Advancing Generalization in Heterogeneous Federated Learning for Real-world Vision Applications

Summary

Debora Caldarola

This thesis focuses on Federated Learning (FL)[1], a machine learning framework that enables collaborative training of a global model across edge devices (*clients*) without compromising user privacy. Unlike traditional centralized approaches that require transferring raw data to a central server, FL operates by exchanging model parameters and performing training locally. Given that most data today originates from edge devices like smartphones and Internet of Things hardware, FL offers a path to leverage this vast and sensible resource while remaining compliant with privacy regulations.

This dissertation addresses **key challenges in deploying FL in real-world scenarios**, where personal habits introduce inherent bias in the local data distributions and users' devices vary widely in computational resources and network reliability. Within this context, optimizing communication efficiency and mitigating the effects of non-uniform data distributions are critical. In particular, this thesis highlights the existence of a significant gap between research on FL and its deployment in real-world complex applications, particularly in the vision domain. The presented works seek to uncover underlying causes of poor model generalization in realistic scenarios and develop an approach that generalizes effectively to the overall data distribution, emphasizing **vision-oriented applications**.

The core contribution of this work lies in leveraging the **geometry of the loss landscape** to understand and mitigate the behavior of models trained in heterogeneous federated scenarios. In particular, recent research trends have identified a link between models' generalization ability and the geometry of their convergence point in the loss landscape, associating poor generalization with sharp minima. However, no prior work has examined the relationship between the geometry of the loss landscape and limited generalization in heterogeneous FL. This thesis explores the hypothesis that convergence to sharp minima may contribute to the poor generalization typical of FL models in heterogeneous settings, and shows that discrepancies between local and global loss surfaces

[1] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.

are influenced by data heterogeneity. By promoting convergence towards globally flat minima with sharpness-aware strategies (*e.g.*, sharpness-aware minimization[2]), substantial improvement in generalization performance can be achieved in several tasks (*e.g.*, large-scale image classification, semantic segmentation, domain generalization), while maintaining communication efficiency. This work not only leads to better performance across diverse data distributions in heterogeneous FL but also opens new avenues for understanding the dynamics of model performance in complex learning environments.

Additionally, to accelerate the speed of convergence and reduce the communication cost, this thesis proposes **novel training orchestration paradigms** that leverage privacy-preserving similarity metrics to facilitate the grouping of similar or dissimilar clients, optimizing collaboration in FL. Differently from previous existing research that often utilizes clustering techniques, this work proposes a new training framework that leverages groups of *dissimilar* clients to reduce communication exchanges while mitigating the models' destructive interference typical of heterogeneous FL environments, resulting in improved model quality and convergence speedup. This approach offers a winning alternative to both the conventional client-server architecture and existing methods that group clients based on similarity to learn cluster-specific models.

Furthermore, the standard FL approach is based on learning a global model that *on average* minimizes all local empirical risks across the entire population. In contrast, personalized FL approaches often learn model parameters specific to each client, aiming for better local performance. However, the former approach may lead to models with limited generalization capabilities, while the latter can result in information loss, as knowledge gained from previous clients is not effectively incorporated. Addressing this challenge, this thesis proposes to leverage Graph Convolutional Neural Networks to enable learning domain-specific information while facilitating knowledge exchange between similar clients.

Finally, with most future data expected from vision-based edge systems, computer vision tasks are notably underrepresented in FL research, partly due to a lack of benchmarks and large-scale federated datasets. To address this gap, this thesis introduces three **novel vision benchmarks for FL**, focusing on semantic segmentation for autonomous driving and collaborative visual place recognition. By demonstrating the effectiveness of state-of-the-art FL algorithms in these diverse and more complex tasks, this work paves the way for their wider adoption across various computer vision domains, broadening the applicability of FL.

[2]Foret, Pierre, et al. "Sharpness-aware minimization for efficiently improving generalization." International Conference on Learning Representations. 2021.