Upping the Game of Taxi Driving in the Age of Uber

Shashi Shekhar Jha, Shih-Fen Cheng, Meghna Lowalekar, Nicholas Wong, Rishikeshan Rajendram
Trong Khiem Tran, Pradeep Varakantham, Nghia Truong Trong, Firmansyah Bin Abd Rahman
Fujitsu-SMU Urban Computing and Engineering (UNiCEN) Corp. Lab
Singapore Management University

Abstract

In most cities, taxis play an important role in providing point-to-point transportation service. If the taxi service is reliable, responsive, and cost-effective, past studies show that taxi-like services can be a viable choice in replacing a significant amount of private cars. However, making taxi services efficient is extremely challenging, mainly due to the fact that taxi drivers are self-interested and they operate with only local information. Although past research has demonstrated how recommendation systems could potentially help taxi drivers in improving their performance, most of these efforts are not feasible in practice. This is mostly due to the lack of both the comprehensive data coverage and an efficient recommendation engine that can scale to tens of thousands of drivers.

In this paper, we propose a comprehensive working platform called the Driver Guidance System (DGS). With real-time citywide taxi data provided by our collaborator in Singapore, we demonstrate how we can combine real-time data analytics and large-scale optimization to create a guidance system that can potentially benefit tens of thousands of taxi drivers. Via a realistic agent-based simulation, we demonstrate that drivers following DGS can significantly improve their performance over ordinary drivers, regardless of the adoption ratios. We have concluded our system designing and building and have recently entered the field trial phase.

Introduction

In most cities, taxis play an important role in providing point-to-point transportation service. If the taxi service is reliable, responsive, and cost-effective, past studies show that taxi-like services (in the literature, such services are usually termed as MoD, *mobility-on-demand*, or MaaS, *mobility-as-a-service*) can be a viable choice in replacing a significant amount of private cars (Ma, Zheng, and Wolfson 2015). However, as pointed out by earlier studies (Ding et al. 2013), making taxi services efficient is extremely challenging, mainly due to the fact that taxi drivers are self-interested and they operate with only local information.

Such lack of coordination has plagued the taxi industry for decades, until the emergence of the ride-hailing industry, which is currently championed by Uber. Uber-like services are more efficient than traditional taxis due to the following reasons: 1) Uber-like services are built on smartphones

for both the drivers and the passengers: compared to proprietary systems employed by most taxi fleets, smartphone-based platforms are much easier to scale up and cheaper to maintain, and can take advantage of the latest sensing technologies embedded in the smartphones, 2) Uber-like services focus predominantly on pre-booked passengers: this greatly reduces uncertainties resulted from having to serve street pickups, 3) Uber-like services fully embrace data science: large-scale and real-time analytics are employed in all aspects of service delivery, including but not limited to dynamic pricing (or surge pricing as it is commonly called), route recommendation, and vehicle dispatch choices, and 4) demand aggregation: taxi markets in most cities are fragmented, thus making it difficult for a passenger to find a taxi even if the passenger intends to book a ride; Uber-like platforms are designed from the beginning to play the role of demand aggregator; although such strategy is expensive to execute initially, the platform owner could enjoy near-monopolistic power once it reaches critical masses in both the drivers' and the riders' markets.

Given all these technological advances, the taxi industry will have to *upgrade* itself by learning from Uber-like services. Although the industry structure and government regulations make it difficult for the taxi industry to follow many of Uber's innovations (such as dynamic pricing or demand aggregation, just to name a few); however, the use of real-time data in improving taxi fleet's efficiency is certainly something the taxi industry is capable of doing. In this paper, we design and implement a scalable framework to achieve exactly this. Recognizing the now dominant position held by Uber-like platforms in most major cities, we have carefully positioned our framework to highlight the market segment that only taxis can satisfy. In most cities, this market segment is pickups on the streets or at the taxi stands. To support this market positioning, we analyze all taxi trips occurred in Singapore from October and November of 2016 (in total there were 33.5 million taxi trips). Of all these trips, 78% are from street pickups or taxi stands, while 22% are from various booking channels. Although this analysis does not cover private vehicles from both the Uber and Grab networks (a Uber-like operator that is competitive in the South East Asia), the huge number of rides coming from street pickups (almost 400,000 per day) indicates that this is still an important channel of demands. The key to keep potential

passengers from leaving this channel will be to make sure wait times stay reasonable. To achieve this, we argue that providing real-time guidance to drivers would be essential and this is the focus of our study.

At the very high level, we design our platform to achieve the following two goals: 1) produce high-quality real-time predictions on passenger demands, and 2) generate personalized guidance for drivers, taking into account the real-time locations of taxis and the predicted demands. In doing this, we make the following contributions:

- We design and implement a real-time demand prediction engine, with accuracy to the street level. Compared to the state-of-the-art approach from the literature, we show that our approach performs significantly better by using a multi-month real-world dataset collected from all taxis in Singapore.

- With the demand prediction engine, we implement a driver's decision support system that is capable of generating personalized guidances for a large number of drivers. The decision support system is based on a multi-step optimization model that globally matches supply and estimated demand.

- We test our system with a real-world dataset of historical traces from around 28,000 taxis in Singapore, and demonstrate that by utilizing the real-time information, our decision support system can significantly reduce empty cruising time required to acquire each trip. As a result, the average daily trips per driver can be increased significantly. This positive result is robust to the fraction of drivers adopting our technology.

Related Work

In the literature, various models have been proposed to provide driving directions to taxi drivers (Yuan et al. 2013; Zhang et al. 2016). Moreira-Matias et al. (2013) propose an approach that combines three different time-series forecasting techniques in order to predict the spatial distribution of passengers for the taxi stands in the city of Porto, Portugal for a time horizon of 30 minutes. They have evaluated their approach using 441 taxis for a particular fleet. Their approach predicts the expected demand at various taxi-stands while no emphasis was given to distribute the available supply based on the expected demand of passengers. Zhang et al. (2015) present a Bayesian learning framework in order to model the taxi drivers' learning and decision-making behaviors. Yuan et al. (2013) propose a recommendation system for both taxi drivers and passengers. They have used a probabilistic model to predict time-dependent behaviors of taxis along with providing a methodology to detect the parking places within the city. Their model uses the current location of the taxis and the time of the day in order to provide a potential parking place. However, they have not considered the real-time distribution of taxis in various states making the model less adaptive to the dynamics of the movement of passengers and taxis. Zhang et al. (2016) present a framework for predicting passenger demand using a hotness parameter derived from the historical demand pattern. This hotness parameter is combined with the current location of the taxi

driver to calculate an attractive score which is used to provide top-k recommendations to the driver.

Qu et al. (2014) discuss a model to recommend a complete driving route to taxi drivers in order to maximize their profits. They represent the taxi trajectories from historical data as a graph and then construct a net profit objective function for evaluating various driving routes. The authors argue that the recommendation of a complete driving route to the drivers is more profitable than recommending a set of prominent location and letting drivers make the decisions. Ge et al. (2010) develop a mobile recommender system for taxi drivers to recommend a sequence of pick-up points such that the recommendation helps the taxi drivers to maximize their revenue while also saving on the amount distance traveled.

Based on our extensive review of the literature (many reviews are not included here in the interest of space), there has been a lot of emphases on developing recommendation models for recommending taxi drivers' with predicted passenger demands. However, the real-time observations and predictions of supply and demand distributions (which is central to a recommendation system) are mostly neglected. This is due to two major difficulties: 1) comprehensive real-time taxi information is rarely available in major cities, and 2) even with real-time data, there is no proven solution that can generate real-time recommendations for tens of thousands of drivers. This is how our work differentiates from the literature: our system is built with real-time streaming of city-wide taxi information, and our system can serve tens of thousands of drivers efficiently and effectively.

The Driver Guidance System for Taxis

Taxi industry is diverse, complicated, and is structured and managed very differently from city to city. To put this paper in context, we build and test our system using Singapore as the testbed. Nonetheless, we believe our framework is general enough to be used in other cities; however, some components might need to be modified or re-calibrated depending on the operating environment.

In this paper, we present the Driver Guidance System (DGS) for taxi drivers. The DGS is designed to be responsive in real-time and easily scalable to tens of thousands of drivers. These requirements are challenging in two aspects: 1) data processing and optimization need to be extremely efficient to satisfy the real-time responsiveness requirement, and 2) to make the DGS scalable, we have to focus not only on speed, but also on effectiveness, i.e., the solution quality should not deteriorate even with high adoption rate. This implies that the system needs to explicitly consider all participating drivers, making the optimization problem more difficult and harder to satisfy the responsiveness requirement.

The DGS is designed to be modular, with three major components: 1) the stream data handler, 2) the real-time demand prediction engine, and 3) the multi-driver recommendation engine. They will be described in detail next.

The Singapore Taxi Industry and Its Dataset

The taxi industry in Singapore is highly regulated and the government regulator decides and approves the fleet sizes

of all fleet operators (there are seven operators, with close to 28,000 taxis at the end of 2016). Only citizens above 30 years of age are eligible to be taxi drivers. To drive a taxi, the prospective driver needs to first pass the vocational license test and medical exam, after which the driver has to rent a taxi from an operator. The rental cost is set by individual operator and usually depends on the brand and the type of the vehicle. The rental cost covers all vehicle-related costs such as insurance, maintenance, road and vehicle tax. The only additional cost the driver needs to pay out of the pocket is fuel. Drivers will keep all revenues earned.

Since the April of 2015, the Land Transport Authority (LTA), which oversees all aspects of land transportation including taxi matters, requested all taxi fleet operators to provide real-time states of taxis under their respective management every 30 seconds. By collaborating with LTA, we have acquired data continuously up to May 2017. This taxi dataset contains the following important information: 1) vehicle ID (an anonymized yet unique ID for each taxi), 2) timestamp when the state is reported, 3) the latitude and longitude of the taxi, and 4) taxi's status, can be either Available, Busy (driver indicates that he is not available for street pickups), Hired, On-Call (responding to booking request), or Offline.

Handling Noisy Stream Data

To support real-time decision support, we design our platform to accept streaming data, assuming that up to 28,000 state updates will be coming in through a private API every 30 seconds. To get rid of most GPS errors, we continuously map all taxis' locations to physical roads using a Hidden-Markov-Model-based map-matching algorithm (Newson and Krumm 2009), and make corrections to the location sensing where necessary. To ensure responsiveness, we keep all taxi-related information within a certain time window required by the map-matching algorithm in memory and we use rolling-based update. This design eliminates redundant computations and significantly reduces stream data processing times.

Demand Prediction Engine

The design of our demand prediction engine is street-based, focusing mainly on demand generation potential for each individual street. The key insight we utilize is to treat each free-cruising taxi as a *demand probe* and update the likelihood of demand generation whenever a free taxi passes by a street. This insight is obtained by analyzing several months of historical taxi dataset, from which we see that except for a few exceptions, the likelihood of *conversion* (i.e., a vacant taxi manages to find a passenger) along a particular street is positively correlated with the elapsed time since last visit by a vacant taxi. This result suggests that for many streets, demand generations do not follow Poisson arrivals (which is a popular model of choice in the literature), as the memoryless property is not valid.

Building on this insight, we proceed to quantify these correlations by executing multilevel logistic regressions for all major streets, where the status of a taxi when it exits the chosen street is treated as the dependent variable (which is binary and can be either hired or vacant), and the elapsed time since last visit by a vacant taxi is the independent variable. The regression is multilevel since we group models based on street, time of the day, and day of the week. The time of the day is discretized into a 30-minute time slot. Formally speaking, the regression follows the equation below:

$$\Pr(\text{HIRED}|\delta_s) = \text{logit}^{-1}(\alpha_{s,t,d} + \beta_{s,t,d}\delta_s), \qquad (1)$$

where s is the street, t is the time-slot of the day and d is the day of the week. α and β are the coefficients of the regression model while δ_s is the elapsed time from the lastest arrival of a taxi in the Available state on the street s. $\Pr(\text{HIRED}|\delta_s)$ signifies the likelihood of finding a passenger on street s by a cruising taxi.

To train the demand prediction model, we use the historical data of street pickups from July 2016 to October 2016. For testing purpose, we take real data from November 2016 and stream its traces as if the data is occurring in real-time. The performance of our prediction model is compared against a Non-Homogeneous Poisson (NHP) model (Moreira-Matias et al. 2012). The NHP model has a time-dependent rate function ($\lambda(t)$). Since the demand of taxis varies with the time of the day, a time-dependent rate function with a cycle of 24 hours (48 time slots in our case) is adopted as a piece-wise linear function.

The following equation describes the Poisson model with time-dependent rate function $\lambda(t)$:

$$\Pr(t_s) = \lambda(t)e^{-\lambda(t)\cdot t_s}, \qquad (2)$$

where t_s is the time from the last trip (street hail) on the street s and t is the current time-slot.

The ROC (receiver operating characteristic) curves of our model and the NHP model for three selected streets are plotted in Figure 1. The ROC curve plots the true positive rate (Sensitivity) versus the false positive rate (1 - Specificity) for a predictor at different threshold settings. The true positive rate is plotted as the y-axis in Figure 1, while the false positive rate is plotted as the x-axis. The black colored curve in the graph shows the ROC curve for the regression model; while the dotted red curve is for the NHP model. The diagonal line ($x = y$) is the line of random guesses. Hence, the points above the diagonal line indicate a better prediction characteristic.

While the ROC plots help us to visually determine the performance of various prediction approaches, we need quantifiable measures to summarize the performance from thousands of streets across all time periods. For this purpose, we adopt the metric of Area Under the Curve (AUC). The AUC measures the area under the ROC curve, and the higher the value, the better the performance. As many streets do not have sufficient data to support our study, we focus on streets that account for close to 80% of street pickup demands (these streets are plotted in Figure 2). To give a visual overview of all the measured AUC values, we plot the AUC values for all highlighted streets in Figure 3, where the X-axis is for the street ID and the Y-axis is for the AUC value.

As can be observed from Figure 3, the predictive power of our regression-based model is consistently better than the NHP model as measured by the AUCs. Further, our model
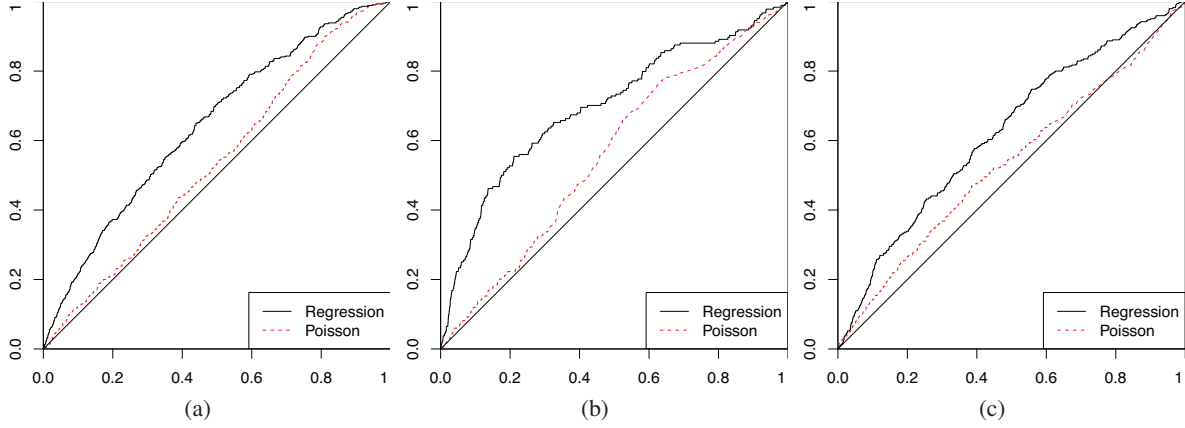
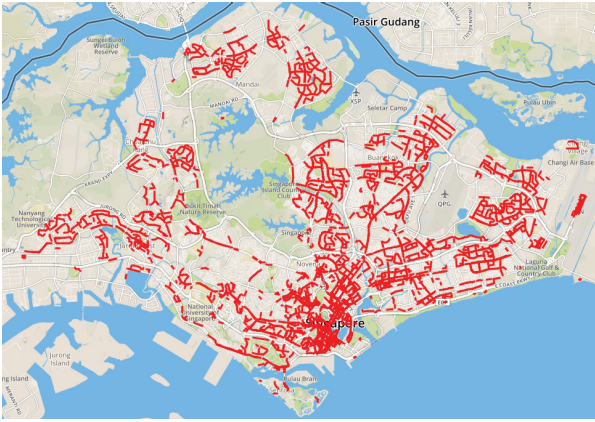Figure 1: Comparing ROC curves: our model versus non-homogeneous Poisson model.

Figure 2: Streets covering close to 80% of pick-up trips in Singapore every month.

Figure 3: Comparing AUCs: our model versus the NHP model. The street IDs are plotted in the x-axis while the AUC values are plotted in the y-axis.

uses the real-time information which makes it more adaptive (NHP uses only historical information).

The output of our model is the likelihood of demand occurrence, however, the required input from our recommendation engine is the *counts* of demand. To estimate the count of expected demand within a time-slot, we have to couple our demand prediction model with an arrival process of vacant taxis. In our prediction engine, we assume that vacant taxis follow a link-dependent and time-dependent Poisson process to arrive at a link in each time slot. The resulting count-based prediction for any link l in time period t is thus computed as:

$$D_l(t) = k_1^l \int_0^T e^{-\lambda_l^t \cdot u} Pr_l^t(HIRED|u)du + k_2^l, \quad (3)$$

where T is the length of a time period, λ_l^t is the Poisson arrival rate of vacant taxis at link l in the time period t, k_1^l and k_2^l are the calibration parameters specific to link l. k_1^l and k_2^l are introduced since we observe empirically that the predicted demand counts are linearly correlated to the observed demands. This relationship is link-dependent.
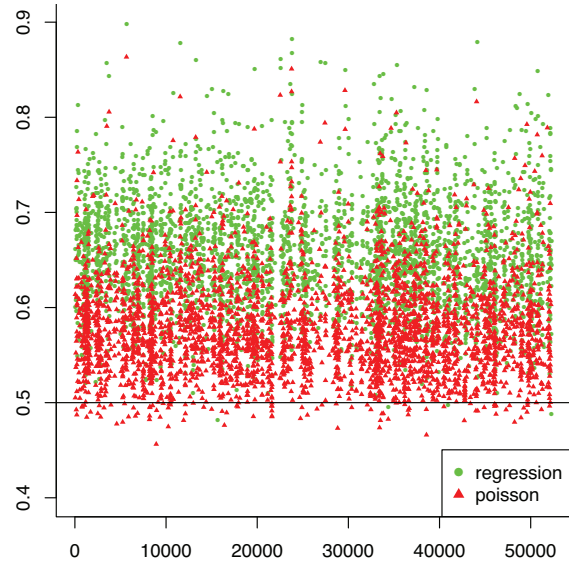
To demonstrate that our prediction engine performs well not just against the NHP model, we compare our engine against a sophisticated ensemble model proposed by Moreira-Matias et al. (2013). The ensemble model contains three forecasting techniques for demand prediction: the NHP model, the weighted NHP model and the Auto-Regressive integrated Moving Average (ARiMA) model. The prediction is generated by linearly combining the outputs from the three models weighted based on their relative errors against the observed demand in previous time-slots. Table 1 shows the Root Mean Square Error (RMSE) for the demand prediction calculated on November 1-3, 2016. As can be observed from the RMSE values in Table 1, our regression-based approach consistently outperforms the ensemble approach by a margin of 10%.

Table 1: The RMSE of our approach and the ensemble approach on selected days.

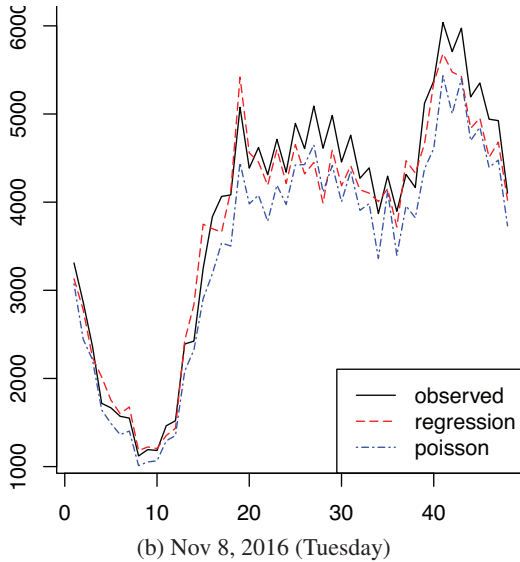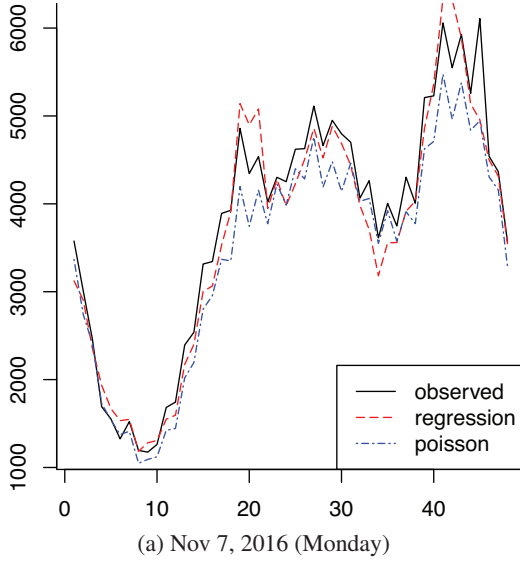| Day	Regression	Ensemble
01-Nov-2016	1.233811	1.326432
02-Nov-2016	1.168554	1.260058
03-Nov-2016	1.200508	1.342625

(a) Nov 7, 2016 (Monday)

(b) Nov 8, 2016 (Tuesday)

Figure 4: Comparison of the prediction performance over the course of two selected days for a link. The time period (30-minute long) is plotted in the x-axis and the demand count is plotted in the y-axis.

To understand the dynamics of the prediction engine, we also examine the performance of the prediction engine over the course of a day for selected links. Two sample plots of predicted demand counts over 48 time periods (each time period is 30 minutes long) for a particular link are shown in Figure 4. From the plots we can see that our approach performs particularly well against the NHP model during the morning and the evening peak hours.

Recommendation Engine

The objective of our recommendation engine is to balance the passenger demand and the taxi supply across the city over multiple time periods. In the previous section, the focus is on designing accurate demand prediction engine. The supply-side information, on the other hand, can be obtained relatively easily from the real-time taxi location data. The challenge that is left to be addressed is an optimization model that can maximize aggregated driver profit in face of a variety of uncertainties. To address this challenge, we adopt a multi-period, multi-driver stochastic recommendation model developed by Lowalekar, Varakantham, and Jaillet (2016). To discretize the decision domain, we divide the whole Singapore into 1km by 1km grid cells, and will use these in computing and conveying recommendations. The objective function defined by the recommendation engine is:

$$
\max \quad \Big(-\sum_{i \in \mathcal{Z}} \sum_{j' \in \mathcal{Z}} Cost^1_{ij'} \cdot u^1_{ij'} \Big)
$$

$$
+ \frac{1}{|\xi^D|} \sum_{k \le |\xi^D|} \sum_{t=1}^{Q} \sum_{i \in \mathcal{Z}} \Big(\sum_{j \in \xi^{D,k}_t} \mathcal{C}^t_{ij} \cdot x^{t,k}_{ij} \Big)
$$

$$
- \frac{1}{|\xi^D|} \sum_{k \le |\xi^D|} \sum_{t=2}^{Q} \sum_{i \in \mathcal{Z}} \Big(\sum_{j' \in \mathcal{Z}} Cost^t_{ij'} \cdot u^{t,k}_{ij'} \Big). \quad (4)
$$

This objective function maximizes the expected profit of all taxi drivers, and contains three components. The first component is the traveling cost resulting from guiding taxis to move without passenger. The second component is the net expected revenue from future demands. The third component is the expected movement cost for future time periods. The later two components are stochastic and it is non-trivial to express the expectations analytically. To make the optimization model tractable, a set of demand samples are chosen to approximate the expectations.

Formally speaking, $\mathcal{Z}$ denotes the set of regions, $\xi^{D,k}_t$ denotes a demand sample set, i.e., set of passenger requests in sample k at timestep t. Each element j of $\xi^{D,k}_t$ corresponds to the tuple $< o_j, d_j, R^{t,k}_j >$, where $R^{t,k}_j$ denotes the number of requests having origin in region o_j and destination in region d_j. $\mathcal{C}^t_{ij}$ denotes the net revenue obtained by assigning a taxi in region i to j^{th} element of $\xi^{D,k}_t$. $Cost_{ij'}$ denotes the cost incurred in moving empty from region i to region j'. The superscripts 1 and t refer to current and future timesteps respectively. Q denotes the planning horizon, i.e., the number of future timesteps for which demand samples are considered. The variable $x^{t,k}_{ij}$ denotes the number of region i taxis [1] assigned to j^{th} element of $\xi^{D,k}_t$ for sam-

---

[1]In case of street pickups, a taxi can be assigned to a request, if and only if, it is present in exactly the same region; but for booking requests, taxis can be assigned from nearby regions.

ple k at timestep t. $u^1_{ij'}$ denotes the number of taxis recommended to move empty from region i to region j' at current timestep. $u^{t,k}_{ij'}$ denotes the number of taxis recommended to move empty from region i to region j' at timestep $t, t > 1$, in sample k.

The recommendation model maximizes the objective defined in equation (4) subjects to the following constraints:

1. At any timestep, for any region, the number of outgoing taxis equals the number of available taxis in the region.

2. At any timestep, for any demand sample, between any pair of regions i and j, the number of requests served is less than the number of i-j requests.

The u^1_{ij} obtained from the solver can be used in deriving guidance to be send out to the drivers in the current timestep. In practice, we execute the multi-step optimization model continuously in order to incorporate latest information.
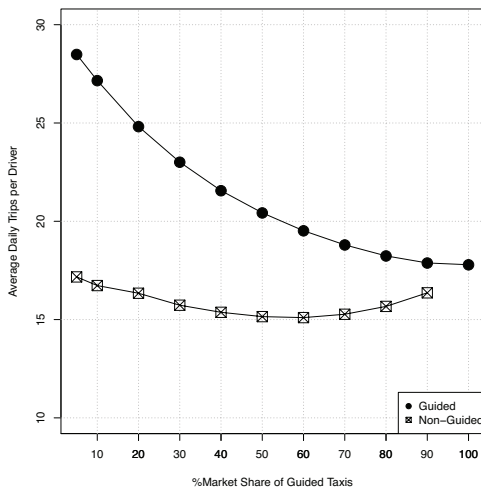

Figure 5: The performance of guided versus non-guided taxi drivers over the market share of the DGS users.

Evaluation

To evaluate the quality of recommendations generated for the taxi-drivers, we simulate the whole framework using an agent-based taxi fleet simulation platform (Cheng and Nguyen 2011) calibrated using the latest taxi driving dataset. To setup the simulation environment, we use 24,000 active taxis in the simulation, and randomly generates passenger demands using the demand profile from November 2, 2016 (a typical Wednesday). The demand profile used for generating the demands is depicted in Figure 7. To evaluate the performance of the DGS under all adoption ratios, we vary the DGS market share from 5% all the way up to 100%. During the simulation, we assume that all DGS-guided drivers will follow our recommendations while the rest non-DGS drivers will follow the empirical roaming patterns derived from the data.

The simulator is configured in a way that it generates streaming data exactly like in the real-world setting. Our recommendation engine is attached to the emulated stream data
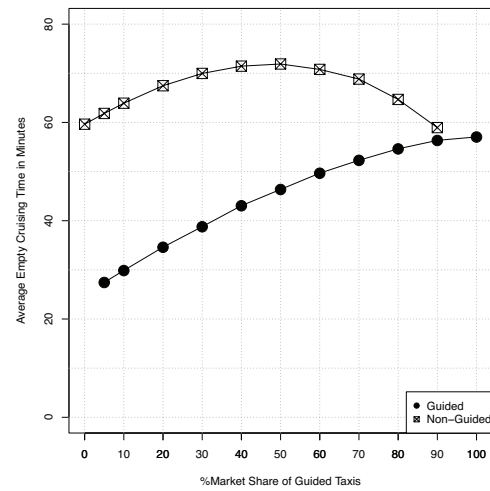

Figure 6: The empty cruising times of guided versus non-guided taxi drivers.
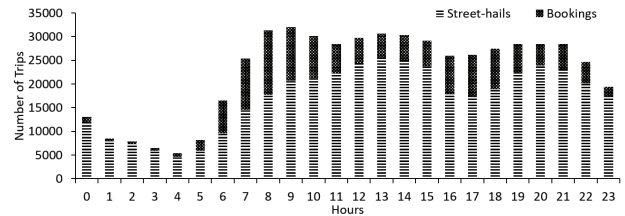

Figure 7: The demand profile with street-hails and booking demands used for the simulation.

and computes recommendations as if it is for real-world consumption. The recommendation engine uses 5 demand samples with 6 timesteps of 5-minute interval.

We measure two performance metrics: the average number of trips per driver and the mean empty cruising time in-between trips. Figure 5 shows the average number of daily trips per driver with market share of guided taxis (DGS taxis) varying from 5% to 100%. As can be seen in the graph in Figure 5, with very low market share of the DGS taxis (at 5%), the average daily trips per drivers is high (28) for the DGS taxis as compared to that of the non-DGS taxis, which is 17. Further, with increase in the market share of the DGS taxis, the average daily trips per driver comes down as more drivers receive guidance resulting in more competition. However, even at 100% market share of the DGS taxis, the average daily trips of DGS taxis is still higher than the highest achievable number by the non-DGS taxis.

Figure 6 shows the empty cruising time in-between trips. Consistent with the number of trips, we can see that guided drivers always have shorter empty cruising time when compared against non-guided drivers.

Delivering Recommendation to the Taxi Drivers

In order to deliver the recommendations to the taxi drivers for decision support, we have developed a mobile phone application (App) which displays the recommended locations

to the taxi drivers over the map of Singapore. The App has been developed for both the Android and the iOS platforms. The decision support framework receives the location of the taxi using the App and provides a preferred region for a taxi driver to find customers. The size of the region is 1km by 1km as stated earlier.

The recommendation engine automatically zooms in and out depending on the distance from the current location to the recommended region. If the taxi is already inside the recommended region, we also highlight the streets based on real-time likelihoods in demand generation. We have recruited the first batch of testing drivers and a 3-month long field trial is currently under way.

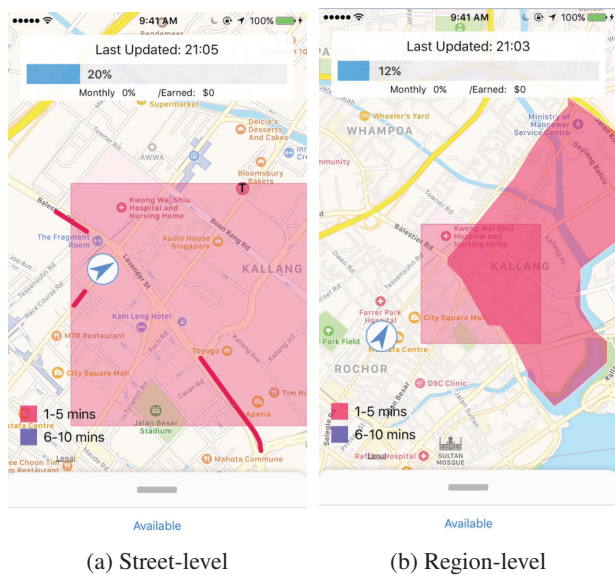The screenshots of the DGS App can be found in Figure 8.

(a) Street-level   (b) Region-level

Figure 8: Mobile App UI displaying different recommendation modes.

Conclusions

In this paper, we propose an efficient and scalable driver guidance system (DGS) for taxi drivers. By working with our collaborator in Singapore, we have obtained both a historical dataset with more than 2 years of taxi operations and a real-time API that allows us to receive status and location updates from all taxis in Singapore continuously. The creation of the DGS is the combined efforts of stream data processing, high-quality demand prediction, and scalable recommendation generation. The outcome, DGS, is the first of its kind, and have demonstrated great potential in a highly realistic simulation environment.

We have entered the initial field trial phase and we expect the trial to continue to scale to thousands of drivers. If the field trial progresses well, it would point out a way for traditional taxi operators to survive the pressure brought in by the disruptors such as Uber and Uber-like services.

References

Cheng, S.-F., and Nguyen, T. D. 2011. TaxiSim: A multiagent simulation platform for evaluating taxi fleet operations. In *2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 14–21.

Ding, Y.; Liu, S.; Pu, J.; and Ni, L. M. 2013. HUNTS: A trajectory recommendation system for effective and efficient hunting of taxi passengers. In *14th IEEE International Conference on Mobile Data Management*, 107–116.

Ge, Y.; Xiong, H.; Tuzhilin, A.; Xiao, K.; Gruteser, M.; and Pazzani, M. 2010. An energy-efficient mobile recommender system. In *16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 899–908.

Lowalekar, M.; Varakantham, P.; and Jaillet, P. 2016. Online spatio-temporal matching in stochastic and dynamic domains. In *30th AAAI Conference on Artificial Intelligence*, 3271–3277.

Ma, S.; Zheng, Y.; and Wolfson, O. 2015. Real-time city-scale taxi ridesharing. *IEEE Transactions on Knowledge and Data Engineering* 27(7):1782–1795.

Moreira-Matias, L.; Gama, J.; Ferreira, M.; and Damas, L. 2012. A predictive model for the passenger demand on a taxi network. In *15th International IEEE Conference on Intelligent Transportation Systems*, 1014–1019.

Moreira-Matias, L.; Gama, J.; Ferreira, M.; Mendes-Moreira, J.; and Damas, L. 2013. Predicting taxi–passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems* 14(3):1393–1402.

Newson, P., and Krumm, J. 2009. Hidden Markov map matching through noise and sparseness. In *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 336–343.

Qu, M.; Zhu, H.; Liu, J.; Liu, G.; and Xiong, H. 2014. A cost-effective recommender system for taxi drivers. In *Proceedings of the 20th ACM SIGKDD International conference on Knowledge discovery and Data mining*, 45–54.

Yuan, N. J.; Zheng, Y.; Zhang, L.; and Xie, X. 2013. T-finder: A recommender system for finding passengers and vacant taxis. *IEEE Transactions on Knowledge and Data Engineering* 25(10):2390–2403.

Zhang, Y.; Li, B.; Krishnan, R.; and Liu, S. 2015. Learning from the offline trace: A case study of the taxi industry. In *36th International Conference on Information Systems*.

Zhang, K.; Feng, Z.; Chen, S.; Huang, K.; and Wang, G. 2016. A framework for passengers demand prediction and recommendation. In *IEEE International Conference on Services Computing*, 340–347.