A data-driven dynamic demand hotspots forecasting framework for on-demand meal delivery platforms

Jingyi Cheng*[1] and Shadi Sharif Azadeh[2]

[1]Transport & Planning, Delft University of Technology, the Netherlands

Short summary

Speed and reliability are the keys to high quality on-demand meal delivery service. The rebalancing of couriers locations according to future demand remains an operational challenge in the industry. This study proposes an adaptive framework to identify and predict the near-future demand hotspots, utilizing the semi real-time predictive information as input. This framework provides demand insights to assist meal delivery platforms in making operational forward-looking resource-demand rebalancing decisions in real time, such as fleet management and demand management. To generate fast and accurate demand forecasting, we incorporate time series features and data-driven machine learning methods to create an adaptive forecasting approach. We create a dynamic demand hotspot clustering algorithm which takes predictive and geographic information as input. In the case study, our predictive forecasting model outperforms the time series and deep learning benchmarks in deterministic forecasting. The hotspots clustering performance is improved by using probabilistic predictive input.
Keywords: short-term demand forecasting; on-demand meal delivery; data-driven methods

1 Introduction

The on-demand meal delivery business has created a global market valued at over 150 billion dollars. However, competition among on-demand meal delivery service platforms (OMDPs), such as DoorDash, Uber, and Grubhub, is intense. Good delivery service quality is essential for maintaining high customer satisfaction, which improves customer loyalty in the long run. Delivery speed and reliability are the key factors in determining service quality. To generate higher profits while providing high-quality service, the platform should operate its limited courier resource efficiently. Various studies have been conducted to enhance the delivery efficiency of OMDPs. Reyes et al. (2018) and Yildiz & Savelsbergh (2019) study the meal delivery routing problem (MDRP) to minimize the travel time. An anticipatory customer assignment strategy is proposed by Ulmer et al. (2021) to study the stochastic dynamic pickup and delivery problem.

To operate efficiently, on-demand meal delivery service platforms should also act proactively in responding to the dynamics of demand in the city. With a limited number of hired couriers, platforms can strategically prioritize assigning couriers to areas with higher predicted demand in the near future. By doing so, more couriers will be available in high-demand areas, reducing the average waiting time for order pickups. To support real-time operational decisions, accurate and fast-to-generate short-term demand predictions of the city are necessary. However, to the best of our knowledge, little attention has been paid to short-term demand forecasting problem of OMDPs. Hess et al. (2021) compare the performance of mainstream forecasting methods in generating deterministic demand predictions per hour. In their case study, an exponential smoothing based model achieves the best performance when a rich amount of historical data is available, while random forecast regression outperforms the rest when the training data is limited.

This study has two main objectives. Firstly, we combine the benefits of both parametric time series and data-driven methods to create a fast-to-generate forecasting approach for the demands arising from different parts of the city in the next 15-minute. Secondly, we propose an adaptive framework to identify and predict the next demand hotspots, utilizing the semi real-time predictive informtion as inputs. Addressing these two objectives, we propose the dynamic demand hotspots forecasting framework, which can be used by the platforms to proactively optimize their fleet and demand management decisions in real-time.

2 Methodology

In this study, we propose the novel dynamic demand hotspots forecasting framework to assist the rebalancing decision-making of on-demand meal delivery platforms by predicting the areas with higher demand in the next 15-minute interval. Our framework consists of two major steps. Firstly, we generate demand predictions for different areas of the city. Then, we generate demand hotspot clustering of the city utilizing the predictive outcomes and the geographic information of these areas.

Compared to the forecasting for a longer time interval, short-term demand predictions are more volatile and prone to fluctuations caused by latent events. Parametric time series models are good at interpreting recent signals for forecasting, but they require manual effort for selecting suitable parameters during training. On the other hand, non-parametric data-driven machine learning models are able to capture the non-linear patterns in data and interpret the interactions between features. To combine the advantages of both sides, we propose an adaptive forecasting approach that recurrently applies lagged-dependent features to generate predictions, based on data-driven machine learning models. Inspired by autoregression analysis, the lagged-dependent features are the previous values of demand, e.g. y^{α}_{t-i} meaning the demand of area α from i time steps ago. In our model, each time step is a 15-minute interval. To distinguish the regular hourly seasonality and the unexpected temporal fluctuations of demand, we choose to include four lagged-dependent features y^{α}_{t-1}, y^{α}_{t-2}, y^{α}_{t-3}, y^{α}_{t-4}, covering the demand information from the recent one hour. And we adopt random forest regression (RF) and eXtreme Gradient Boosting (XGBoost) as the baseline machine learning models. Their adaptive versions we create are called LD-RF and LD-XGBoost respectively.

Although deterministic demand forecasting is the focus of many demand forecasting literature, the determination of future hotspots might be benefit from probabilistic predictive information, which provides distributional insight of demand. The probabilistic predictive information feature $\hat{Y}_t$ we adopt in the dynamic demand hotspot clustering framework is a weighted average of the 25%, 50% and 75% quantile prediction vectors, $\hat{Y}_i = 0.25 \cdot \hat{Y}_{i,0.25} + 0.5 \cdot \hat{Y}_{i,0.50} + 0.25 \cdot \hat{Y}_{i,0.75}$. The quantile prediction vectors are provided by quantile regression forest (QRF) by Meinshausen & Ridgeway (2006) or its adaptive version LD-QRF in this study.

Besides of the predictions, we also include the geographic features latitude and longitude of the area centers as input to the clustering algorithm. Constrained K-means clustering proposed by Bradley et al. (2000) is utilized to generate clusters of areas with a condition of no less than three zones per cluster, where the optimal number of clusters is chosen between 3 to 6 based on the highest mean silhouette coefficient.

3 Results and discussion

Data analysis and feature engineering

Our case study uses the meal order placement data of a city, which is collected from April 1^{st}, 2020 to September 14^{th}, 2020. Each instance contains the placement time, pickup and delivery destination locations. The locations are hashed into hexagonal zones using Uber's H3 geospatial indexing system at a resolution level of 8. The average area per zone is 0.737 km^2. We denote the location of the restaurant as the pickup zone, and the location of delivery destination as the destination zone. Our data covers 20 pickup zones and 50 destination zones.

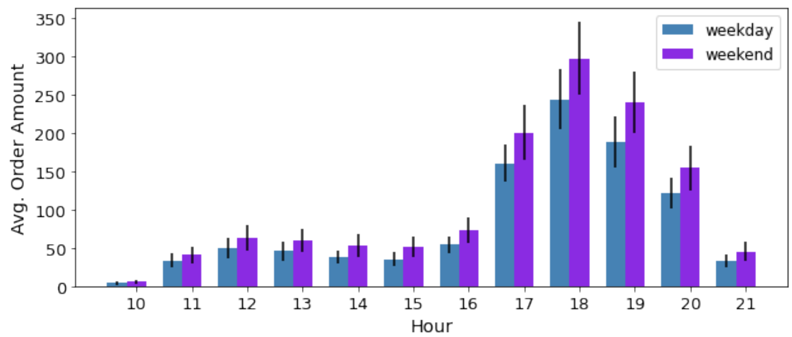


Figure 1: The average total number of received orders per hour for weekdays (i.e. Monday to Thursday) and weekends (i.e. Friday to Sunday), the error bars represent standard deviations.

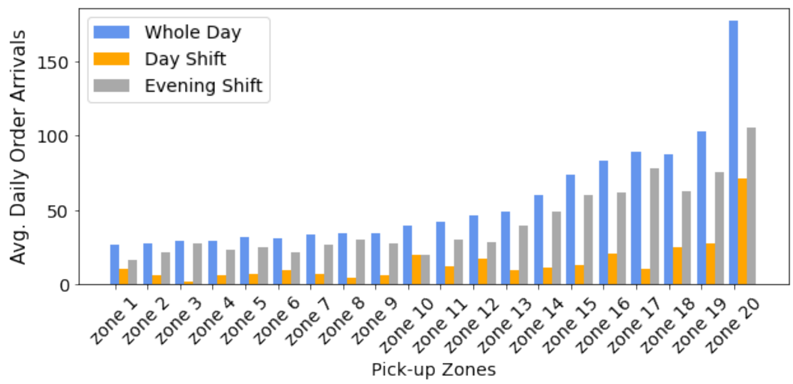


Figure 2: The daily average total number of received orders for each pickup zone at *whole day*, *day shift*, and *evening shift* respectively.

According to our data analysis, all days of the week exhibit a dual-peak demand pattern, which shows the demand level is much higher around dinner time compared to lunch time. Also, Friday, Saturday and Sunday attract more orders than other days in general. To distinguish this demand difference we visualise the average number of orders received per hour during the week and on weekends in Figure 1. We consider Friday to Sunday together as weekend days since the demand on Friday evening is much closer to the patterns of weekend days. A higher demand level is showed among the weekend group than the weekday group. Inspired by the within-day demand pattern, we can further split a day into the *day shift* (10:30-17:00) and the *evening shift* (17:00-21:30). To explore the demand level difference among pickup zones, we compare their daily average number of orders received in Figure 2. We observe a variation in the relative demand levels between the day and the evening shifts, indicating a change of comparative demand levels of the city throughout a day.

In this case study, we are interested in predicting the near future demands in the city, i.e. the aggregated amount of orders coming in the next 15 minutes from all the restaurants in each pickup zone. The per 15-minute demands are the aggregated number orders for each pickup zone according to the 44 different 15-minute time windows each day. We denote this target variable *order demand* of pickup zone α at the i^{th} time interval as y_i^{α}. To capture the within-day and within-week double-seasonal demand patterns in data, we include the temporal features hour and day of a week. A binary variable is adopted to indicate whether the date is a national holiday. Considering that weather also affects the interests of customers in online meal ordering, we include three exogenous weather features, namely the average temperature, precipitation amount and wind speed. The weather features are solely time-dependent, implying that the values measured at the same time are consistent across different zones.

Experimental design, evaluation metrics and baseline models

For model training, we use the first 21 weeks' data (6468 samples) from the processed dataset. And the predictions are generated for the last 1 week's data (308 samples) in a one-step-ahead

rolling window fashion. The predictions of each pickup zone are generated in parallel at each step by the corresponding predictors.
The first forecasting experiment involves performing deterministic demand forecasting for the next 15-minute. To gain an overview of model performance, we report the accuracy of each subset of data by properties 'types of shift' and 'partition of week'. In addition to reporting the mean prediction errors averaged from predictions for all the pickup zones, we are also interested in comparing the individual prediction errors measured per pickup zone to inspect whether certain type of pickup zones are consistently better predicted by a kind of predictor. We apply the root-mean-squared error (RMSE) and mean absolute error (MAE) as the evaluation metrics for deterministic forecasting. To investigate the gains in deterministic forecasting using the data-driven lagged-dependent models, namely lagged-dependent regression forest (LD-RF) and lagged-dependent XGBoost (LD-XGBoost), we include the vanilla regression forest (RF) and XGBoost models as the baselines. Moreover, we also include Trigonometric, Box–Cox, Auto-Regressive-Moving-Average, Trend, and Seasonality (TBATS) model proposed by De Livera et al. (2011) as the conventional time series benchmark for its superiority in capturing complex seasonal patterns and convenience in automatic hyperparameters tuning function. Additionally, we include a two-layer Long short-term memory networks (LSTMs) as the deep learning benchmark, given its similarity in recurrent feature processing to our lagged-dependent methods.
The second forecasting experiment concerns identifying the demand hotspots of the city in the next 15-minute interval using the adaptive clustering framework we propose. The actual and predicted within-cluster median demand values are the medians taken from the clusters generated by actual demand and predictions inputs respectively. The deviation between the actual and predicted within-cluster medians serves as an indicator of predictive clustering performance. Again, RMSE and MAE are applied as error measurements.

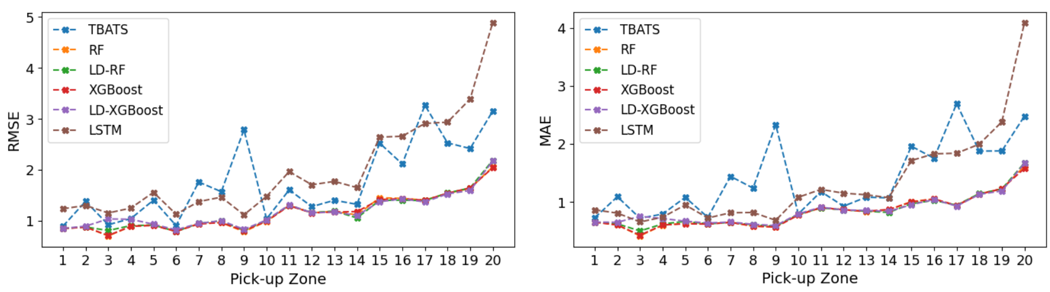
Results of Case Study

The deterministic prediction accuracy of different models are reported in Table 1, measured by the averaged RMSE and MAE. The performance of the vanilla data-driven machine learning models well as their lagged dependent versions are rather close to each other, although XGBoost consistently predicts most accurately for the forecasting of day shift. And all data-driven models outperform the benchmarks TBATS and LSTMs by some margin in terms of prediction errors.
To confirm the statistical significance of prediction errors between models, we perform Diebold-Mariano tests for all pairs of models. Only the forecasting performance difference between RF and XGBoost is tested to be insignificant.

Table 1: The one-week point forecast results of various models trained with 21 weeks' data, measured by MAE and RMSE.

PERIOD	MODEL	WHOLE WEEK		WEEKDAY		WEEKEND	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
Whole Day	**TBATS**	1.762	1.395	1.653	1.373	1.851	1.438
	LSTM	1.975	1.330	1.506	0.961	2.190	1.515
	RF	**1.145**	0.830	1.110	0.851	**1.229**	**0.895**
	LD-RF	1.155	0.838	**1.092**	**0.835**	1.237	0.906
	XGBoost	**1.145**	**0.829**	1.103	0.843	1.235	0.902
	LD-XGBoost	1.174	0.859	1.106	0.849	1.255	0.934
Day Shift	**TBATS**	1.514	1.272	1.613	1.424	1.500	1.257
	LSTM	1.029	0.698	0.551	0.370	1.169	0.815
	RF	0.770	**0.582**	0.664	0.566	0.847	0.655
	LD-RF	0.780	0.591	0.656	0.556	0.865	0.669
	XGBoost	**0.769**	**0.582**	**0.654**	**0.555**	**0.847**	**0.658**
	LD-XGBoost	0.796	0.611	0.680	0.575	0.885	0.697
Evening Shift	**TBATS**	2.040	1.572	1.675	1.336	2.205	1.700
	LSTM	2.799	2.244	2.163	1.719	3.087	2.527
	RF	**1.514**	**1.185**	1.489	1.238	1.607	**1.242**
	LD-RF	1.523	1.197	**1.466**	**1.218**	**1.604**	1.249
	XGBoost	1.515	**1.185**	1.484	1.234	1.619	1.256
	LD-XGBoost	1.549	1.216	1.478	1.220	1.627	1.275

Figure 3 shows the average RMSE and MAE per pickup zone obtained from different forecasting methods. As we analyzed in Figure 2, the average number of orders received per day increase as we move from pickup zone 1 to zone 20. The rise of demand level seems to be positively correlated to the average forecasting errors. And the forecasting performance of benchmarks TBATS and LSTMs are much poorer compared to that of the tree-based methods when applied to the high demand pickup zones.



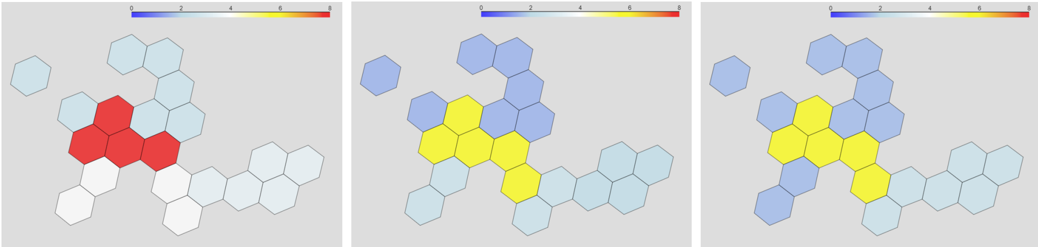
(a) RMSE (b) MAE

Figure 3: The RMSE and MAE of different models measured for each pickup zone at the general (i.e. whole day and whole week) level.

From the deterministic forecasting experiment, RF, LD-RF, XGBoost and LD-XGBoost have shown good performance. Therefore, we continue to apply these four predictors and utilize their deterministic predictions as part of the input for dynamic demand hotspots clustering. Additionally, we include the quantile forecasting approaches QRF and LD-QRF. The vanilla data-driven models are treated as the benchmarks here to evaluate the benefit of using lagged-dependent models as well as applying probabilistic predictive information as input.

Figure 4 shows an example visualization of demand hotspots clustering, where the grids are colored based on the within-cluster average actual/predicted demands. In this example, adaptive predictors

LD-RF and LD-QRF manage to identify the central hotspot, although predicted within-cluster demands are all less than actual.



(a) Actual **(b)** LD-RF **(c)** LD-QRF

Figure 4: Visualization of clustered hotspots for 19:00-19:15 on Monday, September 14^{th} using (a) actual demands, (b) predicted deterministic demands by LD-RF, and (c) predicted quantile demands by LD-QRF respectively.

The RMSE and MAE evaluations of the real and predicted within-cluster median difference are presented in Table 2. On the whole day level, the best RMSE and MAE are obtained by the adaptive predictors using lagged-dependent features, namely the LD-RF, LD-XGBoost and LD-QRF. Also the accuracy measured by MAE suggests the probabilistic predictors to perform better. It means using probabilistic predictive information generate less errors on average, although it is punished by some occasional larger errors.

Table 2: The RMSE and MAE of one-week's predicted within-cluster median demand by dynamic clustering method using deterministic and probabilistic predictions from models trained by 21-week data.

Period	Model	Whole Week		Weekday		Weekend	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
Whole Day	**RF**	0.791	0.810	0.729	0.770	0.867	0.863
	XGBoost	0.795	0.808	0.730	0.762	0.874	0.869
	LD-RF	**0.784**	0.794	**0.721**	0.761	0.861	0.838
	LD-XGBoost	0.788	0.801	0.730	0.759	**0.859**	0.857
	QRF	0.831	0.716	0.781	0.660	0.893	0.791
	LD-QRF	0.820	**0.699**	0.767	**0.648**	0.886	**0.766**
Day Shift	**RF**	**0.657**	0.601	**0.647**	0.551	0.671	0.668
	XGBoost	0.660	0.599	0.648	0.543	0.677	0.673
	LD-RF	0.660	0.601	0.652	0.552	**0.670**	0.667
	LD-XGBoost	0.666	0.614	0.659	0.554	0.675	0.694
	QRF	0.718	**0.483**	0.729	**0.412**	0.713	0.577
	LD-QRF	0.712	0.484	0.727	0.422	0.691	**0.568**
Evening Shift	**RF**	0.952	1.112	0.834	1.086	1.090	1.145
	XGBoost	0.956	1.109	0.834	1.078	1.099	1.151
	LD-RF	**0.935**	1.072	**0.811**	1.063	1.080	1.084
	LD-XGBoost	0.936	1.071	0.822	1.055	**1.070**	1.092
	QRF	0.971	1.054	0.851	1.020	1.112	1.099
	LD-QRF	0.954	**1.008**	0.820	**0.975**	1.109	**1.052**

Again, we perform Diebold-Mariano tests on the within-cluster median predicting outcomes for each pair of models covered in the second experiment. Results show that the forecasting performance between RF and XGBoost, and between XGBoost and LD-XGBoost are not significantly distinguishable from one another. The statistical test is significant for other pairs.

4 Conclusions

This study proposes a dynamic demand hotspots forecasting framework that is able to assist operational decision making for on-demand meal delivery platforms. Through a case study, we show that the performance is improved by using the adaptive data-driven forecasting methods we propose, and the probabilistic predictive feature we create from the quantile demand predictions as input. A further study could focus on generating real-time resource-demand rebalancing decisions by incorporating the predicted insight provided by our framework.

References

Bradley, P. S., Bennett, K. P., & Demiriz, A. (2000). Constrained k-means clustering. *Microsoft Research, Redmond*, *20*(0), 0.

De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*(496), 1513–1527.

Hess, A., Spinler, S., & Winkenbach, M. (2021). Real-time demand forecasting for an urban delivery platform. *Transportation Research Part E: Logistics and Transportation Review*, *145*, 102147.

Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*(6).

Reyes, D., Erera, A., Savelsbergh, M., Sahasrabudhe, S., & O'Neil, R. (2018). The meal delivery routing problem. *Optimization Online*.

Ulmer, M. W., Thomas, B. W., Campbell, A. M., & Woyak, N. (2021). The restaurant meal delivery problem: Dynamic pickup and delivery with deadlines and random ready times. *Transportation Science*, *55*(1), 75–100.

Yildiz, B., & Savelsbergh, M. (2019). Provably high-quality solutions for the meal delivery routing problem. *Transportation Science*, *53*(5), 1372–1388.