Multi-Task Federated Learning for Personalised Deep Neural Networks in Edge Computing: A Critical Analysis

---

1. Introduction

Federated Learning (FL) enables collaborative model training without centralising data, addressing privacy and communication overhead. Yet, non-IID data across devices makes a single global model inadequate for diverse user needs. Personalisation is crucial for tasks like smartphone keyboard predictions, where each user's unique typing style must be captured. To address this, researchers propose multi-task FL approaches, such as Mills, Hu, and Min (2021). They use local batch normalisation parameters for user-specific tuning, improving personalisation while retaining collaborative benefits.

---

2. Research Questions and Motivations

The paper focuses on two primary research questions:

1. *How can federated learning be adapted to effectively handle non-IID data, ensuring improved accuracy for each client?*

2. *How can one incorporate personalisation while still retaining the advantages of federated collaboration?*

Motivations

1. Non-IID Data Complexity - When client data distributions diverge significantly, a single global model may underperform for specific users. Personalisation strategies that adjust certain parameters locally can mitigate this issue.
2. Privacy and Resource Constraints - Regulations and device limitations (e.g. battery and bandwidth) demand methods that minimise data transmission and secure user privacy. FL naturally lends itself to privacy preservation since raw data is never shared.
3. Real-World Applicability - An FL-based approach that can effectively handle varied local distributions is crucial for broader adoption. This will benefit businesses requiring personalisation, such as those involved in recommender systems or healthcare services.

By focusing on client-centric solutions, the paper aims to improve performance for every user without sacrificing the collaborative benefits of FL.

---

3. Contributions

The paper proposes a Multi-Task Federated Learning (MTFL) architecture that maintains a shared global Deep Neural Network (DNN) across all clients but reserves local Batch Normalisation parameters for each client's unique data distribution. The key contributions can be summarised as follows:

1. Local Batch Normalisation for Personalisation
    - The paper isolates means and variances in BN layers, preventing their aggregation across clients. This design ensures each device retains its environment-specific normalisation, this results in effective and lightweight personalisation.

2. User Accuracy (UA) Metric
    - Instead of relying on a single global accuracy metric, the paper proposes UA, which averages individual client accuracies. This metric better reflects performance for users whose data distribution may significantly differ from the global average.

3. Adaptive Momentum-Based Optimisation
    - The paper integrates momentum updates (FedAvg-Adam) to address convergence challenges in highly skewed datasets. By incorporating adaptive learning rates, they reduce the number of required communication rounds, which is critical edge devices with limited resource.

4. Empirical Validation
    - Experiments on the MNIST and CIFAR-10 datasets - They were partitioned to simulate non-IID distributions, demonstrating that local BN improves local performance.
    - Deployment on Raspberry Pi devices: This illustrates real-world feasibility, even under power and bandwidth constraints.

These contributions collectively advance the FL field by demonstrating effective and relatively simple strategy for personalisation that balances performance gains with practical constraints in edge computing environments.

4. Research Methodology and Methods

4.1 Research Methodology

The paper adopts a standard federated averaging (FedAvg) pipeline, in which:

1. Local Training - Each client trains a local copy of the global model on its private dataset.
2. Update Aggregation - The central server collects the locally updated model parameters and aggregates them, by averaging, to form a new global model.
3. Iteration - The updated global model is distributed back to clients for further local training in multiple rounds.

The methodological innovation is that the local BN parameters (means and variances) are excluded from the aggregation step. Each client therefore maintains BN statistics aligned with its own data distribution. All other network parameters are updated collaboratively.

It is noteworthy that recent studies, such as Zhong et al. (2024), have highlighted that relying on mini-batch statistics in BN layers can lead to divergent local gradients in non-IID settings. In comparison to this paper where the authors normalise parameters rather than aggregating BN statistics across clients. This alignment between Mills, Hu, and Min (2021) and Zhong et al. (2024) further reinforces the argument that preserving user-specific BN statistics is crucial for mitigating gradient divergence in non-IID federated learning environments.

4.2 Research Methods

1. Data Splitting and Non-IID Simulation
   - The authors create heterogeneous partitions of the MNIST and CIFAR10 datasets. Some clients receive only a subset of classes, while others experience highly imbalanced label distributions. This setup mimics real-world non-IID conditions.

2. Optimisation - FedAvg-Adam
   - Beyond the standard FedAvg approach, the authors experiment with Adam as the local optimiser, adding adaptive momentum to accelerate convergence and stabilise training under skewed distributions.

3. Performance Evaluation
   - User Accuracy (UA) - They compute client-specific accuracies and average them, rather than using a single global metric.
   - Convergence Speed - The authors measure how quickly the models converge by tracking the number of communication rounds.
   - Edge-Device Deployment - They quantify the training time and resource usage on Raspberry Pis, ensuring that the solution remains viable for constrained hardware.

Together, these methods confirm that local BN personalisation can improve user-level performance while maintaining efficient, model updates, in a private manor.

5. Critique/Evaluation

1. Strengths

    - Lightweight Personalisation – By restricting personalisation to BN parameters, MTFL achieves notable gains in local performance without excessively increasing communication overhead. This approach finds strong support in the work of Liu et al. (2023) on FedDWA, where the use of dynamic weights for server aggregation also highlights that minimal, strategically chosen parameter adaptations can lead to significant improvements in local accuracy. Such evidence aligns with the idea that even modest adjustments, when well-targeted, are sufficient to enhance performance while preserving the global collaboration benefits inherent to federated learning.
    - Meaningful Metric - User Accuracy provides a more transparent view of how individual clients benefit, highlighting the potential real-world impact on specific use cases.
    - Practical Validation - Demonstrations on physical devices enhance the credibility of the proposed approach for real-world edge applications.

2. Limitations

    - Narrow Data Scope – The MNIST and CIFAR10 datasets do not fully represent the complexity of modern tasks, such as large-scale medical imaging or natural language models. The performance of local BN personalisation in deeper architectures (ResNet or Transformers) remains an open question.
    - Hyperparameter Tuning - The paper does not offer exhaustive guidance on tuning local BN momentum, learning rates, or other key parameters. Lack of systematic tuning protocols may hinder broader adoption.
    - Security Considerations - While local BN statistics limit direct data exposure, the system could still be vulnerable to malicious clients updates.

3. Future Directions

    - Scaling to Large Models - Investigating how local BN scales to more complex deep networks and diverse tasks can solidify MTFL's generality.
    - Advanced Communication Strategies - In large deployments, even minimal overhead can scale significantly. Employing sparse updates, adaptive communication intervals, or compression techniques may further optimise resource usage.
    - Robustness and Security - Incorporating adversarial defences and privacy-preserving mechanisms would enhance trust, critical for some applications, such as in medical applications.

Conclusion

By using local Batch Normalisation parameters, this paper effectively tackles the non-IID data challenge in federated learning while preserving collaborative benefits. Enhanced User Accuracy and deployment on edge devices confirm its real-world practicality. However, the limited scope of datasets, minimal hyperparameter tuning guidelines, and security gaps suggest opportunities for further research. Future work may involve scaling to more complex models, refining communication strategies, and integrating stronger adversarial defences to aid in both performance and trust in federated learning deployments.

References

- Mills, J., Hu, J. and Min, G. (2021) - *Multi-task federated learning for personalised deep neural networks in edge computing* - https://arxiv.org/pdf/2007.09236
- Zhong, J., Chen, H.-Y. and Chao, W.-L. (2024) - *Making Batch Normalisation Great in Federated Deep Learning* - https://arxiv.org/abs/2303.06530
- Liu, J., Wu, J., Chen, J., Hu, M., Zhou, Y. and Wu, D. (2023) - *FedDWA: Personalized Federated Learning with Dynamic Weight Adjustment* - https://arxiv.org/pdf/2305.06124