Mid Evaluation Progress Report

Communication-Efficient Learning of Deep Networks from Decentralized Data

Name : S.M.C.R.Siriwardhana
Index No : 210613U

Project Overview

This project aims to enhance the foundational Federated Averaging (FedAvg) algorithm, as implemented in the LEAF framework, by addressing its two key limitations: degraded performance on heterogeneous (non-IID) data and excessive communication overhead. The enhancement strategy is two-fold:

1. **Architectural Enhancement:** Implementing the novel **Kolmogorov-Arnold Networks (KANs)** to replace Multi-Layer Perceptron (MLP) layers, focusing on superior performance and interpretability.
2. **Communication Enhancement:** Integrating **Model Compression** techniques (quantization and pruning) to reduce the size of model updates.

The project has completed Phase 1 (Baseline Analysis) and the initial architectural testing within Phase 2. Preliminary experiments on the MNIST dataset confirmed that a Radial Basis Function (RBF) KAN architecture significantly outperforms both the B-Spline KAN and the MLP baseline, establishing the optimal model foundation for subsequent compression integration.

1. Challenges

1.1. The Federated Learning Imperative

Federated learning (FL) is a pivotal distributed machine learning paradigm that enables collaborative model training on decentralized data while preserving privacy. The central server coordinates a "loose federation" of clients (e.g., mobile devices), each with a private local dataset that is never uploaded to the server. This approach is a direct application of the principle of data minimization, decoupling model training from the need for direct access to raw data and significantly reducing the attack surface for privacy and security risks.

The foundational FedAvg algorithm dramatically reduces communication rounds required for convergence by a factor of 10-100x compared to traditional synchronized stochastic gradient descent (SGD), by allowing clients to perform multiple local training epochs before communicating with the server.

1.2. The Defining Challenges of Federated Learning

The design of FedAvg is a direct response to the unique properties of federated optimization, which differentiate it from traditional distributed optimization in a data center. The two primary challenges addressed by this project are:

- **Communication Efficiency:** This is the most significant bottleneck in the federated setting. Mobile devices are often on slow, expensive, or intermittent connections, making frequent, large-scale data transfers untenable. The underlying assumption of FedAvg is that local computation is "essentially free" compared to the cost of communication, justifying the deliberate trade-off of computation for reduced communication rounds.
- **Statistical Heterogeneity (Non-IID Data):** In a federated network, data on any given client is a function of a particular user's behavior, meaning the local dataset is rarely a representative sample of the overall population. This non-IID nature leads to client drift, causing model bias, slow convergence, and reduced accuracy on underrepresented data subsets.

1.3. Synopsis of Enhancement Methodologies

This project addresses these weaknesses using two complementary methodologies:

1. **Novel Architectural Paradigms (KANs):** The implementation of Kolmogorov-Arnold Networks (KANs), which use learnable, univariate activation functions on the network's edges, offering superior functional approximation and interpretability compared to standard MLPs.
2. **Communication Efficiency (Compression):** The integration of model compression techniques (quantization and pruning) to directly reduce the size of model updates and alleviate the communication bottleneck.

2. Methodology Outline

The project is structured to implement and validate two distinct, complementary enhancement strategies.

2.1. Architectural Enhancement: Integrating Kolmogorov-Arnold Networks (KANs)

Kolmogorov-Arnold Networks (KANs) are a novel architecture that replaces fixed activation functions with learnable, univariate functions, often parameterized as B-splines or Radial Basis Functions (RBFs).

Integration Method (Fed-KAN):
The Fed-KAN approach involves replacing MLP layers in a conventional model with KAN layers.13 The architectural sub-study compared the conventional MLP Baseline, the original B-Spline KAN, and the

RBF KAN, which uses Gaussian Radial Basis Functions (RBFs) for its expressive capabilities.
Benefits and Trade-offs:

- **Superior Performance on Non-IID Data:** Fed-KAN models have shown superior performance to Fed-MLP baselines, particularly on heterogeneous data.
- **Enhanced Interpretability:** KANs' learnable functions can be simplified to reveal underlying symbolic formulas, providing unparalleled interpretability.
- **Computational Bottleneck:** A key drawback is the computational overhead of KANs, as their spline-based functions are not yet optimized for GPU parallelization, challenging FedAvg's assumption of free local computation.

2.2. Communication Enhancement: Integrating Model Compression

Communication efficiency will be achieved through model compression techniques to reduce the size of the model updates exchanged between clients and the server.

Model Compression Techniques:

- **Quantization:** This process reduces the numerical precision of model updates (e.g., from 32-bit floating-point) to lower-bit representations, significantly cutting down message size.
- **Pruning:** Pruning creates sparse models by removing less important parameters. The advanced SpaFL framework, which communicates only small, trainable thresholds instead of full parameters, is being considered for implementation to maximize communication savings.

3. Phase 2 Progress: Mid-Evaluation Results

Initial experiments were conducted on the MNIST dataset, partitioned to simulate federated learning conditions, to identify the optimal KAN architecture.

3.1. Architectural Performance Analysis

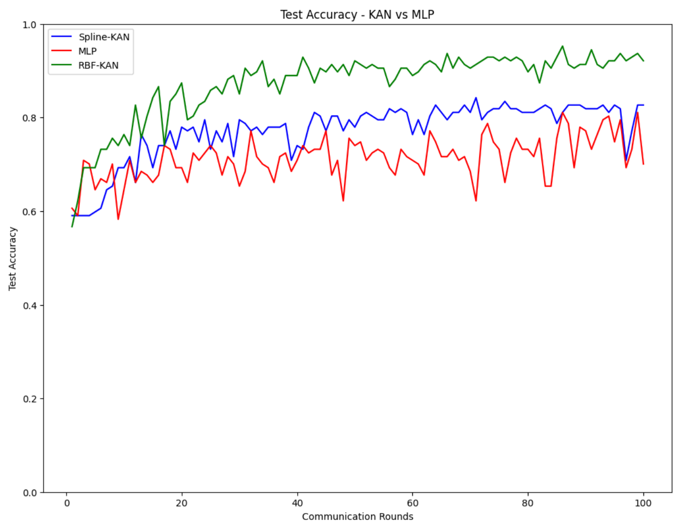
Comparative testing of the three architectural baselines yielded a clear outcome: **the RBF KAN architecture demonstrated superior performance, surpassing both the B-Spline KAN and the conventional MLP baseline.**

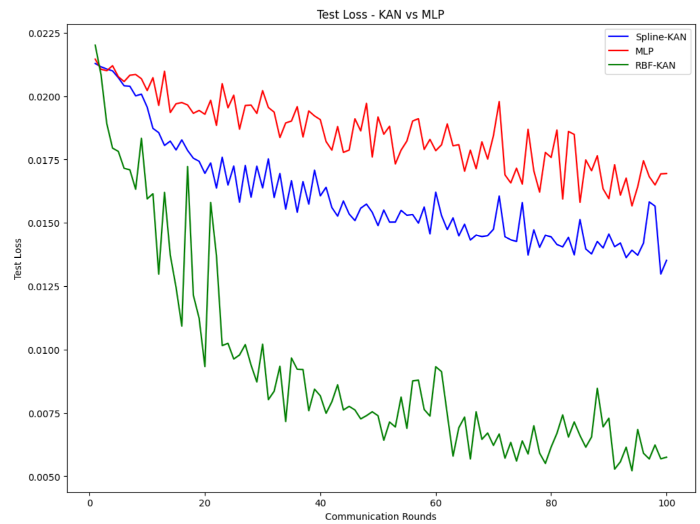
The RBF KAN achieved the highest final test accuracy and the lowest final training loss. Critically, the RBF KAN exhibited a faster initial climb in accuracy and a smoother, more rapid descent of the loss curve, confirming its superior convergence speed in terms of communication rounds compared to the other two architectures.

3.2. Rationale for RBF KAN Superiority

The RBF KAN's superior performance is attributed to the inherent mathematical advantages of Radial Basis Function networks:

- **Robust Function Approximation:** RBF networks excel at robust function approximation, pattern classification, and tolerance to input noise. This capability is crucial in the noisy optimization environment of federated learning where local updates diverge due to non-IID data.
- **Efficiency:** The RBF functional basis provides a robust and efficient foundation for learning features within the KAN structure, proving more effective than the B-Spline formulation in this specific federated setting.





3.3. Next Steps

Based on these results, the **RBF KAN architecture is selected** as the targeted enhancement for the remainder of the project. Phase 2 will now pivot to integrating the communication efficiency enhancement (Model Compression) with the proven RBF KAN model.

4. Conclusions and Future Work

This progress report confirms a successful architectural selection by identifying the RBF KAN as the superior enhancement strategy for this project. By integrating this high-performing architecture with communication compression techniques in the subsequent phase, the project is well-positioned to achieve significant gains over the FedAvg baseline, addressing both performance on heterogeneous data and high communication overhead.

Future work will explore combining these two approaches, such as integrating quantized or pruned RBF KANs into a federated framework, to achieve a system that is simultaneously high-performing, interpretable, and communication-efficient.

References

1. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. a. Y. (2016, February 17). Communication-Efficient Learning of Deep Networks from Decentralized Data. arXiv.org. https://arxiv.org/abs/1602.05629
2. Strategy in Federated Learning by William Lindskog, https://pub.towardsai.net/1-strategy-in-federated-learning-e0e518ada44a
3. Non-IID data in Federated Learning: A Systematic Review with Taxonomy, Metrics, Methods, Frameworks and Future Directions - arXiv, https://arxiv.org/html/2411.12377v1
4. KAN: Kolmogorov–Arnold Networks - OpenReview, https://openreview.net/forum?id=Ozo7qJ5vZi
5. Kolmogorov-Arnold Networks (KANs): A Guide With Implementation | DataCamp, https://www.datacamp.com/tutorial/kolmogorov-arnold-networks
6. FedPAQ: A Communication-Efficient Federated Learning Method with Periodic Averaging and Quantization, https://proceedings.mlr.press/v108/reisizadeh20a.html
7. Model Pruning Enables Efficient Federated Learning on Edge Devices - Shiqiang Wang, https://shiqiang.wang/papers/YJ_SpicyFL2020.pdf
8. A Beginner-friendly Introduction to Kolmogorov Arnold Networks (KAN), https://www.dailydoseofds.com/a-beginner-friendly-introduction-to-kolmogorov-arnold-networks-kan/
9. Understanding Kolmogorov Arnold Networks (KAN) | TDS Archive - Medium, https://medium.com/data-science/understanding-kolmogorov-arnold-networks-kan-e317b1b4d075
10. Convolutional Kolmogorov-Arnold Networks (Convolutional KANs): An Innovative Alternative to the Standard Convolutional Neural Networks (CNNs) - MarkTechPost, https://www.marktechpost.com/2024/06/24/convolutional-kolmogorov-arnold-networks-convolutional-kans-an-innovative-alternative-to-the-standard-convolutional-neural-networks-cnns/
11. Fed-KAN: Federated Learning with Kolmogorov-Arnold Networks for Traffic Prediction, https://arxiv.org/html/2503.00154v1
12. Kolmogorov-Arnold Networks (KAN): Alternative to Multi-Layer Perceptron? - DigitalOcean, https://www.digitalocean.com/community/tutorials/kolmogorov-arnold-networks-kan-revolutionizing-deep-learning
13. Quantization For Federated Learning - Meegle, https://www.meegle.com/en_us/topics/quantization/quantization-for-federated-learning
14. Fed-KAN: Federated Learning with Kolmogorov-Arnold Networks, https://www.researchgate.net/publication/389547173_Fed-KAN_Federated_Learning_with_Kolmogorov-Arnold_Networks_for_Traffic_Prediction