8-16-2024

Mitigating Health Disparities with Fairness-Aware Personalization in Federated Learning

Tongnian Wang
The University of Texas at San Antonio, tongnian.wang@utsa.edu

Yuanxiong Guo
The University of Texas at San Antonio, yuanxiong.guo@utsa.edu

Kim-Kwang Raymond Choo
University of Texas at San Antonio, raymond.choo@utsa.edu

Mitigating Health Disparities with Fairness-Aware Personalization in Federated Learning

TREO Talk Paper

Tongnian Wang
The University of Texas at San Antonio
tongnian.wang@utsa.edu

Yuanxiong Guo
The University of Texas at San Antonio
yuanxiong.guo@utsa.edu

Kim-Kwang Raymond Choo
The University of Texas at San Antonio
raymond.choo@utsa.edu

Abstract

As Artificial Intelligence (AI) proliferates into various aspects of our life, concerns are growing regarding its potential to cause harm by introducing biases across socio-demographic groups. Studies have reported that AI models can generate biased outcomes that disproportionately affect certain socio-demographic groups (Obermeyer et al. 2019; Leslie et al. 2021). Since AI models highly rely on datasets, any disparities in healthcare data among different socio-demographic groups can lead to new health inequalities through data-driven, algorithmic AI models. Therefore, it is essential to develop AI models that are unbiased and perform consistently across all socio-demographic groups to help prevent and mitigate health disparities. However, the healthcare sector faces additional challenges, particularly its reliance on real-world data for AI applications and stringent privacy laws like GDPR and HIPAA that restrict data sharing between entities. These restrictions can create a challenging environment where AI systems, if poorly designed or based on limited data from specific populations, may perpetuate or even worsen discrimination among different socio-demographic groups. To address these issues, Federated Learning (FL) (McMahan et al. 2017) provides a promising solution by enabling entities to collaboratively learn a global AI model without sharing their data. Nevertheless, FL can sometimes increase bias, particularly during the model fusion process where biases in larger datasets may be amplified if these datasets are given more weight. Additionally, differences in data distributions among entities and socio-demographic groups can lead to decreased model performance. If only the global model is shared across entities, there is little opportunity to obtain an accurate and unbiased model.

In this work, we aim to develop a fairness-aware personalization algorithm with FL setting to reduce health disparities while preserving high model accuracy. Unlike traditional FL approaches that rely solely on a global model, our method incorporates representation learning for model decoupling, enabling entities to learn a powerful shared representation with their personalized heads to tailor their model for better performance on their local private datasets. Specifically, our method leverages federated representation learning and formulates a fairness-constrained optimization problem. By reducing reliance on the global model via personalized model in FL, we can potentially mitigate discriminatory impact of bias propagating through FL to all the participating entities.

References

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. 2019. “Dissecting racial bias in an algorithm used to manage the health of populations,” Science (366:6464), pp. 447–453.

Leslie, D., Mazumder, A., Peppin, A., Wolters, M. K., and Hagerty, A. 2021. “Does ‘AI’ Stand for Augmenting Inequality in the Era of Covid-19 Healthcare?,” BMJ, p. n304.

McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. y. 2017. “Communication-Efficient Learning of Deep Networks from Decentralized Data,” in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR, April 10, pp. 1273–1282.