A stepwise interpretable machine learning framework using linear regression (LR) and long short-term memory (LSTM): City-wide demand-side prediction of yellow taxi and for-hire vehicle (FHV) service

Taehooie Kim (Corresponding Author)
Arizona State University, School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Email: taehooie.kim@asu.edu

Shivam Sharda
Arizona State University, School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Email: ssharda@asu.edu

Xuesong Zhou
Arizona State University, School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Email: xzhou74@asu.edu

Ram M. Pendyala
Arizona State University, School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85287-3005
Email: ram.pendyala@asu.edu

Abstract

As app-based ride-hailing services have been widely adopted within existing traditional taxi markets, researchers have been devoted to understand the important factors that influence the demand of the new mobility. Econometric models (EMs) are mainly utilized to interpret the significant factors of the demand, and deep neural networks (DNNs) have been recently used to improve the forecasting performance by capturing complex patterns in the large datasets. However, to mitigate possible (induced) traffic congestion and balance utilization rates for the current taxi drivers, an effective strategy of proactively managing a quota system for both emerging services and regular taxis is still critically needed. This paper aims to systematically design an explainable deep learning model capable of assessing the quota system balancing the demand volumes between two modes. A two-stage interpretable machine learning modeling framework was developed by a linear regression (LR) model, coupled with a neural network layered by long short-term memory (LSTM). The first stage investigates the correlation between the existing taxis and on-demand ride-hailing services while controlling for other explanatory variables. The second stage fulfills the long short-term memory (LSTM) network structure, capturing the residuals from the first estimation stage in order to enhance the forecasting performance. The proposed stepwise modeling approach (LR-LSTM) forecasts the demand of taxi rides, and it is implemented in the application of pick-up demand prediction using New York City (NYC) taxi data. The experiment result indicates that the integrated model can capture the inter-relationships between existing taxis and ride-hailing services as well as identify the influence of additional factors, namely, the day of the week, weather, and holidays. Overall, this modeling approach can be applied to construct an effective active demand management (ADM) for the short-term period as well as a quota control strategy between on-demand ride-hailing services and traditional taxis.

Keywords: On-demand ride hailing; taxi market; short-term demand forecasting; long short-term memory (LSTM); deep neural network (DNN); active demand management (ADM)

1. Introduction

The emergence of on-demand ride hailing services, such as Uber and Lyft, requires transportation planners to better design comprehensive transportation mobility solutions for informed transportation management and policy making. Uber reported a growth in nationwide market share from 9% to 29% whereas existing taxi services observed a decline in their market from 52% to 35% nationwide (Fischer, 2015). In response to the influx of on-demand services, for instance, New York City Rules announced a protective policy for yellow taxi and green taxi drivers in 2018, and the NYC Department of Transportation (DOT) imposed a congestion pricing, "a new electronic toll in place for drivers entering the busiest stretches of Manhattan" (Mckinley and Hu, 2019). In addition, Sanders and Guse (2019) reported that NYC imposed the new regulations to app-based companies and to protect hardworking drivers in Manhattan: at least 69% of the operating cars in Manhattan below 96th St. must serve a passenger, and each company needs to set up a fare rate based on the frequency of carrying passengers.

As a result, the exponential growth of this mobility service and its impact on the transportation network lead to intensive research initiatives. In order to interpret the significant factors affecting the ride-hailing demand, diverse statistical modeling approaches were developed and introduced (Lavieri et al., 2017; Lam and Liu, 2018; Gerte et al., 2018). Forecasting approaches were also conducted to optimize the traffic network, dispersing the fleet of the ride-hailing service or yellow taxis within the network (Laptev et al., 2017; Chen et al., 2017; Zhao et al., 2019). Nevertheless, there are still challenges that need further exploration:

(1) Since the numbers of FHV and yellow taxi drivers are closely correlated, if a cap is imposed for a transportation network company (TNC), what is the expected additional demand to be served by, or more precisely, to be shifted to, yellow taxi drivers?
(2) Does the city traffic management agency need to apply a quota regulation during a rainy day or a snowy day?

A concept of the active demand management (ADM) strategy (FHWA, 2012) with respect to on-demand transit has been incorporated to address similar challenges: proactively optimizing the usage of both taxies and on-demand mobility city-wide (e.g., imposing a day-dependent cap for one type of vehicle (Kamga et al., 2015)). In other words, we aim to understand the dynamics of the competing ride hailing services to mitigate the traffic congestion and maintain the balance of the coexisting system for drivers by analyzing the correlation of both emerging and existing mobility services as well as identifying their own explanatory variables under different situations such as day of the week, holiday, and weather.

The mobility data obtained from NYC is employed as a demonstration use case. A hybrid model developed integrates both multivariate linear regression (LR) and long-short term memory (LSTM), a type of recurrent neural network (RNN). The analytical model in the first layer represents available time series data through a number of interpretable parameters. In the second layer LSTM focuses on designing a data-driven approach to account for the hidden pattern from the first layer's residuals. Combining the two-step process, LR-LSTM predicts taxi demand.

The remainder of this paper is organized as follows. Section 2 presents the literature review examining prior research that explores statistical models and neural network structures. Section 3 describes a real-world data set, namely, New York City (NYC) Taxi Records. In section 4, a stepwise framework is proposed to tackle modeling challenges. The application of the integrated (or coupled) model is demonstrated using the real-world dataset, in section 5. Lastly, section 6 provides conclusion and discusses the estimation results and future work.

2. Literature review and motivation

This review section covers two aspects: (1) statistical approaches to understand demand for ride-hailing services and (2) recent modeling means using statistical modeling techniques as well as data-driven approaches. This is followed by a discussion of the motivations behind our proposed integrated approach.

2.1. Statistical models for demand analysis of ride-hailing services

Researchers used statistical models to calibrate unknown parameters and further interpret the factors influencing the taxi demand within the service area. In particular, Yang and Gonzales (2014) applied multivariate linear regression to estimate taxi trips by identifying significant explanatory variables, to name a few, population, age, education, income, transit accessibility time, and employment. Safikhani et al. (2017) designed the generalized spatial-temporal autoregressive (STAR) to understand spatial and temporal variations in taxi demand in New York City. Lavieri et al. (2017) utilized a multivariate count approach to study the number of trips generated in a specific zone and observed

that different income groups preferred to use the services for different activity purposes. Recently, Lam and Liu (2018) used the discrete choice model to analyze the correlation between dynamic pricing and waiting time in densely populated areas of New York. Gerte et al. (2018) examined the demand for the ride hailing service using a panel based random effects model in order to capture both heteroscedasticity and autocorrelation effects. This study denoted that the highly educated young male group tended to use the ride hailing service frequently.

Overall, the above-mentioned statistical modeling approaches have demonstrated their strength in explaining and predicting the ride-hailing service demand which allows planners to identify significant parameters for informed decision making. Regarding the use of other big data sources, social media data sets have been utilized in a variety of applications to capture individual activity patterns (Gu et al., 2018; Zhang et al., 2017; Hasnat and Hasan, 2018) and detect traffic incidents (Wang et al.,2016; Kuflik et al., 2017). On the other hand, using app-based data (i.e., multi-sourced data with high variances in both location accuracy and time of travel), He and Shen (2015) and Wang et al. (2019) have proposed conceptual frameworks to estimate the impact of the disruptive mobility services on taxi markets.

Nevertheless, the major challenge of statistical approaches is a lack of predictive accuracy, particularly under a complex data environment with different data sources (Altman et al., 1994; Sarle 1994; Paliwal and Kumar, 2009; Karlaftis and Vlahogianni, 2011). A number of case studies (Kumar et al., 2015; Al-Maqaleh et al., 2016; Golshani et al., 2018; Cui et al., 2018) demonstrated improved prediction accuracy by utilizing the neural network structures compared to statistical models.

2.2. Hybrid approaches combining statistical models with deep neural networks (DNNs)

We first focus on papers in the area of causal inference and prediction, as well as two popular DNN modeling tools using convolutional neural network (CNN) and long short-term memory (LSTM). CNN has demonstrated excellent capabilities in the visual data processing where its structure extracts the image features and classifies the images according to the extracted features (LeCun and Bengio, 1995). On the other hand, LSTM specializes in sequential data processing, and its structure stores significant information and forecasts the sequential data (Hochreiter and Schmidhuber, 1997).

Recently, many studies started to integrate more statistically-oriented modeling features in a deep learning framework. For example, Park et al. (2016) applied Bayesian neural network (BNN) to examine the uncertainty of dependent variables and further interpreted the unknown coefficients in traffic demand prediction models. They used an extracted decision tree approach to provide significant explanatory variables based on a pedagogical rule extraction algorithm. A study by Ke et al. (2017) applied the random forest framework to select the exogenous variables, ranking these variables' significance. Additionally, they examined the image intensity from map sequences of travel time rates using CNN and LSTM tools. Along this line, Laptev et al. (2017) utilized Bootstrap and Bayesian techniques with LSTM structure to detect abnormality of data and capture irregular patterns.

In the traffic demand estimation area, a recent study by Chen et al. (2017) employed the ensemble learning approach with the RelifF algorithm to identify the important factors. As a result, they were able to explore ride-splitting behavior of on-demand ride services in real-world data sourced from DiDi. More recently, Zhao et al. (2019) examined heterogeneity in mode-switching behavior by proposing a more systematically defined interpretable machine learning approach. To classify the mode-switching behavior, the model-agnostic interpretation tools were used to study the insights on the switching behaviors.

Overall, we observe two emerging research directions in the field of traffic demand modeling. First, researchers have devoted major efforts to develop hybrid modeling frameworks based on statistical and DNN approaches in order to improve the predictive performance. Second, many research teams have focused on developing interpretable models implemented by feature importance and data extraction algorithms to systematically explore the underlying traffic demand behavior and improve the lack of transparency observed in a nested non-linear structure (Samek et al., 2017; Gunning, 2017). Readers interested in the interpretable machine learning techniques can find more details in prior work (Ribeiro et al., 2016; Lipton, 2018; Molnar, 2020). However, the following challenges still need to be addressed: how to analyze the correlation of both emerging and existing mobility services as well as identify each mode pattern with demand-side factors (i.e., weather, holidays, and day of the week). In order to tackle the challenges, we propose a conceptual interpretable structural decomposition approach as follows:

$$\textit{True demand} = \textit{Linear explanatory components} + \textit{Nonlinear pattern} + \textit{Random fluctuations}$$

To calibrate the different components in this framework, a two-step process is adopted. In the first step the multivariate linear regression (LR) captures the primary linear components, and the estimated coefficients offer

additional insights on each mobility pattern. The second step utilizes the forecasting capabilities of LSTM network to account for the residuals from the first step.

3. Data description

Multiple data sources are utilized in this study: New York City Taxi & Limousine Commission (NYC-TLC) dataset containing daily mobility demand patterns for the two modes (i.e., For-Hire Vehicles (FHVs) and yellow taxi), weekday/weekend, holidays, and weather information available from National Climate Data Center (NCDC).

3.1. New York City (NYC) taxi trip records

The input dataset, NYC-TLC, covers both FHV and yellow taxi trip records collected from 2015 to 2017. In the yellow taxi records the available information includes the pick-up/drop-off events on weekday/weekend, the trip fare, payment type, location ID, and the number of passengers. In the FHV records only the pick-up/drop-off events and location ID are collected. To set up a fair comparison, we explored the data with an overlapping time period for pick-up/drop-off events. The number of pick-up events is used to investigate the dynamic trip demand with respect to FHV and yellow taxi. The study was done in Murray Hill, Manhattan located in Manhattan below 96th Street, and it is one of the highest taxi pick-up locations: the annual average pick-up demand is approximately 5.7 million, and the total trips made are 17 million over the period of three years.

Fig. 1 illustrates the volume of pick-ups and trends between yellow taxi and FHV over three years in Murray Hill, Manhattan. The land use of the study area shows a combination of residential and commercial districts. To be specific, Fig. 1(a) explains the proportion of the annual pick-up volume between the two modes from 2015 to 2017. In addition, Fig. 1(b) displays the weekly distribution of the pick-up volume between yellow taxi and FHV, where the overall demand gradually increases at the beginning of the week and tapers off as the weekend approaches. To further ensure the length of the data between two modes, we aggregate the event time series data in terms of the daily volume, with the day of the week as a binary indicator.

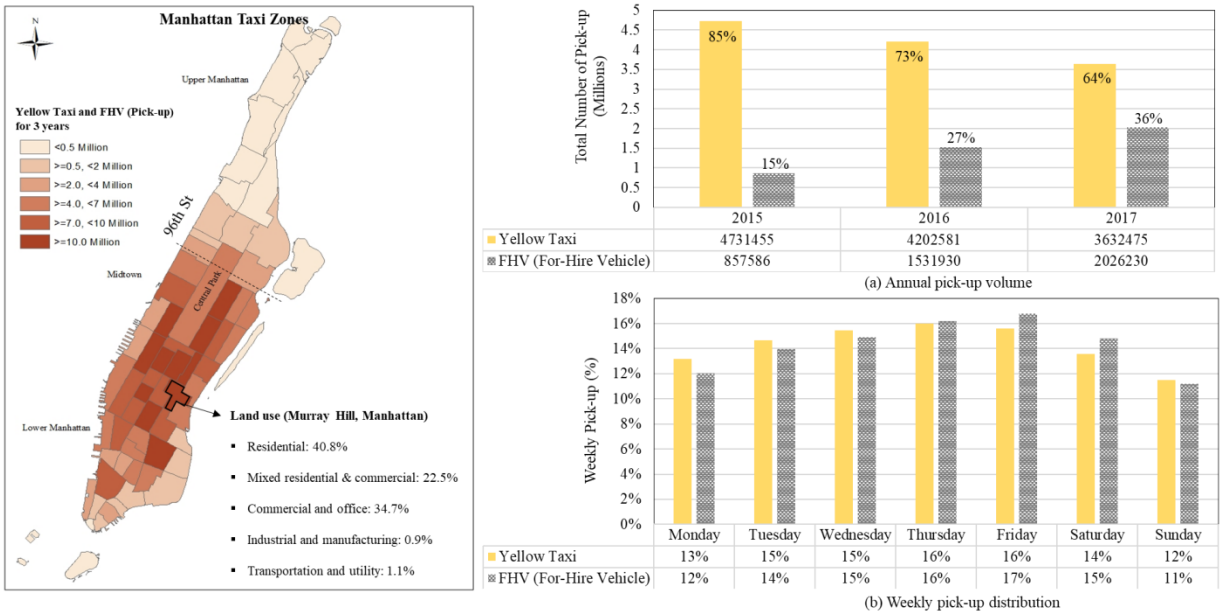
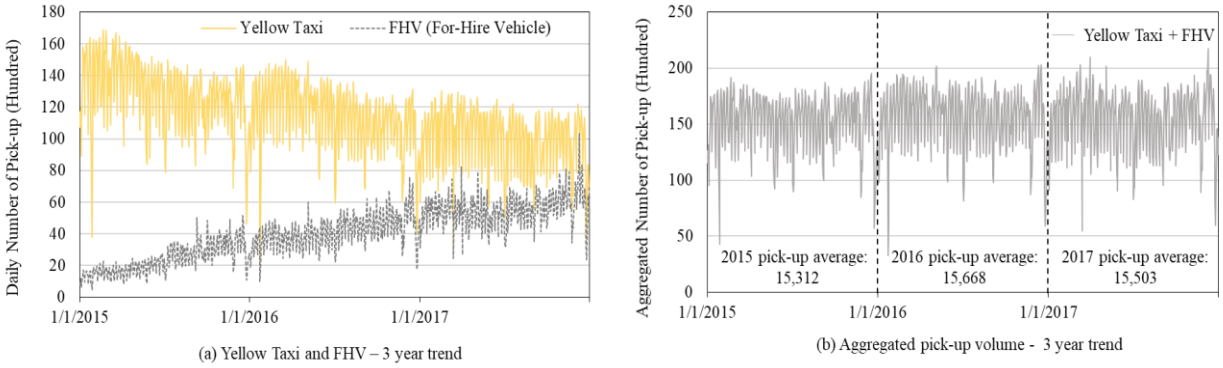


Fig. 1. Illustrations of yellow taxi and FHV pick-ups for Murray Hill, Manhattan

Fig. 2 clearly denotes a mode shifting trend between the yellow taxi and FHV where the transition has stabilized over the years. The percentage of the total trips and the level of usage of yellow taxis have gradually declined as the riders have significantly shifted their demand from yellow taxi to FHV as observed from Fig. 2(a). The total demand of both modes is in fact, very stable, indicating the ride-hailing service has not further induced new demand for the study area as shown in Fig. 2(b).



(a) Yellow Taxi and FHV – 3 year trend

(b) Aggregated pick-up volume - 3 year trend

Fig. 2. Shifts in pick-up demands between yellow taxi and FHV over three years

3.2. Weather and holidays

The prior study by Schneider (2015) and Guo et al. (2018) reported that the weather condition and holidays affect the usage level of yellow taxi and FHV. Accordingly, information collected by the National Climate Data Center (NCDC) is used in this study, which covers the daily average records of both precipitation (i.e., rain or melted snow) and snow in Central Park, Manhattan. Overall, there are 347 rainy days and 22 snowy days within the investigated years.

In addition, holidays observed by the Federal Holiday calendar are identified in the input data such as New Year's Day, Martin Luther King, Jr. Day, George Washington's Birthday, Memorial Day, Independence Day, Labor Day, Columbus Day, Veterans Day, Thanksgiving Day, and Christmas Day.

4. Conceptual modeling framework and the stepwise procedure

This section describes the system control architecture and the process of the stepwise modeling framework. The architecture is explained in terms of its structural equation formula, and the stepwise process is addressed through multivariate linear regression (LR) and Long Short-term Memory (LSTM) neural network.

4.1. System architecture for predicting yellow taxi demand from FHV quota to be controlled

The problem aims to predict the potential trip demand of yellow taxis during the day, with FHV volume as the exogenous input variable, as a function of the other factors such as weather, holidays, day of the week. Fig. 3 displays the architecture composed of the control variables, prediction, and a stepwise approach using two steps (i.e., LR-LSTM). More specifically, we can consider the following use case of the proposed model: the transportation authority plans possible quota for FHV demand, in an effort to create a fair and equitable environment for yellow taxi drivers with reasonable drivers' earning levels.

On day $t-1$, one can use the stepwise model to first estimate yellow taxi demand, briefly denoted as y_{t-1}, using the determined conditions, via LR and then integrate the estimation results with the predicted residuals $\hat{r}_t$ from LSTM, finally forecasting the total pick-up volumes served by yellow taxi, $\hat{y}_t$, at different conditions. On the next day t, we compare the ground truth values with predicted values from our model and then update parameters of LSTM for further use. By setting up a different FHV quota on different days, we hope to meet both system-wide goals of reducing congestion as well as ensuring sufficient taxi utilization rates.

Table 1.
Comparison of characteristics of LR-LSTM with that of different time series models and regression

Model	Predictor	Prediction	Quota control	Level of Interpretability	Prediction Accuracy
ARIMA	y_{t-1}	$\hat{y}_t$	No	Medium	Medium to High
LSTM	y_{t-1}	$\hat{y}_t$	No	Low	High
Regression	X_t	$\hat{y}_t$	Yes	High	Low to Middle
LR-LSTM	X_t, r_{t-1}	$\hat{y}_t, \hat{r}_t$	Yes	High, presented in **5.1**	High, presented in **5.2**

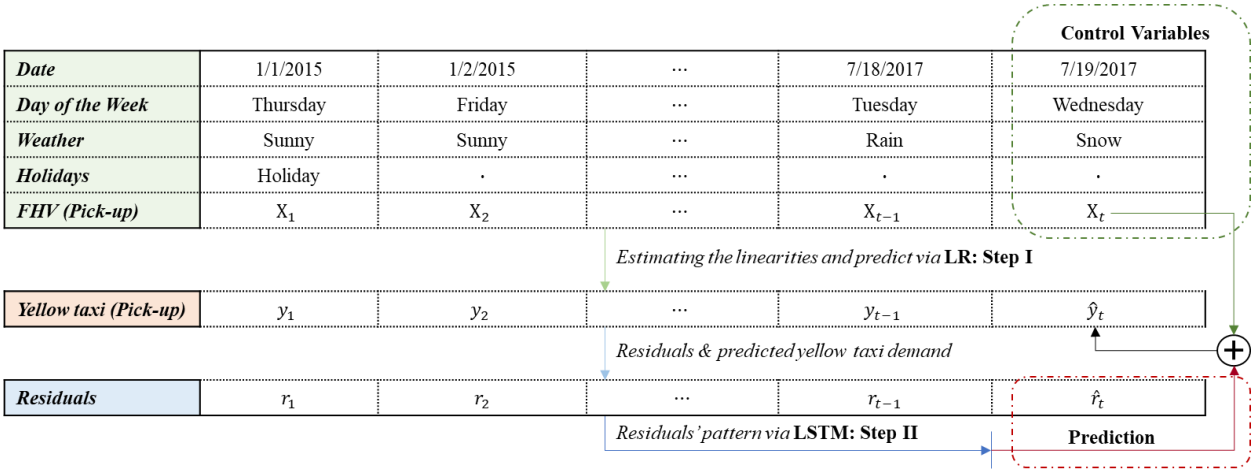


Fig. 3. Illustration of system control architecture to estimate and predict yellow taxi pick-ups

One can further extend this methodology to design personalized incentive schemes in future research, while this study still focuses on controlling the total of FHV or yellow taxi as individual parties. On the other hand, unlike time-series models (e.g., Autoregressive integrated moving average (ARIMA) or long short-term memory (LSTM)) using the historical (correlated) records of yellow taxi as only input data to predict the (unknown) yellow taxi demand, our proposed model seamlessly adds external variables for quota control. Table 1 compares the characteristics of the above models. The stepwise framework starting from Step I, the multivariate linear regression to Step II, the recurrent neural network (RNN) – LSTM is written as follows:

$$\begin{aligned} &\text{Step I:} && y_t = \beta \mathbf{X}_t^{\intercal} \\ &\text{Step II:} && \hat{r}_t = \mathrm{G}(\boldsymbol{\theta}; y_{t-1} - \beta \mathbf{X}_{t-1}^{\intercal}) \\ &\text{Prediction:} && \hat{y}_t = \beta \mathbf{X}_t^{\intercal} + \hat{r}_t \end{aligned} \tag{1}$$

The linearity between trip records of demand y and explanatory coefficients β is captured by $(\beta \mathbf{X}^{\intercal})$, and the residuals r_{t-1}=$(y_{t-1} - \beta \mathbf{X}_{t-1}^{\intercal})$ are fed forward in the residual LSTM function $\mathrm{G}(\cdot)$ to model the nonlinearity effect using the neural network parameters $\boldsymbol{\theta}$. By utilizing both linear and neural network components of β and $\boldsymbol{\theta}$, we can predict the demand of yellow taxi.

Fig. 4 describes the sequential process of Eq. (1) with more details in the following subsection.

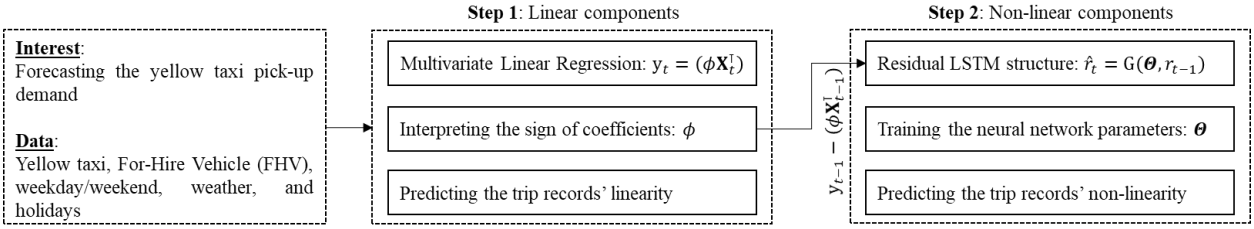


Fig. 4. Stepwise calibration process using LR-LSTM modeling

4.2 Multivariate linear regression for capturing correlations: LR

To determine the linear correlation between the two modes and the demand-side factors, we now address the pursuit with the regression model, as shown in Eq. (2):

$$\hat{y}_{n\times 1} = \mathbf{X}_{n\times p}\hat{\beta}_{p\times 1} = \begin{bmatrix} 1 & \mathrm{X}_{1,1} & \mathrm{X}_{1,2} & \cdots & \mathrm{X}_{1,p-1} \\ 1 & \mathrm{X}_{2,1} & \mathrm{X}_{2,2} & & \mathrm{X}_{2,p-1} \\ & \vdots & & \ddots & \vdots \\ 1 & \mathrm{X}_{n,1} & \mathrm{X}_{n,2} & \cdots & \mathrm{X}_{n,p-1} \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix} \tag{2}$$

$\hat{y}_{n\times1}$ is the computed trip demand of the yellow taxi and n is the total number of measurements. The matrix $\mathbf{X}_{n\times p}$ indicates the explanatory variables: constant, the actual pick-up volumes of FHV, weekday/weekend, weather, and holidays. p denotes the number of explanatory variables, and $\hat{\beta}_{p\times1}$ is the calibrated coefficients corresponding to the independent variables.

The ordinary least squares (OLS) method is utilized to calibrate the coefficients: $\hat{\beta}_{p\times1} = \left(\mathbf{X}_{n\times p}^{\intercal} \cdot \mathbf{X}_{n\times p}\right)^{-1}\mathbf{X}_{n\times p} \cdot y_{n\times1}$. The residuals $r_{n\times1}$ between the estimated demand from OLS and the actual demand are calculated as follows, and LSTM extracts possible nonlinear trends of the residuals to improve the predictive accuracy (Goel and Banerjee, 2017).

$$r_{n\times1} = y_{n\times1} - \hat{y}_{n\times1} \tag{3}$$

4.3 Residual LSTM for capturing non-linear patterns

This subsection describes LSTM-NN approach (Hochreiter and Schmidhuber, 1997) as an extension version of RNN (Rumelhart et al., 1988), after introducing the principle of RNN and its limitation. Then, the residual LSTM is introduced within a computational graph (CG) framework (Olah, 2015; Baydin et al., 2018; Wu et al., 2018; Sun et al., 2019).

4.3.1 Recurrent neural network (RNN)

RNN can be essentially viewed as a non-linear optimization model to minimize the loss function in Eq. (4), where the residual term defined in Eq. (3) follows time dependent variables.

$$\min L_t = \min \sum_{t=1}^{T} (r_t - \hat{r}_t)^2 \tag{4}$$

$$h_t = tanh(W_{rh} r_{t-1} + W_{hh} h_{t-1} + b_h) \tag{5}$$

$$\hat{r}_t = W_{h\hat{r}} h_t + b_h \tag{6}$$

The defined function $tanh(\cdot)$ in Eq. (5) is key to capture non-linear patterns, and Eq. (6) can be refined by different structural forms based on specific tasks (e.g., classification or regression). The parameters, W_{rh}, W_{hh}, and $W_{h\hat{r}}$ are applied at every time step and are shared in the entire structure of RNN. Table 2 lists the required components with definitions at time step t. The above optimization model is numerically solved by using the gradient descent algorithm. As discussed in the machine learning community, long data sequences in RNN architecture could lead to the issue of vanishing gradients for the information of updated parameters. That is, if the updated parameters are fractional values, the carried gradients from the long sequence become insignificant. In order to address this issue, the long short-term memory (LSTM) is implemented in this study (Hochreiter and Schmidhuber, 1997).

Table 2.
Definitions of parameters and variables for RNN components

RNN at time step *t*	Parameters	Definition
Neural Network (NN) parameters to be estimated	W_{rh}	Weight from input layer to hidden state
	W_{hh}	Weight from previous hidden to current hidden state
	$W_{h\hat{r}}$	Weight from hidden state to output
	b_h	Bias at hidden state
	$b_{\hat{r}}$	Bias at output
Variables	r_{t-1}	Input variable (Residuals from LR)
	h_{t-1}	Hidden variable from the previous time step, *t*-1
	h_t	Hidden variable at the current time step, *t*
	$\hat{r}_t$	Output variable (Prediction)
Composite function	$\hat{r}_t = W_{h\hat{r}} h_t + b_{\hat{r}}$	

4.3.2 Long short-term memory (LSTM)

A typical LSTM model consists of four gates (i.e., input, forget, output, and external input gates). The chained structure across the gates transmits not only a hidden state h_t but also previous cell s_t. The shared information of the hidden state and cell state can overcome the vanishing gradient descent effects; the detailed features of LSTM are explained in Olah (2015) and Goodfellow et al. (2016), and the definition of each parameter used is shown in Table 3.

In brief, the first step starts from the forget gate f_t helping to remove the unnecessary information by the sigmoid function $\sigma(\cdot)$, ranging from 0 to 1.

$$f_t = \sigma\left(b_f + W_{rf} r_{t-1} + W_{ff} h_{t-1}\right) \tag{7}$$

where b_f is the forget gate bias term. W_{rf} and W_{ff} are weights, h_{t-1} is the previous hidden state, and r_{t-1} is the time series data at current step $t-1$, defined in Eq. (3) from the LR model. The second step updates and decides the new information to be stored in the cell state through the input gate i_t and the external input gate g_t.

$$i_t = \sigma(b_i + W_{ri} r_{t-1} + W_{ii} h_{t-1}) \tag{8}$$

$$g_t = tanh\left(b_g + W_{rg} r_{t-1} + W_{gg} h_{t-1}\right) \tag{9}$$

The input gate i_t determines a gating value ranging from 0 to 1, and a value of 1 means the input information will be fully stored. $tanh(\cdot)$ generates a vector of new candidate values between -1 and 1 guiding the extent of updating the weights in Eq. (7) and Eq. (8). The multiplication of the input gate i_t and the external gate g_t identifies new significant information, storing it in the cell state, and the third step in Eq. (10) updates the old cell state.

$$s_t = f_t \odot s_{t-1} + i_t \odot g_t \tag{10}$$

Both s_t and s_{t-1} denote new and old cell states produced, and $\odot$ is the Hadamard product, the element-wise products of vectors, matrices, or tensors. The final step is proceeded by the output gate o_t.

$$o_t = \sigma(b_o + W_{ro} r_{t-1} + W_{oo} h_{t-1}) \tag{11}$$

$$h_t = o_t \odot tanh\,(s_t) \tag{12}$$

o_t and h_t are the result of the output gate and a hidden state. Finally, a regression formula defined as the product of the parameter $W_{h\hat{r}}$ and the hidden state h_t with bias term $b_{\hat{r}}$ measures the residual $\hat{r}_t$:

$$\hat{r}_t = W_{h\hat{r}} h_t + b_{\hat{r}} \tag{13}$$

In order to compute the approximate residuals, the nonlinear optimization function applied in LSTM is utilized as follows:

$$\min L_t = \sum_{t=1}^{T} \left((W_{h\hat{r}}(o_t \odot tanh(f_t \odot s_{t-1} + i_t \odot g_t)) + b_{\hat{r}}) - r_t \right)^2 \tag{14}$$

Eq. (14) is the objective function subject to Eqs. (7) to (13) and LSTM neural network parameters are adjusted to minimize the loss L_t. The adjustment process is proceeded by the *Adam* optimizer, the gradient-based stochastic optimization algorithm proposed by Kingma and Ba (2014) with improved computational efficiency for handling a large data set and parameters. The estimating procedure of the neural network parameters conducts the feedforward and backward propagation process. A computational graph (CG) is shown in Fig. 5 to illustrate the process in LSTM in which a feedforward step measures the residuals using the defined functions composed of weights and bias, and a backward propagation step minimizes the loss between the computed residuals and the actual residuals by adjusting the parameters.

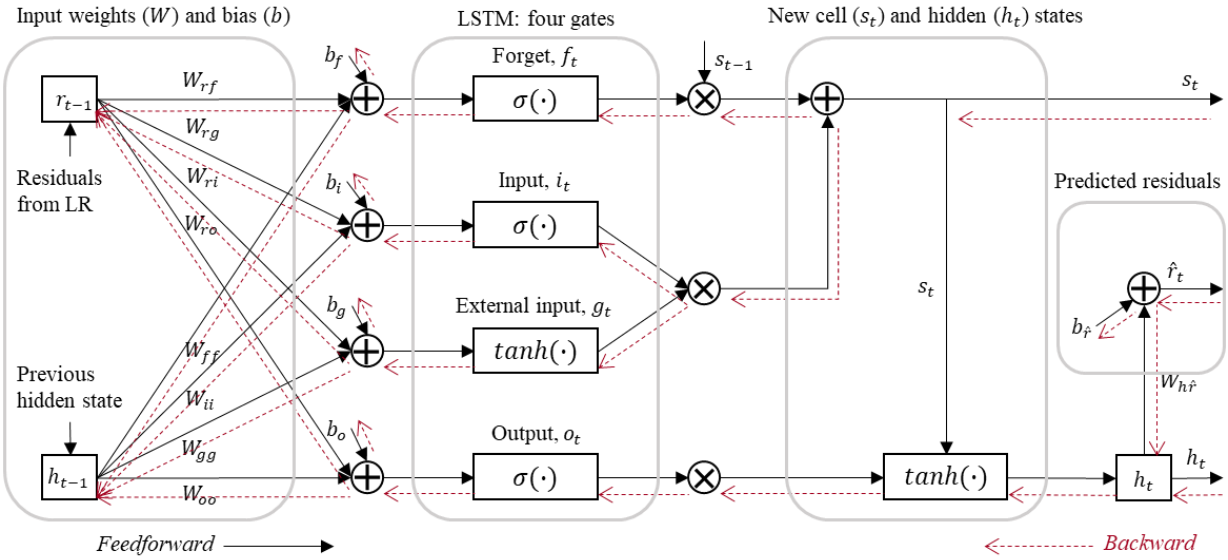


Fig. 5. Illustration of inner LSTM structure: feedforward and backward propagation

Table 3.
Definitions of parameters and variables for LSTM components

LSTM at time step t	Terms	Definition
Neural Network (NN) parameters $\boldsymbol{\theta}$ to be estimated	W_{rf}	Weight from input layer to forget gate
	W_{ff}	Weight from previous forget gate to current forget gate
	W_{ri}	Weight from input to input layer gate
	W_{ii}	Weight from previous input gate to current input gate
	W_{rg}	Weight from input to external input gate
	W_{gg}	Weight from previous external gate to current external gate
	W_{ro}	Weight from input to output gate
	W_{oo}	Weight from previous output gate to current output gate
	$W_{h\hat{r}}$	Weight from hidden state to output
	b_f	Bias at forget gate
	b_i	Bias at input gate
	b_g	Bias at external input gate
	b_o	Bias at output gate
	$b_{\hat{r}}$	Bias at output
Variables	r_{t-1}	Input variable (Residuals from LR)
	h_{t-1}	Hidden variable from the previous time step, t-1
	h_t	Hidden variable at the current time step, t
	s_{t-1}	Cell variable from the previous time step, t-1
	s_t	Cell variable at the current time step, t
	$\hat{r}_t$	Output variable (Prediction)
Composite functions	$\hat{r}_t = W_{h\hat{r}}(o_t \odot tanh(f_t \odot s_{t-1} + i_t \odot g_t)) + b_{\hat{r}}$	

4.4 Integrated model on trip demand forecasting: LR-LSTM

Now we start examining the proposed integrated model LR-LSTM. The linear regression results explain the linearity pattern from variables (i.e., day of the week, holiday, weather, and FHV) so that the pattern learned from the model can help capture the proportional trend on two modes. The residual LSTM generates the unparameterized function to handle the left-over residuals. The mathematical form of the coupled architecture of forecasting the demand is written as:

$$\hat{y}_t = \hat{\beta}\mathbf{X}_t^{\intercal} + W_{h\hat{r}}(o_t \odot tanh(f_t \odot s_{t-1} + i_t \odot g_t)) + b_{\hat{r}} \quad (15)$$

That is, the yellow taxi demand is now estimated by FHV, and the other explanatory variables, together with the predicted residuals. This integrated model can use the regression model to control the pick-up volume of FHV to balance the demand for yellow taxis. Also, the low prediction results caused by traditional extrapolation method can be improved by the residual-oriented model.

Please note that, the input data used for estimating the trip demand is composed of binary variables such as weekday, weekend, and holiday, and the rest of the variables (i.e., FHV, yellow taxi, precipitation, and snow) are continuous, with FHV and yellow taxi values being normalized. In addition, to test the integrated model, the entire dataset is split into the training data, the validation data, and test data as shown in Fig. 6. Using 800 days (85%) out of 931 days, the hyperparameters used in LSTM are tuned. Then, the underlying hyperparameters are determined based on the validation experiments shown in Eq. (14); the number of hidden units used was 10 to 12, the selected batch size was 10, the number of iteration times, training epoch, was up to 500, the learning rate of the *Adam* optimization was 0.001, and the input dimension was 8 (i.e., weekday/weekend and the prior pickup volume residual). It should be noted that if another set of explainable data is available for the model every day, then we re-execute the tuning process and updating the hyperparameters with the corresponding data-receiving frequency.

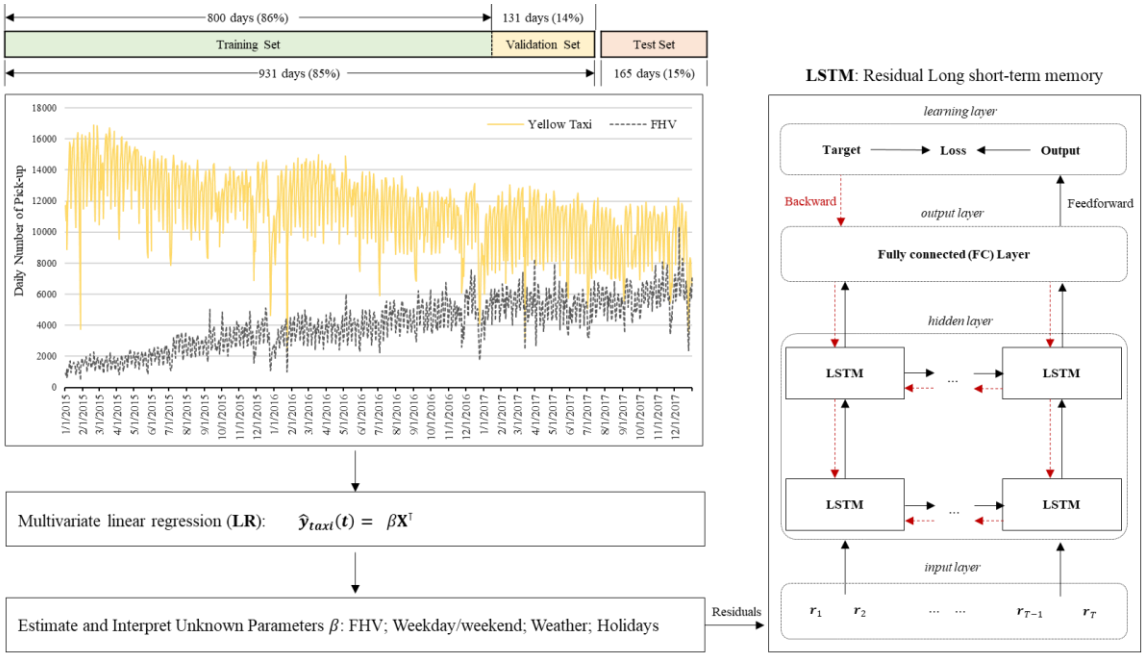


Fig. 6. Hybrid model architecture: two-step system using NYC taxi records

The right end side of Fig. 6 visualizes the two-step process of the coupled LR and LSTM. The *input layer* takes the residual data and transmits the data into the *hidden layer* at the feedforward step denoted as the straight-line arrows in LSTM. Then, a two-layered LSTM expressed in the hidden layer combines the data with the weighted parameters and bias values based on a basic arithmetic operation (e.g., multiplication and addition) and sends the driven values to the *output layer*. In this layer, the fully connected (FC) layer is added to arrange the output results as a one-dimensional structure, extracting the last output value. Lastly, using the mean squared error (MSE) function between the output and target data given by the *learning layer*, LSTM is trained and updated iteratively through backpropagation expressed as the dotted line arrows. The proposed structure is coded by TensorFlow developed by Abadi et al. (2016). To compare the performance to different time-series models, the results of RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) are detailed in the following section.

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_t - \hat{r}_t)^2} \quad (16)$$

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\frac{|r_t - \hat{r}_t|}{r_t} \times 100$$

where r_t is the residuals at time t sent from LR model, and $\hat{r}_t$ is the predicted value by LSTM. T is the total length of the test dataset (i.e., 165 days for the test set).

5. Model estimation results

This section examines the estimation and forecasting results for the pick-up demand. Specifically, the first section discusses individual and collective effects of different factors through the parameters estimated by LR. The second section then focuses on LSTM model validation, a sensitivity analysis of LR-LSTM, and the prediction results conducted by the coupled model. Lastly, we examine the forecasting capability of our proposed integrated model, compared to standard alone models such as regression, ARIMA, and LSTM.

5.1 Trip demand estimation: LR as Local Interpretable Model

The LR model estimates the regular patterns of the yellow taxi and FHV demand, and the results of LR are presented in Table 4. When estimating the yellow taxi demand, the parameter estimates, associated with calendar week, snowy day, holidays, and FHV trips are statistically significant. On the other hand, the estimated FHV demand shows the statistical significance on the aforementioned coefficients as well as rainy day. In addition, the tested variance inflation factor (VIF), a way of measuring multicollinearity, shows a low correlation between independent variables. As shown in Sheather (2009), the generally acceptable range of VIF is less than 5.

According to Table 4, the coefficients corresponding to weekday display the positive sign, indicating that customers are more likely to ride yellow taxis or FHV during the day. On the other hand, as the weekend approaches, the pick-up demand volume of two modes decreases. The estimated day-specific coefficients with respect to Tuesday and Saturday are eliminated due to the statical insignificance and the multicollinearity effect. The model-based calibration findings reflect the expected characteristics of the current data set in Fig. 1(b). The weather conditions affect the level of the taxi and ride-hailing service usage patterns, particularly under a snow event. The negative sign associated with this parameter indicates that the utilization rate of the rides is lower than the regular day. Under the rainy condition, this explanatory factor does not influence the pick-up volume of the taxi statistically. On the one hand, this factor increases the pick-up volume of the ride-hailing services. This noticeable pattern illustrates that travellers prefer to choose the convenient mobility (i.e., app-based ride services) and wait for the service in buildings without being drenched in rain. The coefficients of FHV and yellow taxi can be determined by switching dependent variables in LR. Both coefficients have the expected negative sign, indicating the inverse relationship with the dependent variables used. This interesting finding explains the gradual decline of the yellow taxi demand, whereas FHV demand follows the opposite pattern, and the numerical difference of the coefficients indicates how much the demand of two modes would be changed when increasing the value of either FHV or yellow taxis.

In Table 4, the number of observations is 931 days ranging from 1/1/2015 to 7/21/2017. The goodness of fit measure R^2 denotes that the regression model can explain the variation for the response variable, in terms of an approximate rate at about 57% and 37%. In order to validate if the added predictors truly enhance the interpretability of the model, the adjusted R^2 is measured at 0.565 and 0.366. As the LR model can explain the yellow taxi demand better, we employ the estimated coefficients with respect to the taxi, forecasting the linear pattern. Overall, LR derives the extrapolation results of the yellow taxi volume and then obtains the residuals between the observed data and the estimated results. The statistics and the shape of the residuals derived are plotted in Fig. 7 with the distributed residuals quantified by μ_Y as the mean of residuals and the variance as σ_Y. As noticed in the figure, the distribution of the residuals generally follows a bell-shaped curve similar to the normal distribution.

Table 4.
Multivariate linear regression (LR) estimation results for **yellow taxi** and **FHV** demand

Dependent Variable: **yellow taxi**			*Dependent Variable:* **FHV**		
Predictors	*Coefficients (t-statistics)*	VIF	*Predictors*	*Coefficients (t-statistics)*	VIF
Constant	0.7640 (64.337)	-	Constant	0.9094 (30.625)	-
Monday	-0.0362 (-2.885)	1.786	Monday	-0.0949 (-5.720)	1.343

Tuesday	0.0518 (4.204)	1.720	Tuesday	-	-
Wednesday	0.0951 (7.708)	1.725	Wednesday	0.0663 (4.011)	1.331
Thursday	0.1515 (7.708)	1.729	Thursday	0.1415 (8.375)	1.392
Friday	0.1445 (11.643)	1.744	Friday	0.1535 (9.169)	1.367
Saturday	-	-	Saturday	-	-
Sunday	-0.1754 (-13.970)	1.783	Sunday	-0.2367 (-13.347)	1.534
Snow	-0.8224 (-10.047)	1.021	Snow	-1.2091 (-9.089)	1.161
Rainy day	-	-	Rainy day	0.2334 (-10.113)	1.158
Holiday	-0.3401 (-14.217)	1.058	Holiday	-0.3857 (4.804)	1.119
FHV	-0.3289 (-17.853)	1.155	Yellow Taxi	-0.7628 (-17.763)	1.713
Goodness of fit (number of observations: 931 days)			*Goodness of fit (number of observations: 931 days)*		
R^2(Adj. R^2)	0.573 (0.569)		R^2 (Adj. R^2)	0.372 (0.366)	

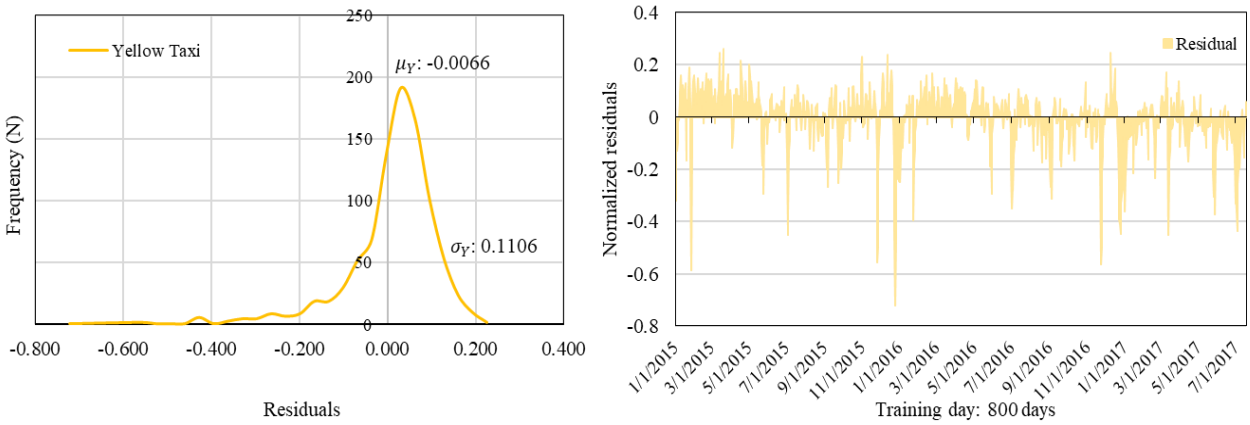


Fig. 7. Yellow taxi residual distribution and measured residuals at each training day

5.2 Validation and prediction: LR-LSTM

Fig. 8(a) displays the learning process of the residual patterns and its evolution of the loss function employing the validation dataset with different prediction time steps (e.g., "step 1" indicates the one-step ahead prediction that forecasts the next one day). As noticed, the lowest error, RMSE, appears in the one-step short-term prediction ("step 1"). In addition, the evolutionary error falls rapidly after 10 training epochs, showing a high convergence speed of the model. Fig. 8(b) further performs the sensitivity analysis for the proposed LSTM model. The test set error measured by RMSE gradually increases when the model predicts longer than the one-step forecasting period. The baseline of RMSE observed in "step1" is 0.1129, and the seven-step "medium term" prediction denotes that the error increases approximately 48% from the baseline.

Using the multivariate linear regression estimators and the residuals measured by LSTM, the number of trips served by yellow taxis at Murray Hill in Manhattan is forecasted in Fig. 8(c). The test dataset covers the time period from 7/22/2017 to 12/31/2017 (i.e., 165 days), predicting the following day. The predicted demand is plotted along the y-axis as the normalized values with the look-ahead window of prediction on the x-axis. Specifically, Fig. 8(c) describes the forecasts of the yellow taxi's daily demand that follow the test pattern (i.e., true records) properly across the entire time horizon. This proposed process can be applicable in predicting the demand of the FHV volume by switching the identification of independent and dependent variables.

5.3 Model performance comparison

In this section we provide a comparison of the proposed model with the different time-series modeling approaches under three different experiments predicting morning peak-hour volumes collected from 6AM to 9AM, evening peak-hour volumes recorded from 4PM to 8PM, and daily volumes of yellow taxis. Evaluating the performance errors (i.e., RMSE and MAPE), we examine the accuracy of LR-LSTM. Autoregressive integrated moving average (ARIMA) and LSTM models are representative of the performance benchmark. The measured RMSE and MAPE presented in Table 6 demonstrates the prediction capability of LR-LSTM. This experiment is prepared by

the hold-out strategy (Refaeilzadeh et al., 2009). Using the training and validation dataset, hyperparameters of each model were tuned in decreasing the error between the observed data and the estimated results.

We examine the performance of forecasting the daily volume of yellow taxis at first. According to the validation result (i.e., 931 days), LR-LSTM displays the lowest RMSE and MAPE values compared to other candidates such that the proposed model shows a better performance on the test dataset (i.e., 165 days); an improvement of 34.2% w.r.t MLR, 19.5% w.r.t ARIMA, and 27.4% w.r.t univariate LSTM. On the other hand, the prediction results of multivariate LSTM are built on the basis of not only the historical yellow taxi data but also the explanatory variables such as weekday/weekend, holidays, and weather. It should be remarked that, although multivariate LSTM (M-LSTM) has an RMSE 4% lower than the LR-LSTM result, M-LSTM does not carry a clear interpretable model structure which prevents it from being used effectively in the active demand management application. In addition, the prediction error of the proposed LR-LSTM model is 15.04% in terms of MAPE, which is consistently lower than the the other models including ARIMA, univariate LSTM and within a similar range as multivariate LSTM.

In order to test whether the developed model is able to forecast well under possible oversaturated conditions (that is passenger demand larger than supply), we also examine the predicted results of the morning and evening peak-hour volumes in Table 6. For the morning peak-hour prediction, LR-LSTM presents the second-best performance on the validation dataset (931 days) as well as on the test dataset (165 days). Interestingly, the multivariate LSTM structure shows the best performance for the validation dataset, but the univariate LSTM model derives the lowest RMSE in the test dataset. In other words, LR-LSTM can maintain the robustness in forecasting validation and test datasets of the morning peak-hour volumes. However, due to the fact that the datasets still include a high variance and other unobserved factors, LR-LSTM and benchmark models display high MAPE errors. For instance, multivariate linear regression (MLR) indicates MAPE of 50%. For the evening peak-hour prediction, LR-LSTM on the other hand demonstrates the robustness of predicting the evening peak-hour volumes and the measured RMSE and MAPE present the second-best result on both datasets. The parameter configuration of studied models is listed as follows:

(a) MLR (Multivariate linear regression): The measured performance is only based on the extrapolation. The prediction on the yellow taxi trips is implemented using the defined coefficients in Table 4 and datasets consisting of the FHV trips, day of the week, weather, and holidays. Similarly, using coefficients estimated by the datasets of the morning and evening peak-hour volumes of the medallion taxi, MLR forecasts the taxis demand and proposes RMSE and MAPE shown in Table 6.

(b) ARIMA: To determine ARIMA model parameters for the morning and evening peak-hour, and the daily demand, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) proposed by Box et al. (2015) are used to construct our ARIMA (2, 1, 1) model for the prediction of the daily volume. That is, two number of lag observations (AR), one number of times called the degree of difference (I), and the one size of moving average window (MA) are used in our experiment. The constructed ARIMA (2, 1, 1) model is represented as follows:

$$\hat{y}_t = y_{t-1} + \alpha_1(y_{t-1} - y_{t-2}) + \alpha_2(y_{t-2} - y_{t-3}) + \beta_1(\varepsilon_{t-1}) + constant \quad (17)$$

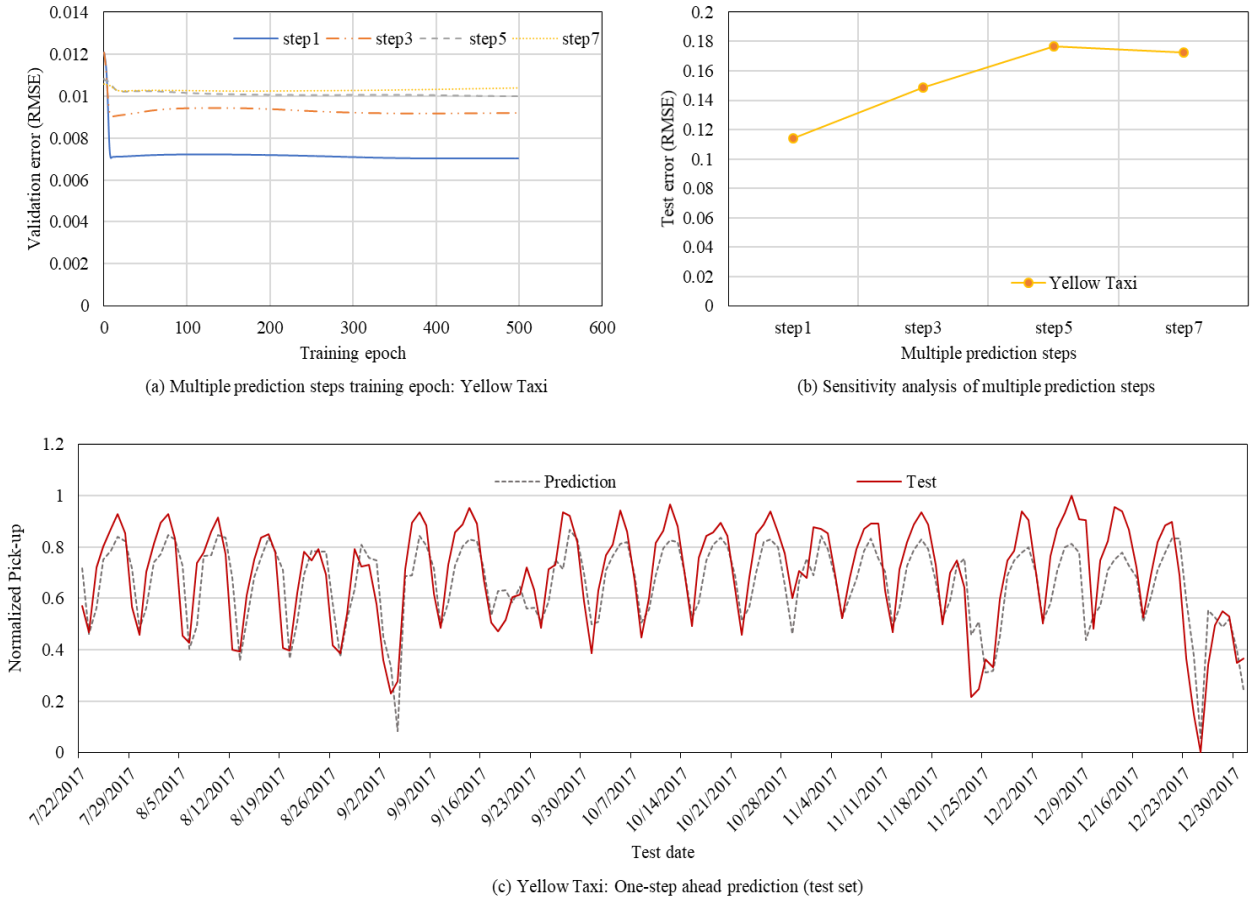
where the predicted yellow taxi demand is calculated by three observed data, namely, y_{t-1}, y_{t-2}, and y_{t-3} with the estimated parameters and constant; both α_1 and α_2 are the AR coefficients, and β_1 is the MA coefficient. Also, the term, ε_{t-1}, indicates the residuals between the observed data and the predicted result, $y_{t-1} - \hat{y}_{t-1}$. Also, in the same manner, the order of the ARIMA model is defined by the proposed method such that the order of ARIMA to forecast the morning peak-hour demand follows ARIMA (2, 1, 3), and the defined configuration of ARIMA for predicting the evening peak-hour is ARIMA (3, 1, 4). The detailed description of calibrated coefficients and the goodness of fit are shown in Table 5. The model information-related criteria, AIC and BIC, informs that the lowest values of AIC/BIC are found in the daily volume prediction of yellow taxis, showing the lowest values of RMSE and MAPE in Table 6.

(c) LSTM: Two different input features are considered for LSTM. Univariate LSTM is trained by the previous time steps of the yellow taxi only. For instance, the taxi demand for 7/20/2017 was only predicted by 7/19/2017 data. Multivariate LSTM employs weekday/weekend, holiday, weather, and the historical time steps of the taxi to train the model. Both LSTM model structures for predicting the daily taxi demand are configured as: 10-12 hidden units are used, the learning rate is set to 0.001, the training epoch covers 500 steps, and the network structure is one input layer, two hidden layers (LSTM), fully connected layer, and output layer. On the one hand, as the time series data (i.e., morning and evening peak-hour volumes of yellow taxis) has the high variance, input data structures of training univariate and multivariate LSTM are adjusted. In other words, instead of predicting the

demand for 7/20/2017 by the previous day data, it is predicted by the seven days of the yellow taxi demand (7/13/2017 to 7/19/2017). With the adjusted input data structure, we determine different configurations for both LSTM models: 15-17 hidden units are employed, the training epoch is between 500 to 1000 steps, and the network structure is similar to the previous network structure except for the number of stacked LSTM layers; 3-4 stacked layers are examined.

Table 5.
ARIMA estimation results for yellow taxi demand

	ARIMA (2, 1, 1)	**ARIMA (2, 1, 3)**	**ARIMA (3, 1, 4)**
Predictors	*Estimated coefficients (t-statistics)*		
Constant	-0.0003 (-1.841)	-0.0002 (-9.416)	-0.0001 (-2.222)
α_1 (Auto Regressive)	0.8095 (26.452)	1.2386 (174.764)	0.8053 (5.058)
α_2 (Auto Regressive)	-0.4119 (-13.521)	-0.9895 (-108.529)	-0.4519 (-2.274)
α_3 (Auto Regressive)	-	-	-0.4358 (-2.713)
β_1 (Moving Average)	-0.9706 (-87.491)	-2.2235 (-142.36)	-1.9283 (-11.299)
β_2 (Moving Average)	-	2.1683 (54.43)	1.5394 (4.037)
β_3 (Moving Average)	-	-0.9448 (-38.402)	-0.3166 (-0.823)
β_4 (Moving Average)	-	-	-0.2832 (-1.656)
	Goodness of fit (number of observations: 931 days)		
Fit Criteria	**Daily**	**Morning Peak-hour**	**Evening Peak-hour**
AIC	-1724.612	-869.926	-1189.187
BIC	-1700.436	-836.079	-1145.670



(a) Multiple prediction steps training epoch: Yellow Taxi

(b) Sensitivity analysis of multiple prediction steps

(c) Yellow Taxi: One-step ahead prediction (test set)

Fig. 8. Training validation, sensitivity analysis, and estimated prediction of yellow taxi pick-up demand

Table 6.
Comparisons of one-step ahead validation and prediction performance with respect to yellow taxi demand

Daily	**Validation Dataset (931 days)**		**Test Dataset (165 days)**	
	RMSE	**MAPE (%)**	**RMSE**	**MAPE (%)**
MLR	0.0998	14.54	0.1715	24.03
ARIMA	0.1114	15.47	0.1404	19.52
Univariate LSTM	0.1110	17.52	0.1556	22.85
Multivariate LSTM	0.0738	9.77	0.1081	14.26
LR-LSTM	0.0620	8.69	0.1129	15.04
Morning (6-9AM)	**Validation Dataset (931 days)**		**Test Dataset (165 days)**	
	RMSE	**MAPE (%)**	**RMSE**	**MAPE (%)**
MLR	0.1886	37.58	0.2058	50.63
ARIMA	0.1608	32.74	0.1791	46.67
Univariate LSTM	0.1480	29.25	0.1499	37.42
Multivariate LSTM	0.1304	22.82	0.1766	41.97
LR-LSTM	0.1478	29.57	0.1655	39.98
Evening (4-8PM)	**Validation Dataset (931 days)**		**Test Dataset (165 days)**	
	RMSE	**MAPE (%)**	**RMSE**	**MAPE (%)**
MLR	0.1240	22.21	0.1237	27.25
ARIMA	0.1272	22.73	0.1714	38.83
Univariate LSTM	0.1300	23.84	0.1233	27.14
Multivariate LSTM	0.1135	20.78	0.1217	27.10
LR-LSTM	0.1189	20.99	0.1228	27.31

6. Conclusions

There are many studies using machine learning models for estimating ridership trends. However, the lack of ability, commonly observed in those data-fitting oriented models, of sensing and understanding the significantly influential factors for transportation service demand could cause difficulties in constructing proper ADM policies and decision tools. This paper proposes a hybrid modeling framework, LR-LSTM, to facilitate the planning effort for balancing the utilization rates of the emerging FHV service and regular taxis. Our developed model formulation integrates the multivariate linear regression (LR) and LSTM to forecast the daily and peak-hour taxi demand. With different data sources being used as explanatory variables such as FHV, weekday/weekend, snow/rain, weather conditions, and holidays, LR is employed to select the statistically important variables and interpret the correlation between the variables. LSTM helps improve the prediction accuracy by capturing the variance that the extrapolation model cannot measure.

Essentially, the proposed methodology shows the capability of maintaining the interpretability as well as predictability, indicating the positive potential of achieving the effective active demand management (ADM) system within a city-wide network. To our limited knowledge, this study represents the first integrated modeling effort of the quota control applications between on-demand ride-hailing services and regular taxis, through a combination of the econometric formulation with the deep learning structure. This methodology is validated based on the real-world experiment obtained from New York City Taxi & Limousine Commission (NYC-TLC), particularly forecasting the pick-up demand of the yellow taxi in Murray Hill, Manhattan. The results of RMSE and MAPE outperform the multivariate linear regression, ARIMA, and univariate LSTM fed exclusively with historical records. With the well calibrated LR-LSTM tool, we can examine different scenarios for policy analysis. For instance, if a policy starts regulating the FHV pick-up demand in New York City, this approach can estimate the impact of the regulation within the network. Overall, by setting up different FHV quota on different days, we can accomplish both system-wide goals of reducing congestion as well as ensuring sufficient taxi utilization rates. More detailed future research along this study can be built on the following premises in the current paper.

(1) A stepwise modeling framework could further refine coupled modeling systems to balance the usage level of the yellow taxi demand and for-hire vehicle (FHV) in the coexisting environment with public transportation systems, so that a proper utilization rate of multiple modes can be achieved.

(2) This structure can further quantify the expected additional demand to be shifted to yellow taxi rides when a cap is imposed for TNC in the real-world environment, and the city traffic management agency can eventually adapt this approach to achieve fair and dynamic quota regulation during a special holiday or weather conditions, thus distributing traffic volumes within the taxi service zones and reducing the likelihood of traffic accidents.
(3) This methodology could examine potential impact on introducing shared micro-mobility technologies (e.g., e-scooters or e-bikes) within pedestrian infrastructure (Harwood, 2020), proposing a process to integrate other app-based mobility services (i.e., FHV) or medallion taxis with the new mobility, particularly in a local community in order to ensure users' safety and accessibility for the mobility service (Pineda, 2019).

Other medium-term future research tasks include exploring additional factors associated with the demand of yellow taxis or ride-hailing services. Particularly, taxi usage patterns recognized by many other explanatory variables such as socio-demographic or the spatial characteristics should be examined by employing either survey-based or social media-based travel behaviour data sets. Then, fusing multi-sourced data (Wang et al., 2019) to mitigate the variance of the time series data is required. Furthermore, constructing different types of neural networks to interpret the prediction results is interesting and beneficial. We can also better quantify the contribution of aforementioned features to the forecasting, selecting critical variables (Ribeiro et al., 2016; Lundberg and Lee, 2017). As an example, we could introduce logistic regression or discrete choice models to handle probabilities of utilizing yellow taxis or ride-hailing services by individuals. This interpretable machine learning techniques could lead to an enhanced conceptual modelling structure:

$$\textit{True demand} = \textit{Interpretable trend pattern} + \textit{Structure deviation} + \textit{Seasonal factors} + \textit{Random noise}$$

Therefore, in terms of broader application of the proposed method, we also see other possibilities of quantifying the impact related to new urban mobility policies, e.g., the regulation of fleet sizes dynamically using the frequency of rides per day in a given time and space and the sustainability of deploying a specific number of vehicles within a service area, as well as ensuring the accessibility of emerging technologies (e.g., shared micro-mobility or ride-hailing services) through income-based payment plans (NACTO, 2019). This data driven analytics approach could shed some light on understanding, modeling and systematically managing the relationship between emerging mobility and existing services, and potentially lead to a more optimized vehicle routing and supply system through pre-trip scheduling and on-line dispatching for autonomous vehicles (Shen et al., 2019; Mao et al., 2020).

Acknowledgements

This research was supported by the Center for Teaching Old Models New Tricks (TOMNET), which is Tier 1 University Transportation Centers sponsored by the US Department of Transportation. The third author is partially supported by NSF Grant No. CMMI 1663657 "Real-time Management of Large Fleets of Self-Driving Vehicles Using Virtual Cyber Tracks".

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)* (pp. 265-283). Software available from https://www.tensorflow.org

Altman, E.I., Marco, G. and Varetto, F., 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of banking & finance, 18(3), pp.505-529.

Al-Maqaleh, B.M., Al-Mansoub, A.A. and Al-Badani, F.N., 2016. Forecasting using artificial neural network and statistics models. International Journal of Education and Management Engineering, 6(3), pp.20-32.

Baydin, A.G., Pearlmutter, B.A., Radul, A.A. and Siskind, J.M., 2018. Automatic differentiation in machine learning: a survey. Journal of machine learning research, 18 (153).

Box, G.E., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M., 2015. Time series analysis: forecasting and control. John Wiley & Sons.

Chen, X.M., Zahiri, M. and Zhang, S., 2017. Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. Transportation Research Part C: Emerging Technologies, 76, pp.51-70.

Cui, Yu, Chuishi Meng, Qing He, and Jing Gao., 2018. Forecasting current and next trip purpose with social media data and Google Places. Transportation Research Part C: Emerging Technologies 97: 159-174.

FHWA. 2012. ATDM Program Brief: An Introduction to Active Transportation and Demand Management. Publication Number: FHWA-HOP-12-032, U.S. Department of Transportation, Federal Highway Administration (FHWA), Washington, DC. 2 p.
Fischer, B., 2015. In Uber vs. taxi cab fight, expense reports offer telling barometer.
Gerte, R., Konduri, K.C. and Eluru, N., 2018. Is there a limit to adoption of dynamic ridesharing systems? evidence from analysis of uber demand data from new york city. Transportation Research Record, 2672(42), pp.127-136.
Goel, H., Melnyk, I. and Banerjee, A., 2017. R2N2: Residual recurrent neural networks for multivariate time series forecasting. arXiv preprint arXiv:1709.03159.
Goodfellow, I., Bengio, Y. and Courville, A., 2016. Sequence Modeling: Recurrent and Recursive Nets. Deep Learning, pp.367-415.
Golshani, N., Shabanpour, R., Mahmoudifard, S.M., Derrible, S. and Mohammadian, A., 2018. Modeling travel mode and timing decisions: Comparison of artificial neural networks and copula-based joint model. Travel Behaviour and Society, 10, pp.21-32.
Gu, Y., Qian, Z.S. and Chen, F., 2016. From Twitter to detector: Real-time traffic incident detection using social media data. Transportation research part C: emerging technologies, *67*, pp.321-342.
Gunning, D. (2017). Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web, 2.
Guo, J., C. Brakewood., C. Liu., and K. Won., 2018. Analysis and comparison of Uber, Taxi, and Uber request via Transit.
Harwood, L., 2020. Shared Micromobility Policies, Permits, and Practices – NCHRP Synthesis 20-05/Topic 52-13
Hasnat, M.M. and Hasan, S., 2018. Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. Transportation Research Part C: Emerging Technologies, 96, pp.38-54.
He, F. and Shen, Z.J.M., 2015. Modeling taxi services with smartphone-based e-hailing applications. Transportation Research Part C: Emerging Technologies, 58, pp.93-106.
Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.
Kamga, C., Yazici, M.A. and Singhal, A., 2015. Analysis of taxi demand and supply in New York City: implications of recent taxi regulations. Transportation Planning and Technology, 38(6), pp.601-625.
Karlaftis, M.G. and Vlahogianni, E.I., 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transportation Research Part C: Emerging Technologies, 19(3), pp.387-399.
Ke, J., Zheng, H., Yang, H. and Chen, X.M., 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transportation Research Part C: Emerging Technologies, 85, pp.591-608.
Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Kuflik, T., Minkov, E., Nocera, S., Grant-Muller, S., Gal-Tzur, A. and Shoor, I., 2017. Automating a framework to extract and analyse transport related social media content: The potential and the challenges. Transportation Research Part C: Emerging Technologies 77: 275-291.
Kumar, R., Aggarwal, R.K. and Sharma, J.D., 2015. Comparison of regression and artificial neural network models for estimation of global solar radiations. Renewable and Sustainable Energy Reviews, 52, pp.1294-1299.
Lam, C. and Liu, M., 2019. Toward Inclusive Mobility: Ridesharing Mitigates Geographical Disparity in Transportation. Available at SSRN 2997190
Laptev, N., Yosinski, J., Li, L.E. and Smyl, S., 2017. Time-series extreme event forecasting with neural networks at uber. In International Conference on Machine Learning (No. 34, pp. 15).
Lavieri, P.S., Dias, F.F., Juri, N.R., Kuhr, J. and Bhat, C.R., 2018. A model of ride-sourcing demand generation and distribution. Transportation Research Record, p.0361198118756628.
LeCun, Y. and Bengio, Y., 1995. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), p.1995.
Lipton, Z.C., 2018. The mythos of model interpretability. Queue, 16(3), pp.31-57.
Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. In Advances in neural information processing systems (pp. 4765-4774).
Mao, C., Liu, Y. and Shen, Z.J.M., 2020. Dispatch of autonomous vehicles for taxi services: A deep reinforcement learning approach. Transportation Research Part C: Emerging Technologies, 115, p.102626.
Mckinley, J. and W. Hu. 2019. Congestion Pricing in Manhattan, First Such Plan in U.S., Is Close to Approval.
Molnar, C., 2020. Interpretable machine learning. Lulu. com.
National Association of City Transportation Officials. 2019. Guidelines for Regulating Shared Micromobility.

Olah, C., 2015. Understanding lstm networks.
Paliwal, M. and Kumar, U.A., 2009. Neural networks and statistical techniques: A review of applications. Expert systems with applications, 36(1), pp.2-17.
Park, H., Haghani, A. and Zhang, X., 2016. Interpretation of Bayesian neural networks for predicting the duration of detected incidents. Journal of Intelligent Transportation Systems, 20(4), pp.385-400.
Pineda, P., 2019. Lime pulls electric scooters from Tempe, citing fees, liability burden
Refaeilzadeh, P., Tang, L. and Liu, H., 2009. Cross-Validation. Encyclopedia of database systems, 5.
Ribeiro, M.T., Singh, S. and Guestrin, C., 2016, August. " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).
Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1988. Learning representations by back-propagating errors. Cognitive modeling, 5(3), p.1.
Safikhani, A., Kamga, C., Mudigonda, S., Faghih, S.S. and Moghimi, B., 2018. Spatial-temporal modelling of yellow taxi demands in New York City using generalized STAR models. International Journal of Forecasting.
Samek, W., Wiegand, T., & Müller, K. R., 2017. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.
Sanders, A. and Guse, C., 2019. NYC to impose some of the world's toughest regulations on Uber and Lyft.
Sarle, Warren S. 1994. Neural networks and statistical models.
Schneider, T., 2015. Analyzing 1.1 billion NYC taxi and Uber trips, with a vengeance.
Sheather, S., 2009. A modern approach to regression with R. Springer Science & Business Media.
Shen, Z.J.M., Feng, B., Mao, C. and Ran, L., 2019. Optimization models for electric vehicle service operations: A literature review. Transportation Research Part B: Methodological.
Sun, J., Guo, J., Wu, X., Zhu, Q., Wu, D., Xian, K. and Zhou, X., 2019. Analyzing the Impact of Traffic Congestion Mitigation: From an Explainable Neural Network Learning Framework to Marginal Effect Analyses. Sensors, 19 (10), p.2254.
Wang, F., Wang, J., Cao, J., Chen, C. and Ban, X.J., 2019. Extracting trips from multi-sourced data for mobility pattern analysis: An app-based data example. Transportation Research Part C: Emerging Technologies 105: 183-202.
Wang, R., Fan, S. and Work, D.B., 2016. Efficient multiple model particle filtering for joint traffic state estimation and incident detection. Transportation Research Part C: Emerging Technologies 71: 521-537.
Wu, X., Guo, J., Xian, K. and Zhou, X., 2018. Hierarchical travel demand estimation using multiple data sources: A forward and backward propagation algorithmic framework on a layered computational graph. Transportation Research Part C: Emerging Technologies, 96, pp.321-346.
Yang, C. and Gonzales, E.J., 2014. Modeling taxi trip demand by time of day in New York City. Transportation Research Record, 2429(1), pp.110-120.
Zhang, Z., He, Q. and Zhu, S., 2017. Potentials of using social media to infer the longitudinal travel behavior: A sequential model-based clustering method. Transportation Research Part C: Emerging Technologies 85: 396-414.
Zhao, X., Yan, X. and Van Hentenryck, P., 2019. Modeling Heterogeneity in Mode-Switching Behavior Under a Mobility-on-Demand Transit System: An Interpretable Machine Learning Approach. arXiv:1902.02904.