Ideias

Renan
seleção de batches <- dados[seleção]
---> quais dados do dispositivo serão utilizados no treinamento. [seleciona forma balanceada]
--> EMD sobre o que é usado no batches

0	1	2	3	4	5	6	7	8	9	
100	250	2500	10	0	0	1800	300	4500	800	-> ruim

batchsize = 128

0	1	2	3	4	5	6	7	8	9	
100	250	250	10	0	0	500	300	500	500	-> bom

batchsize = 128

Em seguida com GAN

0	1	2	3	4	5	6	7	8	9	
600	600	600	600	0	0	600	600	600	600	-> quase ótimo

Renan
Mecanismo de pontuação de teste de modelo
0 1 9
P_global = [250 250 ... 250]

P_global_0
P_global_1
P_global_1_3_5_7_9
P_global_0_2_4_6_8
P_global_0_3_4_7_8

Quantos subconjuntos de P_global precisamos?

- 100, 1000, …

Recebo os modelos locais para a agregação
w_0 -> [pontuação] -> sobre cada subconjunto global
w_1
w_2
w_3
w_4

[pontuação]
O teste sobre vários subconjuntos gerados a partir dos dados de teste do servidor agregador.
A ideia é descobrir qual o padrão de dados que o modelo w_0 … w_4, foi treinado.

Por exemplo:
Se o w_0 foi treinando com um dataset local que só tem lobel dos dígito 5, é provável, que a teste do modelo w_0 recebido pelo servidor agregador em um subconjunto de dados P_global_5 (dados que só tem o dígito 5, gerados do conjunto de teste do servidor agregador) ->implique em alta acurácia.
Conclusão: O servidor agregador descobriu o padrão de dados treinados no modelo w_5, sem revelar os dados privado do dispositivos que treinaou w_5.

Agendamento baseado no EMD (será que é representatividade?)

IV. Formulação do Algoritmo

A. *Seleção de Dispositivos*

Considere S_t como uma fração f_t de N dispositivos e b como um vetor binário indicando a seleção de n_p dispositivos. Dessa forma, a etapa de seleção dos dispositivos participantes em cada rodada de comunicação pode ser formulado como

$$\min \sum_{i \in S_t} \left(\text{EMD}_i + \frac{1}{\text{SINR}_i^{avg}} \right) b_i, \tag{1}$$

onde EMD_i é o valor da métrica EMD do dispositivo, calculada com base na distância da distribuição de seus rótulos com uma distribuição uniforme. Além do mais, $\text{SINR}_i^{avg} = \frac{1}{R} \sum_{j=1}^{R} \text{SINR}_i$ é a média dos valores de SINR para o dispositivo i ao longo de todos os RBs e n_p é o número parcial de dispositivos selecionados para a etapa de escalonamento dos recursos de comunicação. A normalização Min-Max foi utilizada para equalizar as distribuições, garantindo que todos os termos na Equação (1) estejam na mesma escala e contribuam de forma equilibrada. A restrição $\sum_{i=1}^{|S_t|} b_i = n_p$ garante que o número de dispositivos selecionados seja igual a n_p e $b_i \in \{0, 1\}$ indica que b é um vetor binário, onde $b_i = 1$ indica que o dispositivo i foi selecionado do subconjunto S_t.

Estratégias Selecionadas

Enzo

1. Agregação baseada em conhecimento (Knowledge Federated Learning – KFL, Zhixiong Chen – Capítulo 4)

Descrição:
Em vez de compartilhar diretamente os pesos do modelo, cada cliente compartilha **predições de saída (soft labels)** em amostras públicas ou sintéticas. O servidor usa essas informações para treinar um modelo global mais estável, evitando conflitos entre pesos divergentes causados por não-IID, e reduz a necessidade de comunicação

2. CFL (Clustered Federated Learning) com fairness, Albaseer – FL Framework Over Wireless Edge

Descrição:
Agrupa os clientes com distribuições de dados semelhantes em **clusters [como clusterizar?]**, treinando modelos especializados por grupo. Um mecanismo de justiça (fairness) garante que todos os clusters participem do aprendizado. Permite que o modelo global capture variações locais sem ser "puxado" em direções conflitantes. Garante inclusão de todos os perfis de dados, mesmo os minoritários.

3. Agregação com Krum e Mediana (Salama – Decentralised FL))

Descrição:
Substitui a média aritmética tradicional por métodos mais robustos:

- **Krum:** seleciona a atualização mais próxima da maioria,

ignorando outliers.

- **Mediana:** faz agregação coordenada por componente, usando mediana em vez de média.

Filtra atualizações locais extremas que ocorrem naturalmente com dados não-IID, reduzindo os efeitos de modelos enviesados ou mal treinados.

4. Seleção fina de dados locais (Fine-Grained Data Selection, Albaseer – Capítulo sobre otimização de dados e recursos)

Descrição:
Antes de treinar localmente, o dispositivo seleciona apenas as amostras **mais informativas ou com maior impacto esperado** na perda global. Amostras redundantes ou irrelevantes são descartadas. Reduz o viés local causado por dados enviesados e melhora a eficiência energética e de comunicação.

Renan

✅ 1. "A Privacy-Preserved and Efficient Federated Learning Method Based on Important Data Selection" [ideia de não usar todos os dados dos dispositivos]

Estratégia explícita para dados non-IID: O artigo propõe um **algoritmo de seleção de dados locais com base na importância dos dados** em cada terminal. Ao invés de usar todo o conjunto de dados local, apenas **dados mais representativos (de maior importância)** são utilizados em cada iteração, o que:

- Reduz o viés causado por amostras menos informativas.
- Melhora a representatividade das atualizações locais para o modelo global.

- Indiretamente reduz o impacto de distribuições altamente heterogêneas entre os clientes.

Técnica-chave: Algoritmo de seleção de dados importante com otimização multiobjetivo via **NSGA-II**.

2. Fine-Grained Data Selection for Improved Energy Efficiency of Federated Edge Learning

Estratégia adotada:
Este trabalho ataca o problema do non-IID de forma implícita por meio da **seleção de dados em nível fino (fine-grained)**. Em vez de usar todo o conjunto de dados local de um cliente, o algoritmo seleciona **apenas os exemplos mais relevantes** para o treinamento local.

Destaque técnico:

- A exclusão de amostras redundantes ou pouco informativas ajuda a evitar que dados enviesados dominem a atualização do modelo.
- A estratégia melhora a eficiência energética e também contribui para **reduzir o impacto da heterogeneidade dos dados**, embora não realize redistribuição direta.

✅ 6. *An Energy-Aware Multi-Criteria Federated Learning Model for Edge Computing (EaMC-FL)*

Mitigação explícita:

- Realiza **clusterização dos clientes com base na similaridade entre modelos locais [como clusterizar?]**. **[como determinar a quantidade de grupos?]**
- Seleciona representantes informativos de cada cluster para atualização global. **[como selecionar o mais informativo?]**
- Garante diversidade estatística durante o processo de agregação.

DEPOIS DE CLUSTERIZAR?

- **Treina múltiplos modelos, um modelo para cada grupo.**
  - **O algoritmo utiliza as primeiras rodadas para gerar os grupos. Depois treina multiplos modelos.**
- **Usar o mais informativo, para gerar um modelo único (Ensemble Learning).**

O Ensemble Learning se destaca como uma estratégia extremamente eficaz para melhorar o desempenho de modelos de aprendizado de máquina, especialmente em cenários com dados ruidosos, não-lineares ou altamente desbalanceados. No entanto, seu uso deve ser cuidadosamente calibrado, considerando o trade-off entre **acurácia e interpretabilidade**, bem como o custo computacional.

Capítulo 6 — Personalization (Dissertação - Federated Learning)

- **Descrição**:
  - Propõe **personalização local** dos modelos:
    - **Transfer Learning** (finetuning do modelo global no cliente local).
    - **Personalization Vectors**: clientes aprendem vetores de ajuste local.
- **Mitigação de non-IID**:
  - **Sim**:
    - Ao permitir ajustes locais no modelo global, **cada cliente pode adaptar o modelo às suas características únicas**, minimizando o impacto do non-IID.

25. A Hierarchical Knowledge Transfer Framework for Heterogeneous Federated Learning

- **Estratégia:** Proposta do **FedHKT**, uma estrutura de transferência hierárquica de conhecimento que agrupa clientes com distribuições de dados similares em clusters para treinamento colaborativo, e transfere o conhecimento especializado ao servidor, onde é agregado e redistribuído.

5. A Clustered Federated Learning Paradigm with Model Ensemble in O-RAN

- **Estratégia:** Introdução do **Clustered Federated Learning (CFL)** com **ensemble de modelos**. Dispositivos são agrupados em clusters e seus modelos são posteriormente integrados.

8. Ensemble Federated Learning With Non-IID Data in Wireless Networks

- **Estratégia:** Proposta do **Ensemble Federated Learning (EFL)** com **formação de clusters de usuários** baseada em similaridade dos dados, aplicando posteriormente um **ensemble de modelos**.

- **Comentário:** A formação de clusters minimiza a heterogeneidade dentro de cada grupo e o ensemble final combina esses modelos para reduzir o impacto dos dados non-IID no desempenho global.

1. Hierarchical Federated Learning in MEC Networks with Knowledge Distillation

Estratégia explícita para mitigar non-IID:

- O artigo propõe o uso de **Knowledge Distillation** no processo de treinamento local.

- **Como funciona:** durante o treinamento local, o modelo global e um conjunto de modelos históricos regionais (dos edge servers) são usados como fontes de conhecimento para guiar o treinamento dos clientes.

- **Objetivo:** impedir que os modelos locais se desviem do conhecimento global, o que é um problema causado pela deriva de distribuição (distribution drift) em ambientes non-IID, principalmente por mobilidade dos clientes.

Comentário técnico:

- A estratégia é claramente desenhada para **reduzir a deriva entre os dados locais** e o modelo global, sendo uma forma ativa de combate ao impacto da heterogeneidade dos dados ao manter uma "força guia" externa sobre o aprendizado local.

9. Decentralized Federated Learning for Road User Classification in Enhanced V2X Networks

Estratégia explícita para mitigar non-IID:

- Proposta de **Federação Descentralizada** com **troca seletiva de camadas do modelo** (não todos os parâmetros).

- **Como funciona:** os veículos trocam apenas subsetores do modelo, como camadas específicas otimizadas para comunicação e performance, em vez do modelo inteiro.

- **Objetivo:** focar em camadas mais sensíveis à tarefa global, reduzindo o impacto de variações locais causadas por datasets non-IID.

4. Hierarchical Federated Learning with Edge Optimization in Constrained Networks

Estratégia explícita para mitigar non-IID:

- O artigo propõe duas abordagens:

    - **FedSLT (Federated learning with Selected-Layer Transmission):** permite a transmissão parcial do modelo (apenas algumas camadas) para reduzir sobrecarga de comunicação e inconsistências de atualização, útil em cenários com links instáveis.

    - **Edge Aggregation (EA) e Warm-up (WU):** algoritmos de inicialização em servidores de borda que criam modelos personalizados e mais próximos da distribuição local antes da agregação global.