

A Data-Driven Approach to Manage the Curbside Ride-hailing Pick-ups and Drop-offs

Xiaohui Liu¹, Lei Xu², Sean Qian³, Wei Ma^{1*},

¹Civil and Environmental Engineering, The Hong Kong Polytechnic University

²Shenzhen Research Institute of Big Data

³Civil and Environmental Engineering & Heinz College, Carnegie Mellon University

November 30, 2020

Keywords— Curbside management; Double machine learning; Treatment analysis; Ride-hailing services; Spatio-temporal data analysis

1 Introduction

With the proliferation of mobile devices and advancements of accurate positioning services, the ride-hailing service is becoming an indispensable component in mobility systems. Comparing to traditional public transit, passengers who are taking Uber or Lyft can decide on where to pick up and drop off which is rather convenient for themselves. It is reported by the New York City Taxi and Limousine Commission (TLC) that there are approximately 1,000,000 trips every day in the New York. The shape of mobility systems has been completely changed by the convenient, low-cost, and individual-centered ride-hailing services by Transportation Network Companies (TNCs). However, many recent studies have pointed out that the ride-hailing services are contributing to the traffic congestion by exploiting more public resources than driving and public transits (Castiglione et al. 2016, Li et al. 2016, Castiglione et al. 2018, Agarwal et al. 2019). For example, the cruising behaviors of vacant ride-hailing vehicles can densify the urban traffic and hence cause more congestion (Xu et al. 2017). In addition, the traffic can be temporally disturbed or blocked by the pick-ups and drop-offs of the ride-hailing vehicles, especially on one-lane roads (Goodchild et al. 2019). What’s more, if the parking spot of passengers’ pick-ups and drop-offs is on the illegal parking roads, it may lead to some extra traffic accidents, such as the pedestrians hurt by the sudden opening of the taxi doors. Given the impact of cruising behaviors being intensively studied in recent years (Xu et al. 2019, Ban et al. 2019), we notice that, to the best of our knowledge, there are few studies about the impact of ride-hailing pick-ups and drop-offs on traffic congestion and travelers’ behaviors. In view of this, this paper develops a data-driven framework to model the spatio-temporal congestion impact of the ride-hailing vehicles (RVs) pick-ups and drop-offs.

There are many reasons for the congestion caused by pick-ups and drop-offs of ride-hailing vehicles. Firstly, passengers usually choose arbitrary pick-ups and drop-offs locations according to their own preferences. This kind of random parking is harmful to unified road management. Secondly, the pick-ups and drop-offs require the RVs to leave and re-join the traffic stream frequently, which can disturb the traffic and induce extra delays. Thirdly, the RV pick-ups and drop-offs usually happen on the curb spaces, which are originally designed for on-street parking, truck loading, temporary stopping of conventional vehicles (CVs). Due to the limited space on curbside and unexpected high volumes of RV pick-ups and drop-offs, RVs might spill back to the roads and hence cause congestion. Though the impact of RV pick-ups and drop-offs depends on road properties and traffic conditions, the empirical studies have suggested that it is generally negative and significant (Goodchild et al. 2019). Therefore, it is rather essential to explore the impact of pick-ups and drop-offs on the traffic states. It can effectively alleviate traffic jams and reduce

*Corresponding author, +1(412)-708-3668, wei.w.ma@polyu.edu.hk

traffic accidents. However, there is a lack of frameworks to evaluate the impact caused by RV pick-ups and drop-offs.

Currently, the management of RV pick-ups and drop-offs largely depends on practical experiences and heuristics. The RV pick-ups and drop-offs can be viewed as the very-short-term parking, and hence the curbside management can be considered coherently within the framework of parking management (Schaller et al. 2011). For example, some airports (*e.g.* JFK) and hotels have set up RV loading zones, and some airports (*e.g.* LAX) directly ban the curbside pick-ups of RVs. However, the analytical evaluation of different management practices and the optimal management strategies are still lacking due to the lack of theoretical tools.

To summarize, there is a lack of analytical model to quantify and manage the impact of RV pick-ups and drop-offs on the traffic. In view of this, this paper develops a data-driven framework to evaluate and manage the congestion impact of RV pick-ups and drop-offs using the massive traffic speed data and trip record data. Mathematically, the proposed framework builds a Double Machine Learning (DML) model to evaluate how much influence the numbers of pick-ups/drop-offs have on traffic states, and a linear programming is formulated to re-route the RVs to the nearby roads. To demonstrate the effectiveness of the proposed framework, TLC trip record data and INRIX traffic speed data are used together to build the DML and optimization models for the New York city. The preliminary experimental results demonstrate the effectiveness and robustness of the proposed framework.

The remainder of this paper is organized as follows. Section 2 discusses methodology including scenarios clustering, double machine learning and rerouting the pick-ups and drop-offs. Section 3 takes Manhattan as the target city and presents results of the proposed framework. Finally, conclusions are drawn in Section 4.

2 Methodology

In this section, we briefly introduce the learning and managing framework for the RV pick-ups and drop-offs. The proposed framework consists of three components: scenarios clustering, double machine learning model, and pick-up/drop-off rerouting.

The traffic states of a transportation network \mathcal{G} are modeled by using a spatial random variable that evolves over time, $\{F_t^r \in \mathbb{R}^+, t \in \mathbb{T}\}$, where \mathcal{G} is the road network under study consisting of regions \mathcal{R} , r is a region in \mathcal{G} , $r \in \mathcal{R}$, \mathbb{T} is the set of the study periods, and F_t^r is the quantity of measuring traffic state at the region r and time t .

Scenarios clustering.

It is rather important to identify and cluster different traffic scenarios before we begin to train the DML model and make rerouting plans. Because the pick-up/drop-offs and congestion patterns are changing spatio-temporally and are up to the amounts and types of vehicles, the road conditions, surrounding buildings and so on. There is no difference between the behaviors of pick-ups and drop-offs from the perspective of transportation because they can both be regarded as short-term parking. But we can't recognize all these different kinds of scenarios as the same pattern. Otherwise, the dynamics of traffic states will be ignored.

The traffic states at different time periods have both randomness and periodicity (Ma & Qian 2018). This paper aims to identify busy roads and make a plan for rerouting pick-ups and drop-offs so that the traffic congestion can be alleviated. We firstly use the scenarios clustering over time to tackle the temporal dynamics. For example, in one week, the traffic states are various between working days and weekends. Besides, the traffic congestion on Tuesday, Wednesday, Thursday is much heavier than that in other days. What's more, the impact of pick-up/drop-offs is unequal during morning and afternoon peaks. Hence we select the afternoon peaks that last from 16:00 pm to 20:00 pm on Tuesday, Wednesday and Thursday as our target scene. If we can model the busiest scenarios well, the others can be processed by the same method. Secondly, in order to learn the relationship between the traffic states and the total number of

Table 1: Notations of DML

Notations	Variables	Description
F_t^r	Outcome variable	Speed at time t at region r
D_t^r	Treatment variable	Total number of pick-ups and drop-offs at time t at region r
X_t^r	Features	Vector of speed at previous time $\{F_{t-1}^r, F_{t-2}^r, \dots, F_{t-10}^r\}$
W_t^r	Control variable	Precipitation at time t at region r
$\theta(X_t^r)$	Treatment effect	Given X_t^r , how much influence Treatment D_t^r has on Outcome F_t^r

pick-ups and drop-offs, we divide one city into several districts to capture fine-grained transportation patterns. In the next section, we will build the estimating model for each scenario separately.

The Double Machine Learning model.

In the second step, we build the DML model to estimate the effect of pick-ups and drop-offs on the traffic states. It is note that this study attempts to estimate the “treatment effects” of the pick-ups and drop-offs, instead of the correlations among different variables.

This paper aims to identify whether the number of pick-ups and drop-offs is one major reason contributing to the traffic congestion or not. Compared to previous casual inference models, DML not only focuses on modeling non-linear relationship between different variables but also overcomes regularization biases based on arbitrary machine learning methods, including random forests, lasso, ridge, deep neural nets, boosted trees, and various hybrids and ensembles of these methods (Chernozhukov et al. 2018). It is an effective method to capture the casual relationship between the numbers of the pick-up/drop-offs and traffic states.

We will introduce how the DML model achieves unbiased estimation and how this model is used in our research. There are notations used in this model shown as Table 1. In this paper, precipitation W_t^r is considered as control variable because we should explore such casual inference under the same weather condition. Besides, previous speed data is taken as features X_t^r . The target is to estimate $\theta(X_t^r)$ without any bias. If we can achieve this goal, the casual inference between the F_t^r and D_t^r can be captured so that we can make rerouting plans to solve traffic jams.

Firstly, we discuss DML model from partially linear regression as Equation 1 and Equation 2.

$$F_t^r = \theta(X_t^r) \cdot D_t^r + g(X_t^r, W_t^r) + U_t^r, \quad \mathbb{E}[U_t^r | X_t^r, W_t^r] = 0 \quad (1)$$

$$D_t^r = f(X_t^r, W_t^r) + V_t^r, \quad \mathbb{E}[V_t^r | X_t^r, W_t^r] = 0 \quad (2)$$

Where $g(X_t^r, W_t^r)$ and $f(X_t^r, W_t^r)$ trained by using two machine learning approaches are functions of X_t^r and W_t^r . U_t^r and V_t^r are random noises. And θ is the parameter which indicates the treatment effect. Given certain features X_t^r and controls W_t^r , it can interpret how much influence the total numbers of pick-ups and drop-offs have on the traffic states.

In order to estimate $\theta(X_t^r)$, the Equation 1 can be written as Equation 3.

$$F_t^r - \mathbb{E}[F_t^r | X_t^r, W_t^r] = \theta(X_t^r) \cdot (D_t^r - \mathbb{E}[D_t^r | X_t^r, W_t^r]) + U_t^r \quad (3)$$

Then, we train the first stage model shown in Equation 4 and the second stage model as in Equation 5 to represent two expectations in above Equation 3 respectively. We will train these two models by using Random Forest, Gradient Tree Boosting, Support Vector Machines and so on and select two models that perform best among these(Research 2019).

$$q(X_t^r, W_t^r) = \mathbb{E}[F_t^r | X_t^r, W_t^r] \quad (4)$$

$$f(X_t^r, W_t^r) = \mathbb{E}[D_t^r | X_t^r, W_t^r] \quad (5)$$

Next, Equation 3 can be represented by residuals of real value and predicted value shown as Equation 6.

$$F_t^r - q(X_t^r, W_t^r) = \theta(X_t^r) \cdot (D_t^r - f(X_t^r, W_t^r)) + U_t^r \quad (6)$$

We can obtain $\hat{\theta}(X_t^r)$ shown as Equation 7, the estimation value of $\theta(X_t^r)$, by training the final stage model based on linear regression, such as Lasso.

$$\hat{\theta}(X_t^r) = \arg \min_{\theta \in \Theta} \mathbb{E}[(F_t^r - q(X_t^r, W_t^r)) - \theta(X_t^r) \cdot (D_t^r - f(X_t^r, W_t^r))] \quad (7)$$

Finally, we can obtain the predicted value \tilde{F}_t^r of F_t^r as written in Equation 8.

$$\tilde{F}_t^r = \hat{\theta}(X_t^r) \cdot (D_t^r - f(X_t^r, W_t^r)) + q(X_t^r, W_t^r) + U_t^r \quad (8)$$

Rerouting the pick-ups and drop-offs.

Once the impact of the pick-up/drop-offs on the congestion has been learned by the DML model, the average treatment effects of pick-up/drop-offs to the congestion can be extracted from this model. For each time interval t , the solution of rerouting is presented in Equation 9.

$$\begin{aligned} \min_{\tilde{p}_t^r, \tilde{d}_t^r, \Delta_t^p(r, r'), \Delta_t^d(r, r')} \quad & \sum_{r \in \mathcal{R}} c_r(\tilde{F}_t^r) - \sum_{r, r'} c^p(r, r') \Delta_t^p(r, r') - \sum_{r, r'} c^d(r, r') \Delta_t^d(r, r') \\ \text{s.t.} \quad & \hat{\theta}(X_t^r) \cdot (D_t^r - f(X_t^r, W_t^r)) + q(X_t^r, W_t^r) + U_t^r = \tilde{F}_t^r \quad \forall r \\ & p_t^r - \sum_{r' \in \mathcal{N}(r)} \Delta_t^p(r, r') = \tilde{p}_t^r \quad \forall r \\ & d_t^r - \sum_{r' \in \mathcal{N}(r)} \Delta_t^d(r, r') = \tilde{d}_t^r \quad \forall r \\ & p_t^{r'} + \sum_{r \in \mathcal{N}(r')} \Delta_t^p(r, r') = \tilde{p}_t^{r'} \quad \forall r' \\ & d_t^{r'} + \sum_{r \in \mathcal{N}(r')} \Delta_t^d(r, r') = \tilde{d}_t^{r'} \quad \forall r' \\ & 0 \leq \Delta_t^p(r, r') \leq \varepsilon \quad \forall r, r' \\ & 0 \leq \Delta_t^d(r, r') \leq \varepsilon \quad \forall r, r' \end{aligned} \quad (9)$$

where $\tilde{p}_t^r, \tilde{d}_t^r$ are the numbers of pick-up/drop-offs after the re-routing, $c_r(\cdot)$ converts the traffic states to costs, $c^p(r, r')$ and $c^d(r, r')$ are the unit costs of the rerouting from region r to region r' , $\Delta_t^p(r, r')$ and $\Delta_t^d(r, r')$ are the number of vehicles for pick-up/drop-offs to be rerouted from region r to region r' , $\mathcal{N}(r)$ is the neighbors of r , and ε is the maximum number of vehicles to be rerouted. Formulation 9 is a linear programming, hence it can be solved efficiently for large-scale networks.

The objective of Formulation 9 is to minimize the cost of rerouting. This cost contains two parts: (1) some pick-ups and drop-offs that originally happen in region r will be reassigned into this region's neighbor r' . After rerouting, we qualify the effect of rerouting as the cost of current speed state, $c_r(\tilde{F}_t^r)$. (2) Some additional cost $\sum_{r, r'} c^p(r, r') \Delta_t^p(r, r')$ and $\sum_{r, r'} c^d(r, r') \Delta_t^d(r, r')$ will take when some drivers are directed into their neighboring area.

3 Results

To examine the effectiveness of the proposed framework, we use the New York Taxi and Limousine Commission (TLC) trip record data and INRIX traffic speed data to build the proposed model for Manhattan in the New York city. Both the speed data and trip record data are from February 1, 2019 to January 31, 2020. In our experiment, the traffic speed is used to represent the traffic states because it can reflect whether the current road is blocked or not to some extent. And the trip record data contains the pick-up/drop-offs information including the date, the taxi types, current region and so on in Manhattan. The preliminary statistics result is shown in the Figure 1. As we can see, in different regions r , there are varied numbers of pick-ups and drop-offs. Especially for some center regions, the value of pick-ups and drop-offs are much bigger than that in other regions.

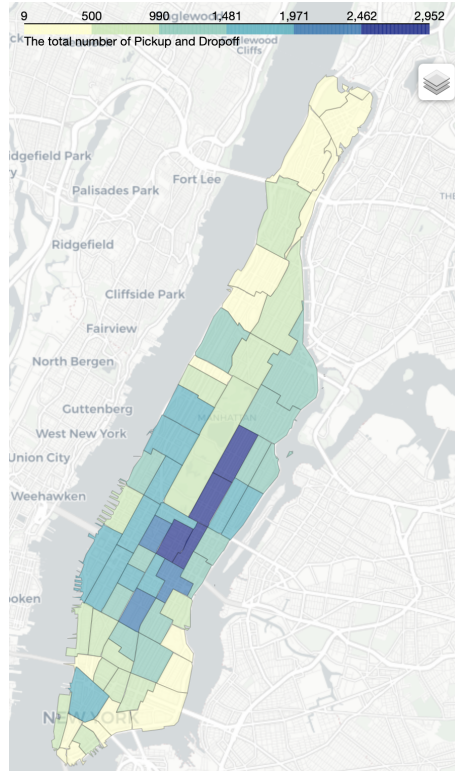


Figure 1: Overview of the distribution of pick-ups and drop-offs in Manhattan district.

The experiments are still on-going, while the preliminary results show that the pick-up/drop-offs have statistically significant impact on the traffic congestion, and the efficiency of the proposed re-routing method is satisfactory.

4 Conclusion

This paper develops a holistic data-driven framework to learn and manage the curbside pick-ups and drop-offs in order to reduce their negative impact to traffic congestion. The New York TLC trip record data and INRIX data are used to examine the proposed framework, and the preliminary results are compelling and satisfactory. The contributions of this paper are summarized as follows:

- It develops a novel data-driven framework to learn the spatio-temporal impact of RV pick-up/drop-offs on the traffic congestion.
- It builds a data-driven framework to re-route the pick-up/drop-offs in order to minimize the congestion induced by the pick-up/drop-offs.
- It examine the proposed framework with real-world data on large-scale networks.

Acknowledgments

The work described in this study was supported by a grant funded by the Hong Kong Polytechnic University (Project No. P0033933). The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein.

References

- Agarwal, S., Mani, D. & Telang, R. (2019), ‘The impact of ride-hailing services on congestion: Evidence from indian cities’, *Available at SSRN 3410623* .
- Ban, X. J., Dessouky, M., Pang, J.-S. & Fan, R. (2019), ‘A general equilibrium model for transportation systems with e-hailing services and flow congestion’, *Transportation Research Part B: Methodological* **129**, 273–304.
- Castiglione, J., Chang, T., Cooper, D., Hobson, J., Logan, W., Young, E., Charlton, B., Wilson, C., Mislove, A., Chen, L. et al. (2016), ‘Tncs today: a profile of san francisco transportation network company activity’, *San Francisco County Transportation Authority (June 2016)* .
- Castiglione, J., Cooper, D., Sana, B., Tischler, D., Chang, T., Erhardt, G. D., Roy, S., Chen, M. & Mucci, A. (2018), ‘Tncs & congestion’.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018), ‘Double/debiased machine learning for treatment and structural parameters’.
- Goodchild, A., MacKenzie, D., Ranjbari, A., Machado, J. & Chiara, G. D. (2019), ‘Curb allocation change project’.
- Li, Z., Hong, Y. & Zhang, Z. (2016), ‘Do ride-sharing services affect traffic congestion? an empirical study of uber entry’, *Social Science Research Network* **2002**, 1–29.
- Ma, W. & Qian, Z. S. (2018), ‘Estimating multi-year 24/7 origin-destination demand using high-granular multi-source traffic data’, *Transportation Research Part C: Emerging Technologies* **96**, 96–121.
- Research, M. (2019), ‘EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation’, <https://github.com/microsoft/EconML>. Version 0.x.
- Schaller, B., Maguire, T., Stein, D., Ng, W. & Blakeley, M. (2011), Parking pricing and curbside management in new york city, Technical report.
- Xu, Z., Chen, Z. & Yin, Y. (2019), ‘Equilibrium analysis of urban traffic networks with ride-sourcing services’, *Available at SSRN 3422294* .
- Xu, Z., Yin, Y. & Zha, L. (2017), ‘Optimal parking provision for ride-sourcing services’, *Transportation Research Part B: Methodological* **105**, 559–578.