Trade the System Efficiency for the Income Equality of Drivers in Rideshare

Yifan Xu$^1$ and Pan Xu$^2$

$^1$Key Lab of CNII, MOE, Southeast University, Nanjing, China
$^2$Department of Computer Science, New Jersey Institute of Technology, Newark, USA

xyf@seu.edu.cn, pxu@njit.edu

Abstract

Several scientific studies have reported the existence of the income gap among rideshare drivers based on demographic factors such as gender, age, race, etc. In this paper, we study the income inequality among rideshare drivers due to discriminative cancellations from riders, and the trade-off between the income inequality (called fairness objective) and the system efficiency (called profit objective). We proposed an online bipartite-matching model where riders are assumed to arrive sequentially following a distribution known in advance. The highlight of our model is the concept of acceptance rate between any pair of driver-rider types, where types are defined based on demographic factors. Specially, we assume each rider can accept or cancel the driver assigned to her, each occurs with a certain probability which reflects the acceptance degree from the rider type towards the driver type. We construct a bi-objective linear program as a valid benchmark and propose two LP-based parameterized online algorithms. Rigorous online competitive ratio analysis is offered to demonstrate the flexibility and efficiency of our online algorithms in balancing the two conflicting goals, promotions of fairness and profit. Experimental results on a real-world dataset are provided as well, which confirm our theoretical predictions.

1 Introduction

Rideshares such as Uber and Lyft have received significant attention among research communities of computer science, operations research, and business, to name a few. One main research topic is the matching policy design of pairing drivers and riders, see, e.g., [Danassis et al., 2019; Ashlagi et al., 2019; Bei and Zhang, 2018; Dickerson et al., 2018; Zhao et al., 2019]. Most of the current work focuses on either the promotion of system efficiency or that of users’ satisfaction or both.

In this paper, we study the fairness among rideshare drivers. There are several reports showing the earning gap among drivers based on their demographic factors such as age, gender and race, see, e.g., [Cook et al., 2018; Rosenblat et al., 2016]. In particular, [Hinchliffe, 2017] has reported that “Black Uber and Lyft drivers earned $13.96 an hour compared to the $16.08 average for all other drivers” and “Women drivers reported earning an average of $14.26 per hour, compared to $16.61 for men”. The wage gap among drivers from different demographic groups is partially due to the discriminative cancellations from riders, which can be well spotted especially during off-peak hours when the number of riders is comparable or even less than that of drivers. Note that in rideshares like Uber and Lyft, after a driver accepts a rider: (1) all sensitive information of the driver such as name and photo will be accessible to the rider and (2) riders can cancel the driver for the first two minutes free of charge [Dough, 2019]. This makes the discriminative cancellations from riders technically possible and economically worry-free.

We aim to address the income disparity among drivers due to discriminative cancellations from riders and its trade-off with system efficiency. Note that the two goals, promoting the group-level income equality among drivers and the system efficiency, are somewhat conflicting. Consider the off-peak hours for example, when riders are kinds of scarce resources. To maximize the system efficiency, rideshares like Uber should please riders by assigning them to their “favorite” drivers. This can effectively reduce any possible cancellations from riders and thus, minimize the risk of driving away riders to other rivals like Lyft. This measure, however, will offer those drivers “popular” among riders much more chances of getting orders than others and as a result, hurt the group-level income equality greatly.

In this paper, we propose two parameterized matching policies, which can smoothly trade-off the above two goals with provable performances. We adopt the online-matching based model to capture the dynamics in rideshare, as commonly used before [Dickerson et al., 2018; Zhao et al., 2019]. Assume a bipartite graph $G = (U, V, E)$ where $U$ and $V$ represent the sets of types of offline drivers and online requests, respectively. Each driver type represents a specific demographic group (defined by gender, age, race, etc.) with a given location, while each request type represents a specific demographic group with a given starting and ending location.

There is an edge $f = (u, v)$ if the driver (of type) $u$ is capable of serving the request (of type) $v$ (e.g., the distance...
between them is below a given threshold). The online phase consists of $T$ rounds and in each round, a request $v \in V$ arrives dynamically. Upon its arrival an immediate and irrevocable decision is required: either reject $v$ or assign it to a neighboring driver in $U$. We assume each $u$ has a matching capacity of $B_u \in \mathbb{Z}^+$, which captures the number of drivers belonging to the type $u$. Additionally, we have the following key assumptions in the model.

**Arrivals of online requests.** We consider a finite time horizon $T$ (known to the algorithm). For each time-step or round $t \in [T] \equiv \{1, 2, \ldots, T\}$, a request of type $v$ will be sampled (or $v$ arrives) from a known distribution $\{q_v\}$ such that $\sum_{v \in V} q_v = 1$. Note that the sampling process is independent and identical across the online $T$ rounds. For each $v$, let $r_v = T \cdot q_v$, which is called the arrival rate of request $v$ with $\sum_{v \in V} r_v = T$. Our arrival assumption is commonly called the known identical independent distributions (KIID). This is mainly inspired from the fact that we can often learn the arrival distribution from historical logs [Yao et al., 2018; Li et al., 2018]. KIID is widely used in many practical applications of online matching markets including rideshare and crowdsourcing [Zhao et al., 2019; Dickerson et al., 2018; Singer and Mittal, 2013; Singla and Krause, 2013].

**Edge existence probabilities.** Each edge $f = (u, v)$ is associated with an existence probability $p_f \in (0, 1]$, which captures the statistical acceptance rate of a request of type $v$ toward a driver of type $u$. The random process goes as follows. Once we assign $u$ to $v$, we observe an immediate random outcome of the existence, which is present (i.e., $v$ accepts $u$) with probability $p_f$ and not ($v$ cancels $u$) otherwise. We assume that (1) the randomness associated with the edge existence is independent across all edges; (2) the values $\{p_f\}$ are given as part of the input. The first assumption is motivated by individual choice and the second from the fact that historical logs can be used to compute such statistics with high precision.

**Patience of requests.** Each request $v$ is associated with patience $\Delta_v \in \mathbb{Z}^+$, which captures an upper bound of unsuccessful assignments the request $v$ can tolerate before leaving the platform. Under patience constraints, we can dispatch each request $v$ to at most $\Delta_v$ different drivers. Observe that we cannot broadcast $v$ to a set of at most $\Delta_v$ different drivers simultaneously. Instead, we should assign $v$ to at most $\Delta_v$ distinct drivers (maybe of the same type though) in a sequential manner until either $v$ accepts one or $v$ leaves the system after running out of patience. We refer to this as the online probing process (OPP). Note that OPP starts immediately after a request $v$ arrives if $v$ is not rejected by the algorithm, and ends within one single round before the next request arrives.

We say an assignment $f = (u, v)$ is successful if $u$ is assigned to $v$, and $v$ accepts $u$ which occurs with probability $p_f$. Assume that the platform will gain a profit $w_f$ from a successful assignment $f = (u, v)$ (we call a match then). For a given policy ALG, let $\mathcal{M}$ be the set of (possibly random) successful assignments; we interchangeably use the term matching to denote this set $\mathcal{M}$. Inspired by the work of [Nanda et al., 2019; Lesmana et al., 2019], we define two objectives, namely profit and fairness, which capture the system efficiency and group-level income equality among drivers, respectively.

**Profit:** The expected total profit over all matches obtained by the platform, which is defined as $\mathbb{E}[\sum_{f \in \mathcal{M}} w_f]$.

**Fairness:** Let $\mathcal{M}_u$ be the set of edges in $\mathcal{M}$ incident to $u$. Define the fairness achieved by ALG over all driver types as $\min_{u \in U} \frac{\mathbb{E}||\mathcal{M}_u||}{B_u}$.

### 1.1 Preliminaries and Main Contributions

**Competitive ratio.** The competitive ratio is a commonly-used metric to evaluate the performance of online algorithms. Consider an online maximization problem for example. Let $\text{ALG}(\mathcal{I}) = \mathbb{E}_{I \sim \mathcal{I}}[\text{ALG}(I)]$ denote the expected performance of ALG on an input $\mathcal{I}$, where the expectation is taken over the random arrival sequence $I$. Let $\text{OPT}(\mathcal{I}) = \mathbb{E}[\text{OPT}(I)]$ denote the expected offline optimal, where $\text{OPT}(I)$ refers to the optimal value after we observe the full arrival sequence $I$. Then, competitive ratio is defined as $\min_{\mathcal{I}} \frac{\text{ALG}(\mathcal{I})}{\text{OPT}(\mathcal{I})}$. It is a common technique to use an LP to upper bound the $\text{OPT}(\mathcal{I})$ (called the benchmark LP) and hence get a valid lower bound on the target competitive ratio. In our paper, we conduct online competitive ratio analysis on both objectives.

**Main Contributions.** Our contributions can be summarized in the following three aspects. First, we propose a new online matching based model to address the income inequality among drivers from different demographic groups and its trade-off with the system efficiency in rideshare. Second, we present a robust theoretical analysis for our model. We first construct a bi-objective linear program (LP-(1) and LP-(2)), which is proved to offer valid upper bounds for the respective maximum profit and fairness in the offline optimal. Then, we propose LP-based parameterized online algorithms WarmUp and AttenAlg with provable performances on both objectives. We say an online algorithm achieves an $(\alpha, \beta)$-competitive ratio if it achieves competitive ratios $\alpha$ and $\beta$ on the profit and fairness against benchmarks LP-(1) and LP-(2), respectively. Results in Theorems 2 and 3 suggest that AttenAlg can achieve a nearly optimal ratio on each single objective either fairness or profit, though there is some space of improvement left for the summation of both ratios.

**Theorem 1.** $\text{WarmUp}(\alpha, \beta)$ achieves a competitive ratio at least $\begin{pmatrix} \alpha \cdot \frac{1-1/e^2}{2} \cdot \beta \cdot \frac{1-1/e^2}{2} \end{pmatrix}$ simultaneously on the profit and fairness for any $\alpha, \beta > 0$ with $\alpha + \beta \leq 1$.

**Theorem 2.** $\text{AttenAlg}(\alpha, \beta)$ achieves a competitive ratio at least $\begin{pmatrix} \alpha \cdot \frac{1}{e+1}, \beta \cdot \frac{1}{e+1} \end{pmatrix}$ $\sim$ $(0.46 \cdot \alpha, 0.46 \cdot \beta)$ simultaneously on the profit and fairness for any $\alpha, \beta > 0$ with $\alpha + \beta \leq 1$.

**Theorem 3.** No algorithm can achieve an $(\alpha, \beta)$-competitive ratio simultaneously on the profit and fairness with $\alpha + \beta > 1$ or $\alpha > 0.51$ or $\beta > 0.51$ using LP-(1) and LP-(2) as benchmarks.

Last, we test our model and algorithms on a real dataset collected from a large on-demand taxi dispatching platform.
Experimental results confirm our theoretical predictions and demonstrate the flexibility of our algorithms in tradeoffing the two conflicting objectives and their efficiency compared to natural heuristics.

2 Related Work

Here is a few recent work addressing the fairness issue in rideshares. [Sühr et al., 2019] proposed two notions of amortized fairness for fair distribution of income among drivers, one is related to absolute income equality, while the other is averaged income equality over active time. [Lesmana et al., 2013; Aggarwal et al., 2014] considered nearly the same two objectives as proposed in this paper. Note that both of the aforementioned work considered an essential offline setting in the way that all arrivals of online requests are known in advance by considering a short time window. Additionally, both ignore the potential cancellations from riders, and assume each rider will accept the assigned driver surely (i.e., all \(p_f = 1\)). [Nanda et al., 2019] studied an interesting “dual” setting to us. They focused on the peak hours and examined the fairness on the rider side due to discriminative cancellations from drivers.

Our model technically belongs to a more general optimization paradigm, called Multi-Objective Optimization. Here are a few theoretical work which studied the design of approximation or online algorithms to achieve a bi-criterion approximation and/or online competitive ratios, see, e.g., [Ravi et al., 1993; Korula et al., 2013; Fata et al., 2016]. The work of [Brubach et al., 2020; Fata et al., 2019] have the closest setting to us: each edge has an independent existence probability and each vertex from the offline and/or online side has a patience constraint on it. However, all investigated one single objective: maximization of the total profit over all matched edges.

3 Valid Benchmarks for Profit and Fairness

We first present our benchmark LPs and then an LP-based parameterized algorithm. For each edge \(f = (u, v)\), let \(x_f\) be the expected number of probes on edge \(f\) (i.e., assignments of \(v\) to \(u\) but not necessarily matches) in the offline optimal. For each \(u (v)\), let \(E_u (E_v)\) be the set of neighboring edges incident to \(u (v)\). Consider the following bi-objective LP.

\[
\begin{align*}
\max & \quad \sum_{f} w_f x_f p_f \\
\max & \quad \min_{u \in U} \sum_{f \in E_u} x_f p_f \\
\text{s.t.} & \quad \sum_{f \in E_u} x_f p_f \leq B_u \quad \forall u \in U \quad (3) \\
& \quad \sum_{f \in E_v} x_f \leq \Delta_v \cdot r_v \quad \forall v \in V \quad (4) \\
& \quad \sum_{f \in E_v} x_f p_f \leq r_v \quad \forall v \in V \quad (5) \\
& \quad 0 \leq x_f \leq r_v \quad \forall f \in E_v 
\end{align*}
\]

Let LP-(1) and LP-(2) denote the two LPs with the respective objectives (1) and (2), each with Constraints (3), (4), (5), (6). Note that we can rewrite Objective (2) as a linear one like \(\max \eta \) with additional linear constraints as \(\eta \leq \sum_{f \in E_v} x_f p_f / B_u\) for all \(u \in U\). For presentation convenience, we keep the current compact version. The validity of LP-(1) and LP-(2) as benchmarks for our two objectives can be seen in the following lemma.

**Lemma 1.** LP-(1) and LP-(2) are valid benchmarks for the two respective objectives, profit and fairness. In other words, the optimal values to LP-(1) and LP-(2) are valid upper bounds for the expected profit and fairness achieved by the offline optimal, respectively.

**Proof.** We can verify that objective functions (1) and (2) each captures the exact expected profit and fairness achieved by the offline optimal by the linearity of expectation. To prove the validity of the benchmark for each objective, it suffices to show the feasibility of all constraints for any given offline optimal. Recall that for each edge \(f\), \(x_f\) denotes the expected number of probes on \(f\) (i.e., assignments of \(u\) to \(v\) but not necessarily matches) in the offline optimal. Constraint (3) is valid since each driver \(u\) has a matching capacity of \(B_u\). Note that the expected arrivals of \(v\) during the whole online phase is \(r_v\) and \(v\) can be probed at most \(\Delta_v\) times upon each online arrival. Thus, the expected number of total probes and matches over all edges incident to \(v\) should be no more than \(r_v \Delta_v\) and \(r_v\), respectively. This rationalizes Constraints (4) and (5). The last constraint is valid, since for each edge, the expected number of probes should be no more than that of arrivals. Therefore, we justify the feasibility of all constraints for any given offline optimal.

4 LP-based Parameterized Algorithms

The following lemma suggests that for any online algorithm ALG, the worst-case scenario (i.e., the instance on which ALG achieves the lowest competitive ratio) arrives when each driver type has a unit matching capacity. The proof is deferred to the full version.

**Lemma 2.** Let ALG be an online algorithm achieving an \((\alpha, \beta)\)-competitive ratio on instances with unit matching capacity (i.e., all \(B_u = 1\)). We can twist ALG to ALG' such that ALG' achieves at least an \((\alpha, \beta)\)-competitive ratio on instances with general integral matching capacities.

From Lemma 2, we assume unit capacity for all driver types throughout this paper w.l.o.g. In the following, we will present a warm-up algorithm (WarmUp) and then another refined algorithm (AttenAlg), which can be viewed as a polished version of WarmUp with simulation-based attention techniques. The main idea of AttenAlg is primarily inspired by the work [Brubach et al., 2020]. Both WarmUp and AttenAlg invoke the following dependent rounding techniques (denoted by GKPS) introduced by [Gandhi et al., 2006]. For simplicity, we state a simplified version of GKPS tailored to star graphs which suffices in our paper.

Recall that \(E_v\) is the set of edges incident to \(v\) in the compatibale graph \(G\). GKPS is such a dependent rounding technique that takes as input a fractional vector \(z = \{z_f, f \in E_v, z_f \in [0, 1]\}\) on \(E_v\), and output a random binary vector \(Z = \{Z_f, f \in E_v\}\), which satisfied the following properties.

1. **Marginal distribution:** \(E[Z_f] = z_f\) for all \(f \in E_v\);
2. **Degree preservation:** \(\Pr[\sum_{f \in E_v} Z_f \leq \sum_{f \in E_v} z_f] = 1\);

3. https://tinyurl.com/yd9ucv8p
Algorithm 1 Sub-Routine $SR(x)$

1. Apply GKPS to the fractional vector $z$ and let $Z$ be the random binary vector output.
2. Choose a random permutation $\pi$ over $E_v$.
3. Follow the order $\pi$ to process each $f = (u, v) \in E_v$ until $v$ is matched:
4. if $Z_f = 1$ and $u$ is available then
   - Probe the edge $f$ (i.e., assign $v$ to $u$).
5. else
   - Skip to the next one.

4.1 The First Algorithm Warm-Up $(\alpha, \beta)$

Let an online vertex $v$ arrive at $t$. Our job is to probe at most $\Delta_v$ edges in $E_v$ until $v$ is matched. Let $z$ be a given fractional solution on $E_v$. Warm-Up $(\alpha, \beta)$ invokes the following procedures (denoted by $SR(z)$) as a subroutine during an online round: first, it first selects a set $S_v$ of at most $\Delta_v$ edges from $E_v$ in a random way guided by a given fractional vector $x$ on $E_v$ and then follows a random order to process all edges in $S_v$ one by one. The details of $SR$ are stated in Algorithm 1.

Based on $SR$, the main idea of Warm-Up $(\alpha, \beta)$ is as simple as follows: each round when an online vertex $v$ arrives, it invokes $SR(x^v)$ and $SR(y^v)$ with probabilities $\alpha$ and $\beta$ respectively. Recall that $x^v$ and $y^v$ are the scaled optimal solutions to LP-(1) and LP-(2) restricted to $E_v$, each has a total sum at most $\Delta_v$. Thus, when we run $SR(x^v)$ or $SR(y^v)$ after $v$ arrives online, we will probe at most $\Delta_v$ edges incident to $v$ since the final rounded binary vector has at most $\Delta_v$ ones due to Property of Degree Preservation in the dependent rounding. The details of Warm-Up $(\alpha, \beta)$ are as follows.

We conduct an edge-by-edge analysis. It would suffice to show that each $f$ is probed with probability at least $z_f^* \cdot \alpha \cdot (1 - 1/\epsilon)^2$ and $y_f^* \cdot \beta \cdot (1 - 1/\epsilon)^2$ in Warm-Up $(\alpha, \beta)$. Then, by linearity of expectation, we can get Corollary 1. Focus on a given $u$ and a time $t \in [T]$. Let $SF_{u,t}$ be the event that $u$ is available at (the beginning of) $t$.

**Lemma 3.** For any given $u$ and $t \in [T]$, we have $\Pr[SF_{u,t}] \geq \left(1 - \frac{1}{T}\right)^{t-1}$.

**Proof.** Recall that we assume w.l.o.g. that each $B_u=1$ due to Lemma 2. For each given $t < T$ and $f = (u, v) \in E_v$, let $X_{f,t}$ indicate if $v$ arrives at time $t$; $Y_{f,t}$ indicate if $f$ is probed during round $t$; $Z_{f,t}$ indicate if $f$ is present when probed. Note that in each subroutine of $SR(x^v)$ and $SR(y^v)$ after $v$ arrives, $f$ will be probed only when the final rounded vector has the entry one on $f$. Thus we claim that $\mathbb{E}[Y_{f,t}] \leq \alpha z_f^*/r_v + \beta y_f^*/r_v$, due to Property of Marginal Distribution in dependent rounding and statements of Warm-Up $(\alpha, \beta)$. Thus,

\[
\Pr[SF_{u,t}] = \prod_{t < t} \Pr\left[\sum_{f \in E_u} X_{f,t} Y_{f,t} Z_{f,t} = 0\right] = \prod_{t < t} \left(1 - \Pr\left[\sum_{f \in E_u} X_{f,t} Y_{f,t} Z_{f,t} \geq 1\right]\right) \\
= \prod_{t < t} \left(1 - \sum_{f \in E_u} \frac{r_v}{\mathbb{E}} \left(\alpha z_f^*/r_v + \beta y_f^*/r_v\right) p_f\right) = \prod_{t < t} \left(1 - \frac{1}{T}\right)^{t-1}
\]

Now assume $SF_{u,t}$ occurs (i.e., $u$ is available at $t$). Consider a given $f \in \mathbb{E}$ and let $1_{f,t}$ indicate if $f$ is probed during round $t$ in Warm-Up $(\alpha, \beta)$. Notice that $1_{f,t}$ occurs if (1) $v$ arrives at time $t$ and (2) $f$ is probed either in $SR(x^v)$ or $SR(y^v)$.

**Lemma 4.** $\Pr[1_{f,t}\mid SF_{u,t}] \geq \frac{z_f^*}{T}$, $\Pr[1_{f,t}\mid SF_{u,t}] \geq \frac{y_f^*}{T}$.

**Proof.** We focus on the first inequality and try to show that $f$ is probed at $t$ in $SR(x^v)$ with probability at least $\frac{z_f^*}{T}$ (including the probability of its online arrival). Observe that events $v$ arrives at time $t$ and Warm-Up $(\alpha, \beta)$ runs the subroutine $x^v$ both happen with probability $\frac{z_f^*}{T}$. Let $X_f^*$ be the rounded binary vector from $x^v$ and we use $Y_f^*$ to denote its entry on $f$. Let $E_{a \rightarrow b}$ be the set of edges in $E_v$ excluding $f = (u, v)$. For each $f' \in E_{a \rightarrow b}$, let $Y_{f'}$ indicate if $f'$ falls before $f$ in the random order $\pi$ and $Z_{f'}$ indicate if $f'$ is present when probed. Thus we have

\[
\Pr[1_{f,t}\mid SF_{u,t}] \geq \frac{z_f^*}{T}
\]

\[
\Pr[1_{f,t}\mid SF_{u,t}] \geq \frac{y_f^*}{T}
\]
1 \leq x^*_f / r_v \text{ due to negative correlation in dependent rounding and (2) } E[Y_f] = 1/2, E[Z_f] = p_f. \text{ Inequality (13) follows from the fact } \sum_{f \in E, x^*_f \neq r_v} x^*_f p_f \leq r_v \text{ due to Constraint (5). Following a similar analysis, we can prove the second part.} \qed

Now we have all ingredients to prove the main Theorem 1.

\textbf{Proof.} Consider a given } f = (u, v) \in E, \text{ let } \kappa_f \text{ and } \kappa_f^y \text{ be the expected number of successful probes of } f \text{ in } \text{SR}(x^*) \text{ and } \text{SR}(y^*) \text{ respectively. Here a probe of } f = (u, v) \text{ is successful iff } u \text{ is available when we assign } v \text{ to } u \text{ (but no necessarily means } f \text{ is present).}

\[
\kappa_f \geq \sum_{t=1}^{T} \Pr[\text{SR}(u,v)] \Pr[1_{f,t}] \text{SR}(u,v) \\
\geq \sum_{t=1}^{T} \left(1 - \frac{1}{2}\right) t^{-1} \frac{\alpha x^*_f}{2^T} \sim \alpha x^*_f (1 - 1/e)
\]

The last term is obtained after taking } T \to \infty. \text{ Similarly, we can show that } \kappa_f^y \geq \frac{\beta y^*_f (1 - 1/e)}{2}.

Let } \text{Profit}(\alpha, \beta) \text{ be the expected total profit obtained by WarmUp(\alpha, \beta). \text{ By linearity of expectation, we have}}

\[
\text{Profit}(\alpha, \beta) \geq \frac{(1 - 1/e)^2}{2} \sum_{f \in E} x^*_f p_w c_w.
\]

We know the expected profit in offline optimal is upper bounded by \[\sum_{f \in E} x^*_f p_w c_w.\] Thus we claim that WarmUp(\alpha, \beta) achieves a ratio at least \[\alpha (1 - 1/e)/2\] on the profit. Similarly, we can argue that WarmUp(\alpha, \beta) achieves a ratio at least \[\beta (1 - 1/e)/2\] on the fairness. \qed

\subsection{The Second Algorithm AttenAlg(\alpha, \beta)}

Inspired by [Brubach et al., 2020], we can improve at least the theoretical performance of WarmUp with attenuation techniques applied to edges and (offline) vertices. The motivation behind is very simple. Note that edges in \[E_v\] are competing for each other since we have to stop probing whenever \[v\] is matched. Thus, attenuating those edges which win the higher chance of probing over others can potentially boost the worst-case performance.

Let } \{T, \mu_t, u;v \in [T]\} \text{ be such a series that is defined as } \gamma_1 = 1, \mu_1 = 1 - \gamma_2/2, \gamma_{t+1} = \gamma_t (1 - \mu_t) / T, \text{ and } E_{v; t} \text{ be the set of available edges } f = (u, v) \in E_v \text{ at time } t \text{ (i.e., } u \text{ is available at } t). \text{ The formal description of AttenAlg is stated in Algorithm 3. We defer the proofs of Theorems 2 and 3 to the full version}.^2

\section{Experiments}

\subsection{Data Preprocessing}

We use the New York City yellow cabs dataset\(^3\), which is collected during the year of 2013. Although the demographics of the drivers and riders are not recorded in the original dataset, we synthesize the racial demographics for riders and drivers in a similar way to [Nanda et al., 2019]. To simplify the demonstration, we consider a single demographic factor of the race only, which takes two possible options between “disadvantaged” (D) or “advantaged” (A). We set the ratio of D to A to be 1 : 2 among riders, which roughly matches the racial demographics of NYC [Review, 2019]. Similarly, we set the ratio of D to A among drivers to be 1 : 2 [Hall and Krueger, 2017]. The acceptance rates among the four possible driver-ri-der pairs (based on race status only), (A,A), (A,D), (D,A), (D,D), are set to be 0.6, 0.1, 0.1 and 0.3, respectively. These probabilities are then scaled up by a factor \[\eta\] such that \[p_f = \eta \cdot (1 - \eta) \cdot p_f.\] In our experiments we set \[\eta = 0.5.\] Note that we can apply our model straightforwardly to the case when the real-world distribution of \[\{p_f\}\] values is known or can be learned. We collect records during the off-peak period when a lot of drivers are on the road while the requests are relatively lower than peak hours. On January 31, 20,701 trips were completed in the off-peak hour (4–5 PM), compared to 35,109 trips in the peak hour (7–8 PM). We focus on longitude and latitude ranging from (−73, −75) and (40.4, 40.95) respectively. We partition the area into 40 \times 11 grids with equal size.

We construct the compatibility graph \[G = (U, V, E)\] as follows. Each } u \in U \text{ represents a driver type which has attributes of the starting location and race. Each } v \in V \text{ represents a request type which has attributes of the starting location, ending location, and race. We downsample from all driver and request types such that } |U| = 57 \text{ and } |V| = 134. \text{ For each driver type } u, \text{ we assign its capacity } B_u \text{ with a random value uniformly sampled from [1, B] where we vary } B \in \{10, 15, 25\}. \text{ For each request type } v, \text{ we sample a random patience value } \Delta_v \text{ uniformly from [1,2] and a random arrival rate } r_v \sim N(5, 1) \text{ (Normal distribution), and then set } T = \sum_{v \in V} r_v. \text{ We add an edge } f = (u, v) \text{ if the Manhattan distance between starting location of request type } v \text{ and the location of driver type } u \text{ is not larger than } 1. \text{ The profit } w_f \text{ for each } f \text{ is defined as the normalized trip length of the request type } v \text{ such that } 0 \leq w_f \leq 1.

\textbf{Algorithms.} We test the WarmUp(\alpha, \beta) with } \alpha + \beta = 1 \text{ against two natural heuristic baselines, namely Greedy-P (short for Greedy-Profit) and Greedy-F (short for Greedy-Fairness)\(^4\). Suppose a request type of } v \text{ arrives at time } t. \text{ Recall that } E_v \text{ is the set of neighboring edges incident to } v \text{ (i.e., the set of assignments feasible to } v). \text{ Let } E_v^* \subseteq E_v \text{ be the set of available assignments } f = (u, v) \text{ such that there

\[\text{Algorithm 3 AttenAlg(\alpha, \beta)}\]

\begin{algorithmic}[1]
\State \text{for } t = 1, 2, \ldots, T \text{ do }
\State \text{Apply vertex-attenuation such that each } u \in U \text{ is available at } t \text{ with probability equal to } \gamma_t.
\State \text{Let } v \text{ arrive at time } t.
\State \text{With probability } \alpha, \text{ Run } \text{SR}(x^*). \text{ Apply edge-attenuation such that each edge } f \in E_{v,t} \text{ is probed in } \text{SR}(x^*) \text{ with probability equal to } \mu_f x^*_f / r_v.
\State \text{With probability } \beta, \text{ Run } \text{SR}(y^*). \text{ Apply edge-attenuation such that each edge } f \in E_{v,t} \text{ is probed in } \text{SR}(y^*) \text{ with probability equal to } \mu_f y^*_f / r_v.
\State \text{With probability } 1 - \alpha - \beta, \text{ reject } v.
\end{algorithmic}

\(^3\)http://www.andresmh.com/nyctaxitrips/

\(^4\)A future direction is to consider a hybrid version of Greedy-P and Greedy-F, which optimizes the two objectives simultaneously.
In contrast, Greedy-F will repeat greedily selecting an available assignment \( f \in E_v \) with the maximum weight \( w_fp_f \) over \( E_v \) (breaking ties arbitrarily) until either \( v \) accepts a driver or \( v \) runs out of patience. In contrast, Greedy-F will repeat greedily selecting an available \( f = (u^*, v) \in E_v \) with \( u^* \) having the least matching rate before either \( v \) accepts a driver or leaves the system. Note that we use LP-(1) and LP-(2) as the default benchmarks for profit and fairness, respectively.

5.2 Results and Discussions

Figure 1 shows the results of competitive ratios for the proposed algorithm with different values of \( \alpha \) with \( \beta = 1 - \alpha \). We can observe that the profit and fairness competitive ratios of WarmUp always stay above the theoretical lower bounds (in dotted lines), as predicted in Theorem 1. The gaps between performances and lower bounds suggest that theoretical worst scenarios occur rarely in the real world. Note that when \( B = 25 \) and \( \alpha = 1 \) as shown in Figure 1(c), the lower bound is tight and matches the fairness performance.

Figure 2 shows the profit and fairness performances of WarmUp compared to Greedy-P and Greedy-F. Here are a few interesting observations. (1) As for profit, Greedy-P can always beat Greedy-F but not necessarily for WarmUp. The advantage of Greedy-P over WarmUp becomes more apparent when \( B \) is large and less when \( B \) is small. Note that in our experiment, the expected total number of arrivals of riders is fixed and therefore, \( B \) directly controls the degree of imbalance between drivers and riders. When \( B \) is larger, we have more available drivers compared to riders and thus, Greedy-P will outperform all the rest for profit. When \( B \) is small, however, we really need to carefully design the policy to boost profit. That’s why WarmUp becomes dominant. (2) As for fairness, Greedy-F seemingly can always dominate the rest, though WarmUp shows high flexibility in the fairness performance. WarmUp shows a relatively low sensitivity toward the first parameter \( \alpha \) for profit while high sensitivity toward the second parameter \( \beta \) for fairness: the latter becomes particularly obvious when \( B \) is large.

6 Conclusion

In this paper, we present a flexible approach for matching requests to drivers to balance the two conflicting goals, maximizations of income equality among all rideshare drivers and the total revenue earned by the system. Our proposed approach allows the policy designer to specify how fair and how profitable they want the system to be via two separate parameters. Extensive experimental results on the real-world dataset show that our proposed approaches not only are far above the theoretical lower bounds but also can smoothly tradeoff the two objectives between the two natural heuristics. Our work opens a few directions for future research. The most direct one is to shorten the gap between the sum of ratios of profit and fairness achieved by AttenAlg (which is 0.46). It will be interesting to give a tighter online analysis than what are presented here or offer a sharper hardness result which suggests the sum of the two ratios should be much lower than 1.

Acknowledgments

Pan Xu was partially supported by NSF CRII Award IIS-1948157. The authors would like to thank the anonymous reviewers for their helpful feedback.
References


