- I. AQG^[1]: Adaptive Searching for More Efficient Quantization Precision
 - Adjusted the quantization precision adaptively based on the client's updates.



Searching Strategy

- II. AQUILA^[2]: Mathematical Optimization
 - Minimized compression error.
 - Developed an optimal quantization precision strategy.
 - Significantly reduced communication costs.



Mao Y, Zhao Z, Yan G, et al. Communication-efficient federated learning with adaptive quantization[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2022, 13(4): 1-26.
Zhao Z, Mao Y, Shi Z, et al. AQUILA: Communication efficient federated learning with adaptive quantization in device selection strategy[J]. IEEE Transactions on Mobile Computing, 2023.

Research Experience

III. SAFARI^[3]: Sparsification Strategy under Limited and Unreliable Communications

- Provided theoretical analysis of sparse model similarity under bounded data dissimilarity.
- Achieved fast and robust convergence with 60% of the weights pruned and 80% of the client updates lost.

- Stochastic Ouantization Rotation-based Quantizatio $s(x_u, x_v) := \|x_u - x_v\|$ 4.1 Unreliable 3.1 Quantization Lattice-based Ouantizatio networks 4. Communication 1-bit (Sign) Ouantization Model update A Client A environment uantized Compressed Sensi 4.2 Noisy fading channels bsent model I Adaptive Sparsification Model update I Client B 3.2 Sparsification Bidirectional Sparsification Central Server 3. Communication Global loss function Layer-wise Sparsification Efficiency Model update C Time consumption Data-additional KD Client C 3.3 Knowledge 5. Communication Energy consumption Distillation resource allocation Data-free KD Client selection Updated global Reduce Communication Rounds Hybrid objective 10 20 30 3.4 Other methods Client Index Low-rank Matrix Approximation
- IV. Survey of Communication Challenges in Federated Learning^[4]

SAFARI: System Architecture

Sparse Model Similarity Matrix

Survey Problems and Structure

[3] Mao Y, Zhao Z, Yang M, et al. Safari: Sparsity-enabled federated learning with limited and unreliable communications[J]. IEEE Transactions on Mobile Computing, 2023.
[4] Zhao Z, Mao Y, Liu Y, et al. Towards efficient communications in federated learning: A contemporary survey[J]. Journal of the Franklin Institute, 2023, 360(12): 8669-8703.

Yuzhu Mao (TBSI, Tsinghua University)

Research Experience

- V. FL-TAC^[5]: Multi-Task Fine-Tuning of Large Pretrained Models
 - Trained task-specific low-rank adapters for downstream task adaptation.
 - Achieved enhanced task performance with reduced communication cost.







Heterogeneous Task Distribution Serv

Server-Client Adapter Transmission

Server-side Adapter Aggregation

- VI. Adaptive Parameter-Efficient Fine-Tuning (Ongoing Work)
 - Achieved efficient fine-tuning through an adaptive resource allocation strategy.
 - Achieved effective fine-tuning by optimizing the cost-generalization trade-off.

[5] Ping S*, Mao Y*, Liu Y, et al. FL-TAC: Enhanced fine-tuning in federated learning via low-rank, task-specific adapter clustering[C]. ICLR 2024 Workshop on Large Language Model (LLM) Agents.