Fairness-Aware Demand Prediction for New Mobility

An Yan, Bill Howe

yanan15@uw.edu, billhowe@uw.edu

Information School, University of Washington, Seattle, WA, USA

1. Overview: Unfairness in Urban Mobility

Problem:
New transportation services routinely discriminate by race, etc. [1,2]. Advanced predictive models are reinforcing this inequity by training on biased data. Existing fairness metrics are not applicable in this continuous spatio-temporal setting.

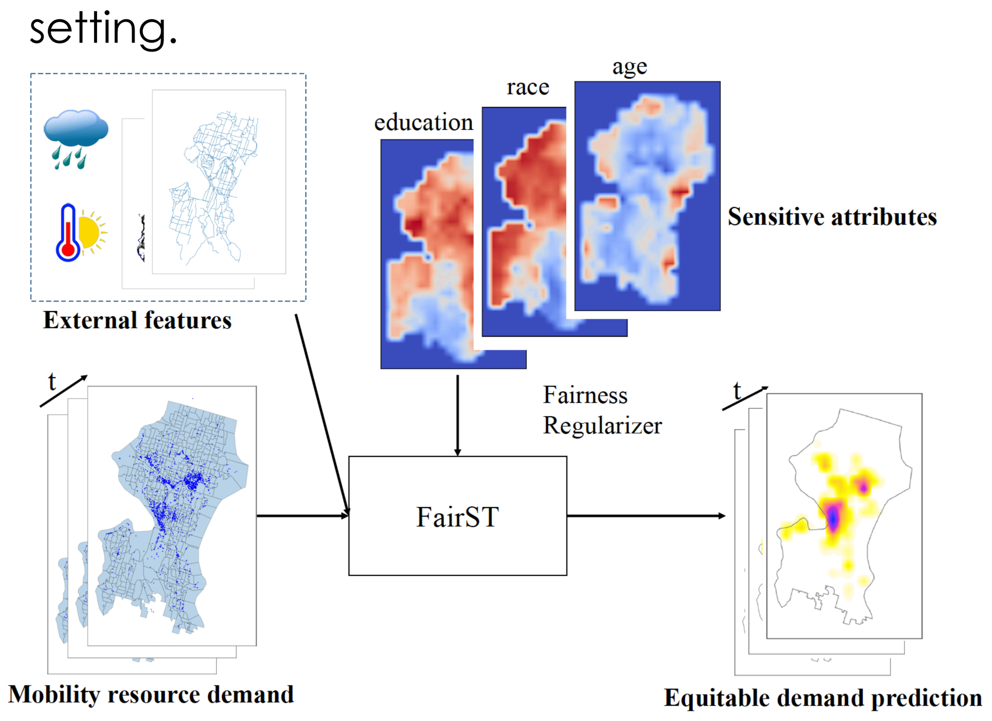


Figure 1: FairST overview

FairST is based on deep neural networks. It not only models the spatial-temporal dynamics of mobility system, but also makes equitable predictions by incorporating a fairness regularizer that encourages equal prediction between groups defined by, for example, race, age, or education level.

For more details, please refer to our paper: Fairness-Aware Demand Prediction for New Mobility, AAAI, 2020.

Our contributions: We are the first to consider fairness in machine learning in spatial-temporal settings. Specifically, we 1) propose a mobility resource demand prediction algorithm based on 3D CNN; 2) We propose two fairness metrics for urban mobility; 3) We propose fairness regularizers for deep networks in spatial-temporal settings; and 4) We evaluate our method using two real-world datasets. Our experiments show that our method effectively reduces the fairness gap by more than 80% while achieving better accuracy than state-of-the-art fairness-oblivious models.

2. Use Cases

A. Datasets

	Seattle Bikeshare	RideAustin
Time range	Oct. 1, 2017 to Oct. 31, 2018	Aug. 1, 2016 to Apr. 13, 2017
# of trips	~ 1,600,000	~ 1,400,000
1D features	temperature, sea level pressure, and precipitation	temperature, sea level pressure, and precipitation
2D features	bike lanes, slopes, etc.	road network, POI, etc.

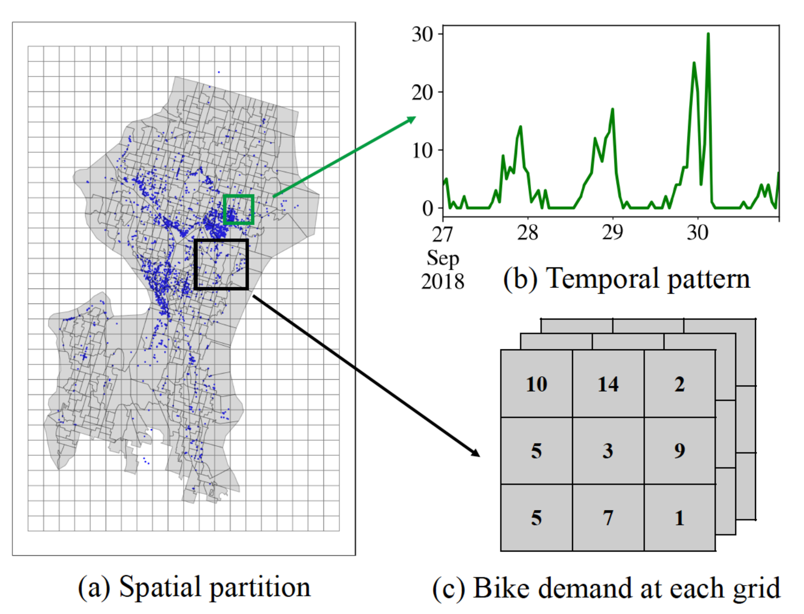


Figure 2: Data preprocessing for Seattle Bikeshare

B. Our goal

We aim to build fair models to forecast next time step demand for mobility resource for a city based on the demand of previous time steps. For both Seattle bikeshare and RideAustin, we aim to predict next hour demand based on the demand of the last 7 days (168 hours).

3. Model Architecture

We design a three-stream prediction framework based on 1D, 2D, and 3D CNN to 1) capture the spatio-temporal context, and 2) include external features to help with accuracy. We use a submodel consists of 3D convolution layers to learn from 3D historical demand, a submodel with 1D convolution layers to learn from 1D time series features, and a submodel with 2D convolution layers to extract information from 2D urban features.

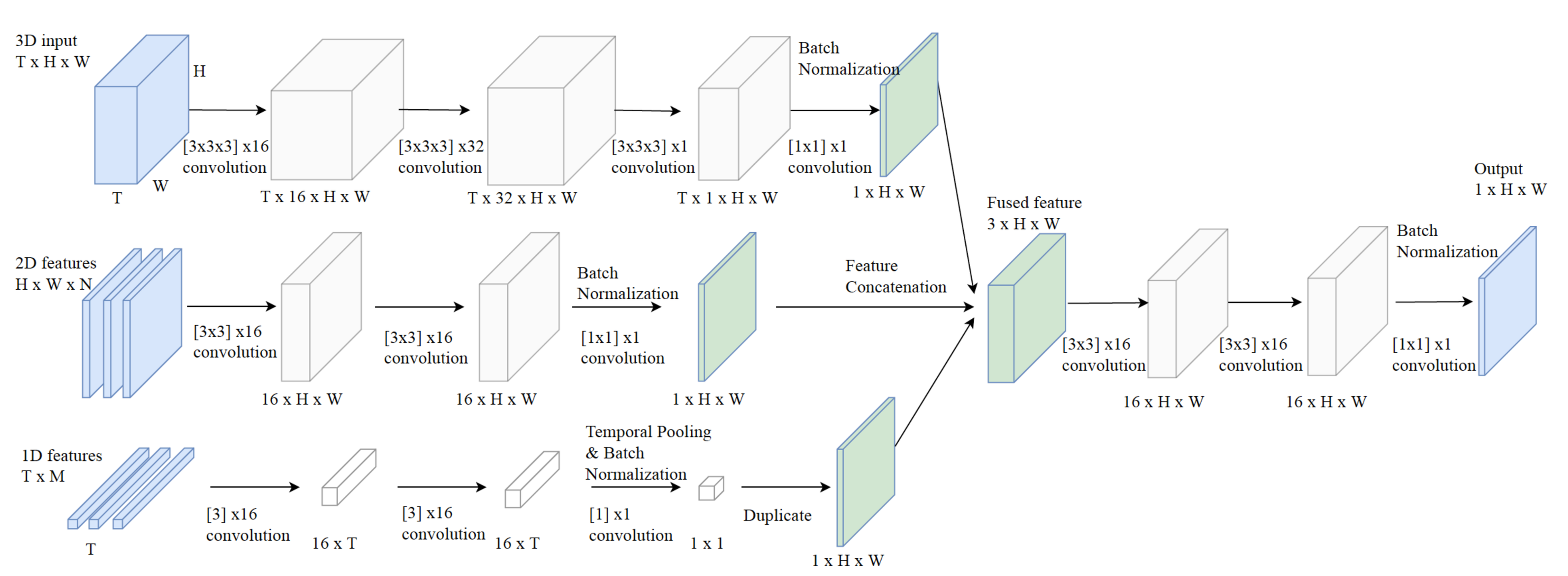


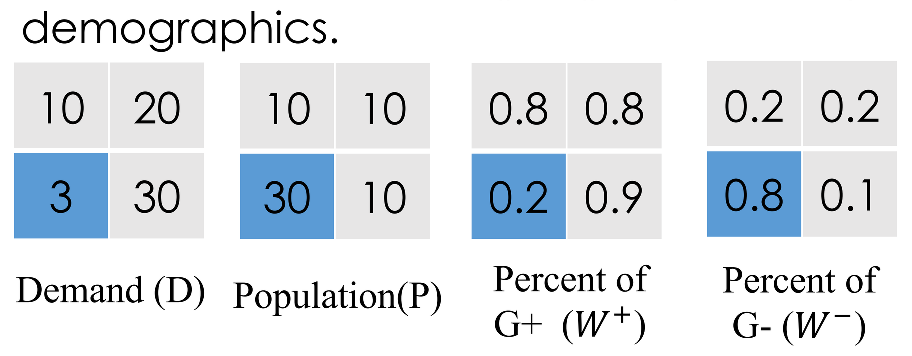
Figure 3: A three-stream network architecture

This work is funded by NSF project: "BIGDATA: F: Collaborative Research: Foundations of Responsible Data Management". Award # 1740996.

4. Fairness Metrics and Regularizers

A. Fairness Metrics

We consider fairness as individuals of different demographic groups receiving equal resources. We propose two fairness metrics: **a Region-based Fairness Gap (RFG)** and an **Individual-based Fairness Gap (IFG).** Both RFG and IFG measure the gap between mean per capita demand across two groups (G+ and G-) over a certain period of time. However, for RFG, everyone that lives in the same region is assigned the same group label, whereas IFG assigns group labels proportionally based on the region's demographics.



$$\mathbf{RFG} = \frac{\sum_{i \in G+} D_i}{\sum_{i \in G+} P_i} - \frac{\sum_{j \in G-} D_j}{\sum_{j \in G-} P_j}$$

$$\mathbf{IFG} = \frac{\sum D_i * W_i^+}{\sum P_i * W_i^+} - \frac{\sum D_i * W_i^-}{\sum P_i * W_i^-}$$

B. Fairness Regularizers

Based on the RFG and IFG, we define two fairness loss terms, **Region-based Fairness loss (RF loss)** and **Individual-based Fairness loss (IF loss)** to incorporate fairness into training. The overall loss function of the model is a weighted sum of an accuracy loss and a fairness loss. We use MAE as accuracy loss. The overall loss is defined as

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

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

5. Key Result: Improved Fairness without Loss of Accuracy

A. Takeaways

- **It works:** Using FairST fairness regularizers can reduce unfairness by at least 80% while improving accuracy over state-of-the-art approaches.
- **Why it works:** Fairness regularizers "decorrelate" the predicted demand from the protected attribute (race, gender, etc.) to avoid an obvious (but non-generalizable) unfair model.
- **How it works:** We can balance the tradeoffs between fairness and accuracy with a parameter λ. λ = 0 means ignoring fairness.

Table 1: FairST compared to baselines for Seattle bikeshare (Single attribute, i.e. race)

	λ	MAE	RFG	IFG	Spearman's rho
Ground Truth	/	/	112.568	38.969	0.016
HA	/	0.484	194.454	79.906	0.565
ARIMA	/	0.538	319.032	129.447	0.569
LSTM	/	0.468	280.685	116.023	0.522
ConvLSTM	0.00	0.432	74.485	22.907	0.210**
3D CNN	0.00	0.408	100.878	31.915	0.091
FairST	0.00	0.382	83.127	25.073	0.168**
FairST + RF	0.02	**0.379**	79.570	24.694	0.144**
FairST + RF	0.50	0.404	**10.627**	**3.363**	-0.076
FairST + IF	0.20	**0.379**	63.130	15.281	0.085
FairST + IF	1.50	0.406	38.473	4.902	**-0.070**

Table 2: FairST compared to baselines for RideAustin (Single attribute, i.e. race)

	λ	MAE	RFG	IFG	Spearman's rho
Ground Truth	/	/	80.120	59.742	0.120*
HA	/	0.662	48.457	33.550	0.118*
ARIMA	/	0.597	82.587	61.457	0.117*
LSTM	/	0.570	61.329	42.101	0.073
ConvLSTM	0.00	0.567	66.428	46.534	0.073
3D CNN	0.00	0.532	62.004	48.713	0.051
FairST	0.00	0.472	76.340	54.274	0.073
FairST + RF	0.05	0.475	56.703	49.092	**-0.034**
FairST + RF	0.80	0.524	**0.347**	32.436	-0.059
FairST + IF	0.06	**0.463**	67.358	50.357	0.131*
FairST + IF	1.20	0.515	-27.397	**9.473**	-0.100

B. Fairness vs. Accuracy Tradeoffs

Fig. 4 shows that as λ increases, accuracy decreases and fairness increases, indicating that both IF and RF regularizers **consistently** help the model to approach equity on multiple sensitive attributes without sacrificing too much accuracy.

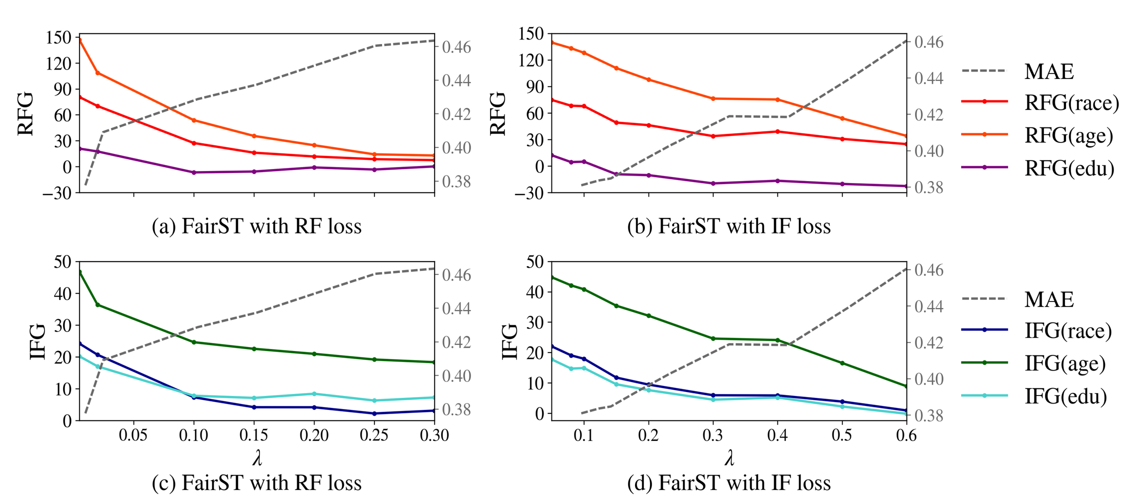


Figure 4: λ vs. fairness loss. (a) and (c) show the results of FairST with RF regularizer. (b) and (d) show the results of FairST with IF regularizer for Seattle Bikeshare on three sensitive attributes, i.e., race, age, and education level.

Fig 5. shows that 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.

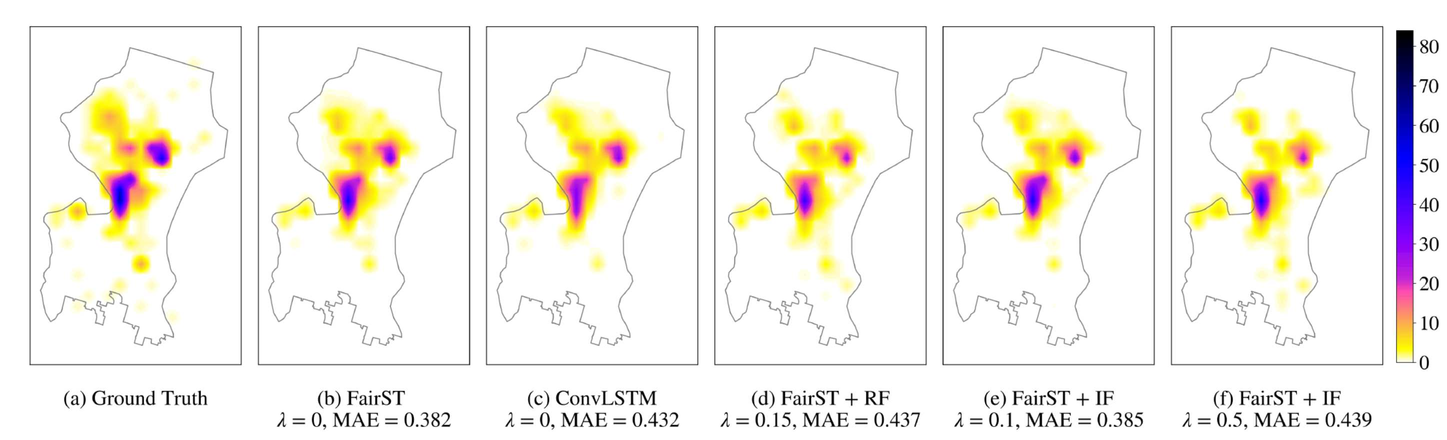


Figure 5: Ground truth vs. predictions heat maps for September 27, 2018 16:00 pm - 17:00 pm. (d), (e), (f) are the predictions from FairST using RF or IF regularizer on multiple sensitive attributes.

Key References:

[1] Anne Elizabeth Brown. 2018. Ridehail revolution: Ridehail travel and equity in Los Angeles. Ph.D. Dissertation. UCLA.

[2] Yanbo Ge, et al., . 2016. Racial and gender discrimination in transportation network companies. National Bureau of Economic Research.

[3] Xingjian, S. H. I., et al., Convolutional LSTM network: A machine learning approach for precipitation nowcasting. NIPS(2015).

[4] Du Tran, et al., Learning Spatiotemporal Features with 3D Convolutional Networks. ICCV (2015).

[5] Toon Calders, et al., Controlling attribute effect in linear regression. ICDM (2013).