Sample Title

Anonymous Author(s)

Figure 1: Seattle Mariners at Spring Training, 2010.

Abstract

CCS Concepts

• **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Keywords

Do, Not, Use, This, Code, Put, the, Correct, Terms, for, Your, Paper

ACM Reference Format:
Anonymous Author(s). 2018. Sample Title. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

2 Preliminary

In this section, we define the fairness-driven trajectory editing problem and highlight the research challenges.

2.1 Problem Definition

We consider a trajectory dataset

3 Methodology

Given the set of trajectories $\mathcal{T} \triangleq \{\tau_i\}_{i=0}^{N}$, we seek to design an editing approach so that the edited trajectories are able to i) reflect fair service coverage in the city and ii) maintain personal decision-making preferences of each taxi driver and iii) maintain high service quality in terms of minimizing the waiting time for services, that is,

$$
\begin{aligned}
\max_{\mathcal{T}'} \quad & \alpha_1 F_{causal} + \alpha_2 F_{st} + \alpha_3 F_{fidelity} + \alpha_4 F_{quality} \\
\text{s.t.:} \quad & \\
& ||\tau' - \tau||_\infty \leq \epsilon \quad \forall \tau \in \mathcal{T} \text{(subtle edits constraint)} \\
& ||\tau' - \tau||_0 \leq \eta \quad \forall \tau \in \mathcal{T} \text{(limited modifications constraint)} \\
& ||\mathcal{T}' - \mathcal{T}||_0 \leq \zeta \text{(limited modifications on the dataset constraint)} \\
& Discriminator\_confidence(\tau') \geq 0 \text{(authenticity constraint)}
\end{aligned}
$$

The trajectory modification algorithm will iterate through the set of trajectories $\mathcal{T}$ and evaluate each trajectory's contribution to the fairness of taxi service coverage via the F_{causal} and F_{st} objective terms. Fidelity to taxi driver decision-making preferences will be enforced via a ST-SiameseNet discriminator. High service quality will be maintained by constraining the extent of each modification, the number of modifications made per trajectory, and the number of modifications made to the dataset as a whole. []

3.1 F_{st} Formulation

3.1.1 Spatial-temporal Fairness – Service Rate. To assess the quantity of taxi service in each area, we adapt the service rate concept introduced by Su et al [1]. In the FAMAIL data model, the city is partitioned into equal area grid cells (replacing the irregular Traffic Analysis Zones (TAZ) in Su et al.'s work) for ease of analyzing taxi drivers' decision-making behaviors. Each taxi's GPS trajectory (with timestamps and occupancy status) is processed to identify trip origins (pickup events) and trip destinations (drop-off events) as follows: We need also to provide why it makes sense to focus on grid cells instead of business district My suggestion would be to use the same reasoning as the cGAIL paper: "Map gridding. For the ease of analyzing taxi drivers' decision-making behaviors, we partition the city into small equal side-length grid cells..."

(1) Identify Pickups and Drop-offs: We detect a pickup when a vehicle's occupancy indicator switches from 0 to 1, and

Conference acronym 'XX, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

a drop-off when it switches from 1 to 0. The latitude and longitude at these transition points are mapped to the corresponding grid cell.

(2) Count Trips per Cell: For a chosen time period p (e.g. one month or a specific time-of-day window), count the total number of pickups and drop-offs occurring in each grid cell i. Denote by O_i^p the number of trips originating (pickups) in cell i during period p, and D_i^p the number of trips with destinations (drop-offs) in cell i during period p.

(3) Normalize by Fleet Size and Duration: Direct trip counts are influenced by the number of operating taxis and the length of the sampling period. Building on Su et al., we normalize by the number of active taxis in an $n \times n$ grid of cells surrounding cell i and length (in days or hours) of the period p (both n and p are treated as hyperparameters). Let N_n^p be the total number of taxis active in period p, and T^p be the number of days or hours in period p. We define the **arrival service rate** (ASR) and the **departure service rate** (DSR) for cell i in period p as:

$$ASR_i^p = \frac{D_i^p}{N_n^p\, T^p}, \qquad DSR_i^p = \frac{O_i^p}{N_n^p\, T^p} \tag{1}$$

Each of these is a rate of trips per taxi per time period (in days or hours) in grid cell i. This formulation is adapted from Su et al. with one simplification: we omit the term A_i because all grid cells are equal in area (unlike heterogeneous TAZ sizes). Thus, ASR and DSR measure how frequently an average taxi serves cell i (arriving or departing) in period p, controlling for fleet size and time span.

3.1.2 Spatial-temporal Fairness – Gini Coefficient. Using the service rates, we asses spatial fairness via the Gini coefficient, a standard measure of inequality. For a given period p, let $\{ASR_i^p : i = 1, ..., n\}$ be the set of arrival service rates across all n grid cells, and similarly $\{DSR_i^p\}$ for departures. We compute separate Gini coefficients for the arrival distribution and departure distribution, denoted G_a^p and G_d^p respectively. The Gini coefficient is defined as:

$$G \triangleq 1 + \frac{1}{n} - \frac{2}{n^2\, \bar{x}} \sum_{i=1}^{n} (n - i + 1)\, x_{(i)}, \tag{2}$$

where $x_{(i)}$ is the i-th smallest value among $\{x_1, ..., x_n\}$, n is the number of grid cells covering the city, and $\bar{x}$ is the mean of all x_i. In our case, the x_i represents a service rate (arrival or departing) for cell i. To obtain G_a^p, we substitute $x_i = ASR_i^p$, and for G_d^p, we substitute $x_i = DSR_i^p$. Intuitively, $G = 0$ indicates perfect equality, while $G = 1$ indicates maximal inequality. A higher G_a^p means taxi drop-offs are unevenly distributed across space in period p, and similarly with G_d^p for pick-ups.

Monthly and Intra-Day Segments: We can evaluate these Gini-based fairness metrics over different temporal scopes to capture both long-term and short-term variations in spatial inequality. In particular, we consider:

- Monthly Periods: For each month in the dataset, we compute G_a and G_d using all trips that occurred during that month. This yields a sequence of monthly Gini coefficients $\{G_a^{(\text{Jan})}, G_a^{(\text{Feb})}, \ldots\}$ (and likewise for G_d), allowing us to observe trends or changes in spatial fairness from month to month.
- Intra-Day Periods: Within each day, we analyze finer time segments (e.g. peak morning hours, midday, evening, late night). For each such time window (aggregated across entire study duration), we compute G_a and G_d. This reveals how spatial inequality may differ by time of day.

3.1.3 Spatial-Temporal Fairness Term (F_{st}) in the Objective. Using the above metrics, we define a unified spatial fairness score F_{st} to be used in the trajectory editing objective function. This term aggregates the spatial inequality measurements across the chosen time periods, ensuring that fairness is considered over both space and time. First, we define **per-period spatial-temporal fairness** as the compliment of the Gini coefficients for period p. Specifically, let

$$F_{\text{spatial}}^p = 1 - \frac{1}{2}\left(G_a^p + G_d^p\right). \tag{3}$$

Here, $F_{\text{spatial}}^p \in [0, 1]$, where a higher value means more equitable service distribution in period p.

Let P be the set of time periods under consideration. To capture overall spatio-temporal fairness, we aggregate F_{spatial}^p across all periods of interest $p \in P$. We define:

$$F_{st} \triangleq \frac{1}{P} \sum_{p \in P} F_{\text{spatial}}^p = 1 - \frac{1}{2P} \sum_{p \in P} \left(G_a^p + G_d^p\right). \tag{4}$$

This F_{st} is our proposed Spatial Fairness term in the trajectory modification objective function. It produces a single scalar value that increases when spatial inequality is reduced in each period p. **Maximizing F_{st} therefore promotes more uniform taxi service distribution citywide**.

3.1.4 Supply-Demand Definitions in Spatio-Temporal Context. **Demand $D_{i,p}$**: We define demand as the number of taxi pickup events occurring in grid cell i during period p. This value is directly given by the `number_of_pickups` in the `latest_volume_pickups.json` dataset for each cell-time pair. Formally $D_{i,p}$ denotes demand at cell i and time p, then:

$$D_{i,p} = \text{number\_of\_pickups}_{i,p} \tag{5}$$

the count of trips originating in cell i during period p. This represents the legitimate demand for service in that spatio-temporal context.

Supply ($S_{i,p}$): We approximate supply as the availability of taxis in the vicinity of cell i during time p. (using the traffic_volume feature from `latest_volume_pickups.json` or by re-calculating this metric using the raw dataset) We calculate this using our RAW TAXI GPS DATASET as the number of taxi trajectory points or transits through a cell in that interval by summing over an $n \times n$ neighborhood of cells around i to capture local taxi presence. For a chosen neighborhood size n (e.g. 5×5), let $\mathcal{N}_n(i)$ be the set of cells in an $n \times n$ square centered on i (including i itself). Then the supply at i, p is computed as:

$$S_{i,p} = \sum_{j \in \mathcal{N}_n(i)} \text{traffic\_volume}_{i,p}. \tag{6}$$

In other words, $S_{i,p}$ aggregates the number of taxis present in cell i and its surrounding $(n-1)\times(n-1)$ neighbors during period p. For example, with $n=5$, this includes the 2 cells in each direction around i, forming a 5×5 block. (At city boundaries, $\mathcal{N}_n(i)$ is truncated accordingly.) This neighborhood-based supply proxy assumes taxis in adjacent cells can respond to demand in cell i, aligning the supply measure with service availability in that area.

Supply–Demand Ratio ($Y_{i,p}$): Using the above, we define the service supply–demand ratio for cell i in period p as:

$$Y_{i,p} = \frac{S_{i,p}}{D_{i,p}}, \tag{7}$$

i.e. the number of taxis available per demand unit in that spatio-temporal context. This ratio $Y_{i,p}$ serves as our *outcome of interest* for causal fairness analysis, indicating how well supply meets demand. A higher $Y_{i,p}$ (>1) means an abundance of taxis relative to pickups (potential oversupply), whereas a lower $Y_{i,p}$ (<1) indicates that demand outstrips local supply (potential undersupply).

Handling edge cases: If $D_{i,p}=0$ (no pickups in cell i during p), $Y_{i,p}$ is undefined (division by zero). In practice we might exclude such cases from ratio calculations, since no actual demand occurred. (Notably, if $D_{i,p}=0$ but $S_{i,p}>0$, it indicates taxis were present with no demand – an oversupply scenario we might address through the fairness metric below by treating the *expected* fair supply in that cell as zero.)

3.1.5 *Causal Fairness Score Formulation.*

The goal of the causal fairness term is to reward configurations in which service levels (as measured by $Y_{i,p}$) are explained by demand rather than spatio-temporal context (such as regional income levels or time of day). In a fair system, variance in $Y_{i,p}$ across the city should be largely explained by $D_{i,p}$. We formalize this by isolating the part of the supply–demand ratio $Y_{i,p}$ that is not explained by demand.

Demand-Expected Service Ratio ($g(d)$): We define $g(d)=\mathbb{E}[Y \mid D=d]$ as the expected supply–demand ratio given demand d. This function models the baseline level of service we would expect for a given demand level across the dataset. It can be derived empirically by regressing $Y_{i,p}$ on $D_{i,p}$. We assume $g(d)$ to be a monotonic or smooth function learned from the data:

$$g(D_{i,p}) = \mathbb{E}[Y_{i,p} \mid D_{i,p}] \tag{8}$$

This isolates the component of service that is explained by demand.

3.1.6 *Per-Period Fairness Score F^p_{causal}.*

For each period p, we compute a fairness score based on how much of the variance in supply–demand ratios is explained by demand, rather than location or time. Let $\mathcal{I}_p=\{i : D_{i,p}>0\}$ be the set of active cells in period p. The causal fairness score for that period is defined as:

$$F^p_{\text{causal}} = \frac{\mathrm{Var}_p(g(D_{i,p}))}{\mathrm{Var}_p(Y_{i,p})} \tag{9}$$

where $F^p_{\text{causal}} \in [0,1]$ and $\mathrm{Var}_p(\cdot)$ is the empirical variance across all cells in $\mathcal{I}_p$. This is a demand-conditioned coefficient of determination: it quantifies what fraction of the variation in $Y_{i,p}$ is explainable by $D_{i,p}$.

Interpretation: A higher F^p_{causal} means demand accounts for most of the variation in service levels, implying a causally fair system. If $F^p_{\text{causal}}=1$, demand fully determines supply; if $F^p_{\text{causal}}=0$, context dominates.

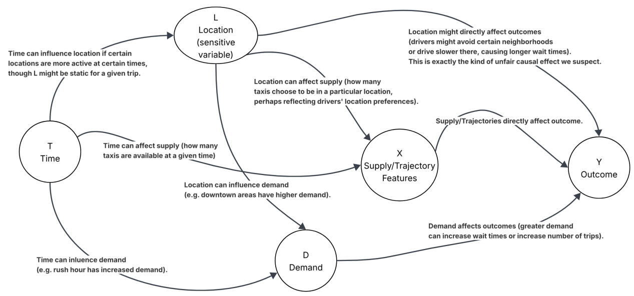

Figure 2: (preliminary!) Causal Graph: Currently in use during the development of the Causal Fairness objective term.

3.1.7 *Aggregating to a Single Causal Fairness Term(F_{causal}).*

Finally, we aggregate the period-specific scores to obtain one scalar fairness metric for the entire dataset or evaluation window. If we consider a total of P periods (e.g. $p=1,\ldots,P$ could index all hourly slots in the data), the overall causal fairness term is the average of the per-period values:

$$F_{\text{causal}} = \frac{1}{P}\sum_{p\in P} F^p_{\text{causal}}. \tag{10}$$

where P is the set of all defined time periods (e.g., hours, days, or segments). This aggregation ensures that fairness is promoted consistently over the full spatio-temporal domain. A high F_{causal} indicates that service provision is aligned with demand across both space and time.

This score is structured to be maximized, aligning it with other fairness terms such as F_{st}. It supports trajectory editing that promotes service consistency wherever demand is present.

4 Evaluation

References

[1] Rongxiang Su, Zhixiang Fang, Hong Xu, and Lian Huang. 2018. Uncovering Spatial Inequality in Taxi Services in the Context of a Subsidy War among E-Hailing Apps. *ISPRS International Journal of Geo-Information* 7, 6 (2018). doi:10.3390/ijgi7060230

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009