**Title: Sparse Updates as a Regularizer in Federated Learning: An Experimental Study on Non-IID Data**

**Authors:**

* Your Name (Student ID)
* Your Teammate's Name (Student ID)

**Abstract**

Federated Learning (FL) presents a privacy-preserving paradigm for collaborative model training but is significantly challenged by the statistical heterogeneity of client data. This work investigates the impact of extreme Non-IID data on the performance of fine-tuning a pre-trained Vision Transformer (DINO ViT-S/16) on the CIFAR-100 dataset. We show that standard Federated Averaging (FedAvg) fails to converge, achieving only ~6.5% accuracy compared to a ~77% accuracy in an ideal IID setting. To address this, we explore sparse fine-tuning via gradient masking as a form of model editing. Our key finding is that a simple **Random Mask** acts as a powerful regularizer, improving accuracy to ~13%, whereas a more strategic **Least-Sensitive Mask** fails to provide benefits (6.31% accuracy). This suggests that in high-drift scenarios, strong, generalized regularization is more critical than targeted knowledge preservation.

**1. Introduction**

Federated Learning (FL) enables collaborative machine learning across decentralized devices without centralizing user data, offering significant privacy advantages. The most common algorithm, Federated Averaging (FedAvg), involves rounds of local client training followed by the aggregation of model updates at a central server.

While promising, FL's performance is often hindered by the real-world challenge of **statistical heterogeneity**, where data is not independent and identically distributed (Non-IID) across clients. This can lead to "client drift," where local models diverge significantly, and their conflicting updates degrade the performance of the global model upon aggregation.

Our project investigates this challenge by fine-tuning a large, pre-trained Vision Transformer (DINO ViT) on CIFAR-100 in a simulated, yet severe, Non-IID environment. We explore **model editing**, specifically sparse fine-tuning via gradient masking, as a potential solution to mitigate client drift. We compare standard FedAvg against two sparse update strategies to determine their effectiveness.

**2. Methods**

**2.1. Dataset and Model**

* **Dataset:** We use the **CIFAR-100** dataset, which contains 100 classes. The data was split into a 40,000-image training set, a 10,000-image validation set, and a 10,000-image test set.
* **Model:** Our base model is the **DINO ViT-S/16**, a powerful pre-trained Vision Transformer, to which we added a linear classification head for the 100-class problem.

**2.2. Federated Learning Configuration**

* **Algorithm:** Federated Averaging (FedAvg).
* **Parameters:** Our setup consisted of K=100 clients, with a participation rate of C=0.1 (10 clients per round). Each client performed J=4 local training epochs per round.

**2.3. Data Distribution Scenarios**

To measure the impact of data heterogeneity, we used two data distribution strategies:

* **IID:** Each client receives a random, balanced shard of the training data.
* **Non-IID:** Each client receives data from only **2 classes (Nc=2)**, creating a severe and challenging data distribution scenario.

**2.4. Sparse Update Strategies (Model Editing)**

We implemented two gradient masking techniques:

1. **Least-Sensitive Mask:** We first calculate the diagonal of the Fisher Information matrix as a proxy for parameter sensitivity. A static mask is then created to freeze the 70% most sensitive weights, allowing updates only on the remaining 30%.
2. **Random Mask:** For each client in each round, a new mask is generated, randomly freezing 70% of the gradients.

**3. Experiments and Results**

Our experiments were designed to first establish benchmarks and then test our model editing solutions against the Non-IID challenge.

**3.1. Benchmarks: Centralized and Federated IID Training**

We first established two key benchmarks:

* **Centralized Training:** Achieved a test accuracy of **~34%**, representing the upper bound for a standard, non-federated approach.
* **Federated IID Training (No Mask):** In an ideal federated scenario with IID data, our model performed exceptionally well, reaching a test accuracy of **~77%**. This confirms our FL implementation is robust.

**3.2. The Impact of Non-IID Data**

When switching to the Non-IID setting, performance collapsed.

* **Federated Non-IID (No Mask):** The accuracy plummeted to **~6.5%**. This stark drop from the IID result (77%) clearly demonstrates the severity of the client drift problem.

**3.3. Evaluating Model Editing as a Solution**

We then applied our two masking strategies to the challenging Non-IID setting. The results are summarized in the table below and the plot in Figure 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment** | **Data** | **Model Editing Strategy** | **Final Test Accuracy (%)** |
| FL IID Baseline | IID | None | ~77.0 |
| FL Non-IID Baseline | Non-IID | None | ~6.5 |
| FL Non-IID with Strategic Mask | Non-IID | **Least-Sensitive** | 6.31 |
| FL Non-IID with Random Mask | Non-IID | **Random** | **~13.0** |

(Here, you should insert your final comparison plot that shows the IID, Non-IID, and Non-IID with masks results)

Figure 1: Validation accuracy comparison across different federated learning strategies.

**3.4. Discussion**

Our most significant finding is that the simple **Random Mask** more than doubled the accuracy of the Non-IID baseline, while the more strategic **Least-Sensitive Mask** provided no benefit. We hypothesize that this is because the random mask acts as a powerful **regularizer**. By preventing clients from overfitting to their extremely biased local data, it forces the learning of more general features, leading to more stable and effective aggregation at the server. The failure of the least-sensitive mask suggests that simply protecting old knowledge is insufficient when client drift is this severe.

**4. Conclusion**

This project successfully demonstrated the critical challenge of Non-IID data in Federated Learning. We showed that standard FedAvg, while effective in an IID setting, fails in a realistic heterogeneous environment. Our investigation into model editing revealed that a simple, non-strategic approach—a random gradient mask—was surprisingly effective. We conclude that this technique's success comes from its role as a strong regularizer, which is more crucial for mitigating client drift than more complex, knowledge-preserving methods in this setting. For future work, we propose combining this successful random masking technique with an intelligent client selection strategy, such as prioritizing clients with the highest local loss, to further improve performance.