Quantifying fairness of federated learning LPPM models
Amina Ben Salem, Besma Khalfoun, Sonia Ben Mokhtar, Afra Mashhadi

To cite this version:

HAL Id: hal-03703623
https://hal.archives-ouvertes.fr/hal-03703623
Submitted on 24 Jun 2022
Poster: Quantifying Fairness of Federated Learning LPPM Models

Amina Ben Salem
Besma Khalfoun
Sonia Ben Mokhtar
INSa Lyon
Lyon, France
{firstname}.{last-name}@insa-lyon.fr

Afra Mashhadi
University of Washington
USA
mashhadi@uw.edu

ABSTRACT

Despite the great potential offered by Artificial Intelligence in the context of smart mobility, it comes with the greater challenge of preserving the privacy of users. Federated Learning (FL) has gained popularity as a privacy-friendly approach, however, an equally important aspect rarely addressed in the literature is its fairness. In this work we audit a FL-based privacy-preserving model. We use Entropy to determine similarity within the system’s input data and compare its value against that of the output to detect unfair treatment.

KEYWORDS

Privacy, Fairness, Federated Learning

1 INTRODUCTION

Understanding human mobility based on location-based data generated by smartphone devices has become a fundamental part of urban and environmental planning in cities. Through the collection of these geo-traces, it has become possible for the scientific community and policy-makers to model citizens’ daily commutes using crowd-sensed car-share data [5], city bicycles [8] and RFID cards [13], or to build predictive algorithms to estimate people’s flows [15] for traffic management and community resources [3]. However, location data is highly sensitive in terms of privacy as it can reveal a great level of information about individuals.

Recently, a new set of works have been proposed that have leveraged decentralized methods to tackle privacy concerns. For example, the mobile crowd-sensing community has started to explore alternatives and possibilities of a paradigm shift that would decouple the data collection and analysis from a centralized approach to a distributed setting by moving towards Federated Learning [4, 11]. In Federated Learning (FL) end-devices train their own models using locally preserved training data while sharing the benefits of a global aggregated model across all clients [12]. Other approaches have proposed FL models to automatically assign the best-suited location privacy-preserving methods (LPPM) [6]. While substantial research progress has been made in the area of privacy, little attention has been given to auditing the fairness of these black-box models which are orchestrated in a decentralized setting. Existing algorithms for fairness auditing are designed under the assumption of centralized settings and operate with the assumption that they have unrestricted access to the model [1, 2].

Motivated by these gaps, in this poster, we audit the fairness of a FL model designed for preserving the privacy of individuals. In order to do so, we first implement a set of metrics for measuring and evaluating fairness in the context of spatial-temporal FL models as previously proposed by [10]. We then audit a FL model for enhancing the location privacy of users, namely EDEN [6]. EDEN is a FL model that automatically selects the best Location Privacy-Preserving Method (LPPM) and its corresponding configuration without sending raw geo-located traces outside the user’s device. In this work, we treat EDEN as a black box, and to assess the outcome of EDEN on traces we rely on pre and post entropy of trajectories as we detail next.

2 BACKGROUND AND DEFINITION

Literature on fairness in machine learning strives to avoid the fact that the decision made by automated systems and algorithms are skewed toward the advantaged groups or individuals, by examining fairness from two perspectives of group based and individual based fairness. Individual fairness claims that similar individuals should be treated similarly regarding their specific task. In most cases, the difficulty with individual fairness lies in the notion of measuring similarity. To measure individual fairness in the context of spatial temporal applications, we need two sets of definitions corresponding to the similarity between users’ trajectories, and the similarity of the outcome of the FL model. In this work we consider that individuals who are similar in terms of their mobility, should receive an equal privacy gain from EDEN. We define the similarity of trajectories by borrowing tenets from mobility literature and measure entropy of users as a measure of their maximum predictability. In this paper, we define entropy as a measure of Shannon Entropy ($E_h$). A larger entropy indicates greater disorder, and consequently reduces the predictability of an individual’s movements. We define entropy following notion in [9, 14] and measure ($E_h$) as:

$$E_h = -\sum_{i=1}^{n} P(x_i) \log_2 [P(x_i)]$$

(1)
where \( n \) is the length of probability vector, \( P(x_i) \) is the probability of visiting location \( x_i \) considering only spatial pattern.

To assess the fairness of EDEN, we hypothesize that user traces with similar entropy should receive similar privacy gain. As we treat EDEN as a black-box model, we assess privacy gain as the entropy of the traces after EDEN has been applied. In an ideal setting, we expect the entropy to increase for all the users (i.e., predictability to decrease). To measure individual fairness we thus compare the entropy of similar user traces to their output by EDEN and we study in detail the percentage of users for whom the entropy decreases after applying EDEN. We refer to this group as the disadvantaged group.

3 PRELIMINARY RESULTS

We evaluate the fairness of EDEN on three mobility datasets but we illustrate the results of only one dataset: MDC [7] due to the lack of space. Figure 1 presents the different levels of pre and post entropies for EDEN’s LPPMs schemes for the MDC dataset. We observe that EDEN increases the entropy of most users. Indeed the cases where we find the outcome of EDEN to disadvantage users (decrease their entropy) are 3% for MDC. We next study the fairness for those traces that correspond to the disadvantaged group.

![Figure 1: Entropy level of each LPPM schema on raw traces and post-EDEN traces of MDC dataset.](image)

Figure 2 presents the entropy decline for the disadvantaged users for the MDC dataset. As we can see the users with lower entropy prior to applying EDEN receive a relatively less decline in their post-EDEN entropy as well as smaller variation. In this plot, the size of each box presents the fairness as measured by the difference in outcome after applying EDEN. That is users who initially had lower predictability (high pre-EDEN entropy) exhibit a larger variation in their post-entropy after applying EDEN, corresponding to different treatments. Likewise, users with low entropy (highly predictable patterns) receive a similar outcome from EDEN.

4 CONCLUSION

In summary, in this paper, we have studied the trade-off between privacy and fairness and have presented our early results in auditing a black-box privacy-preserving model, EDEN, on two real-life datasets. We have shown that while EDEN increases the overall entropy of users (decreasing their predictability), for a very small percentage of users it fails to achieve fairness. Our future directions include designing and implementing our methodology under an automatic framework that could audit any black-box FL privacy model by intercepting the input and output of the model.

REFERENCES