Paper Title:

FedAudio: A Federated Learning Benchmark for Audio Tasks

Paper Link:

https://arxiv.org/pdf/2210.15707.pdf

1 Summary

1.1 Motivation

FedAudio is motivated by the escalating concerns surrounding data privacy, particularly in handling personal audio data. With the widespread adoption of consumer applications like Amazon Alexa, Google Assistant, and Apple Siri, the potential disclosure of sensitive information through audio data has become a critical issue. The paper addresses this pressing concern by leveraging federated learning (FL) as a privacy-preserving solution.

The motivation is rooted in the absence of specialized FL benchmarks for audio-related tasks, despite the proliferation of FL benchmarks in other domains. While FL has gained attention for its ability to mitigate data privacy issues, existing benchmarks have not adequately encompassed audio data and tasks. The motivation for FedAudio is to fill this gap by introducing a benchmark tailored explicitly for federated audio tasks.

1.2 Contribution

The primary contribution of the paper lies in the development of FedAudio, a benchmark explicitly designed for audio-related tasks in federated learning. Unlike existing FL benchmarks, FedAudio includes four representative datasets covering keyword spotting, speech emotion recognition, and sound event classification. A key differentiator is the introduction of data noises and label errors to emulate real-world challenges in deploying FL systems.

1.3 Methodology

The methodology involves the selection of four diverse audio datasets aligned with FL use cases. These datasets are intentionally non-IID distributed by speaker or actor ID, providing a more realistic representation. The introduction of data noises and label errors adds a layer of complexity, making FedAudio a robust benchmark for evaluating FL algorithms.

1.4 Conclusion

FedAudio aims to catalyze new FL research in the acoustic and speech research community. The benchmark results, including the impact of client sample ratio, FL optimizers, and data heterogeneity, provide insights into the challenges and performance variations in federated audio tasks.

2 Limitations

2.1 First Limitation

One limitation of FedAudio lies in the specificity of the chosen datasets and tasks. The study's reliance on particular datasets and tasks may restrict the generalizability of the findings. Variations in input data characteristics or different audio-related tasks might yield different outcomes, making it essential to interpret the results within the scope of the selected benchmarks.

2.2 Second Limitation

While FedAudio makes significant strides in addressing the lack of specialized FL benchmarks for audio tasks, it acknowledges the finite scope of tools and datasets included. The study recognizes the possibility of other audio-related tasks or datasets exhibiting different behaviors or requiring distinct considerations. This limitation prompts the acknowledgment that FedAudio may not cover the entire landscape of audio-related FL challenges and applications.

3 Synthesis

FedAudio addresses a crucial need for specialized FL benchmarks in audio tasks, acknowledging the widespread use of consumer devices collecting audio data. The paper's findings not only contribute to advancing FL research in the acoustic and speech domain but also pave the way for more inclusive and representative models. The introduction of real-world challenges in the form of data noises and label errors enriches the benchmark's utility, making it a valuable resource for researchers and developers. The potential societal impact is significant, considering the increasing reliance on audio-related applications and the need for privacy-preserving solutions. FedAudio serves as a stepping stone towards building more robust and unbiased FL models for audio tasks.