ENTROPY-AIDED PROMPT FEDERATED LEARNING

Jiaming Pei
School of Computer Science
The University of Sydney
jpei0906@uni.sydney.edu.au

Haotian Wu
School of Aerospace, Mechanical and Mechatronic Engineering
The University of Sydney
hawu3107@uni.sydney.edu.au

ABSTRACT

To overcome the impact of low-quality prompts on the training effect of prompt-Fl, we introduce the information entropy method to calculate the probability distribution of the model output to guide the local client to generate higher-quality prompts. Experimental results show that our method performs well on multiple datasets in both IID and non-IID settings. In addition, our method also leads significantly in training time, showing its efficiency in practical applications.

1 INTRODUCTION

Prompt-based Federated Learning (PromptFL) Guo et al. (2023) still encounters key limitations in practical applications. Notably, the quality of generated prompts is crucial in PromptFL; however, it often varies significantly, directly impacting the effectiveness of model training. Across different clients, due to variations in data distribution and the size of local data sets, the generated prompts can exhibit significant differences in information richness and feature representation. This inconsistency in prompt quality may prevent the global model from fully capturing the key features of the data from all clients, especially in non-IID data settings Zhuang et al. (2023); Qi et al. (2023). To address the issues caused by the uneven quality of prompts in PromptFL, our method adopts an innovative communication strategy based on information entropy. In this strategy, we no longer rely on prompts directly generated by each client. Instead, clients calculate the information entropy of their model outputs and transmit these entropy values as key information to the server. Information entropy, as a metric quantifying model uncertainty and data diversity, provides the server with a more consistent and reliable way to assess the data quality and model performance of each client. This approach enables the server to more accurately identify which clients may require additional attention or support, thereby allowing for targeted adjustments in the training process of the global model. Different from Tent Wang et al. (2020) which minimizes entropy during test time to adapt model parameters, our method takes a distinct route. Rather than minimizing entropy, we employ information entropy as a metric to quantify model uncertainty and data diversity. Not only does this enhance the efficiency of model training, but it also reduces the risk of training outcomes being affected by uneven data distributions by decreasing reliance on the original quality of prompts.

2 METHODOLOGY

In our federated learning (FL) system, N clients collaborate to update a global model, denoted as Θ. Following the approach of prompt-based FL methods, our system initially transmits prompts derived from each client's private local dataset to the server (Step ①). The server then generates synthetic data corresponding to these prompts using a foundational diffusion model, which is employed to train Θ (Step ②). Subsequently, the updated global model Θ is distributed back to the clients. Each client, denoted as client n, calculates the model's entropy by analyzing the probability distribution of its outputs. In tasks like image classification, the model is assumed to output a probability distribution for each input, indicating the likelihood of various output categories. The information entropy, $H(x)$, is computed using the formula $H(x) = -\sum_{i=1}^{I} p(x_i) \log p(x_i)$, where I represents the number of classes, and $p(x_i)$ is the probability of the model predicting the i-th class (Step ③). Diverging from traditional FL methods, our approach involves uploading the calculated information entropy $H(x)$ to the server, significantly reducing communication overhead. The server identifies clients with entropy below the median and prompts them to generate high-quality prompts

Table 1: Comparison results of four federated learing methods under IID and Non-IID settings.

Metrics	Method	Clipart	Infogragph	Caltech101	Quickdraw
Accuracy under IID setting (/%)	FedAvg	80.91	56.24	76.47	68.91
	Centralized	81.21	60.74	80.14	70.14
	PromptFL	88.23	60.14	88.11	75.36
	Tent	86.16	60.11	86.13	72.12
	ours	90.11	68.36	91.52	79.61
Accuracy under non-IID setting (/%)	FedAvg	70.12	41.9	70.58	62.14
	Centralized	73.45	50.14	76.94	62.39
	PromptFL	85.49	57.12	72.14	70.58
	Tent	84.16	60.11	83.13	69.12
	ours	89.35	65.41	88.14	76.78
Training time under non-IID Setting (/s)	FedAvg	7087	7187	8074	8194
	Centralized	6817	7045	7946	8045
	PromptFL	2410	2444	2482	2349
	Tent	82388	2311	2401	2231
	ours	1964	1924	2046	1973

The degree of non-IID is 0.2 Li et al. (2022)

for the next round of communication. In contrast, clients with higher entropy continue their local training. This process is repeated in multiple rounds until the global model Θ reaches convergence.

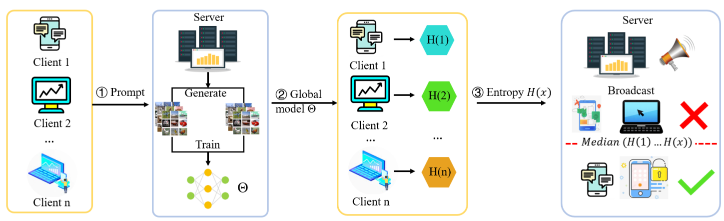

Figure 1: Overview of our framework in one-round communication.

3 Experiment and Results

To outperform the superiority of our method, we compared it with baseline and benchmark method like FedAvg McMahan et al. (2017), Centralized ML, Tent and PromptFL Guo et al. (2023) under four novel data sets Dillon (2006); Dur (2014); Cheema et al. (2012); Bansal et al. (2021). Experimental results show that in the IID environment, the accuracy of our method on each data set exceeds other methods. In the non-IID environment, this advantage is more obvious, and it is also significantly ahead in training time. This significant performance improvement is mainly attributed to the fact that our method significantly reduces communication overhead by transmitting information entropy instead of model parameters or gradients, which is especially advantageous when processing non-IID data. In addition, our method calculates information entropy by analyzing the probability distribution of the model output, thereby effectively guiding the iterative update of the model and improving the accuracy and training efficiency of the model on diverse data sets.

4 Conclusion

Our research calculates information entropy by analyzing the probability distribution of model output, and our method not only reduces the communication burden but also enhances data privacy protection. Furthermore, by guiding model updates in a targeted manner, our approach further improves the learning efficiency and performance of the global model. These advantages make our method particularly suitable for application scenarios with high data diversity and limited communication resources.

URM Statement

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2024 Tiny Papers Track.

References

Monika Bansal, Munish Kumar, Monika Sachdeva, and Ajay Mittal. Transfer learning for image classification using vgg19: Caltech-101 image data set. *Journal of ambient intelligence and humanized computing*, pp. 1–12, 2021.

Salman Cheema, Sumit Gulwani, and Joseph LaViola. Quickdraw: improving drawing experience for geometric diagrams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1037–1064, 2012.

George L Dillon. Clipart images as commonsense categories. *Visual communication*, 5(3):287–306, 2006.

Banu Inanc Uyan Dur. Data visualization and infographics in visual communication design education at the age of information. *Journal of arts and humanities*, 3(5):39–50, 2014.

Tao Guo, Song Guo, Junxiao Wang, Xueyang Tang, and Wenchao Xu. Promptfl: Let federated participants cooperatively learn prompts instead of models-federated learning in age of foundation model. *IEEE Transactions on Mobile Computing*, 2023.

Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. Federated learning on non-iid data silos: An experimental study. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pp. 965–978. IEEE, 2022.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. PMLR, 2017.

Zhuang Qi, Lei Meng, Zitan Chen, Han Hu, Hui Lin, and Xiangxu Meng. Cross-silo prototypical calibration for federated learning with non-iid data. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 3099–3107, 2023.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020.

Weiming Zhuang, Chen Chen, and Lingjuan Lyu. When foundation model meets federated learning: Motivations, challenges, and future directions. *arXiv preprint arXiv:2306.15546*, 2023.