NATIONAL INSTITUTE FOR CONGESTION REDUCTION

FINAL REPORT
JANUARY 2024

Dynamic Prediction of Shared Micromobility Usage with Multi-Task Learning

Yujie Guo
Ying Chen
Yu Zhang

National Institute for Congestion Reduction
University of South Florida
Center for Urban Transportation Research | University of South Florida

NICR
NATIONAL INSTITUTE FOR CONGESTION REDUCTION

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Dynamic Prediction of Shared Micromobility Usage with Multi-Task Learning

Prepared by

Yujie Guo
Department of Civil and Environmental Engineering
University of South Florida

Ying Chen
Department of Civil and Environmental Engineering
Northwestern University

Yu Zhang
Department of Civil and Environmental Engineering
University of South Florida

Prepared for

National Institute for Congestion Reduction
University of South Florida
Center for Urban Transportation Research

4202 E. Fowler Avenue, ENG030, Tampa, FL 33620-5375
nicr@usf.edu

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Dynamic Prediction of Shared Micromobility Usage with Multi-Task Learning		5. Report Date January 31, 2024
		6. Performing Organization Code
7. Author(s) Yujie Guo, Ying Chen, and Yu Zhang		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL 33620 Department of Civil and Environmental Engineering, Northwestern University, Evanston, IL 60208		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 69A3551947136; #79075-31
12. Sponsoring Organization Name and Address U.S. Department of Transportation University Transportation Centers 1200 New Jersey Avenue, SE Washington, DC 20590 United States National Institute for Congestion Reduction 4202 E. Fowler Avenue Tampa, FL 33620-5375 United States		13. Type of Report and Period Covered Final Report, [August 1, 2022 – January 31, 2024]
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Taxis and Transportation Network Companies (TNCs) are important components of the urban transportation system. An accurate short-term forecast of passenger demand can help operators better allocate taxi or TNC services to achieve supply–demand balance in real time. As a result, drivers can improve the efficiency of passenger pick-ups, thereby reducing traffic congestion and contributing to the overall sustainability of the program. Previous research has proposed sophisticated machine learning and neural-network-based models to predict the short-term demand for taxi or TNC services. However, few of them jointly consider both modes, even though the short-term demand for taxis and TNCs is closely related. By enabling information sharing between the two modes, it is possible to reduce the prediction errors for both. To improve the prediction accuracy for both modes, this study proposes a multi-task learning (MTL) model that jointly predicts the short-term demand for taxis and TNCs. The model adopts a gating mechanism that selectively shares information between the two modes to avoid negative transfer. Additionally, the model captures the second-order spatial dependency of demand by applying a graph convolutional network. To test the effectiveness of the technique, this study uses taxi and TNC demand data from Manhattan, New York, as a case study. The prediction accuracy of single-task learning and multi-task learning models are compared, and the results show that the multi-task learning approach outperforms single-task learning and benchmark models.

17. Key Words		18. Distribution Statement	
Shared mobility, machine learning, demand forecast, sustainability			
19. Security Classification (of this report) Unclassified.	20. Security Classification (of this page) Unclassified.	21. No. of Pages 27	22. Price

Table of Contents

Figures

Tables

Executive Summary

This report summarizes a study on predicting the short-term demand for taxis and Transportation Network Companies (TNCs) using a multi-task learning (MTL) model. The study focuses on the Manhattan area and explores the correlation between taxi and TNC demand at different temporal levels. The MTL model, which incorporates a gating mechanism for information sharing, outperforms single-task learning models and other baseline models. The study also investigates the spatial dependency of the demand model and finds that considering the interaction of spatial dependency improves the model's performance. The findings have practical implications for TNC companies, suggesting the potential benefits of sharing information in a model to enhance forecasting accuracy and improve efficiency in resource allocation. The report suggests several potential research directions, including exploring advanced MTL techniques, investigating the modeling of transportation modes with weaker correlations, testing the generalizability of the proposed technique in different cities or for different tasks, and studying the impact of improved demand forecasting on traffic congestion. Overall, the study demonstrates the effectiveness of the MTL model in predicting short-term demand for taxis and TNCs and provides avenues for further research in the field of demand forecasting and traffic management.

Chapter 1. Introduction

Taxi and Transportation Network Company (TNC) services play essential roles in the urban transportation system. Taxis offer traditional ride-hailing services, while TNCs such as Uber connect drivers and riders using Internet-based mobile technology. The ridership of TNCs has grown rapidly. For example, Uber's ridership increased from 3.79 billion in 2017 to 6.3 billion in 2021 (Iqbal, 2023). Thank to efficient driver–passenger matching technology and a more flexible pricing model, the capacity utilization rate of TNCs (the fraction of the time/mile in which a driver takes fare-paying passengers) is much higher than that of taxis (Cramer & Krueger, 2016). Due to the popularity and innovative business model of TNCs, the ridership of taxis has lost ground in many cities. For example, in 2016, the number of TNC trips made was 12 times that of taxi trips in San Francisco (SFCTA, 2007).

Improving short-term demand forecasting for both TNCs and taxis has positive impacts on sustainability. With accurate prediction, operators can assign the right number of vehicles at the right time to reduce the idle time of drivers and waiting time for passengers, leading to an improved capacity utilization ratio. The capacity utilization ratio is seldom revealed by TNCs or taxi companies, but a study has shown its value ranged between 43.5% and 51.7% for selected cities in the United States between 2013 and 2015 (Cramer & Krueger, 2016). Improving the utilization ratio could potentially help address traffic congestion problems, improve traffic speeds, and reduce traffic emissions.

Various studies have developed short-term demand forecasting models for transportation modes. Some studies have used traditional time-series forecasting models such as autoregressive integrated moving average (ARIMA) and its variants to predict traffic demand by capturing the temporal correlation of data (Li et al., 2012; Moreira-Matias et al., 2013; Shekhar & Williams, 2007). Recently, with the advent of higher computing power and the popularization of AI technologies, many studies have applied neural network models to demand forecasting (Ke et al., 2017; Jin et al., 2020; Yoa et al., 2019).

The majority of studies mentioned earlier are based on single-task learning, where the model predicts for one transportation mode only. Recently, multi-task learning (MTL) has garnered significant attention in the AI domain, as it enables different tasks to share information, thereby enhancing the prediction accuracy. Some transportation demand forecasting research has also adopted this technique (Bai et al., 2019; Ke et al., 2021; Li et al., 2020; Liang et al., 2023; Liu et al., 2021). However, the majority of these studies allow information sharing between tasks without controlling for "negative transfer", which is common and could reduce the effectiveness of multi-task learning.

The demand patterns for TNCs and taxis are closely correlated. For example, in New York City, TNCs and taxis show similar spatial–temporal patterns (Poulsen et al., 2016; Zhao et al., 2020), and their demand is correlated with the same set of land use and sociodemographic factors (Correa, 2017). Given such correlated patterns, leveraging information sharing between the two modes could potentially improve the demand forecasting accuracy. The idea of incorporating TNC information into a taxi demand forecasting model has been experimented with, which showed an improvement in the model prediction accuracy (Zhao et al., 2020). However, this study did not embed taxi information into a TNC model, and there have been no studies developing a multi-task learning model to simultaneously predict the demand for these two modes. In New York City, Yellow Cabs can be hailed on the Uber app (Hu et al., 2023), a partnership that provides an opportunity for data sharing and modeling, which could potentially improve the demand forecasting accuracy for both the taxi and TNC modes.

To capture the spatial dependency of demand, previous studies utilized first-order relationships to construct spatial graphs, such as considering the distance between two zones (Li et al., 2017; Yu et al., 2021) or determining whether two zones are neighbors (Geng et al., 2019). These researchers commonly assume that zones closer to each other or in proximity have a stronger relationship. This research explores a higher order of spatial dependency, which could capture more comprehensively the interaction of spatial relationships, which has not been explored in the literature yet.

To fill in the research gaps, this study proposes a multi-task learning approach to forecast the demand for taxis and TNCs simultaneously. The model adopts a gating mechanism that selectively shares information between the two modes to avoid the negative transfer that commonly occurs in MTL. In addition, the model also captures the second-order spatial dependency of the demand by applying a graph convolutional network. The contribution of this study is threefold:

- The evolving shared mobility sector longs for better demand prediction for different formats of sharing services. This study proposes a multi-task learning model to predict the demand for the TNC and taxi modes simultaneously to meet these needs.
- This study explores methodological improvements to increase the prediction accuracy. The techniques considered include a gating mechanism to mitigate the negative transfer between the two modes and spatial embedding, capturing the interaction of spatial dependency.
- Extensive experiments are conducted using actual taxi and TNC trip data from Manhattan, NYC. The experimental results show that the proposed modeling approach outperforms the single-task learning model and other benchmark learning models.

Chapter 2. Literature Review

This section reviews the research-related methods for capturing spatiotemporal dependency and multi-task learning.

2.1. Modeling Spatial – Temporal Dependency of Transportation Demand

Utilizing telematics technology, taxi companies can collect detailed data for each trip, including pick-up/drop-off zones, start/end time, and trip trajectories. This produces a vast amount of data, which has attracted significant research interest. Researchers have explored the data in various ways, including analyzing spatial–temporal demand patterns (Bischoff et al., 2015; Dong et al., 2019; Nuzzolo et al., 2018), exploring the impact of urban structures (e.g., land use patterns, access to different transportation modes, etc.) on taxi demand (Yang et al., 2018; Nuzzolo et al., 2019), building short-term demand forecasting models (Kuang et al., 2019; Luo et al., 2020), and developing models for the visual querying of taxi trip data (Ferreira et al., 2013). These efforts help us understand travel behaviors and support evidence-based policymaking.

To conduct short-term demand forecasting, capturing the spatial–temporal correlation is essential. For a neural-network-based forecasting model, a common approach is to stack spatial layers and temporal layers in the models, and this approach has been adopted in (Ke et al. 2017; Lu et al., 2021).

Modeling spatial dependency can improve the prediction accuracy (Geng et al., 2019, Xu et al., 2019; Yao et al., 2016). There are generally two techniques used to capture the spatial dependency of zones: convolutional neural networks (CNNs) and graph neural networks (GNNs). The first approach requires the study area to be partitioned into regular grids, such as image pixels, and the demand of a zone is analogous to the value of a pixel in an image (Ke et al. 2017). The second approach can handle non-Euclidean structural data such as friendship networks, transportation networks, etc. Due to their flexibility and potentially better performance, graph neural networks have become more popular in recent years (Defferrard et al., 2016; Kipf & Welling 2016). To construct a graph of a transportation network, nodes are defined as zonal areas, and edges are defined in various ways depending on the specific definition. Edges can be defined based on whether two zones have traffic flow (Xu et al., 2019), whether two zones are spatial neighbors, whether zones are connected by major roads or have simi-lar POIs (Geng et al., 2019), the distance between nodes (Li et al., 2017; Yu et al, 2017), etc. However, in the existing literature, the aforementioned definitions are limited to first-order spatial dependency.

To capture the temporal dependency of transportation demand, popular techniques include Long Short-Term Memory (LSTM) (Xu et al., 2017) and Gated Recurrent Units (GRUs) (Jin et al., 2020). Compared to LSTM, GRUs have a lighter computing burden and still achieve comparable performance. Historical time steps might contribute differently to the forecasting of the next time stamp. Hence, the attention method could be applied to extract historical time steps that are important to demand. The attention mechanism has been shown to be effective in improving the demand forecasting accuracy (Geng et al., 2019; Jin et al., 2020; Liu et al., 2029; Zhao et al., 2023).

2.2. Multi-Task Learning

MTL involves learning multiple related tasks simultaneously to improve the generalization performance of the forecasting model. In the transportation demand forecasting domain, the application of MTL is thriving. Some studies have applied MTL to predicting different tasks for one transportation mode. A task includes predicting the demand in a zone (Luo et al, 2020; Zhang et al., 2019) or predicting pick-up/drop-off (Kuang et al., 2019), etc. Though these studies show the benefits of MTL in improving the prediction accuracy, they simply share information between tasks without differentiating between positive and negative information. There were also a few studies we found that applied MTL to predicting the demand for multiple transportation modes. For example, one study developed a knowledge adaptation module that boosted the prediction of transportation modes with fewer stations (e.g., ferries) by adapting the demand pattern from station intensive modes (e.g., buses). The model results show that MTL improves the demand forecasting performance for modes with fewer stations (Li et al., 2020). Another study we found looked into demand prediction for the subway and TNCs (Liang et al., 2020).

In an MTL model, sharing parameters between tasks is not always successful; if shared tasks are not closely related or information is shared too extensively, this can affect the model performance. This phenomenon is called "negative transfer" and is common in applications such as natural language processing Ruder et al., 2019) and computer vision (Strezoski et al., 2019). To reduce the negative influence of task sharing, some MTL studies have attempted to answer questions on which layers to share, what parameters to share, how to address implicit or explicit task relationships, and how to define the im-portance of tasks (Misra et al., 2016; Ruder et al., 2019; Strezoski et al., 2019). These MTL approaches share full or partial features between tasks without discerning their helpfulness, while gated MTL (Xiao et al., 2018) adopts a gating mechanism called a Gated Sharing Unit that can filter the feature flows between tasks and greatly reduce task inference.

Chapter 3. Methodology

This section defines the problem of demand forecasting for taxis and TNCs, introduces a single-task learning model that can be used for predicting taxi or TNC demand, and also describes the gating mechanism that is used to build the multi-task learning model.

3.1. Preliminary: Problem Definition

The demand forecasting problem aims to predict the demand for multiple transportation modes M for the study areas A at the time interval t + 1 given historical demands until time interval t, where A = {a1, a2, …, an} is denoted as the set of areas; M = {m1, m2, … mj} is the set of transportation modes; the set of time sequences is denoted as I = {1, 2 …, t, …, T}; and the historical time sequence can be defined as $\boldsymbol{l}_t = \{t-k, t-k+1, \dots, t\}$, where k is a recall factor. Mathematically, the problem can be defined as

$$\boldsymbol{y}_{t+1}^{\boldsymbol{A},\boldsymbol{M}} = F(\boldsymbol{y}_{t-k}^{A,M}, \dots \boldsymbol{y}_{t}^{A,M}) \tag{1}$$

where $\boldsymbol{y}_{t+1}^{\boldsymbol{A},\boldsymbol{M}}$ is the transportation demand for areas A and modes M at t + 1, and F(·) is the forecasting function with inputs on historical passenger demand for the transportation modes. The following section firstly describes the single-task learning model, and then introduces the MTL model that is built based on the single-task learning model.

3.2. Single-Task Learning Model

A single-task learning model stacks GCN and LSTM layers and incorporates an attention layer to enhance the forecasting accuracy. The model is designed to capture the spatial–temporal dependency of transportation demand and is composed of six layers, as shown in Figure 1.

The first layer is an input layer, with the input being the historical demand for taxis or TNCs. The input is then passed to the GCN layer to capture the spatial dependency, and its output is then passed to LSTM to capture the temporal dependency. The fourth layer is an attention layer, which assigns different weights to the output from LSTM. Higher weights are assigned to the outputs that are more correlated with our prediction. The fifth layer is a fully connected layer, and the last layer is an output layer. Next, we will explain each layer in more detail.

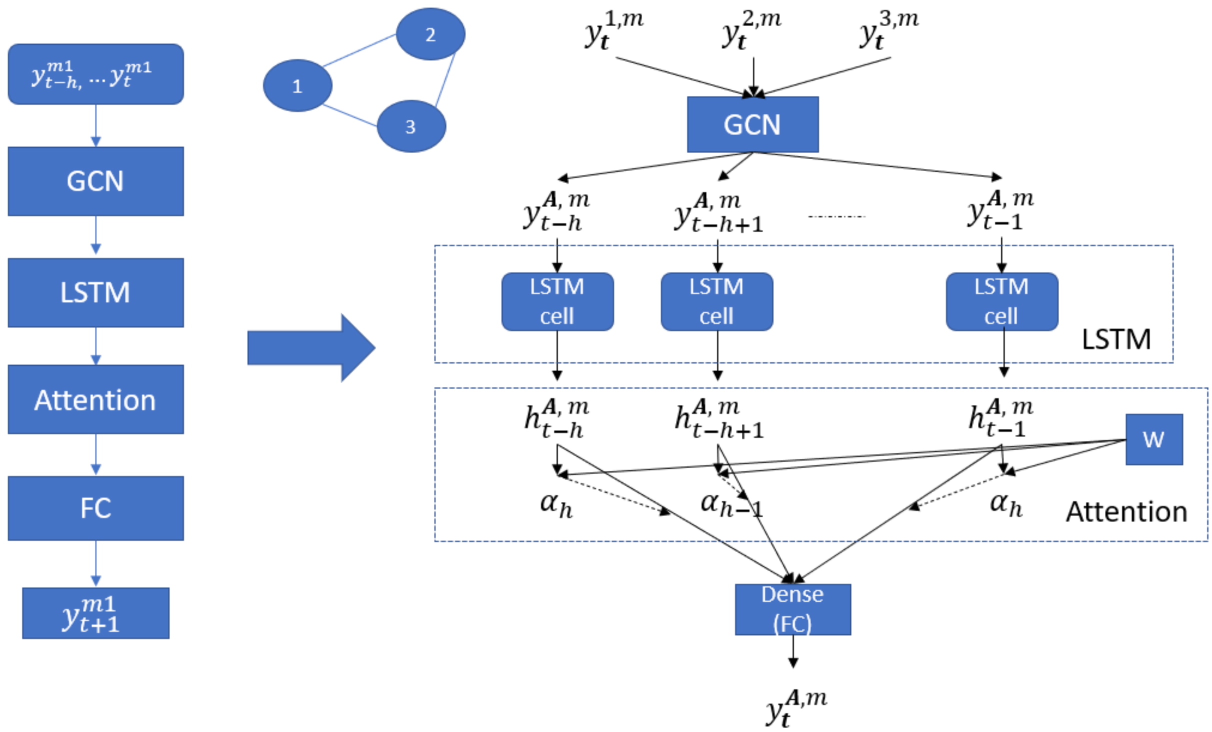

Figure 1. Single-task learning model (base model).

Input Layer: The input layer is the input for the model. It is the historical demand for taxis or TNCs for different zones.

Graph Convolutional Layer: In the context of transportation, a network can be depicted as a graph, with nodes representing various entities like taxi zones, communities, neighborhoods, traffic analysis zones, or census tracts and edges indicating relationships between nodes (e.g., neighboring taxi zones). The signal of a node refers to the historical demand from its corresponding zone. A graph convolutional network (GCN) (Kipf & Welling, 2016) works by smoothing a node's signal through the transformation and aggregation of the demand data from its neighboring nodes (e.g., nearby taxi zones). The graph convolutional layer is defined as

$$\boldsymbol{Z} = \widetilde{\boldsymbol{A}}\boldsymbol{X}\boldsymbol{W} \tag{2}$$

where $\boldsymbol{Z} \in \boldsymbol{R}^{N\times D}$ is the output from the GCN, $\widetilde{\boldsymbol{A}} \in \boldsymbol{R}^{N\times N}$ is the normalized adjacency matrix with self-loops, $\boldsymbol{X} \in \boldsymbol{R}^{N\times K}$ is the input for the GCN, and $\boldsymbol{W} \in \boldsymbol{R}^{K\times D}$ is the learned weight. A GCN structure is adopted in this study to capture the spatial dependency between zones.

Long Short-Term Memory Layer: LSTM is adopted in this study to capture temporal dependency. LSTM has been popularly used in demand forecasting research, such as the studies (Luo et al., 2020; Yao et al., 2018). An LSTM cell has the structure shown in Figure 2. Each cell has inputs of $\boldsymbol{x}_t$, a hidden state $\boldsymbol{h}_{t-1}$, and a cell state $\boldsymbol{c}_{t-1}$ and outputs the hidden state $\boldsymbol{h}_t$ as the final output or as the input to the next cell and $\boldsymbol{c}_t$ to the next cell state. The structure within the cell has the ability to decide what information to store or throw away for cell state c; it continues to update based on different time steps and finally decides the output. A detailed explanation of LSTM is provided in (Olah, 2020). The formulation for the computation in an LSTM cell is shown in Figure 2 and explained thereafter.

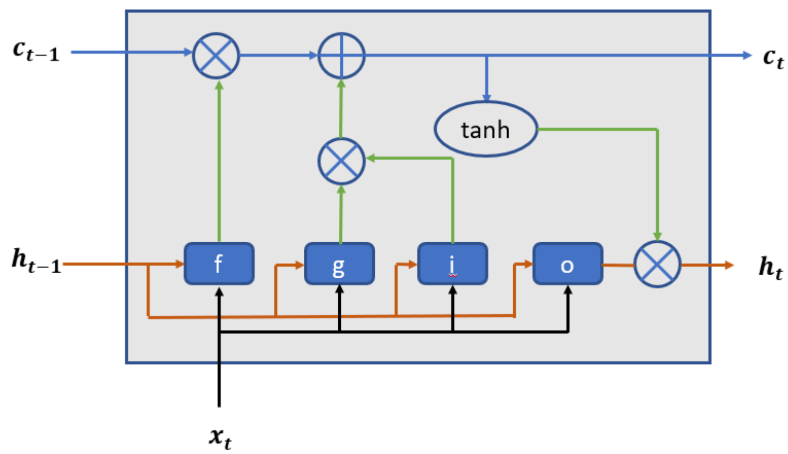

Figure 2. LSTM cell structure.

$$f = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

$$i = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

$$g = tanh(w_g \cdot [h_{t-1}, x_t] + b_g) \quad (5)$$

$$o = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$c_t = f \odot c_{t-1} + i \odot g \quad (7)$$

$$h_t = o \odot \tanh(c_t) \quad (8)$$

where σ is a sigmoid function that is given by $\sigma(x) = \frac{1}{1+e^x}$. It outputs values between 0 and 1, which controls the flow of information. The cross symbol in the figure refers to multiplication, and the plus symbol is a merge function that outputs the sum of the inputs. $\odot$ is elementwise multiplication. $\boldsymbol{w}_f$, $\boldsymbol{w}_g$, $\boldsymbol{w}_i$, $\boldsymbol{w}_o$, $\boldsymbol{b}_f$, $\boldsymbol{b}_g$, $\boldsymbol{b}_i$, and $\boldsymbol{b}_o$ are trainable parameters.

Attention Layer: The attention mechanism applied in this study is from Yang et al. (Yang et al., 2016), which had success in dealing with sequence learning tasks. Mathematically, the attention method is defined as:

$$u_i = \tanh(W_w h_i + b_w) \quad (9)$$

$$\alpha_i = \frac{\exp(u_i{}^T u_w)}{\sum_{i \in l_t} \exp(u_i{}^T u_w)} \quad (10)$$

$$\widetilde{h_t} = \sum_{i \in l_t} \alpha_i h_i \quad (11)$$

where the hidden output $\boldsymbol{h}_i = \{h_{t-k}, \dots h_t\}$ is fed into Equation (9) to obtain $\boldsymbol{u}_i$ as the hidden representation of $\boldsymbol{h}_i$. Then, the importance of the time step (t – k) is measured as the similarity between $\boldsymbol{u}_i$ and $\boldsymbol{u}_w$, which is normalized to $\boldsymbol{\alpha}_i$ using a softmax function. Finally, the output $\widetilde{\boldsymbol{h}_t}$ is the weighted sum of the hidden representation of $\boldsymbol{h}_i$. $\boldsymbol{u}_w$ in Equation (10) functions as the high-level representation of the "important time step". During training, $\boldsymbol{u}_w$, $\boldsymbol{W}_w$, and $\boldsymbol{b}_w$ are randomly initialized and jointly learned in the model.

Fully Connected Layer: FC refers to the fully connected layer or dense layer in a neural network. This layer has the number of neurons that is equal to the number of zones for forecasting. The neurons are connected to every neuron in the preceding layer.

Output Layer: The output layer generates the future demand (e.g., next hour) of different zones.

3.3. Multi-Task Learning Model

To build a multi-task learning model, this study adopts a "gating" mechanism called a Gated Sharing Unit (GSU) (Xia et al., 2018) as shown in Figure 3b. A GSU allows the model to filter features from other tasks and select those that are useful to the task; it avoids harmful feature inference if two feature maps are concatenated directly. The overall architecture of the model is depicted in Figure 3a.

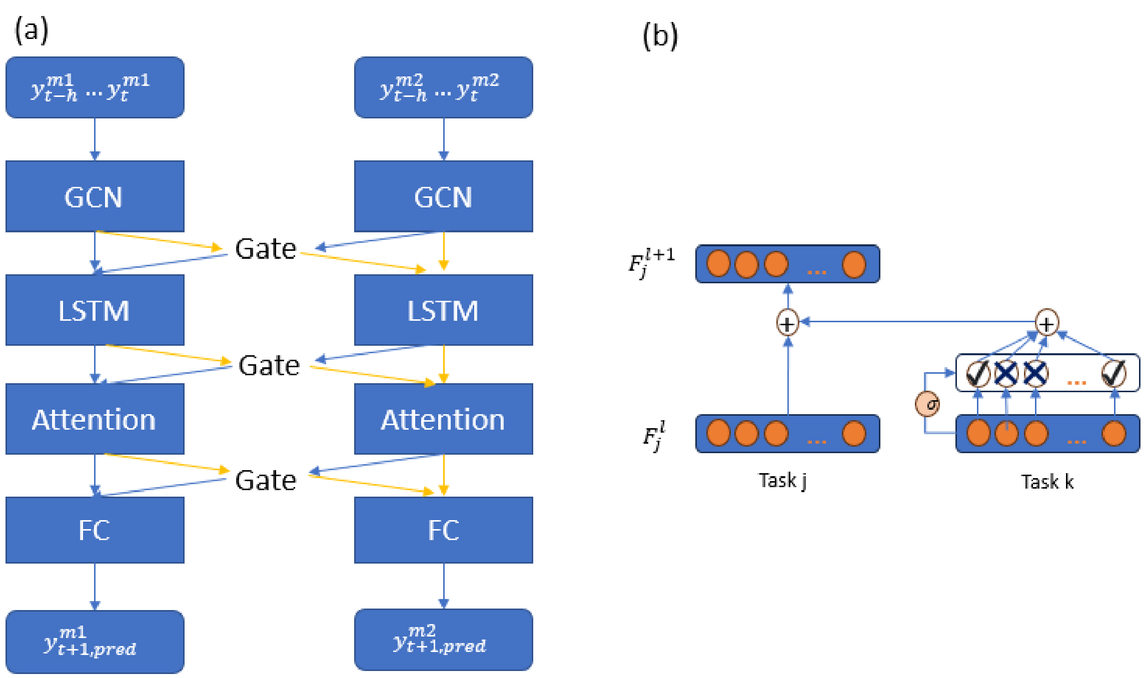

Figure 3. (a) Multi-task learning model, (b) Gated Sharing Unit (Xiao et al., 2018).

There are two steps involving a GSU. Assume that there are two modes, j and k. The first step computes how much information will be merged from mode k to mode j. For this purpose, a gate is inserted to select the useful features from mode k, which is calculated using

$$\boldsymbol{g}_{jk}^{l} = \sigma(\boldsymbol{W}_{jk}^{l} \cdot \boldsymbol{F}_{k}^{l} + \boldsymbol{b}_{jk}^{l}) \tag{12}$$

where l is the level of the layers, and σ is a sigmoid function that guarantees the values of g are bounded between 0 and 1. $\boldsymbol{W}_{jk}^{l}$ is the weight that will be trained, $\boldsymbol{F}_{k}^{l}$ are the output parameters of mode k in layer l, and $\boldsymbol{b}_{jk}^{l}$ is the bias term. $\boldsymbol{g}_{jk}^{l}$ is a vector. The gate controls how much information from mode k at layer l will be passed to mode j. As shown in Figure 3b, the check mark indicates that more information from the preceding neuro will contribute to task j, while the cross mark indicates less information contribution.

The second step computes the merge of features between mode j and mode k. It can be calculated using the following equation:

$$\boldsymbol{F}_j^{l+1} = \sum_{k \neq j} \boldsymbol{g}_{jk}^{l} \odot \boldsymbol{F}_k^{l} + \boldsymbol{F}_j^{l} \tag{13}$$

where $\odot$ denotes elementwise multiplication. This formula outputs the fused parameters $\boldsymbol{F}_j^{l+1}$. From this equation, the features from mode j are directly passed to the next layer, and the features from mode k are merged into mode j after filtering using the gate $\boldsymbol{g}_{jk}^{l}$.

Chapter 4. Experiments and Model Performance Evaluation

This section describes the experiment settings and presents a performance evaluation of the proposed models.

4.1. Study Area

4.1.1. Study Site Selection and Data Preprocessing

This study selected Manhattan, New York, as the case study area, as both Yellow Cabs and TNCs service that area, and trip data are publicly accessible from the NYC Taxi & Limousine Commission (NYC Taxi & Limousine Commission, 2023). One-year trip data for 2018 are retrieved for the study, including Yellow Cab and For-Hire Vehicle (FHV) trip data. The FHV data includes Uber, Lyft, and other platforms that allow passengers to use apps to request trip services.

For both transportation modes, information such as trip pick-up zone, drop-off zone, pick-up time, and drop-off time is selected from the dataset. To remove erroneous trip records, trips are filtered according to travel time and travel distance. The minimum travel time duration is set to 1 min, with the maximum set to 2 h and the minimum travel distance set to be greater than 0.2 miles. This results in a dataset containing 87.1 million records of taxi trips and 99.3 million records of TNC trips within Manhattan. The trip data are aggregated at the hourly level, with each zone representing the hourly demand for taxi and TNC services. In total, the processed dataset comprises 59 taxi zones (features) and 8760 time steps (365 days × 24 h per day).

4.1.2. Data Analysis

Figure 4 illustrates the aggregated monthly, daily, and hourly trips for taxi and TNC services, while Figure 5 shows the total trip counts for both transportation modes. The correlation coefficient in Figure 4 is calculated using Pearson's correlation method. As depicted in Figure 4a, there is an inverse relationship between taxi and TNC demand on a monthly basis—TNC demand displays a rising trend while taxi demand declines. This suggests a competitive dynamic between TNC and taxi services in the Manhattan area. At the aggregate trip level (Figure 5a), a seasonal pattern emerges, with both taxi and TNC trips peaking in popularity during October and March and decreasing during the summer and winter months. The daily (Figure 4b) and hourly (Figure 4c,d) patterns reveal a similar temporal demand pattern for both TNC and taxi services, as indicated by the positive correlation coefficient. The similarity in demand patterns is also evident in the total demand analysis (Figure 5b–d). The strong correlation between the two modes suggests MTL is a suitable approach to jointly modeling taxi and TNC demand. Examining Figure 4c,d, it is observed that the TNC demand generally tends to be slightly higher, with both modes following a comparable hourly trend. However, there are instances where the taxi demand equals or exceeds the TNC demand at certain hours, indicating temporal volatility at a finer granularity. This volatility may introduce noise into the MTL approach if information sharing between the modes is simply uniform. Therefore, it is crucial for the MTL model to selectively filter unnecessary information for effective information exchange.

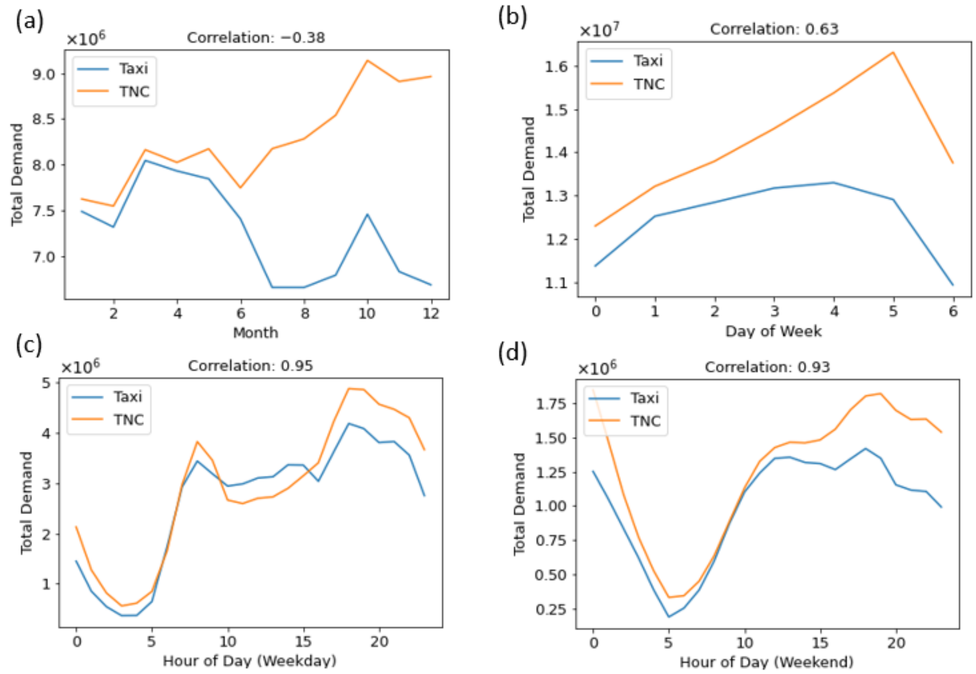

Figure 4. Demand correlation of taxis and TNCs at different temporal levels. (a) Monthly demand correlation, (b) Daily correlation for day of the week, (c) Hourly correlation for weekdays, (d) Hourly correlation for weekends.

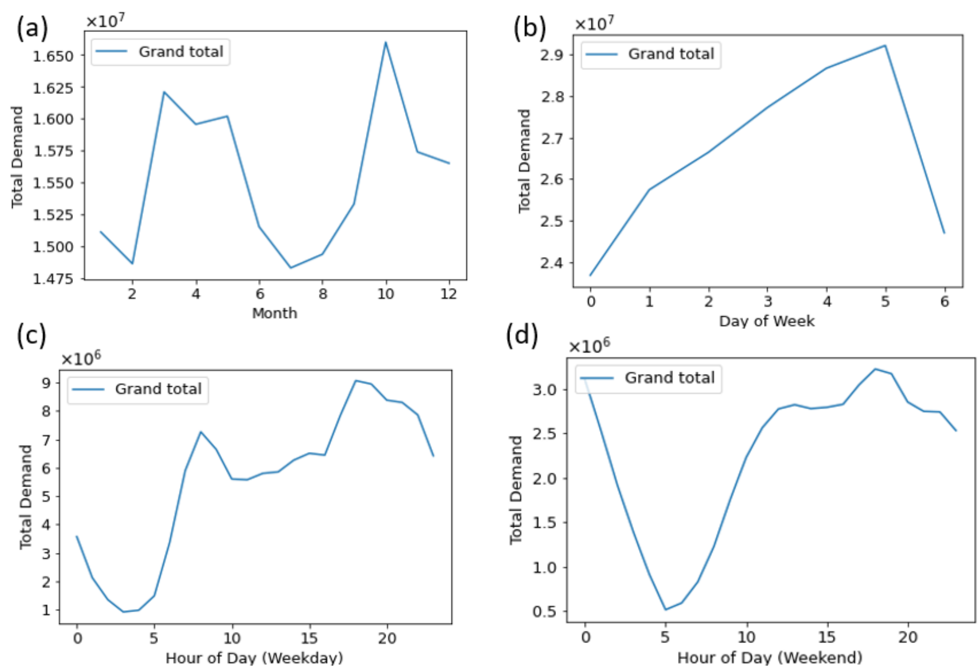

Figure 5. Total temporal demand for taxis and TNCs. (a) Total monthly demand, (b) Total daily demand for day of the week, (c) Total hourly demand on weekdays, (d) Total hourly demand for weekends.

The above analysis is conducted for all study areas. A similar hourly pattern is also identified at the local level, as shown in Figure 6. The correlation coefficient is computed for each taxi zone at the hourly level and is positive for all zones, as shown in Figure 5, which suggests the close short-term demand correlation between taxis and TNCs. Thus, sharing the information between the two modes could potentially be beneficial to the model.

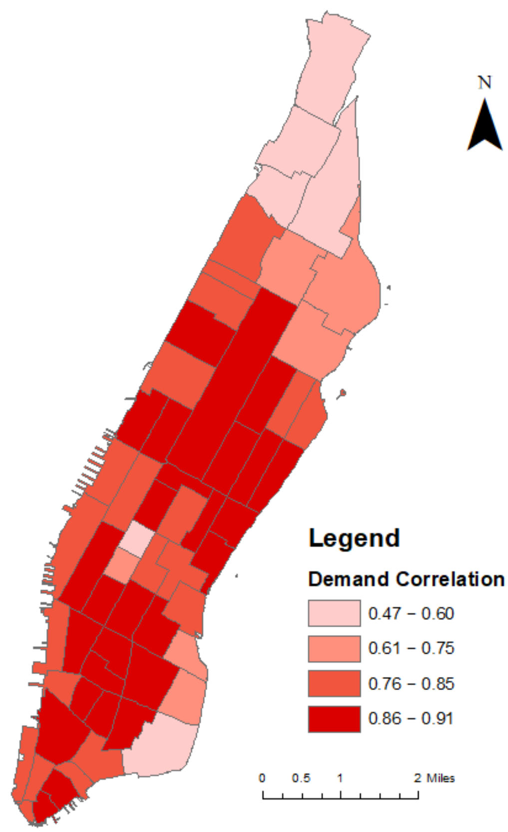

Figure 6. Demand correlation for taxis and TNCs.

4.2. Model Training

The 2018 one-year dataset included 8760 time steps, with the data arranged sequentially by time. The first 85% (approximately 310 days) is used for training, and the remaining 15% is used for testing. The "looking back" time step is set as 12, which means 12 h historical demand is used to forecast the demand for the next time step (next hour).

The constructed model has four layers: the GCN, LSTM, attention, and dense layers (as illustrated in Figures 1 and 3). TensorFlow 2.1 (Abadi et al., 2016), an open-source library renowned for training neural network models, is utilized for training the model. Training stops after the training loss is higher than the minimum training loss for five consecutive epochs. The model is implemented using the Python programming language, and the hardware used for model training includes an Intel(R) Core(TM) i7-9750H CPU with 16 GB of RAM.

4.3. Model Evaluation

4.3.1. Description of the Baseline Models and the Proposed Models in the Experiment

To demonstrate the performance of the proposed MTL model, besides the single-task learning model, several popular time-series models are also selected for comparison. The baseline models include the following:

- ARIMA: An autoregressive integrated moving average model, a statistical model widely used for time-series forecasting.
- MLP: A multi-layer perception, the most basic neural network. In this study, a three-layer neural network is used, which includes an input layer, a dense layer, and an output layer.
- XGBoost: eXtreme Gradient Boosting, which applies boosting to a tree-based machine learning model—widely known as an efficient model that solves data science problems accurately (Chen et al., 2016).

To test the effectiveness of MTL and the interaction of spatial dependency, we also compare models that do not consider spatial dependency, considering first-order spatial dependency, and considering the interaction of spatial dependency. Each variation of single-task learning and MTL is built. Specifically, we have the models listed as below:

- Single-task learning (without a GCN): Single-task learning model shown in Figure 1 without a GCN layer.
- Multi-task learning (without a GCN): MTL model shown in Figure 3 without a GCN layer. The Gated Sharing Unit is applied after the LSTM layer.
- Single-task learning (GCN-Distance): Single-task learning model shown in Figure 1, with graph edge defined as the inverse distance between zones.
- Multi-task learning (GCN-Distance): MTL model shown in Figure 3, with graph edge defined as the inverse distance between zones.
- Single task learning (GCN-Neighbor): Single-task learning model shown in Figure 1, with graph edge defined as 1 if two zones share boundaries and 0 otherwise.
- Multi-task learning (GCN-Neighbor): MTL model shown in Figure 3, with graph edge defined as 1 if two zones share boundaries and 0 otherwise.
- Single task learning (GCN-Interaction): Single-task learning model shown in Figure 1, with graph edge defined as the product of inverse distance dependency and neighbor dependency.
- Multi-task learning (GCN-Interaction): MTL model shown in Figure 3, with graph edge defined as the product of inverse distance dependency and neighbor dependency.

4.3.2. Evaluation Metrics

To compare the performance of these models, this study adopted two metrics which are popularly used for regression tasks—the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)—given by

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

where y_i is the ground truth value and $\hat{y}_i$ is the predicted value.

4.4. Results and Discussions

This section compares the performance of the baseline models and the proposed models. Table 1 shows the results of the evaluation metrics for the different models. Overall, the proposed models have a better performance compared to the baseline models, the MTL models beat the single-task learning models, and considering spatial interactions brings additional benefits. Looking into the details of the model evaluation, it is interesting to see that the XGBoost model has an RMSE of 36.9 and an MAE of 21.8 for the taxi mode, which is slightly better than the single-task learning model without a GCN, suggesting that the XGBoost model, in this case, performs very well for time-series forecasting. When a gating unit is applied (MTL), the multi-task learning model without a GCN performs better than the single-task learning model without a GCN and XGBoost, suggesting the effectiveness of parameter sharing in MTL. Table 1 also shows the model performance when the distance dependency or neighbor dependency is considered, and their prediction accuracy outperforms the models that do not consider a GCN. When the distance dependency is captured, the taxi prediction errors are lower than those for the model capturing neighbor dependency, but the model's TNC prediction errors are a bit higher. Finally, we also test the model that considers the interaction between distance dependency and neighbor dependency. As Table 1 shows, the model performance further improves, and, again, the MTL model outperforms the single-task learning model, which makes MTL (GCN-Interaction) the best model.

Table 1. Model performance comparison among different methods.

	Taxi		TNC	
	RMSE	**MAE**	**RMSE**	**MAE**
ARIMA	54.1	32.6	56.3	37.1
MLP	47.9	30.0	49.5	34.4
XGBoost	36.9	21.8	41.0	26.1
Single-task learning (without a GCN)	37.7	22.6	41.0	27.2
Multi-task learning (without a GCN)	36.1	21.8	39.5	26.1
Single-task learning (GCN-Distance)	36.7	21.8	40.2	26.2
Multi-task learning (GCN-Distance)	35.8	21.5	40.5	25.9
Single-task learning (GCN-Neighbor)	37.4	22.2	39.8	26.1
Multi-task learning (GCN-Neighbor)	36.5	21.9	39.4	25.5
Single-task learning (GCN-Interaction)	35.8	21.1	38.2	25.0
Multi-task learning (GCN-Interaction)	**34.7**	**20.9**	**37.2**	**24.2**

To visualize the model's performance, a random sample of the predicted and actual demand for one day (24 time stamps) from the test data was taken. Figure 7a shows the forecasted demand and real demand for taxis averaged across all zones, with a similar representation for TNCs shown in Figure 7b. Both figures reveal a close match between the forecasted and observed demand, indicating the good performance of the model.

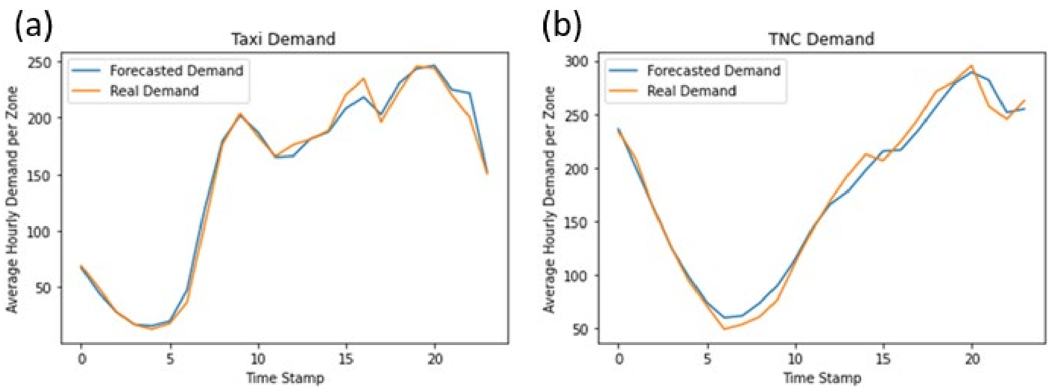

Figure 7. Real and forecasted demand for (a) taxis, (b) TNCs.

Chapter 5. Conclusions

This study develops a multi-task learning model for predicting the short-term demand of taxis and TNCs. The study selects Manhattan as the case study area and explores the short-term and long-term demand correlation for taxis and TNCs. At the short-term (hourly) level, the demand for taxis and TNCs presents similar patterns, which indicates it could be beneficial to share information between the two modes in a model. The developed multi-task learning model employs a gating mechanism that selectively shares information across the two modes. The experimental results and a model performance comparison show that MTL outperforms single-task learning and other baseline models. This study also investigates the spatial dependency of the demand model, and considering the interaction of spatial dependency outperforms the first-order dependency that is commonly used in the literature.

Given the effectiveness of the methodology, TNC companies can leverage this technique to enhance their forecasting accuracy, leading to various improvements in resource allocation efficiency. For example, by accurately predicting spikes in demand within specific areas, TNC companies can strategically deploy TNC or taxi drivers to minimize the wait time for passengers. Additionally, short-term demand forecasting also facilitates the anticipation of traffic congestion in particular areas, enabling TNCs to optimize their routes. From a traffic management standpoint, integrating predictions of demand for taxi and TNC services into existing intelligence transportation systems can effectively contribute to reducing traffic congestion and enhancing the re-liability of transportation options.

Several potential research directions could be extended from this study. First, while this study applies effective MTL techniques, it would be worthwhile to explore other advanced MTL techniques, such as gradient surgery (Yu et al., 2020), to test whether the prediction errors can be further reduced. A summary of the MTL literature is available in (Crawshaw, 2020). Second, some transportation modes may exhibit weaker correlations but still have significant implications, such as shared e-scooters and TNCs, which have a competing relationship (Gua & Zhang, 2021). Investigating whether MTL can effectively model these modes would be an interesting avenue of research. Third, the GSU technique could be tested with data from different cities or for different tasks (e.g., traffic flow, TNC/taxi forecasting) to demonstrate its generalizability. Fourth, while improved demand forecasting can benefit route planning, the impact of this forecasting on traffic congestion remains a question worth exploring.

References

Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv 2016, arXiv:1603.04467.

Bai, L.; Yao, L.; Kanhere, S.S.; Yang, Z.; Chu, J.; Wang, X. Passenger demand forecasting with multi-task convolutional recurrent neural networks. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Macau, China, 14–17 April 2019; pp. 29–42.

Bischoff, J.; Maciejewski, M.; Sohr, A. Analysis of Berlin's taxi services by exploring GPS traces. In Proceedings of the 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Budapest, Hungary, 3–5 June 2015; pp. 209–215.

Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

Correa, D.; Xie, K.; Ozbay, K. Exploring the taxi and Uber demand in New York City: An empirical analysis and spatial modeling. In Proceedings of the 96th Annual Meeting of the Transportation Research Board, Washington, DC, USA, 8–12 January 2017.

Cramer, J.; Krueger, A.B. Disruptive change in the taxi business: The case of Uber. Am. Econ. Rev. 2016, 106, 177–182.

Crawshaw, M. Multi-task learning with deep neural networks: A survey. arXiv 2020, arXiv:2009.09796.

Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 3837–3845.

Dong, X.; Zhang, M.; Zhang, S.; Shen, X.; Hu, B. The analysis of urban taxi operation efficiency based on GPS trajectory big data. Phys. A Stat. Mech. Its Appl. 2019, 528, 121456.

Ferreira, N.; Poco, J.; Vo, H.T.; Freire, J.; Silva, C.T. Visual exploration of big spatio-temporal urban data: A study of New York City taxi trips. IEEE Trans. Vis. Comput. Graph. 2013, 19, 2149–2158.

Geng, X.; Li, Y.; Wang, L.; Zhang, L.; Yang, Q.; Ye, J.; Liu, Y. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 3656–3663.

Guo, Y.; Zhang, Y. Understanding factors influencing shared e-scooter usage and its impact on auto mode substitution. Transp. Res. Part D Transp. Environ. 2021, 99, 102991.

Hu, W.; Browning, K.; Zraick, K. Uber Partners with Yellow Taxi Companies in N.Y.C. Available online: https://www.nytimes.com/2022/03/24/business/uber-new-york-taxis.html (accessed on 30 March 2023).

Iqbal, M. Uber Revenue and Usage Statistics. Available online: https://www.businessofapps.com/data/uber-statistics/ (accessed on 28 March 2023).

Jin, G.; Cui, Y.; Zeng, L.; Tang, H.; Feng, Y.; Huang, J. Urban ride-hailing demand prediction with multiple spatio-temporal information fusion network. Transp. Res. Part C Emerg. Technol. 2020, 117, 102665.

Ke, J.; Feng, S.; Zhu, Z.; Yang, H.; Ye, J. Joint predictions of multi-modal ride-hailing demands: A deep multi-task multi-graph learning-based approach. Transp. Res. Part C Emerg. Technol. 2021, 127, 103063. https://doi.org/10.1016/j.trc.2021.103063.

Ke, J.; Zheng, H.; Yang, H.; Chen, X.M. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transp. Res. Part C Emerg. Technol. 2017, 85, 591–608.

Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. arXiv 2016, arXiv:1609.02907.
Kuang, L.; Yan, X.; Tan, X.; Li, S.; Yang, X. Predicting taxi demand based on 3D convolutional neural network and multi-task learning. Remote Sens. 2019, 11, 1265.
Li, C.; Bai, L.; Liu, W.; Yao, L.; Waller, S.T. Knowledge adaption for demand prediction based on multi-task memory neural network. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Online, 19–23 October 2020; pp. 715–724.
Li, X.; Pan, G.; Wu, Z.; Qi, G.; Li, S.; Zhang, D.; Zhang, W.; Wang, Z. Prediction of urban human mobility using large-scale taxi traces and its applications. Front. Comput. Sci. 2012, 6, 111–121.
Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. arXiv 2017, arXiv:1707.01926.
Liang, J.; Tang, J.; Gao, F.; Wang, Z.; Huang, H. On region-level travel demand forecasting using multi-task adaptive graph attention network. Inf. Sci. 2023, 622, 161–177.
Liang, Y.; Huang, G.; Zhao, Z. Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach. Transp. Res. Part C Emerg. Technol. 2022, 140, 103731.
Liu, H.; Wu, Q.; Zhuang, F.; Lu, X.; Dou, D.; Xiong, H. Community-Aware Multi-Task Transportation Demand Prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021.
Liu, Y.; Liu, Z.; Lyu, C.; Ye, J. Attention-based deep ensemble net for large-scale online taxi-hailing demand prediction. IEEE Trans. Intell. Transp. Syst. 2019, 21, 4798–4807.
Luo, H.; Cai, J.; Zhang, K.; Xie, R.; Zheng, L. A multi-task deep learning model for short-term taxi demand forecasting considering spatiotemporal dependences. J. Traffic Transp. Eng. 2020, 8, 83–94.
Misra, I.; Shrivastava, A.; Gupta, A.; Hebert, M. Cross-stitch networks for multi-task learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3994–4003.
Moreira-Matias, L.; Gama, J.; Ferreira, M.; Mendes-Moreira, J.; Damas, L. Predicting taxi–passenger demand using streaming data. IEEE Trans. Intell. Transp. Syst. 2013, 14, 1393–1402.
NYC Taxi & Limousine Commission. TLC Trip Record Data; NYC Taxi & Limousine Commission: New York, NY, USA, 2023.
Nuzzolo, A.; Comi, A.; Papa, E.; Polimeni, A. Understanding taxi travel demand patterns through Floating Car Data. In Proceedings of the Conference on Sustainable Urban Mobility, Skiathos Island, Greece, 24–25 May 2018; Springer: Cham, Switzerland, 2018; pp. 445–452.
Nuzzolo, A.; Comi, A.; Polimeni, A. Exploring on-demand service use in large urban areas: The case of Rome. Arch. Transp. 2019, 50, 77–90.
Olah, C. Understanding LSTM Networks. Available online: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed on 1 April 2020).
Poulsen, L.K.; Dekkers, D.; Wagenaar, N.; Snijders, W.; Lewinsky, B.; Dan, T.; Lange, M.; Kogenhop, M. Using Big Data to Optimize Real-Time Operations in On-Demand Ride-Hailing Services. Transp. Res. Procedia 2021, 52, 225–233.
Poulsen, L.K.; Dekkers, D.; Wagenaar, N.; Snijders, W.; Lewinsky, B.; Dan, T.; Lange, M.; Kogenhop, M. Using Big Data to Optimize Real-Time Operations in On-Demand Ride-Hailing Services. Transp. Res. Procedia 2021, 52, 225–233.
Rahman, S.; Wei, S. Analyzing Demand for Ride-hailing Services in New York City: A Machine Learning Approach. Transp. Res. Rec. 2019, 2673, 12–22.
Rana, R.; Hu, X.; Shahabi, C.; Song, Y. Spatio-temporal demand prediction for ride-hailing services using multi-task learning. ACM Trans. Intell. Syst. Technol. 2021, 12, 1–20.

RideGuru. How Much Do Uber Drivers Make? Available online: https://ride.guru/content/newsroom/how-much-do-uber-drivers-make (accessed on 28 March 2023).
Sarker, M.N.I.; Wu, M.; Chan, A.P.C. Moving beyond cars: A review of ride-hailing apps in China. Travel Behav. Soc. 2021, 25, 149–161.
Sen, S.; Ustun, B.; Zhang, J.; Zhang, W.; Pavone, M.; Akkiraju, R. Contextual exploration of New York City's ride-hailing demand using online learning. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 2113–2122.
Tang, J.; Liang, J.; Huang, J.; Wang, Z.; Gao, F.; Huang, H. A multi-task deep learning model for short-term taxi demand prediction. Remote Sens. 2019, 11, 1265.
Tang, J.; Liang, J.; Zhang, H.; Zhang, L.; Wang, Z. A multi-task convolutional neural network for predicting taxi demand hotspots. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 3135–3144.
Tian, Y.; Pan, Y.; Zhao, Y.; Xie, X.; Yang, D. Trajectory-based taxi demand prediction using multi-task learning. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 3085–3092.
Wang, Z.; Tang, J.; Gao, F.; Zhang, H.; Liang, J. A hybrid deep learning model for taxi demand prediction considering spatial and temporal dependencies. Sensors 2019, 19, 1700.
Wang, Z.; Tang, J.; Liang, J.; Zhang, H.; Huang, H.; Gao, F. Predicting taxi demand hotspots with convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 3135–3144.
Wu, X.; Yan, H.; Xu, R.; Zhang, X.; Chen, X.; Wang, F. Spatio-Temporal Demand Forecasting for Ride-Hailing Services using Multi-Task Learning. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Online, 10–13 December 2020; pp. 1919–1927.
Xia, F.; Liu, X.; He, X.; Wang, L.; Liu, W. Heterogeneous graph attention network for ride-hailing demand prediction. IEEE Trans. Intell. Transp. Syst. 2020, 22, 3122–3130.
Xu, Y.; Jiang, Z.; Ma, Z.; Liu, Y.; Liu, Y.; Shi, X.; Zhou, X. A spatiotemporal deep learning approach for traffic flow prediction using taxi trajectory data. IEEE Trans. Intell. Transp. Syst. 2020, 22, 3835–3845.
Yan, X.; Kuang, L.; Tan, X.; Li, S.; Yang, X. Predicting taxi demand based on 3D convolutional neural network and multi-task learning. Remote Sens. 2019, 11, 1265.
Yu, H.; Wu, Z.; Wang, W.; Wang, W.; Jiang, M.; Chen, Z.; Yang, C. Real-time taxi demand prediction using historical GPS data from a ride-hailing platform. Transp. Res. Part C Emerg. Technol. 2020, 117, 102665.
Zhang, L.; Zhang, J.; Du, B.; Wei, S. Deep spatio-temporal residual networks for demand prediction in ride-hailing service. In Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–4 October 2018; pp. 47–56.
Zhang, M.; Dong, X.; Zhang, S.; Shen, X.; Hu, B. Passenger Demand Forecasting with Multi-Task Convolutional Recurrent Neural Networks. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Macau, China, 14–17 April 2019; pp. 29–42.
Zhang, S.; Zheng, H.; Yu, J.; Jiang, Y.; Yang, H. Multi-Task Learning-Based Multi-Modal Travel Demand Prediction with Graph Attention Network. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 953–962.
Zheng, S.; Wang, J.; Huang, C. Estimating time-of-day and day-of-week variations of taxi travel demand using GPS trajectory data. ISPRS Int. J. Geo-Inf. 2018, 7, 146.
Zhou, B.; Jin, Y.; Zeng, L.; Feng, Y.; Hu, X.; Huang, J. Forecasting ride-hailing demand with multi-task learning. arXiv 2018, arXiv:1807.03165.

The National Institute for Congestion Reduction (NICR) will emerge as a national leader in providing multimodal congestion reduction strategies through real-world deployments that leverage advances in technology, big data science and innovative transportation options to optimize the efficiency and reliability of the transportation system for all users. Our efficient and effective delivery of an integrated research, education, workforce development and technology transfer program will be a model for the nation.

www.nicr.usf.edu