The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Fairness-Aware Demand Prediction for New Mobility

An Yan, Bill Howe
Information School, University of Washington
Seattle, WA, 98105
{yanan15, billhowe}@uw.edu

Abstract

Emerging transportation modes, including car-sharing, bike-sharing, and ride-hailing, are transforming urban mobility yet have been shown to reinforce socioeconomic inequity. These services rely on accurate demand prediction, but the demand data on which these models are trained reflect biases around demographics, socioeconomic conditions, and entrenched geographic patterns. To address these biases and improve fairness, we present FairST, a fairness-aware demand prediction model for spatiotemporal urban applications, with emphasis on new mobility. We use 1D (time-varying, space-constant), 2D (space-varying, time-constant) and 3D (both time- and space-varying) convolutional branches to integrate heterogeneous features, while including fairness metrics as a form of regularization to improve equity across demographic groups. We propose two spatiotemporal fairness metrics, region-based fairness gap (RFG), applicable when demographic information is provided as a constant for a region, and individual-based fairness gap (IFG), applicable when a continuous distribution of demographic information is available. Experimental results on bike share and ride share datasets show that FairST can reduce inequity in demand prediction for multiple sensitive attributes (i.e. race, age, and education level), while achieving better accuracy than even state-of-the-art fairness-oblivious methods.

Introduction

New mobility refers to emerging transportation modes including car-sharing, bike-sharing, and ride-hailing (Goldman and Gorham 2006). These new mobility services provide technology-based, on-demand, and affordable alternatives to traditional transportation services. Supply and demand in new mobility systems are difficult to model due to complex dependencies on traffic patterns, weather, human behavior, socioeconomic conditions, and more. These services therefore rely crucially on accurate and high-resolution demand models, trained on a variety of relevant datasets, to guide resource optimization and maximize system utility (Bell and Smyl 2018; Mooney et al. 2019).

But accuracy can be misleading; these models may overfit to strong biases in the source data related to socioeconomic conditions and demographics. For example, low ridership in poor, black neighborhoods is not necessarily (or even typically) an indication of low demand (Brown 2018; Ge et al. 2016). City governments have a mandate to deliver a transportation system that benefits all citizens, particularly for historically underrepresented groups, such that an individual's access to resources as allocated by an algorithm should not be dependent on sensitive attributes such as race and age. Any demand prediction model should therefore temper accuracy (with respect to prior data) with fairness (given known biases in that data). Any fairness-agnostic demand prediction model is at risk of reinforcing inequitable access to transportation services.

In this paper, we propose a model for fairness-aware demand prediction for new mobility systems, which extends our previous work introducing the concept (Yan and Howe 2019). We consider a multi-stream network design that integrates data from multiple sources and maximizes accuracy. We then introduce novel fairness metrics for spatial-temporal settings, and design a corresponding regularizer for each.

Modeling mobility resource demand is challenging because of the complex spatial and temporal patterns it exhibits, as well as the many external factors that influence it (Li, Zheng, and Yang 2018). We address this challenge by using a three-stream architecture: a submodel that built upon 3D convolutional neural network (CNN) to capture spatial-temporal correlations within the mobility demand; a 2D CNN submodel to learn information from spatial features such as road network; and a 1D CNN submodel to include information from time series such as city-wide rainfall. The three submodels are then fused together to produce the final prediction.

To incorporate fairness into the prediction model, we need to define "fairness" in the context of mobility demand prediction. Although there has been intensive research in developing fairness metrics for credit scoring, online advertising, employment, etc. (Zliobaite 2015; Hardt, Price, and Srebro 2016), most of the existing metrics are inapplicable in spatial-temporal settings (Yan and Howe 2019). We consider fairness as the requirement that individuals of different demographic groups receives equal amount of mo-

bility resource. This notion is inspired by *group fairness* (Dwork et al. 2012) that requires the advantaged group and the disadvantaged group receive similar predicted outcomes. It also aligns with *vertical equity*, a concept in transportation literature that requires the policies to favor the disadvantaged groups (Delbosc and Currie 2011). Depending on how we assign group labels (i.e., advantaged or disadvantaged) to a geographic region, we propose two metrics: *region-based fairness gap (RFG)* and *individual-based fairness gap (IFG)*. Both measure the gap between mean per capita demand across groups over a period of time with respect to a sensitive attribute (e.g., race). However, RFG assumes that a categorical group label (e.g., white) is assigned to the entire region; while IFG allows numeric values (e.g., percentage white) based on demographics. In this sense, IFG is a finer-grained metric than RFG.

Based on RFG and IFG, we propose two corresponding regularizers to enforce fairness in the prediction model. Fairness regularizers serve as additional terms in the loss function, encouraging the model to achieve accurate and equitable prediction between groups defined by one or more sensitive attributes at the same time.

We name our approach FairST, a **Fair**ness-aware **S**patial-**T**emporal model. FairST can be naturally extended to other scenarios that involve spatial-temporal modeling and have fairness concerns such as crime incidence prediction. We summarize our main contributions as follows:

- We propose a new mobility resource demand prediction algorithm based on 3D CNN to model the temporal and spatial dependencies. The proposed algorithm adopts a three-stream architecture to integrate exogenous features with various dimensions.
- We propose *region-based fairness gap (RFG)* and *individual-based fairness gap (IFG)* to measure the gap between mean per capita demand across groups over a certain period of time. RFG focuses on discrete sensitive attributes while IFG deals with continuous attributes.
- We adapt these metrics for use as fairness regularizers for deep neural networks in spatial-temporal settings, allowing neural models to learn fair predictions for both single and multiple sensitive attributes.
- We evaluate our method using two real mobility datasets. Our experiments demonstrate that our method effectively closes the fairness gaps while achieving better accuracy than state-of-the-art fairness-oblivious models.

Related Work

Equity in New Mobility Systems. A number of studies indicate that in North America, advantaged groups have more access to docked bikeshare than disadvantaged groups (Hosford and Winters 2018; Ursaki and Aultman-Hall 2016). In examining access equity of dockless bikes in Seattle, Mooney et al. (Mooney et al. 2019) found that more college-educated and higher-income residents have access to more bikes. Overall, current literature suggests that disparities exist in the access of bikeshare systems. The equity of ride-hailing services is less clear. Although some studies found that service quality is not necessarily associated with the income or minority fraction of pickup locations (Wang and Mu 2018), the findings from some other studies suggest the otherwise (Stark and Diakopoulos 2016; Ge et al. 2016; Brown 2018). Existing works focus mostly on assessing equity based on the outcomes of deployed systems, we argue that approaches for preventing unequal resource distribution or dynamically correcting unfairness are lacking.

Spatial-temporal Prediction. Accurate demand prediction is an essential step towards effective resource allocation (e.g., bike rebalancing and ride dispatch) strategies. Early work adopted time series methods such as ARIMA or classical machine learning algorithms to predict mobility resource demand (Vogel, Greiser, and Mattfeld 2011; Li et al. 2015). Recently, deep neural networks have become popular for modeling spatial-temporal data (Wang et al. 2017). Recurrent Neural Networks (RNN) can capture temporal dependencies (Xu, Ji, and Liu 2018) and Convolutional Neural Networks (CNN) can capture spatial structures (Yao et al. 2018). Therefore, researchers use variants of RNNs and CNNs to model spatial-temporal problems (Zhang et al. 2018). ConvLSTM adopts a LSTM network structure, but replaced fully connected nodes with convolutional structures in state transitions, achieving the advantages of CNNs and RNNs (Xingjian et al. 2015). StepDeep is a network based on 3D CNN to predict spatial-temporal urban events. It achieved better accuracy than other methods including DeepSD (Shen et al. 2018; Wang et al. 2017).

No existing work in modeling urban resource demand considers fairness in their solutions. FairST builds on the state of the art 3D CNN and uses fair regularizers to guide the model to make equitable spatial-temporal prediction.

Fairness in Machine Learning. Studies on fairness in machine learning focus on identifying and removing bias in the outcome with respect to some sensitive group (e.g., race) (Hutchinson and Mitchell 2019). Many competing fairness metrics have been proposed for classification. Individual fairness states that similar individuals should be treated similarly. Group fairness is better aligned with most legal and practical definitions, arguing that members of a disadvantaged group should receive similar treatment to an advantaged group, by experiencing similar predicted outcomes (Dwork et al. 2012). The majority of fairness research focuses on classification settings rather than regression settings (Komiyama et al. 2018). Metrics for classification involve discrete probabilities and are difficult to adapt directly to regression settings. Calders et al. proposed using equal means as a fairness metric in linear regression. Fairness was incorporated through constraints in loss functions (Calders et al. 2013). Berk et al. developed a series of fairness regularizers for linear and logistic regression. They used group fairness and individual fairness analogs in regression settings (Berk et al. 2017). Our proposed method was inspired by Berk et al.'s work, but the metrics and the formulation of the loss function are novel, as is the spatial-temporal setting. Moreover, current studies focus on enforcing fairness with respect to one single attribute at a time. We demonstrated that fairness with respect to multiple sensitive attributes can be achieved together in one model.

Use Cases

In this section we describe the datasets, data preprocessing, and problem formulation for our two mobility use cases.

Datasets

Mobility data. We obtained **Seattle dockless bikeshare data** from the Transportation Data Collaborative operated by the University of Washington. The data spans from October 1, 2017 to October 31, 2018, including over 1.6 million trips. We use the number of pickup as a proxy for demand as there is no ground truth for "true demand". **RideAustin**[1] is a ride-hailing service operating in Austin, Texas. Rides data is openly available[2]. The data used in this paper spans from August 1, 2016 to April 13, 2017, including over 1.4 million trips. We use the number of rides as a proxy for demand.

Socioeconomic data. Socioeconomic data including population, race, age (under or over 65), and education level for Seattle and Austin at the block group level were obtained from the SimplyAnalytics database (SimplyAnalytics 2018).

Weather features. Previous studies show that weather conditions are associated with bike demand and ride requests, and can be helpful for prediction (Li et al. 2015; Shen et al. 2018; Wang et al. 2017). We obtained hourly weather data for Seattle and Austin from the National Centers for Environmental Information (NCEI) [3]. We included city-level air temperature, sea level pressure, and precipitation as features for prediction. They are all 1D time series as they do not have spatial variations.

Urban features. Urban forms are associated with the access and usage of new mobility systems (Wang and Mu 2018). We collected 2D features such as bike lanes and steep slopes for Seattle as they may be associated with bikeshare demand (McNeil et al. 2017; Li et al. 2018). Likewise, we collected features such as road network and Point of Interest that were suggested by the literature for RideAustin demand prediction (Wang et al. 2017; Shen et al. 2018). These urban datasets are all openly available [4].

Data Preparation

We partition the study area into equal-sized squares. The size of the squares is 1km by 1km for Seattle bikeshare and 2km by 2km for RideAustin. The purpose of square grid partitioning is to prepare the data as a tensor that CNN based models can take. Figure 1 illustrates the method that we used to process the Seattle bikeshare dataset. The RideAustin dataset was processed in the same way. We transformed 2D urban data to grid representation using the count or the total length of features within each grid. We calculated socioeconomic attributes for each grid using proportional allocation based on area.

Problem Definition

Given the historical observations of resource demand of a city, we aim to make equitable demand prediction for the next time step. For Seattle bikeshare and RideAustin, we aim to predict hourly demand based on the demand of the last 7 days (168 hours). The prediction problem is similar to predicting next frame based on the previous 168 frames in a video. We generated slices of 169 hours for training and prediction (168 hours for training and to predict the next 1 hour). For Seattle bikeshare, we use the data from October 2017 to August 2018 for training and the data from September to October, 2018 for testing. The training data contains 8040 temporal slices and the test data contains 1464 temporal slices. For RideAustin, we use the data from August 2016 to February 2017 for training and the data from March to April 2017 for testing. The training data contains 5088 temporal slices and the test data contains 1056 slices.

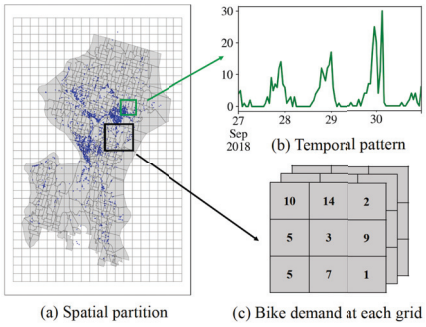


Figure 1: Data preprocessing. (a) We partition a city into square grids. (b) Each grid has a time series of mobility resource demand. (c) Each hour is akin to a frame in a video, with each grid cell as a pixel whose value is the demand.

Model and Fairness Metrics

In this section, we detail our spatial-temporal model architecture and describe our proposed fairness metrics and corresponding fairness regularizers.

Model Architecture

3D convolutions The core building block of FairST is a 3D convolutional network to model spatial-temporal prediction. The input is a 3D tensor, with spatial information modeled as a 2D heat map (typically of demand) and temporal variation is modeled in the third dimension. 3D convolutions then can capture both spatial and temporal dynamics, emphasizing locality (Ji et al. 2013).

A three-stream network architecture. We propose a generic framework for predicting mobility resource demand. It relies on 3D CNN to capture the spatio-temporal context, and submodel fusion to include exogenous features of various dimensions(Yan and Howe 2019). We use a submodel that consists of 3D convolution layers to learn from 3D historical demand (and potentially other space- and time-varying features, if available), a submodel with 1D convolution layers to learn from features that vary only with time over typical spatial scales (e.g., region-scale weather), and a submodel with 2D convolution layers to learn from features that vary only with space over short time scales (e.g.,

[1] http://www.rideaustin.com/

[2] https://data.world/ride-austin/ride-austin-june-6-april-13

[3] https://www.ncei.noaa.gov/access/search/index

[4] https://data.seattle.gov/ and https://data.austintexas.gov/

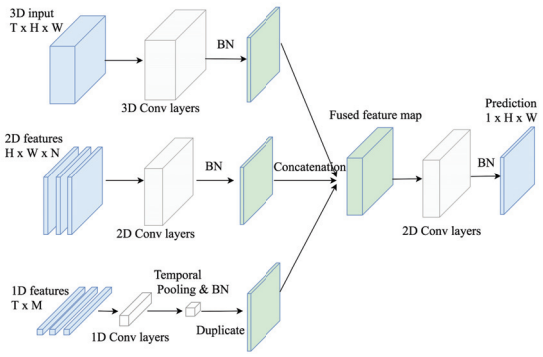


Figure 2: A generic network architecture for predicting new mobility resource demand (Yan and Howe 2019). T, H, W are the number of time steps, height of input, and width of input, respectively. N and M are the number of 2D and 1D features, respectively. BN:Batch Normalization.

topography, road networks). The outputs of all submodels were fused together, on top of which additional 2D convolutional layers were applied to achieve the final prediction (See Figure 2). Compared to fusing all features before being fed to a single network, this strategy has two main advantages: 1) Integrating semantically related features into one submodel can potentially reinforce the effectiveness of one another (Zheng et al. 2014). In our case, 1D features represent mutually correlated meteorological information, and 2D features reflect the time-invariant characteristics of the city. 2) Fusing all features early, at the dataset level, requires all features to have the same shape, meaning that 1D and 2D features must be replicated in the "missing" dimensions to make them 3D. This redundancy brings unnecessary computation overhead and wasted model capacity.

The first submodel uses 3D convolutions and takes the resource demand history as input. The submodel consists of three 3D convolutional layers, followed by a 2D convolutional layer, as shown in Figure 2. The number of filters of 3D convolutional layers are 16, 32, and 1, respectively. We use $3 \times 3 \times 3$ filters because it is the size that has shown to be effective in previous studies (Tran et al. 2015). We use padding to ensure the layer outputs are of the same size as inputs. The third 3D convolutional layer uses 1 filter to achieve temporal pooling. Finally, a 2D convolution layer is used to integrate information from previous layers and output the feature map for submodel fusion. We keep the model light-weight and skip spatial pooling to avoid deconvolution operations afterwards, which can be prone to overfitting in small training sets (Fu et al. 2017).

The second model is based on two 2D convolution layers. The number of filters of 2D convolutional layers are 16 and 16, respectively. The third model is based on three 1D convolutional layers. The number of filters of 2D convolutional layers are 16, 16, and 1, respectively. The size of the output of the third 1D convolutional layer is 1×1. It is duplicated to the size of the city to be fused with the outputs of the other two streams.

Training objectives. The loss function is a weighted sum of an accuracy loss and a fairness loss (fairness regularizer). The loss function is defined as

$$L = L_{accuracy} + \lambda L_{fairness} \tag{1}$$

where $L_{accuracy}$ is the Mean Absolute Error (MAE), $L_{fairness}$ is the fairness loss, and λ is the weight for the fairness loss. We detail the fairness regularizers in the following section.

Fairness Metrics and Regularizers

Informed by group fairness in machine learning literature (Dwork et al. 2012) and *vertical equity* in the transportation literature (Delbosc and Currie 2011), we cast fair prediction in the mobility setting as adjusting demand to reduce variation in per capita resource demand across groups. This definition assumes that variance in demand across groups (e.g., race) is due to differences in access, advertising, technology, or other factors that reflect societal inequity.

Given this approach to fairness, we propose two fairness metrics: a Region-based Fairness Gap (RFG) and an Individual-based Fairness Gap (IFG) (Yan and Howe 2019). Both measure the gap between mean per capita demand across two groups over a certain period of time. RFG is used when each geographic region is assigned a categorical group label (e.g., Caucasian). IFG is used when demographic distribution information is available, such that the sensitive attribute is numeric (e.g., percentage of Caucasian). In this paper we assume a rectilinear grid partitioning, but these two metrics can be used for any customized partitioning (e.g., zip codes or census tracts.)

Intuition. RFG draws upon the idea that people live in the same region share similar public facilities and economic status, so they may have similar commute patterns and demand for transport resources. For example, a white person may live in a predominately black community, but she frequents the same bus stops and grocery stores as her neighbors. Therefore, when assessing mobility resource demand equity, policies to distribute resources may primarily consider the majority group. In fact, it is a common practice in Transportation Equity Analysis to treat a region homogeneously (Wang et al. 2018). However, we caution that discretization of the sensitive attributes by a threshold itself is biased, as the minority population in a region may be underrepresented. In practice, we can assign each region the group label (e.g., race) with the highest population, or some criteria defined by local governments.

Notation. We now introduce notation.

- Let s_i be the ith square region of the study area $\mathcal{S}$.
- Let p_i denote the population of square region s_i divided by the total population of the city.
- Let $\hat{y}_{i,t}$ and $y_{i,t}$ denote the predicted demand and ground truth demand for region s_i at time t, respectively.
- Let $E_T[\hat{y}_{i,t}]$ denote the average predicted value for the ith square region in $\mathcal{S}$ over time period T.

Region-based Fairness Gap (RFG). We now formally define RFG. We assign one group label, either G^+ (the advantaged group) or G^- (the disadvantaged group) with respect to a single sensitive attribute A (e.g., race) to a region s_i. The RFG between two demographic groups defined by A over time T is defined as:

$$RFG = \frac{\sum_{i \in G^+} E_T[\hat{y}_{i,t}]}{\sum_{i \in G^+} p_i} - \frac{\sum_{j \in G^-} E_T[\hat{y}_{j,t}]}{\sum_{j \in G^-} p_j} \quad (2)$$

The first term can be interpreted as the per capita demand for group G^+ averaged over T. The denominator is the total population (normalized) of G^+. Likewise, the second term is the mean per capita demand in group G^- over T.

Individual-based Fairness Gap (IFG). For region s_i, let w_i^+ (e.g., 30%) and w_i^- (e.g., 70%) be its percentage of people in the advantaged group and the disadvantaged group regarding A, respectively. IFG assumes that given the predicted demand, the number of resources a group will get is proportional to the population percentage of that group. For example, if the predicted demand for bikeshare is 100 bikes for a region and the percentage of white people is 30%, then the demand that allocated to the Caucasian group is 30 bikes. IFG between two demographic groups defined by A over time T is defined as:

$$IFG = \frac{\sum_{i \in \mathcal{S}} E_T[\hat{y}_{i,t}] w_i^+}{\sum_{i \in \mathcal{S}} p_i w_i^+} - \frac{\sum_{j \in \mathcal{S}} E_T[\hat{y}_{j,t}] w_j^-}{\sum_{j \in \mathcal{S}} p_j w_j^-} \quad (3)$$

The first term is the predicted per capita demand allocated to the advantaged group averaged over T. The second term can be interpreted similarly.

Fairness regularizers. We define two loss terms, Region-based Fairness loss and Individual-based Fairness loss, that correspond to RFG and IFG, respectively.

The *Region-based Fairness loss (RF loss)* at time t is defined as

$$L_{RF}(t) = \frac{1}{\sum_{i \in \mathcal{S}} y_{i,t}} \left| \frac{\sum_{i \in G^+} \hat{y}_{i,t}}{\sum_{i \in G^+} p_i} - \frac{\sum_{j \in G^-} \hat{y}_{j,t}}{\sum_{j \in G^-} p_j} \right| \quad (4)$$

The first term is the estimated per capita demand in group G^+ at time t. Likewise, the second term is for group G^-. $\sum_{i \in \mathcal{S}} y_{i,t}$ is a normalizing factor.

The Individual-based Fairness loss (IF loss) at time t is defined as

$$L_{IF}(t) = \frac{1}{\sum_{i \in \mathcal{S}} y_{i,t}} \left| \frac{\sum_{i \in \mathcal{S}} \hat{y}_{i,t} w_i^+}{\sum_{i \in \mathcal{S}} p_i w_i^+} - \frac{\sum_{j \in \mathcal{S}} \hat{y}_{j,t} w_j^-}{\sum_{j \in \mathcal{S}} p_j w_j^-} \right| \quad (5)$$

The first term is the estimated per capita demand for advantaged group at time t. Likewise, the second term is for disadvantaged group.

Multiple sensitive attributes can be represented together in one loss function as the weighed sum of fairness loss of each attribute. The composite loss is defined as

$$L_{fairness}(t) = \sum_{a=1}^{A} \lambda_a L_{fairness_{(a,t)}} \quad (6)$$

where λ_a is the weight term for the ath attribute and $L_{fairness_(a,t)}$ is the fairness loss.

Experiments

Using Seattle dockless bikeshare data and RideAustin data, we first compare FairST without fairness loss ($\lambda = 0$) with state-of-the-art spatial-temporal models in terms of prediction accuracy. We then incorporate RF loss and IF loss into FairST, comparing against other existing fairness regularizers on a single attribute (i.e. race). Finally, we integrate the fairness losses for race, age, and education into FairST to evaluate its capability of reducing unfairness for multiple attributes in one shot.

Implementation

To implement FairST, we train 200 epochs for Seattle bikeshare and 350 epochs for RideAustin using Adam optimizer with a batch size of 32. The learning rate starts at 0.005 and decays every 5,000 steps with a rate of 0.96. To implement Region-based Fairness loss, we assign each region a label for each attribute. The label is determined based on the mean statistics of the city. For example, if the percentage of college-educated people in Seattle if 53.48%, then regions with more than 53.48% college-educated people will be labeled as college-educated group. The same method is used for discretizing race and education level.

Baseline Models

To evaluate the prediction accuracy of our method, we compare FairST with several other baselines: 1) **Historical Average (HA)**. We compute $\hat{y}_{i,t}$ using the mean values of all previous observations at location s_i at the same time of the day and the same day of the week. 2) **Autoregressive Integrated Moving Average Model (ARIMA)**. ARIMA is a commonly used time series model. We develop an independent ARIMA model for each individual grid cell. 3) **Long short-term memory Network (LSTM)** (Gers, Schmidhuber, and Cummins 2000). LSTM is a variant of RNN that can learn long-term temporal dependencies. We train the LSTM model individually for each grid. 4) **ConvLSTM** (Xingjian et al. 2015). The ConvLSTM can capture both spatial and temporal dependencies in one network. We also compare FairST with various 3D CNN models: a **3D CNN** model that is equivalent to FairST without external features; a **3D CNN + 1D** model that consists of a 3D CNN submodel and a 1D CNN submodel; and a **3D CNN + 2D** model that consists of a 3D CNN submodel and a 2D CNN submodel.

Baseline Fairness Regularizers

We compare RF loss and IF loss with two other existing fairness losses.

Equal Means Loss (EM Loss). Equal Means loss enforces the mean prediction to be the same for different groups (Calders et al. 2013). It is defined as:

Table 1: FairST compared to baselines for predicting Seattle bikeshare demand (multiple attributes). ** means correlation is significant at the 0.05 level. * means correlation is significant at the 0.01 level. SR:Spearman's rho.

	λ	MAE	RFG (race)	RFG (age)	RFG (edu)	IFG (race)	IFG (age)	IFG (edu)	SR (race)	SR (age)	SR (edu)
Ground Truth	/	/	112.568	160.089	37.471	38.969	51.338	30.053	0.016	0.174**	0.338**
HA	/	0.484	194.454	49.494	193.477	79.906	17.641	54.692	0.565**	0.477**	0.500**
ARIMA	/	0.538	319.032	62.793	319.648	129.447	28.170	90.505	0.569**	0.463**	0.489**
LSTM	/	0.468	280.685	61.437	277.938	116.023	23.778	79.162	0.522**	0.441**	0.425**
ConvLSTM	0.000	0.432	74.485	139.666	19.934	22.907	44.459	19.101	0.210**	0.355**	0.324**
3D CNN	0.000	0.408	100.878	169.240	38.873	31.915	53.133	26.851	0.091	0.256**	0.394**
3D CNN + 1D	0.000	0.387	88.587	153.625	19.802	26.791	49.058	20.691	0.291**	0.376**	0.077
3D CNN + 2D	0.000	0.378	93.299	157.025	33.946	28.661	49.792	24.457	0.158**	0.246**	0.370**
FairST	0.000	0.382	83.127	147.437	23.400	25.073	47.403	20.885	0.168**	0.191**	0.328**
FairST + RF	0.005	**0.377**	80.565	146.665	20.855	24.168	46.732	20.184	0.111*	0.262**	0.348**
FairST + RF	0.150	0.437	16.140	35.562	-5.712	4.199	22.543	7.112	-0.019	0.107*	0.321**
FairST + RF	0.250	0.460	**8.650**	**14.242**	**-3.364**	2.226	19.178	6.299	**0.011**	**0.090**	0.231**
FairST + IF	0.100	0.385	67.695	128.010	4.905	17.927	40.811	14.874	0.099	0.231**	**0.347****
FairST + IF	0.150	0.394	49.075	110.725	-9.322	11.738	35.410	9.529	0.030	0.181**	0.385**
FairST + IF	0.500	0.439	30.668	53.896	-20.291	3.823	16.536	2.200	0.117*	0.222**	0.085
FairST + IF	0.600	0.460	24.753	34.011	-22.700	**0.898**	**8.855**	**-0.185**	0.060	0.158**	**-0.055**

$$L_{EM}(t) = \frac{1}{\sum_{i \in \mathcal{S}} y_{i,t}} \left| \frac{\sum_{i \in G^+} \hat{z}_{i,t}}{n^+} - \frac{\sum_{j \in G^-} \hat{z}_{j,t}}{n^-} \right| \quad (7)$$

where p_i is the population of region s_i divided by the city total population. $\hat{z}_{i,t} = \frac{\hat{y}_{i,t}}{p_i}$, denoting the predicted per capita demand. n^+ and n^- denote the number of advantaged and the disadvantaged regions, respectively.

Pairwise Fairness Loss (Pairwise Loss). Berk et al. defined a fairness regularizer that corresponds to group fairness (Berk et al. 2017). Comparisons of predictions across groups are based on cross pairs $i \in G^+$ and $j \in G^-$.

$$L_{PF}(t) = \frac{1}{\sum_{i \in \mathcal{S}} y_{i,t}} \left(\frac{1}{n^+ n^-} \sum_{\substack{i \in G^+ \\ j \in G^-}} d(z_{i,t}, z_{j,t})(\hat{z}_{i,t} - \hat{z}_{j,t}) \right)^2 \quad (8)$$

$$d(z_{i,t}, z_{j,t}) = e^{-(z_{i,t} - z_{j,t})^2} \quad (9)$$

The model will increase the penalty as the difference between $\hat{z}_{i,t}$ and $\hat{z}_{j,t}$ increases, weighted by a similarity function $d(z_{i,t}, z_{j,t})$.

Evaluation Metrics

Prediction accuracy of all models is evaluated with **MAE** and **RMSE** (Root Mean Square Error). We evaluate the fairness of prediction outcomes using **RFG** and **IFG**, but we also consider the correlation between the ranked demand and the proportion of the disadvantaged group via **Spearman's rho** (Hauke and Kossowski 2011). That is, we are considering that city planners are interested in assessing whether the regions with the highest demand also happen to be the advantaged neighborhoods. A positive Spearman's rho suggests disparities in demand.

Results and Discussion

In this section, we show that proposed fairness regularizers give better performance than baseline regularizers, and that FairST is able to achieve better accuracy and less inequity than baseline models.

Prediction Accuracy

We compare prediction accuracy of our model with baselines. Table 1 and Table 3 show the accuracy of all models on the Seattle bikeshare data and the RideAustin data, respectively. It is observed that the 3D CNN based methods (i.e., 3D CNN, 3D CNN + 1D, 3D CNN + 2D, and FairST without fairness) proposed by this paper achieve higher prediction accuracy than the other methods. Furthermore, the incorporation of external features improves accuracy in both Seattle bikeshare and RideAustin cases.

Fair Prediction: Single Attribute

We tested the effectiveness of Region-based Fairness loss (RF) and Individual-based Fairness loss (IF) on a single attribute, i.e. race on two datasets. We compared the results with FairST trained with Equal Means loss (Equal Means) and Pairwise loss (Pairwise).

Table 2: FairST compared to baselines for Seattle bikeshare demand prediction (single attribute). SR = Spearman's rho. ** means correlation is significant at the 0.05 level. * means correlation is significant at the 0.01 level.

	λ	MAE	RMSE	RFG (race)	IFG (race)	SR (race)
ConvLSTM	0.00	0.43	1.94	74.49	22.91	0.21**
3D CNN	0.00	0.41	1.74	100.88	31.92	0.09
FairST	0.00	0.38	1.70	83.13	25.07	0.17**
FairST + RF	0.02	**0.38**	**1.67**	79.57	24.69	0.14**
FairST + RF	0.50	0.40	1.78	**10.63**	**3.36**	-0.08
FairST + IF	0.20	**0.38**	1.68	63.13	15.28	0.09
FairST + IF	1.50	0.41	1.79	38.47	4.90	**-0.07**

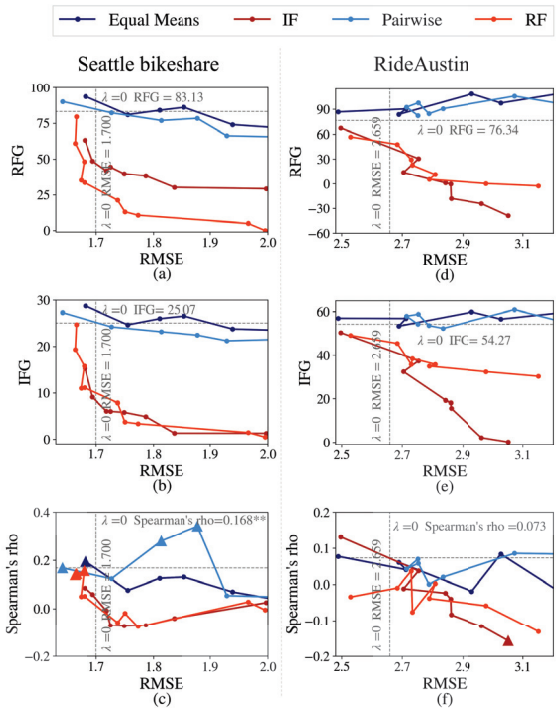


Figure 3: (a), (b), and (c) show the relationship between RMSE vs. RFG, IFG, and Spearman's rho, respectively for Seattle bikeshare. (d), (e), and (f) show the results of RideAustin. Triangles in (c) and (f) represent statistical significance (p-value < 0.01).

Figure 3 (a), (b), (d), and (e) show that RF and IF regularizers can reduce RF and IF gaps consistently and effectively. The use of fairness loss *improves* the accuracy over the baseline ($\lambda = 0$) for small values of λ, possibly due to a regularizing effect. Figure 3 (c) and (f) show the fairness of models evaluated by Spearman's rho. Overall, the use of RF loss or IF loss helps to "decorrelate" the predicted demand and race. For example, in Seattle bikeshare case, FairST ($\lambda = 0$) would lead to an unfair prediction (see Table 2). That is, there is a positive correlation (Spearman's rho = 0.168, p-value < 0.01) between the predicted demand and the percent of Caucasian. As Figure 3 (c) shows, with an IF or a RF regularizer, the Spearman's rho was brought to near zero, and no longer significant. In contrast, Spearman's coefficients of models with an Equal Means or a pairwise regularizer stay positive and sometimes show significantly positive correlation between the prediction and race. In the RideAustin case (see Figure 3 (f)), the predicted outcome of FairST ($\lambda = 0$) does not show a significant correlation with race. The Spearman's coefficients of models with an IF or a RF regularizer decrease and stay below zero, while the patterns of models with an Equal Means or a pairwise regularizer are less clear.

Table 2 and Table 3 offer insights on the effectiveness of RF and IF regularizer in bringing down fairness gaps while keeping higher accuracy, compared to baselines. For example, compared to 3D CNN, RF regularizer brings down

Table 3: FairST compared to baselines for RideAustin demand prediction (single attribute). SR:Spearman's rho. GT:Ground truth. ** means correlation is significant at the 0.05 level. * means correlation is significant at the 0.01 level.

	λ	MAE	RMSE	RFG (race)	IFG (race)	SR (race)
GT	/	/	/	80.12	59.74	0.12*
HA	/	0.66	4.39	48.46	33.55	0.12*
ARIMA	/	0.60	3.34	82.59	61.46	0.12*
LSTM	/	0.57	4.26	61.33	42.10	-0.05
ConvLSTM	0.00	0.57	4.03	66.43	46.53	0.12
3D CNN	0.00	0.53	3.14	62.00	48.71	0.05
3D CNN+1D	0.00	0.48	2.70	69.13	51.05	0.10
3D CNN+2D	0.00	0.48	2.78	71.31	50.63	0.09
FairST	0.00	0.47	2.66	76.34	54.27	0.07
FairST + RF	0.05	0.48	2.53	56.70	49.09	**-0.03**
FairST + RF	0.80	0.52	2.98	**0.35**	32.44	-0.06
FairST + IF	0.06	**0.46**	**2.50**	67.36	50.36	0.13*
FairST + IF	1.20	0.52	2.71	-27.40	**9.47**	-0.10

99.5% RFG (from 62.00 to 0.35) and IF regularizer brings down 80.5% IFG (from 48.71 to 9.47) while keeping maintaining MAE and RMSE (3).

In summary, in the single sensitive attribute scenario, FairST is able to achieve an accuracy better than the state-of-the-art spatio-temporal models while reducing more than 80% of fairness gap.

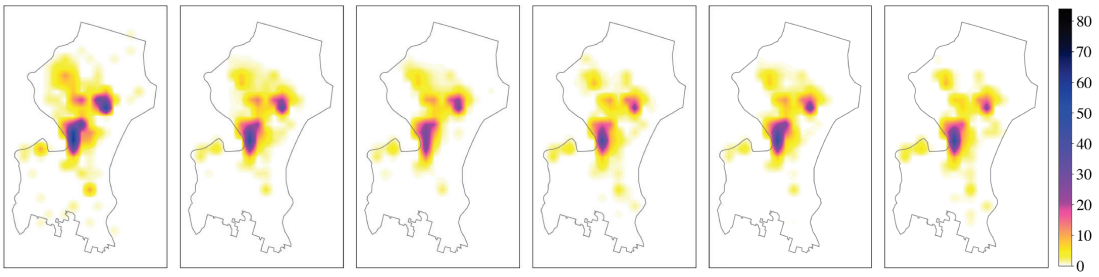
Fair Prediction: Multiple Attributes

Having demonstrated the effectiveness of closing fairness gaps with IF and RF regularizers on a single sensitive attribute, we now turn to multiple sensitive attributes. We conduct two experiments on Seattle bikeshare dataset using RF loss and IF loss, respectively according to Equation 6. We set λ_a to be 1.0 for all three attributes, i.e. race, age, and education level.

Figure 5 shows the results of FairST with RF ((a) and (c)) and IF regularizer ((b) and (d)) evaluated using RFG and IFG. Overall, as λ increases, accuracy decreases and fairness increases, indicating that both regularizers consistently help the model to approach equity on multiple sensitive attributes without sacrificing too much accuracy.

We now step back to compare FairST and baselines in terms of accuracy and fairness. Table 1 shows the results of FairST with RF regularizer and IF regularizer, denoted by FairST + RF and FairST + IF, with different λs. As can be observed, ground truth shows demand gaps between groups, indicating that there were more bikes picked up by whites, young people and college-educated people than the others. There are also significant positive correlations between demand and sensitive attributes (age and education level) as indicated by Spearman's coefficients.

Compared to ground truth, all baseline models without fairness consideration amplify inequality in terms of one or more metrics. LSTM achieves good accuracy but drastically enlarges fairness gaps of race and education. ConvLSTM shows better fairness than all baselines in terms of IFG and RFG, but gives higher Spearman's coefficients for race and age than 3D CNN model. FairST with IF or RF regularizer can help reducing inequity in terms of all metrics. For exam-



(a) Ground Truth | (b) FairST $\lambda = 0$, MAE = 0.382 | (c) ConvLSTM $\lambda = 0$, MAE = 0.432 | (d) FairST + RF $\lambda = 0.15$, MAE = 0.437 | (e) FairST + IF $\lambda = 0.1$, MAE = 0.385 | (f) FairST + IF $\lambda = 0.5$, MAE = 0.439

Figure 4: Ground truth vs. predictions heat maps for September 27, 2018 16:00 pm - 17:00 pm. (d), (e), (f) are the predictions from FairST using RF or IF regularizier on multiple sensitive attributes.

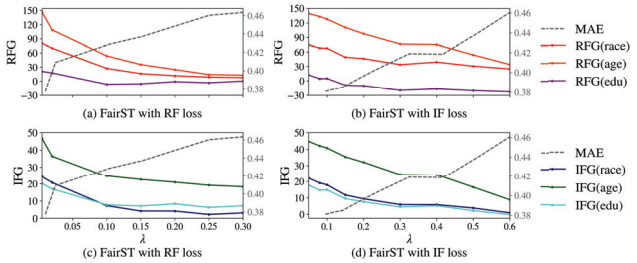


Figure 5: λ vs. fairness loss. (a) and (c) show the results of FairST with RF regularizer. (b) and (d) show the results of FairST with IF regularizer.

ple, compared to ConvLSTM, FairST + RF ($\lambda = 0.15$) and FairST + IF ($\lambda = 0.5$) show comparable accuracy but better fairness in terms of all fairness metrics. FairST + IF ($\lambda = 0.15$) outperforms 3D CNN in both accuracy and fairness.

To understand better how FairST achieves fairness, we visualize the predictions from five different settings as illustrated in Figure 4. We choose Figure 4 (d) and (f) because their MAEs are similar to that of ConvLSTM in Figure 4 (c), so that we can visually compare how these three models distribute demand. All five models are capable of learning spatial-temporal dependencies. FairST ($\lambda = 0$) accurately highlights the hot spots. Compared to ConvLSTM, FairSTs are better at capturing fragmented details around major hot spots. Adding fairness regularizers to FairST preserved the major hot spots but "re-weighted" some values in place and "redistributed" demand from some neighborhoods to others. For example, compared to Figure 4 (b) which does not consider fairness, Figure 4 (d) and Figure 4 (f) tend to capture more demand from the south part of the city where the disadvantaged population concentrates, and less demand from the northwest part of city which is dominated by the wealthy and well-educated population.

To summarize, in multiple sensitive attributes scenario, FairST is able to reduce unfairness for all three attributes consistently. With selected regularizer weight, FairST outperforms baseline models in both accuracy and fairness.

Conclusion

In this paper, we introduced FairST, a fairness-aware spatio-temporal model for predicting new mobility resource demand. Building on 3D convolutional neural network, the three-stream framework offers a generic approach to capture spatial-temporal correlations of dynamic new mobility systems and simultaneously utilize various external features. We proposed two fairness metrics that measure equity gaps between social groups for urban mobility systems. Experiments on two real-world datasets demonstrate that FairST is able to close more than 80% of fairness gap for a single sensitive attribute and at the same time achieve *better* accuracy than state-of-the-art but fairness-oblivious methods. Further experiments show that FairST is able to reduce unfairness for multiple attributes, outperforming baselines in both accuracy and fairness.

References

Bell, F., and Smyl, S. 2018. Forecasting at uber: An introduction. *Uber Engineering*.

Berk, R.; Heidari, H.; Jabbari, S.; Joseph, M.; Kearns, M. J.; Morgenstern, J.; Neel, S.; and Roth, A. 2017. A convex framework for fair regression. *CoRR* abs/1706.02409.

Brown, A. E. 2018. *Ridehail revolution: Ridehail travel and equity in Los Angeles*. Ph.D. Dissertation, UCLA.

Calders, T.; Karim, A.; Kamiran, F.; Ali, W.; and Zhang, X. 2013. Controlling attribute effect in linear regression. In *ICDM*, 71–80. IEEE.

Delbosc, A., and Currie, G. 2011. Using lorenz curves to assess public transport equity. *Journal of Transport Geography* 19(6):1252–1259.

Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; and Zemel, R. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, 214–226. New York, NY, USA: ACM.

Fu, C.-Y.; Liu, W.; Ranga, A.; Tyagi, A.; and Berg, A. C. 2017. Dssd: Deconvolutional single shot detector. *arXiv preprint arXiv:1701.06659*.

Ge, Y.; Knittel, C. R.; MacKenzie, D.; and Zoepf, S. 2016. Racial and gender discrimination in transportation network

companies. Technical report, National Bureau of Economic Research.

Gers, F.; Schmidhuber, J.; and Cummins, F. 2000. Learning to forget: continual prediction with lstm. *Neural computation* 12(10):2451.

Goldman, T., and Gorham, R. 2006. Sustainable urban transport: Four innovative directions. *Technology in society* 28(1-2):261–273.

Hardt, M.; Price, E.; and Srebro, N. 2016. Equality of opportunity in supervised learning. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, 3323–3331. USA: Curran Associates Inc.

Hauke, J., and Kossowski, T. 2011. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae* 30(2):87–93.

Hosford, K., and Winters, M. 2018. Who are public bicycle share programs serving? an evaluation of the equity of spatial access to bicycle share service areas in canadian cities. *Transportation research record* 0361198118783107.

Hutchinson, B., and Mitchell, M. 2019. 50 years of test (un)fairness: Lessons for machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, 49–58. New York, NY, USA: ACM.

Ji, S.; Xu, W.; Yang, M.; and Yu, K. 2013. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence* 35(1):221–231.

Komiyama, J.; Takeda, A.; Honda, J.; and Shimao, H. 2018. Nonconvex optimization for regression with fairness constraints. In Dy, J., and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 2737–2746. Stockholmsmässan, Stockholm Sweden: PMLR.

Li, Y.; Zheng, Y.; Zhang, H.; and Chen, L. 2015. Traffic prediction in a bike-sharing system. In *SIGSPATIAL '15*, 33:1–33:10. New York, NY, USA: ACM.

Li, X.; Zhang, Y.; Sun, L.; and Liu, Q. 2018. Free-floating bike sharing in jiangsu: Users' behaviors and influencing factors. *Energies* 11(7):1664.

Li, Y.; Zheng, Y.; and Yang, Q. 2018. Dynamic bike reposition: A spatio-temporal reinforcement learning approach. In *KDD'18*, KDD '18, 1724–1733. New York, NY, USA: ACM.

McNeil, N.; Dill, J.; MacArthur, J.; Broach, J.; Howland, S.; et al. 2017. Breaking barriers to bike share: Insights from residents of traditionally underserved neighborhoods. Technical report, National Institute for Transportation and Communities.

Mooney, S. J.; Hosford, K.; Howe, B.; Yan, A.; Winters, M.; Bassok, A.; and Hirsch, J. A. 2019. Freedom from the station: Spatial equity in access to dockless bike share. *Journal of Transport Geography* 74:91–96.

Shen, B.; Liang, X.; Ouyang, Y.; Liu, M.; Zheng, W.; and Carley, K. M. 2018. Stepdeep: A novel spatial-temporal mobility event prediction framework based on deep neural network. In *KDD*, 724–733. ACM.

SimplyAnalytics. 2018. Easi/mri census us.

Stark, J., and Diakopoulos, N. 2016. Uber seems to offer better service in areas with more white people. that raises some tough questions. *The Washington Post*.

Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; and Paluri, M. 2015. Learning spatiotemporal features with 3d convolutional networks. In *ICCV*, 4489–4497. Washington, DC, USA: IEEE Computer Society.

Ursaki, J., and Aultman-Hall, L. 2016. Quantifying the equity of bikeshare access in us cities. In *95th Annual Meeting of the Transportation Research Board, Washington, DC*.

Vogel, P.; Greiser, T.; and Mattfeld, D. C. 2011. Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences* 20:514–523.

Wang, M., and Mu, L. 2018. Spatial disparities of uber accessibility: An exploratory analysis in atlanta, usa. *Computers, Environment and Urban Systems* 67:169–175.

Wang, D.; Cao, W.; Li, J.; and Ye, J. 2017. Deepsd: supply-demand prediction for online car-hailing services using deep neural networks. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 243–254. IEEE.

Wang, Q.; Phillips, N. E.; Small, M. L.; and Sampson, R. J. 2018. Urban mobility and neighborhood isolation in america's 50 largest cities. *Proceedings of the National Academy of Sciences* 115(30):7735–7740.

Xingjian, S.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *NIPS*, 802–810.

Xu, C.; Ji, J.; and Liu, P. 2018. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets. *Transportation Research Part C: Emerging Technologies* 95:47–60.

Yan, A., and Howe, B. 2019. Fairst: Equitable spatial and temporal demand prediction for new mobility systems. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 552–555. ACM.

Yao, H.; Wu, F.; Ke, J.; Tang, X.; Jia, Y.; Lu, S.; Gong, P.; Ye, J.; and Li, Z. 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In *AAAI*.

Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X.; and Li, T. 2018. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence* 259:147 – 166.

Zheng, Y.; Capra, L.; Wolfson, O.; and Yang, H. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5(3):38.

Zliobaite, I. 2015. A survey on measuring indirect discrimination in machine learning. *arXiv preprint arXiv:1511.00148*.