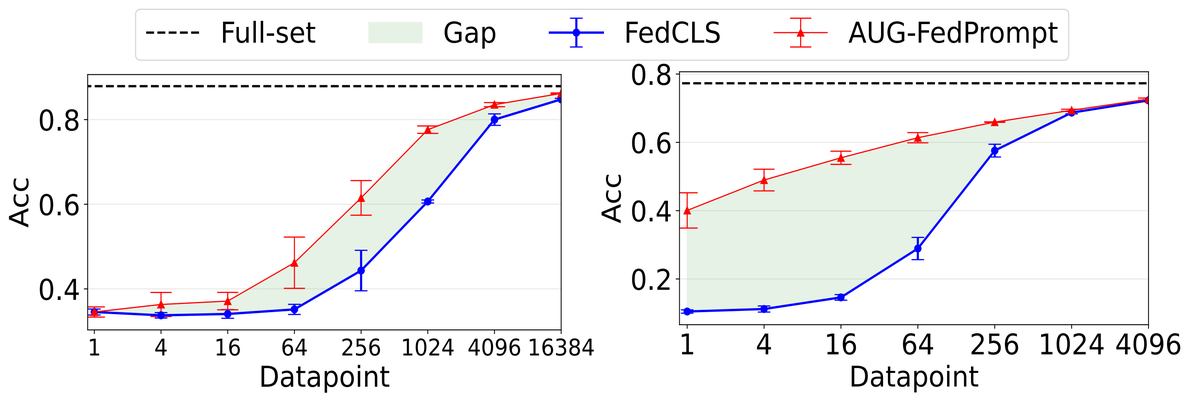
- Pseudo labeling ⟶ Lack and skewness of labels
- Prompt Learning ⟶ Improve the quality of pseudo labels

Workflow:

Clients with labeled training data conduct local prompt learning without sharing their data. They send the fine-tuned models to the cloud for federated average aggregating. The aggregated model is distributed to all clients, even those who did not participate in the last round of training, that means they do not have enough labeled data. These clients can leverage the received model to do pseudo labeling on their unlabeled data. They add the data with the highest confidence as training samples for next iteration.

Experiments

Dataset	Prompt	Train	Test
AGNEWS [20]	a (____) b	120,000	7.600
MNLI [24]	“a” ? ‖ ____, “b”	392,702	9,815
YAHOO [20]	[Category:] a ____ b	1,400,000	60,000
YELP-F [20]	It was ____. a	650,000	50,000



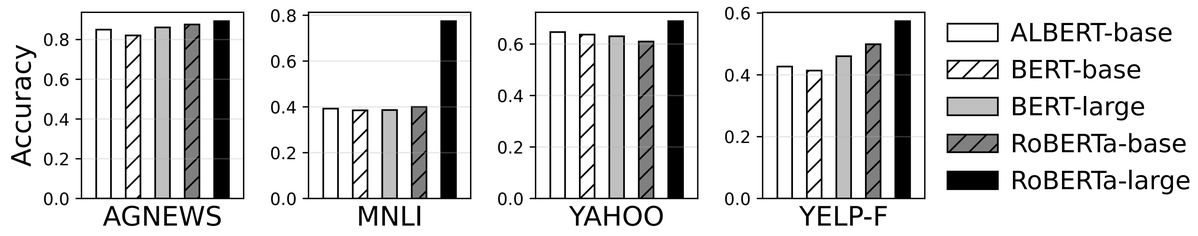
(a) MNLI (b) YAHOO

- AUG-FedPrompt enjoys up to 50%, 25%, 55%, 38% accuracy improvement separately for 4 datasets.
- For a usable accuracy, AUG-FedPrompt saves up to 99% training data compared to full-set federated finetuning.

Dataset		AGNEWS	MNLI	YAHOO	YELP-F
Uniform	FedCLS	66.1±12.8	60.1±0.4	57.6±1.9	54.0±0.1
	FedPrompt	**87.0**±0.8	**77.6**±0.8	**66.0**±0.1	**61.9**±0.7
Skewed	FedCLS	64.8±3.1	37.7±5.6	24.4±10.3	38.3±8.8
	FedPrompt	68.4±2.4	42.4±5.8	41.8±4.3	51.2±1.8
	w/ augment	**90.2**±0.5	**75.7**±1.2	**66.9**±1.1	**58.2**±2.4

AUG-FedPrompt shows competitive performance under various federated few-shot learning settings, regardless of uniform or skewed label distribution.

Limitations



AUG-FedPrompt prefers large language models.

Model	ALBERT-base [29]	BERT-base [1]	BERT-large [1]	RoBERTa-base [25]	RoBERTa-large [25]
Memory (GB)	3.7	5.4	OOM (9.8)	5.8	OOM (10.4)
Latency (s)	1.4	1.9	~7.8	2.1	~8.1
Param. (M)	11.7	109.5	334.9	124.6	355.3

Finetuning these ‘behemoths’ can be extremely resource-intensive.

Conclusion

This work explores a crucial but less explored issue: data labels can be scarce in federated learning. Our system AUG-FedPrompt shows competitive performance under various federated few-shot learning settings, requiring less than 0.1% data to be manually labeled. In the future, we will improve its resource efficiency to make it more practical.

Contact: Dongqi Cai (cdq@bupt.edu.cn)