Title: Understanding ride-sourcing drivers' customer-search behavior

Junji Urata[a c], Jintao Ke[b], Zhengtian Xu[c 1], Guojun Wu[d], Yafeng Yin[c e], Hai Yang[b], Jieping Ye[f]

[a] *Department of Civil Engineering, University of Tokyo, Japan*

[b] *Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*

[c] *Department of Civil and Environmental Engineering, University of Michigan, United States*

[d] *Data Science Program, Worcester Polytechnic Institute, United States*

[e] *Department of Industrial and Operations Engineering, University of Michigan, United States*

[f] *AI Labs, Didi Chuxing, Beijing, China*

Keywords: Ride-sourcing service, Customer-search, Driver behavior, Dynamic discrete choice

Introduction

With the boom of mobile internet technology, ride-sourcing services now become increasingly popular for intra-city travels. The great success of ride-sourcing services has triggered heated discussions on the analysis and management of such on-demand ride-hailing systems. However, less attention has drawn on the behavior of ride-sourcing drivers, partially due to the confidentiality of service data. By virtue of comprehensive empirical data from Didi Chuxing, this paper thus aims to comprehend the behavior of ride-sourcing drivers in customer-search.

Many empirical studies have been carried out to investigate the customer-search behavior of taxi drivers, who shares a significant similarity with drivers in ride-sourcing markets. A group of researchers from Hong Kong first applied the Multinomial Logit (MNL) model to capture strategic zonal choices of Hong Kong taxi drivers' customer-search (see, e.g., Wong et al. 2014, 2015). They proposed a cell-based Logit-opportunity model to tackle the local customer-search behavior of taxis by considering the opportunities along search paths. Recently, Tang et al. (2019) argued that between different destination choices of vacant taxis, there are substantial overlaps in paths, which invalidate the use of MNL models. Instead, they proposed a mixed path size Logit-based customer-search model and tested its effectiveness in predicting routing choices over the trajectory of 36,000 taxis in Beijing. Although these static search models substantially facilitate empirical calibration, they fall short in dealing with highly time-varying market conditions and the time effects in drivers' choices. Zheng et al. (2018) modeled vacant taxi drivers' anticipatory behavior by using a time-dependent framework. But their study focused on the one-shot decision choices of taxi drivers between the urban areas and condensed-demand areas, such as airports and railway stations, and is unsuitable for behavioral calibration. One of the major difficulties in calibrating the search behavior is that drivers' trajectories do not fully reflect their real preferences. It is common for drivers to get matched to passengers before reaching the actual cruising destinations, especially in app-based ride-hailing markets. Sometimes, drivers do not even have specific search destinations in mind. Therefore, we are in need of a behavioral model that can cope with the intense market variations and identify drivers' latent search patterns with modeling differentiations.

To fill the research gap, this paper develops a dynamic discrete choice model to investigate the customer-search behavior of ride-sourcing drivers when they are idle. The model translates market conditions, including both supply and demand information, into a spatiotemporal continuum of opportunity values to idle drivers. Specifically, our model differentiates three universal modes of search movements, as staying motionless, cruising around without

[1] Corresponding author. E-mail address: xzt@umich.edu (Z. Xu)

target areas, or repositioning towards specific destinations, respectively. By calibrating the model, we shed light on ride-sourcing drivers' behavioral concerns over various factors underlying their customer-search decisions, which can be insightful for the platform's supply management.

Methodology

A dynamic discrete choice model with absorbing Markov chains is introduced to describe the customer-search processes of ride-sourcing drivers over hexagonal cells. We materialize the process into a cross-nested structure, where drivers sequentially make zonal-transition decisions with a probability of getting matched at each stage. Figure 1 sketches the nest structure of an idle driver's sequential decision-making process. During the process, a driver at each stage can choose to either stay for a while in the current zone or reposition to one of the adjacent zones, based on the expected utility of each choice. The upper nest indicates the scenarios of different zonal search patterns, either to stay motionless, cruise nearby without specific destinations, or reposition towards hotspot areas.

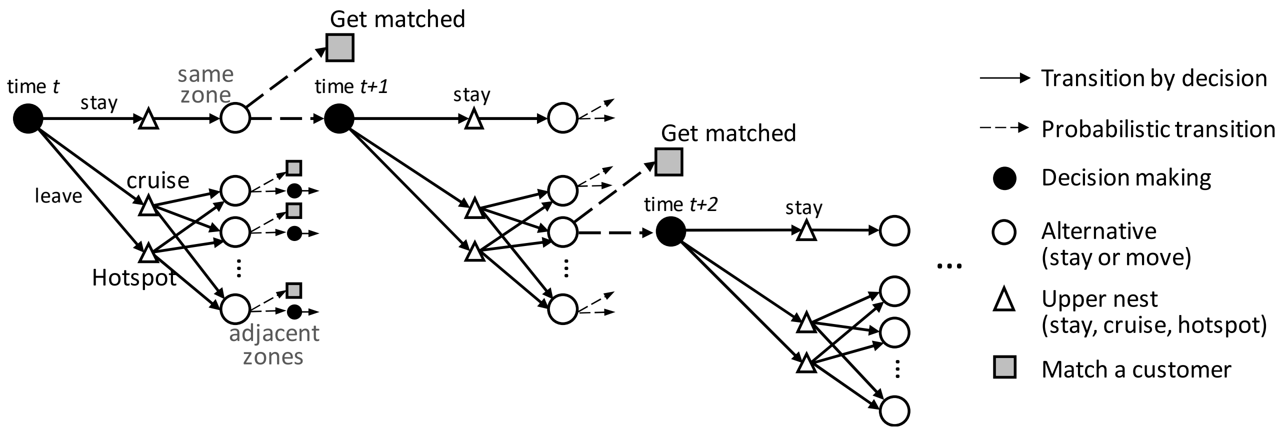

Figure 1: Decision making processes of ride-sourcing drivers

The expected value of our proposed model with cross-nested structure is given by

$$V^T(z_t) = \frac{1}{\mu_l} \ln \sum_{l \in C'(z_t)} \exp\big(\mu_l v'_l(l|z_t)\big) \tag{5}$$

where l is the choice in the upper nest, $\frac{1}{\mu_l}$ is a scale parameter of upper nest that reflects the degree of independence among the unobserved portions of utility for alternatives in an upper nest. The log-sum (inclusive) utility v'_l can be further specified as

$$v'_l(l|z_t) = \frac{1}{\mu_n} \ln \sum_{z_{t+1} \in C(z_t, l)} \left(\alpha_{z_t l} \exp\left(v(z_{t+1}|z_t) + \rho_{z_t} \beta V^T(z_{t+1})\right)\right)^{\mu_n} \tag{6}$$

where $\frac{1}{\mu_n}$ is a scale parameter of the random utility of different alternatives, and $\alpha_{z_t l}$ is an allocation parameter that reflects the likelihood of alternative z_t being a member of nest l with $\alpha_{z_t l} \geq 0$ and $\sum_l \alpha_{z_t l} = 1$ (Train, 2009). Besides, the choice probability is given by

$$P^T(z_{t+1}|z_t) = \sum_{\forall l} P^T(l|z_t) P^T(z_{t+1}|l) =$$
$$\sum_{\forall l} \frac{\exp \mu_l v'_l(l|z_t)}{\sum_{l' \in C'(z_t)} \exp \mu_l v'_{l'}(l'|z_t)} \frac{\left(\alpha_{z_t l} \exp\left(v(z_{t+1}|z_t) + \rho_{z_t} \beta V^T(z_{t+1})\right)\right)^{\mu_n}}{\sum_{z_{t+1} \in C(z_t, l)} \left(\alpha_{z_t l} \exp\left(v(z_{t+1}|z_t) + \rho_{z_t} \beta V^T(z_{t+1})\right)\right)^{\mu_n}}, (\mu_n \geq \mu_l \geq 1). \tag{7}$$

We estimate the behavioral parameters $\Theta = (\theta, \beta, \mu, \alpha)$ using maximum-likelihood method to understand drivers' customer-search behavior.

Results

Two datasets are collected from 10 weekdays in August 2017 at a mid-sized city in China. The first dataset contains the information of around 5 million orders. The order information is then aggregated and averaged by hexagonal cells and 4-min time intervals to produce explanatory market variables, including trip fare (TF), number of requested orders (NR), pickup time (PT), probability of being matched (MP: $1 - \rho_{z_t}$), distance to the nearest hotspot (DH), and duration of drivers' stay in the zone (ST). Two hotspot areas are identified in the city using aggregate demand statistics. The second dataset contains the GPS trajectories of 32,000 active drivers, including 15,000 full-time drivers, 8,000 diligent part-time drivers, and 9,000 normal part-time drivers. This dataset is processed to generate idle drivers' zonal search sequences.

All the explanatory variables and search trajectories are fed into our dynamic discrete choice model to estimate the effects of various factors in drivers' customer-search.l Table 1 shows the coefficient estimates by different driver groups. In general, the coefficient estimates confirm that idle drivers prefer to search in areas with higher number of requests, higher probability of being matched, and higher per-trip fares. Besides, drivers tend to cruise towards the hotspot areas nearby, and are more active searching customers while inside the areas with higher number of requests.

Table 1. Coefficient estimates by different driver groups (full-day, T = 2)

Driver type	Full-time		Diligent part-time		Normal part-time	
	Param.	t-Stat.	Param.	t-Stat.	Param.	t-Stat.
Utility						
TF	0.02	2.10*	0.09	4.76**	0.01	0.29
NR	0.08	36.77**	0.07	15.92**	0.06	13.97**
PT	-0.02	-3.03**	0.08	5.71**	0.08	5.53**
MP	1.09	18.98**	0.79	7.07**	0.71	6.35**
SZ	1.71	187.97**	1.86	90.83**	1.77	95.41**
Allocation						
DH	-6.10	-6.69**	-11.86	-3.31**	-13.96	-3.56**
NR	0.41	4.96**	0.37	3.72**	0.33	2.52**
ST	-6.42	-5.84**	-12.86	-2.71**	-16.41	-3.29**
discount β	0.33	23.02**	0.20	6.30**	0.23	7.01**
scale μ_n	1.75	65.31**	1.88	29.19**	1.86	30.74**
Observations		165410		45532		38246
LL(0)		-184763		-50732		-42316
Final LL		-112181		-28153		-24835
CV LL		-109698		-28126		-24075
Adjusted ρ^2		0.39		0.45		0.41

Notes: ** = significant at 0.01, * = significant at 0.05.

Figure 2 presents the coefficient estimates on various factors for full-time drivers in different time periods of a day. As shown by the figure, full-time drivers during the morning peak show higher willingness to move compared to the other periods, and prefer the areas with higher numbers of requests. In contrast, during the evening peak, drivers are more sensitive about the probability of being matched. Their decisions are more farsighted taking into account the market conditions in further-away areas. Such search preferences continue into the night period but weaken down. During the midnight when ride requests appear to be spatially sparse, idle drivers appear to favor the areas with long-distance orders and earn higher per-order profits therein.

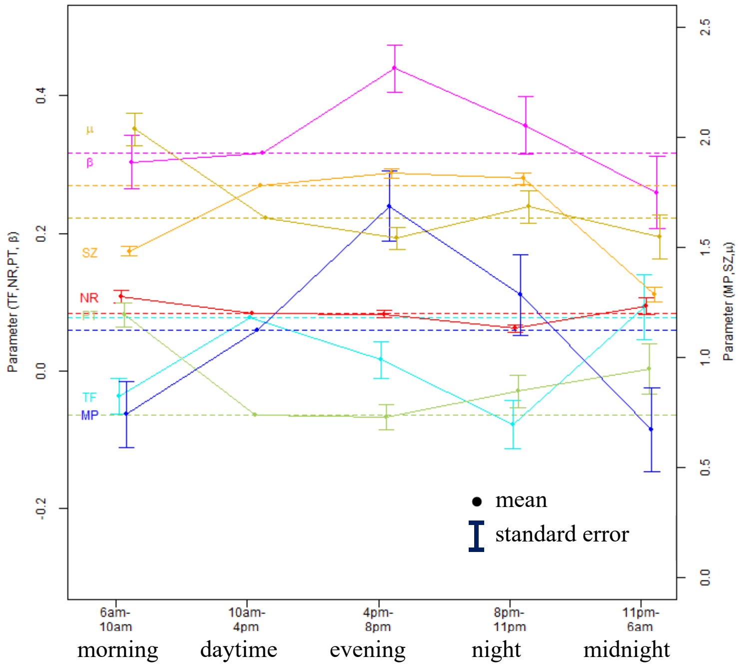

Figure 2: Full-time drivers' sensitivity on various factors in customer-search across different time of a day (The period 10am - 4pm is set as the base for comparison)

Conclusion

To the best of our knowledge, this paper is one of the first attempts to investigate ride-sourcing drivers' customer-search behavior. A dynamic discrete choice model is proposed to rationalize the time-dependent search movements of idle drivers within the spatial market. Based on large-scale datasets from real-world operations, we calibrate drivers' time-variant sensitivity to various factors in their decision making when being idle. The accuracy of our model calibration as well as further explorations and discussions on drivers' latent preferences in cruising or repositioning will be presented in the conference meeting.

References

Tang, J., Wang, Y., Hao, W., Liu, F., Huang, H., and Wang, Y. (2019). A mixed path size logit-based taxi customer-search model considering spatio-temporal factors in route choice. *IEEE Transactions on Intelligent Transportation Systems*.

Train, K. E. (2009). Discrete choice methods with simulation. *Cambridge university press*.

Wong, R.C.P., Szeto, W.Y. and Wong, S.C. (2014). A cell-based logit-opportunity taxi customer-search model.

Transportation Research Part C: Emerging Technologies, 48, 84-96.
Wong, R. C., Szeto, W., and Wong, S. (2015). Sequential logit approach to modeling the customer-search decisions of taxi drivers. *Asian Transport Studies*, 3(4): 398–415.
Zheng, Z., Rasouli, S., and Timmermans, H. (2018). Modeling taxi driver anticipatory behavior. *Computers, Environment and Urban Systems*, 69, 133-141.