Reducing Discrimination in Learning Algorithms for Social Good in Sociotechnical Systems

Katelyn Morrison
SCI, University of Pittsburgh, United States
katelyn.morrison@pitt.edu

Abstract

Sociotechnical systems within cities are now equipped with machine learning algorithms in hopes to increase efficiency and effectiveness by modeling and predicting trends. Machine learning algorithms have been applied in these domains to address challenges such as balancing the distribution of bikes throughout a city and identifying demand hotspots for ride sharing drivers. These algorithms applied to challenges in sociotechnical systems have exacerbated social inequalities due to previous bias in datasets or the lack of data for marginalized communities. In this paper, I will address how smart mobility initiatives in cities use machine learning to address challenges and how these algorithms unintentionally discriminate against features such as socioeconomic status to motivate the importance of algorithmic fairness. Using the bike sharing program in Pittsburgh, I will present a position on how discrimination can be eliminated from the pipeline.

1 Introduction

As smart cities emerge, sociotechnical systems are starting to eliminate traditional systems such as replacing modern transportation systems with state-of-art technologies. Citizens can now interact with these systems through hardware and software leaving behind traces of their data [Rangwala, 2018]. An example of a sociotechnical system within a city is a transportation system [Levine, 2020] - most recently evolving as smart mobility systems. Smart mobility initiatives are spreading from city to city, introducing dynamic, on-demand bike sharing programs, ride hailing programs, and car sharing programs [Yan and Howe, 2020].

Emerging smart mobility systems are supported by a variety of Information and Communication Technologies (ICT) [Benevolo *et al.*, 2016] which have allowed for the possibility of collecting copious amounts of data on citizens mobility patterns. Citizens interact with navigation apps, bus routing apps, ride sharing apps, and other transit related apps numerous times throughout their day to navigate the city. These citizens play an important role within several systems in the city including the transportation system and the future governance related to it. Their data makes up a large dataset that is then used in learning algorithms to assist urban planners, policy makers, and government officials in evaluating policies or making decisions. In terms of smart mobility initiatives, the data helps decision makers evaluate where to redistribute resources and when to eliminate resources resources in an area with little demand [Yan and Howe, 2020; Luo *et al.*, 2020]. Nonetheless, these complex, dynamic systems collecting copious amounts of data, and the citizens providing the data, are rapidly becoming important agents in the decision making and policy making processes.

The fairness and availability of smart mobility initiatives and the learning algorithms applied to them within a city are important to assess given that smart mobility initiatives are motivated by providing a cheap, yet dynamic, on-demand mode of transportation for citizens [Yan and Howe, 2020]. Initiatives such as bike sharing programs have the potential to increase the overall quality of life and health of citizens [Mooney *et al.*, 2019]. Other initiatives such as car sharing on ride hailing have the potential to reduce traffic congestion which ultimately helps reduce the CO_2 emissions [Luo *et al.*, 2020]. Furthermore, these affordable, dynamic, on-demand modes of transportation help connect citizens to the job market and economy within cities. However, these smart mobility initiatives only fulfill their mission if they are equally accessible to all social classes and all neighborhoods within a city. Using previous usage data, learning algorithms have been deployed in an attempt to complement the system. Unfortunately, learning algorithms applied to these initiatives tend to intensify any social inequalities that are present in the previous usage data [Chui *et al.*, 2019].

2 Related Work

The awareness of algorithmic bias and lack of fairness in learning algorithms is not novel, but the notion of fairness in learning algorithms applied to urban systems has been recently explored. Recent research has highlighted the lack of previously acknowledging the impact these learning algorithms have had on mobility systems [Yan and Howe, 2020]. There are several works considering novel techniques to predict demand of smart mobility systems or identify hotspots at a given space and time, but as [Yan and Howe, 2020] mentioned, ”No existing work in modeling urban resource demand considers fairness in their solutions". The following works that consider or show a lack of considering fairness within their algorithm provide a strong motivation that learning algorithms may need to be re-evaluated in the context of urban and social systems.

2.1 Applications in Sociotechnical Systems

[Yan and Howe, 2020] develops a model that represents the demand prediction of bikes in a bike sharing program or demand hotspots of rides for a ride hailing driver. Specifically, this model is based on three convolutional neural networks (CNN) combined into one to model the demand of smart mobility resources. They use regularizers in their loss function to ensure fairness while training and testing the model. In their work, they define fairness, "... as the requirement that individuals of different demographic groups receives equal amount of mobility resources".

Another study looks at the expansion of electric vehicle sharing systems and uses a data-driven approach to predict the demand of the system currently, in the near future, and in the long term [Luo *et al.*, 2020]. They use features from a station while considering the global spatial and local temporal attributes to a given system's station network. Their demand prediction model is made up of Graph Convolutional Neural Networks (GCN), but do not strongly consider fairness of the prediction given that they are incorporating historical knowledge.

Two recent studies evaluate fairness, but in a slightly different context. Switching from the customer's, or user's, perspective to the provider's, or driver's, perspective, these two studies focus on the fairness of the drivers' income or opportunity to pick up passengers [Rong *et al.*, 2019; Lesmana *et al.*, 2019]. [Rong *et al.*, 2019] specifically studies this problem in the context of Changsha, China to optimize the dispatching of taxis to predicted hotspots fairly. [Lesmana *et al.*, 2019] focuses more broadly on the ride hailing system instead of specifically taxis and evaluate their contributions on data from New York City. They both want to optimize the system's efficiency while incorporating the fairness of revenue among drivers and acknowledge its potential to be applied in other domains relating to sociotechnical systems.

2.2 Fairness in Learning Algorithms

Learning algorithms used in the domain of smart mobility have provided significance assistance to decision makers and urban planners. However, individuals outside the domain of smart mobility or outside the domain of machine learning may be unaware of unintended consequences that occur when deploying these algorithms. The fairness in supervised learning algorithms are specifically worth investigating given that the data being used in the context of smart mobility mainly consists of labeled historical or spatial-temporal user data.

[Yan and Howe, 2020] achieves fairness in their demand prediction algorithm by using fairness regularizers in their deep convolutional neural network. They created two regularizers, one for region based and one for individual based that were dependent on group labels.

[Hardt *et al.*, 2016] provides a definition of how to remove any discrimination that was learned by the predictor. They also give more power to decision makers interacting with the model. They provide a framework to construct classifiers that are deemed fair that are modified in the post-processing step to minimize loss.

2.3 Evaluating Accessibility in Systems

Several studies have evaluated and investigated the accessibility, fairness, and discrimination present within current smart mobility systems. Many of these studies have leveraged geospatial and spatial-temporal analysis techniques to reveal current and past trends.

A case study of Pittsburgh's HealthyRide bike sharing program evaluates the fairness, or accessibility, of the current distribution of bike stations with future works to perform spatial-temporal analysis to identify the growth of stations relative to neighborhoods' socioeconomic status over time [Morrison, 2020]. It is discovered when the study was conducted in April 2020 that 95% of the bike stations were placed in neighborhoods that were considered to have little to no evidence of poor housing. The poor housing feature was used as an indicator of socioeconomic status in this study. Fairness overall was determined by using three features: capacity of bike stations, location of bikes stations, and the percentage of poor housing in a neighborhoods.

Two studies in Manizales, Colombia investigate the social accessibility of the public bike sharing system, *Manizales En Bici*. which requires the bikes to be returned and retrieved from a station. An important characteristic of Manizales it that it is split into sectors, neighborhoods more generally, that are grouped by socioeconomic status. One study investigates the problem of unequal distribution of the bikes by calculating the minimum travel time to get around the city in comparison to socioeconomic status and geography [CARDONA *et al.*, 2017]. The other study uses ArcGIS software [Zuluaga Garcia and others, 2017]. Overall, both of these studies agree that the locations of the *Manizales En Bici* stations were more accessible or available to individual of higher socioeconomic status.

A separate study evaluated a dockless bike sharing program, where the bikes do not require a station to be returned to [Mooney *et al.*, 2019]. This study investigates the bike sharing program in Seattle and discovered that while no neighborhood was being exclusively avoided, the richer neighborhoods tended to have more bikes available. Their conclusion from the studied pointed out that dockless smart mobility systems can potentially offer more equity than other models.

3 Future Works

In smart mobility systems, it is common to use supervised learning algorithms to predict demand, identify demand hotspots, and the best location to distribute resources in the present day, near future, and in the long term to help decision makers and urban planners. These algorithms are motivated to help mitigate the exploitation vs. exploration problem that bike sharing programs are encountered with. Considering the Pittsburgh HealthyRide bike sharing program, I will identify a novel approach to identify demand hotspots and prime locations while remaining impartial to certain features such as socioeconomic status, race, and geographic location.

Pittsburgh's bike sharing program has 100 bike stations which have been placed based on 6 criteria defined by the HealthyRide organization. They define a good location for a station to be within $1/4$ mile from another station, on public land, not obstructing other public utilities, not hidden in allies or side streets, has a certain amount of square footage to build the station, and has access to sunlight [HealthyRide, 2020]. Their current plan for expansion is undetermined and they are asking for users to fill out a feedback form. Instead of soliciting feedback in this case, I propose using Bayesian Optimization for model-based reinforcement learning to explore new areas for stations followed by Bayesian inference to update the prior and posterior based on evidence that results from the exploration.

In the case of bike sharing programs, the agent is a planner and they are trying to optimize the location to put a bike station or the location to put more bikes. It is costly and takes time to build the bike stations, so the planners want to make sure that they will build the station in a location that will have a high demand. The planner is faced with an exploration vs. exploitation trade off. An agent (i.e. planner) would be rewarded every time it explores uncharted neighborhoods or neighborhoods with very little historical user data. These rewards are to encourage the planner to explore and identify regions that do not have any source of historical data to contribute to evidence of demand. The policies for the agent (i.e the planner) will represent the probabilities for each action that can be taken (i.e. the next neighborhoods a planner should construct a station at). These probabilities would be calculated and updated by using an acquisition function.

The value function can be defined as the posterior using historical user data (i.e. the evidence), the prior belief (i.e prior likelihood of demand), and the current likelihood given the location the planer chooses. Currently, the bike sharing programs place bikes and stations in regions where they have a high certainty that demand will be high. If they ignore this high certainty, they will eventually place bikes or stations in regions that could have otherwise been excluded or discriminated against.

This approach allows planners to consider evidence of past user demand for certain locations and times. It also allows for the demand of station locations in previously unexplored regions to be identified. This approach forces the bike sharing programs to explore new regions and potentially identify a new cohort of users for their system.

4 Conclusion

When considering future work in fairness-aware learning algorithms applied in sociotechnical systems, researchers should avoid developing and deploying algorithms that adhere to "fairness through unawareness" [Hardt *et al.*, 2016]. Identifying alternatives to the current demand and hot spot prediction algorithms are imperative to increasing the accessibility of smart mobility options to all citizens. Identifying and evaluating current and novel approaches to implement learning algorithms within sociotechnical systems for social good is only started to be addressed. There is a lot of work in the field of fairness in machine learning and a lot of work in deploying learning algorithms in sociotechnical systems to achieve efficiency, but these two have remained domain agnostic of each other. Future work should consider novel approaches that combine that novel techniques from both domains to reduce discrimination. I identify a approach using Bayesian optimization for model-based reinforcement learning to help planners identify locations to place smart mobility stations (i.e car sharing stations or bike sharing stations). From the continuous user data, Bayesian inference can be used to help predict new demand as the planners reevaluaute locations.

References

[Benevolo *et al.*, 2016] Clara Benevolo, Renata Paola Dameri, and Beatrice D'auria. Smart mobility in smart city. In *Empowering Organizations*, pages 13–28. Springer, 2016.

[CARDONA *et al.*, 2017] Mateo CARDONA, Juan ZULUAGA, and Diego ESCOBAR. Análisis de la red de ciclorutas de manizales (colombia) a partir de criterios de accesibilidad territorial urbana y cobertura de estratos socioeconómicos. *Revista Espacios*, 38(28), 2017.

[Chui *et al.*, 2019] Michael Chui, Martin Harrysson, James Manyika, Roger Roberts, Rita Chung, Pieter Nel, and Ashley van Heteren. Applying artificial intelligence for social good, Nov 2019.

[Hardt *et al.*, 2016] Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. *CoRR*, abs/1610.02413, 2016.

[HealthyRide, 2020] HealthyRide. Station planning, 2020.

[Lesmana *et al.*, 2019] Nixie S Lesmana, Xuan Zhang, and Xiaohui Bei. Balancing efficiency and fairness in on-demand ridesourcing. In *Advances in Neural Information Processing Systems*, pages 5309–5319, 2019.

[Levine, 2020] Ben Levine. Data-driven research aims to solve first/last mile problem. March 2020.

[Luo *et al.*, 2020] Man Luo, Bowen Du, Konstantin Klemmer, Hongming Zhu, Hakan Ferhatosmanoglu, and Hongkai Wen. D3p: Data-driven demand prediction for fast expanding electric vehicle sharing systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–21, 2020.

[Mooney *et al.*, 2019] Stephen J Mooney, Kate Hosford, Bill Howe, An Yan, Meghan Winters, Alon Bassok, and Jana A Hirsch. Freedom from the station: Spatial equity in access to dockless bike share. *Journal of transport geography*, 74:91–96, 2019.

[Morrison, 2020] Katelyn Morrison. Mapping bike sharing stations and neighborhood housing conditions, Apr 2020.

[Rangwala, 2018] Harsh Rangwala. Socio-technical systems, Dec 2018.

[Rong *et al.*, 2019] Huigui Rong, Qun Zhang, Xun Zhou, Hongbo Jiang, Da Cao, and Keqin Li. Tesla: A centralized taxi dispatching approach to optimizing revenue efficiency with global fairness. 2019.

[Yan and Howe, 2020] An Yan and Bill Howe. Fairness-aware demand prediction for new mobility. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1079–1087, 2020.

[Zuluaga Garcia and others, 2017] Juan David Zuluaga Garcia et al. *Propuesta metodologica para el diagnostico y planificacion urbana de una red de ciclorrutas. Caso estudio: Manizales*. PhD thesis, Universidad Nacional de Colombia-Sede Manizales, 2017.