CMT-Net: A Mutual Transition Aware Framework for Taxicab Pick-ups and Drop-offs Co-Prediction

Yudong Zhang[1,2], Binwu Wang[1,2], Ziyang Shan[1,2], Zhengyang Zhou[1,2†], Yang Wang[1,2§]
[1] University of Science and Technology of China (USTC), Hefei, China
[2] Suzhou Institute for Advanced Research, USTC, Suzhou, China
{zyd2020,wbw1995,shanzy,zzy0929}@mail.ustc.edu.cn,angyan@ustc.edu.cn*

ABSTRACT

With increasing population of modern cities, accurate estimation of regional passenger demands is critical to online taxicab services as such platforms aim at a reformation of taxicab scheduling for a more efficient order dispatching. Though great efforts have been made on passenger demand predictions, existing works still have the following shortcomings: i) they mostly performed based on uniform grid partition, which results in the imbalance of demand volumes among regions and even non-vehicle regions in such partition, ii) none of previous demand forecasting efforts have highlighted the important mutual influences between pick-ups and drop-offs, which are of great significance for taxicab scheduling. To this end, we first devise a multi-kernel based clustering to achieve a taxicab-behavior and geographic-aware sub-region partition, hence a more balanced and compact regional division is obtained. Subsequently, we emphasize the essential factors with regard to mutual transition quantification in taxicab predictions, then propose a Transfer-LSTM and an Origin-Destination-based transition matrix to respectively capture the drop-to-pick and pick-to-drop spatiotemporal transition patterns. Hence, a novel mutual-transition-aware co-prediction framework is devised by capturing complex spatiotemporal interactions between pick-ups and drop-offs. Extensive experiments on two real-world taxicab datasets demonstrate our co-prediction framework is superior to state-of-the-art methods, thus providing novel perspectives to urban human mobility understanding and transition-based taxicab scheduling.

CCS CONCEPTS

• **Information systems** → **Spatial-temporal systems**; **Data mining**.

KEYWORDS

Taxicab demand prediction, Spatiotemporal data mining, Cluster-based partition, Mutual transition aware

Prof Yang Wang is the corresponding author, Zhengyang Zhou is the joint corresponding author.

WSDM '22, February 21–25, 2022, Tempe, AZ, USA

ACM ISBN 978-1-4503-9132-0/22/02…$15.00
https://doi.org/10.1145/3488560.3498394

ACM Reference Format:
Yudong Zhang[1,2], Binwu Wang[1,2], Ziyang Shan[1,2], Zhengyang Zhou[1,2†], Yang Wang[1,2§]. 2022. CMT-Net: A Mutual Transition Aware Framework for Taxicab Pick-ups and Drop-offs Co-Prediction. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM '22), February 21–25, 2022, Tempe, AZ, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3488560.3498394

1 INTRODUCTION

The advanced online taxicab calling platforms including Uber[17], CAOCAO [2] can greatly facilitate people's urban travels. However, under the fast expansion of metropolises, some weaknesses in these data-driven taxicab platforms become increasingly prominent, such as overlong waiting for passengers and business missing for drivers due to their imperfect demand predictions and irrational order-driver matches [23, 28]. Therefore, there exists an increasing demand to promote the prediction accuracy of both taxicab pick-ups and drop-offs, to further improve the delivery rate of passengers' service demands, the overall revenue of taxicab drivers as well as the efficiency of urban travels.

Great efforts [1, 3, 8, 14, 16, 19, 21, 22, 24, 26] have been achieved in taxicab business prediction with machine or deep learning methods. At very first, previous works focus on a single mission of forecasting pick-up demands in a specific region. And recently, pioneering works [4, 9, 12] further provide the citywide regional businesses by respectively capturing both inner- and inter-region global correlations of taxicab running with Contextualized Spatial-Temporal Network (CSTN) [12], SpatioTemporal Encoder-Decoder Residual Multi-Graph Convolutional network (ST-ED-RMGC) [9] and Multi-view Localized Correlation learning (MLC-PPF) [4]. Nevertheless, all these previous works focus on the prediction of pick-up demands, and the issue of regional taxicab drop-offs prediction is rarely considered in previous research.

However, the awareness of drop-off events is of great significance for taxicab scheduling. In particular, for the demand side, the number of drop-offs in a specific region during a future time period can directly influence the number of vacant taxicabs which are available for scheduling to satisfy online calling taxicab demands and the number of pick-up demands in subsequent time periods. For the supply side, the taxicab drivers prefer to cruise for very short distances around the last drop-off points to look for new passengers picking up, which influences the future pick-up demands correspondingly. In a specific region, by cooperating the predicted numbers of drop-offs and pick-ups during a future time period, online taxicab calling platforms can accurately schedule taxicabs to maximally satisfy taxicab service demands and reduce the number

of vacant taxicabs as many as possible, hence achieve the cooperative target of enhancing the delivery rate of passengers' service demands and the overall revenue of taxicab drivers. Consequently, the co-prediction of taxicab pick-ups and drop-offs is essential for efficient and accurate citywide taxicab scheduling.

Given the significance of predicting both the numbers of pick-up demands and drop-off events, an ultramodern work [10] has taken an initial step toward addressing this challenge. This work uses a 3D-ResNet and Long-Short Term Memory (LSTM) integrated network to respectively capture the spatial and temporal dependencies of both pick-ups and drop-offs by considering co-predictions as two independent missions. However, in this paper, we have discovered some interesting mutual transition patterns between pick-ups and drop-offs. To be detailed, Figure 1 illustrates the spatiotemporal running patterns of urban taxicabs during the traffic rush hours in both morning and afternoon. First, as illustrated in Figure 1 (a) and (b), this kind of layout-driven tidal patterns determines that the numbers of drop-offs of different regions are impacted by the numbers of pick-ups of some other specific regions. Second, in Figure 1 (c) and (d), these tidal patterns also determine that the numbers of drop-offs of different regions can likewise influence the numbers of pick-ups of some surrounding districts, as taxicab drivers tend to look for new passengers around the last drop-off points. Hence, in this work, inspired by such transition correlations, we propose to enable a mutual-transition-aware co-prediction framework, and it will reasonably contribute to more reliable and accurate forecasting.

Moreover, recent literature [30, 31] reveals that proper spatial division is also of significance for balanced and efficient spatiotemporal forecasting. Previous works have used the conventional grid-based divisions or administrative divisions. However, in most cases, these methods cannot adapt to the natural but irregular layout of urban areas. For instance, grid divisions will inevitably cover the useless car-free sub-regions of lakes, rivers and parks, which directly results in the sparsity of elements in some sub-regions, and hence reduces the prediction accuracy. Regarding administrative divisions, they mostly lead to too coarse-grained forecasting which cannot support effective decision-making to fine-grained local taxicab scheduling. Therefore, an effective spatial partition method is highly demanded to simultaneously satisfy natural fine-grained forecasting and alleviate spatial sparsity challenges by overcoming the heterogeneous spatial layouts of geographical factors.

In this paper, we propose a novel mobility prediction Network of Co-prediction with Mutual-Transition awareness (CMT-Net), for jointly predicting citywide pick-ups and drop-offs by simultaneously addressing the above challenges. Specifically, to alleviate the sparsity issue and incorporate diversified factors for clustering learning, we devise a multi-kernel K-means clustering to adaptively partition the urban area into irregular sub-regions, by leveraging all existing GPS points and clustered pick-ups and drop-offs numbers. The clustered existing taxicab GPS points can reasonably exclude those car-free sub-regions and equilibrate the assigned taxicab arrival numbers for each sub-region. And the multiple kernels in the clustering method can essentially be the diversified proximity metric, including both taxicab behaviors and geographical proximity, for measuring point-wise distances. We then construct the clustered sub-regions to an undirected graph based on the Euclidean distances and taxicab trips relevance among regions. Second, to explicitly consider both periodical taxicab demand patterns and the particular mutual spatiotemporal transition patterns between pick-ups and drop-offs, we propose our mutual-transition aware joint prediction framework. In particular, to obtain diversified and semantic spatial embeddings for demand forecasting, we take advantage of the multi-head neighboring spatial embedding mechanism. To extract the mutual transitions among pick-ups and drop-offs, we design a Transfer-LSTM to learn the underlying transfers from drop-ups to pick-offs with a gated filtering mechanism, and design a multi-granularity pick-drop Origin-Destination (OD) matrix to extract the pick-ups transferring to drop-offs. With our integrated model, the mutual-transition information can be extracted, and the predicted numbers of both pick-ups and drop-offs can be jointly optimized with our dual objectives. Experiments show that our proposal can improve the prediction accuracies of pick-ups and drop-offs numbers by 14.12% and 13.19% (RMSE) on New York City, 6.44% and 8.28% (RMSE) on Suzhou.

The key contributions of this paper are summarized as follows:

- To the best of our knowledge, this is the first paper which reveals the mutual transitions among taxicab pick-ups and drop-offs, and this is also the first work which aims at jointly predicting both pick-ups and drop-offs in a citywide range by taking the mutual transition correlations into account with a specifically designed mutual transition aware framework.
- We propose a multi-kernel K-means clustering, which incorporates a hybrid proximity measurement considering both taxicab behaviors and geographical proximity, on district partition instead of traditional grid partition, maintaining the consistency of the regional volume, and alleviating the sparsity issue caused by geographical layouts.
- A novel mutual-transition-aware model is designed to capture bilateral transfer patterns between pick-ups and drop-offs, with a gated filter incorporated Transfer-LSTM to learn drop-to-pick transfers and an OD-based pick-drop transition matrix to quantify the potential pick-to-drop transitions.

The rest of this paper is organized as follows. Section 2 introduces the preliminaries and formalizes the problem. Section 3 investigates the proposed CMT-Net framework. Section 4 presents empirical studies and Section 5 concludes the paper.

2 PROBLEM DEFINITION

In this section, we formally define some basic concepts as well as the problem studied in this paper.

Definition 1 (**Original trip**). *An original taxicab trip τ contains the location and temporal information of its corresponding pick-ups and drop-offs events, i.e., $\tau = \{p_\tau, d_\tau\}$ where p_τ and d_τ correspond to the pick-ups and drop-offs respectively. Here $p_\tau = \{t_{p_\tau}, l_{p_\tau}\}$ where t_{p_τ} and l_{p_τ} are the time and location of the pick-ups. Accordingly, $d_\tau = \{t_{d_\tau}, l_{d_\tau}\}$ indicates the time and location of the drop-offs.*

Definition 2 (**Region semantic graph**). *Based on the locations of all pick-ups and drop-offs, the entire urban area can be partitioned into K irregular sub-regions, i.e., $\{v_0, v_1, \cdots, v_{K-1}\}$ where $v_i (0 \le i \le K-1)$ denotes the i-th sub-region. The partitioned sub-regions can be modeled as an undirected graph $\mathcal{G}_t(\mathcal{V}, \mathcal{E})$ where vertex*

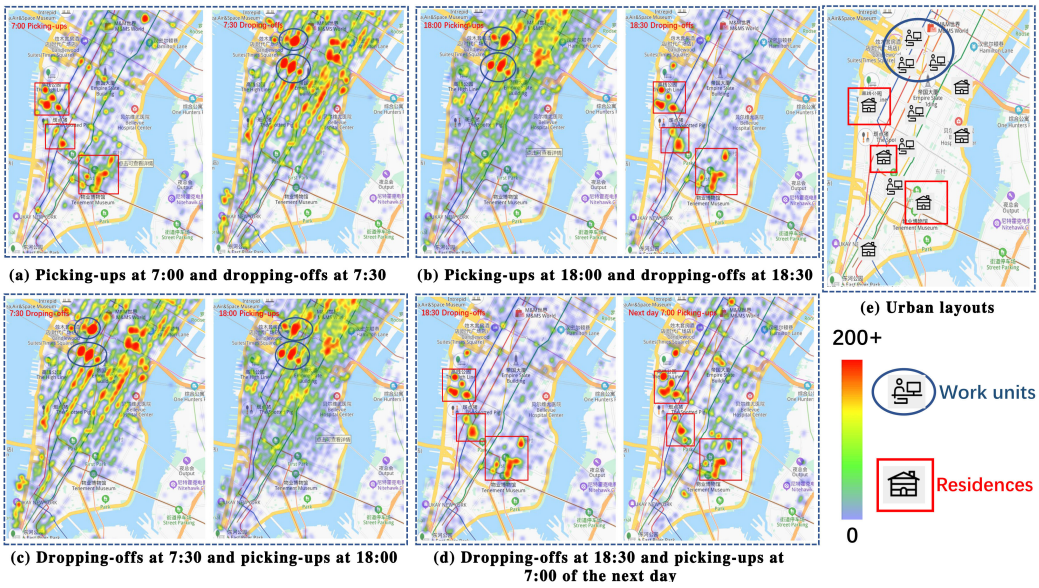


Figure 1: Layout-driven spatiotemporal tidal running patterns and correlations of the urban taxicab in Manhattan: distributions of pick-ups at 7:00 and 18:00 are correlated with distributions of drop-offs at 18:30 of the last day and 7:30 respectively and distributions of drop-offs at 7:30 and 18:30 are associated with distributions of pick-ups 30 minutes ago

$v_i \in \mathcal{V}(0 \le i \le K-1)$ is the i-th sub-region and edge $e_{ij} \in \mathcal{E}$ represents the connectivity between v_i and v_j. The detailed calculation of the connectivity between sub-regions will be described by Equation 6 and 7 in Subsection 3.1, corresponding to the undirected edge between region v_i and v_j.

Definition 3 (**Regional pick-ups and drop-offs**). *Given a sub-region v_i, the pick-ups $\mathcal{P}_t^{v_i}$ and drop-offs $\mathcal{D}_t^{v_i}$ within this sub-region during time interval Δ after time t can be defined as*

$$\begin{cases} \mathcal{P}_t^{v_i} = \{p_\tau \,|t_{p_\tau} \in [t, t+\Delta) \cap l_{p_\tau} \in v_i\} \\ \mathcal{D}_t^{v_i} = \{d_\tau \,|t_{d_\tau} \in [t, t+\Delta) \cap l_{d_\tau} \in v_i\} \end{cases} \tag{1}$$

and the historical pick-ups set $\mathbb{P}_t^{v_i}$ and drop-offs set $\mathbb{D}_t^{v_i}$ of sub-region v_i since time t can be denotes as

$$\begin{cases} \mathbb{P}_t^{v_i} = \left\{\mathcal{P}_{t-(m-1)\Delta}^{v_i}, \mathcal{P}_{t-(m-2)\Delta}^{v_i}, \cdots, \mathcal{P}_t^{v_i}\right\} \\ \mathbb{D}_t^{v_i} = \left\{\mathcal{D}_{t-(m-1)\Delta}^{v_i}, \mathcal{D}_{t-(m-2)\Delta}^{v_i}, \cdots, \mathcal{D}_t^{v_i}\right\} \end{cases} \tag{2}$$

where m is the number of historical time intervals.

Definition 4 (**pick-ups and drop-offs co-predictions**). *Given region division $\{v_0, v_1, \cdots, v_{K-1}\}$ and the historical pick-ups and drop-offs of all sub-regions, i.e., $\{\mathbb{P}_t^{v_0}, \cdots, \mathbb{P}_t^{v_{K-1}}\}$ and $\{\mathbb{D}_t^{v_0}, \cdots, \mathbb{D}_t^{v_{K-1}}\}$, the target is to predict $\left|\mathcal{P}_{t+\Delta}^{v_i}\right|$ and $\left|\mathcal{D}_{t+\Delta}^{v_i}\right|$ for any region $v_i (0 \le i \le K-1)$.*

3 PROPOSED CMT-NET FRAMEWORK

In this section, we first propose a clustered-based sub-region partition, and then present the detailed mutual-transition-aware pick-ups and drop-offs co-prediction framework, CMT-Net.

3.1 Cluster-based Sub-region Partition and Semantic Graph Construction

Plenty of spatiotemporal learning frameworks usually employ the traditional grid division methods for region partition, however, the regular grids will bring car-free sub-regions such as lake, river, and park, which will break the natural shape of urban areas and consequently increase the data sparsity issue in grid-based spatial distributions, and pose great challenges into fully capturing semantically correlations among adjacent sub-regions. In order to solve the above problems, we propose a multi-kernel K-means clustering to aggregate the spatial taxicab points (located in different latitude or longitude), which takes the role of partitioning the whole urban area into different irregular but semantically correlated sub-regions. We first aggregate all point locations into N tiny regular grids with small enough side length r, and we consider pick-up and drop-off patterns as well as the geographical proximity for clustering. Hence, for each tiny grid, we construct the feature vector $e_i = \{Nums_P, Nums_D, Latitude, Longitude\} \in R^{d_s}$, where $d_s = 4$ (4-dimensional spatial feature). $Nums_P$, $Nums_D$ represent the total number of pick-ups and drop-offs in the whole training set, respectively. We use $E = \{e_0, e_1, \ldots\ldots, e_{N-1}\}$ to represent the citywide feature set. The kernel function can be expressed as:

$$K(e_i, e_j) = \phi_\gamma(e_i)^T \phi_\gamma(e_j) = \sum_{n=1}^{d_s} \gamma_n^2 \mathcal{K}_n(e_i, e_j) \tag{3}$$

where $\mathcal{K}(e_i, e_j)$ represents the kernel of Gaussian distance between e_i and e_j, $\phi_\gamma(e_i) = [\gamma_1\phi_1(e_i)^T, ..., \gamma_{d_s}\phi_{d_s}(e_i)^T]^T$, and γ_i is the i-th base kernel. $\phi_i(\cdot)$ maps e_i onto the i-th reproducing kernel Hilbert space. As result, we can get a kernel matrix $\mathcal{K}_\gamma$ that calculated by

applying $\mathcal{K}(\cdot,\cdot)$ to E. We have the following objective [13]:

$$\min_{V,\gamma} tr(\mathcal{K}_\gamma(I_N - V^T V)) \tag{4}$$

where $V \in R^{N\times K}$ is the clustering matrix, and satisfies $V^T V = I_K$. The identity matrices I_N and I_K are of size N and K, where K is the number of clusters. We alternately optimize the clustering results V and kernel basis γ and finally reach the semantic regional partition.

With the above clustering, we can obtain the semantic partition of sub-regions $\{v_0, v_1, \cdots, v_{K-1}\}$. In particular, our multi-kernel based clustering partition can be deemed as multi-dimensional proximity learning considering both geographical locations and the clustered pick-ups and drop-offs numbers, which can be interpreted as the summarized taxicab demand preferences and urban traveling patterns in the dataset. Hence, the clustering-based partition can naturally exclude the car-free regions like lakes and parks, which alleviates the spatial sparsity issues in grid-based partitions. We will evaluate the effectiveness of our proposed clustering mechanism in experiments.

Given that, we can further construct the K-node semantic graph. Specifically, our semantic graph is a static graph, and the semantic adjacency matrix $\mathcal{A}$ can be calculated by:

$$\mathcal{A} = \begin{bmatrix} \alpha_{00} & \cdots & \alpha_{0(K-1)} \\ \vdots & \ddots & \vdots \\ \alpha_{(K-1)0} & \cdots & \alpha_{(K-1)(K-1)} \end{bmatrix} \tag{5}$$

where $\alpha_{ij}(0 \le i, j \le K-1)$ indicates the undirected semantic connection between sub-region v_i and v_j and can be calculated by

$$\alpha_{ij} = \begin{cases} 1, & if\ \rho_{ij} \ge 0.5 \\ 0, & otherwise \end{cases} \tag{6}$$

and here we have

$$\rho_{ij} = \frac{\delta_{max} - \delta_{ij}}{2(\delta_{max} - \delta_{min})} + \frac{c_{ij} - c_{min}}{2(c_{max} - c_{min})} \tag{7}$$

Notice here δ_{max} and δ_{min} indicate the maximum and minimum Euclidean distances between every sub-region pair, c_{max} and c_{min} correspond to the maximum and minimum taxicab transition volumes between two sub-regions respectively. Specifically, δ_{ij} equals to the Euclidean distance between v_i and v_j, and c_{ij} is the total number of taxicab trips that start from one region and end at another in all historical periods. Given Equation 7, we have $\rho_{ij} \in [0,1]\,(0 \le i, j \le k-1)$, and we define that there exist an undirected semantic edge from v_i to v_j if and only if $\rho_{ij} \ge 0.5$.

3.2 CMT-Net for Pick-up and Drop-off Co-Prediction

In this subsection, we introduce the detailed architecture and design of CMT-Net. It is known that the traffics can be determined by a self-periodical and tendency-aware prediction, as well as the mutual transitions, where the latter is especially studied in this paper. Hence, we design two LSTM-based sequence learning to extract the self-temporal patterns and two mutual-transition modules to explicitly capture transitions between pick-ups and drop-offs. The predictions of pick-ups and drop-offs are both considered as the main tasks of this multi-task learning framework. The overall solution is demonstrated in Figure 2.

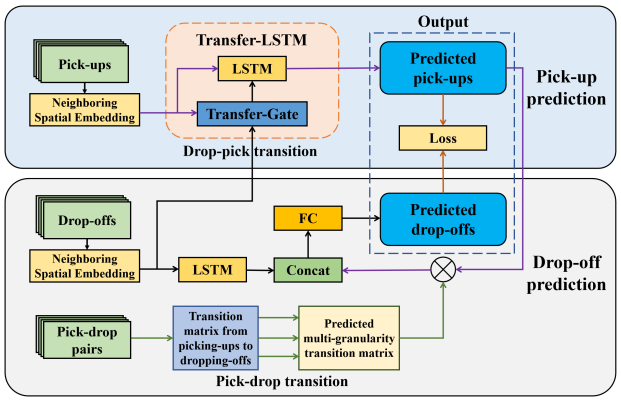


Figure 2: Framework overview of CMT-Net

3.2.1 Neighboring spatial embedding. The spatial contexts and correlations of regional pick-ups or drop-offs are essential for the co-predictions. To incorporate such spatial and semantic contexts, we here design a novel neighboring spatial embedding mechanism to extract the neighboring correlations of pick-ups and drop-offs. By incorporating with the semantic graph, this neighboring spatial embedding mechanism can significantly enlarge its horizon to involve more effective correlations. Specifically, regarding a specific region v_i, for time t to $t-(m-1)\Delta$, we employ the Graph Attention Networks (GATs) [18] to calculate the pick-ups or drop-offs correlations between region v_i and v_i's neighboring regions within the corresponding semantic graphs $\mathcal{G}_{t-k\Delta}(\mathcal{V},\mathcal{E})(k \in \{0,1,\cdots,(m-1)\})$ of the m previous time points. Given time $t-k\Delta(k \in \{0,1,\cdots,(m-1)\})$ and region v_i, assuming v_j is directly connected to v_i in $\mathcal{G}_{t-k\Delta}(\mathcal{V},\mathcal{E})$, the attention coefficient between v_i and v_j can be calculated by

$$\chi_{t-k\Delta}(v_i, v_j) = \gamma_1 \omega_{v_i} \mathcal{F}^{v_i}_{t-k\Delta} + \gamma_2 \omega_{v_j} \mathcal{F}^{v_j}_{t-k\Delta} \tag{8}$$

where $\mathcal{F}^{v}_{t-k\Delta}$ is the feature set of region v which includes $\mathcal{P}^{v}_{t-k\Delta}$ or $\mathcal{D}^{v}_{t-k\Delta}$, average speed of region v during time interval $[t-(k+1)\Delta, t-k\Delta)$, and average travel time of all taxi businesses in $\mathcal{F}^{v}_{t-k\Delta}$. Here, $\omega_v \in \mathbb{R}^{1\times \mathcal{F}^{v}_{t-k\Delta}}$ is a learnable transformation to transform the inputted feature set of region v to a one-dimension embedded feature, and parameters γ_1 and γ_2 are optimized by a single-layer feed forward neural network. Next, the calculated attention coefficient can be regularized by

$$\begin{aligned} \mathrm{X}_{t-k\Delta}(v_i, v_j) &= \mathrm{Softmax}[\chi_{t-k\Delta}(v_i, v_j)] \\ &= \frac{\exp[\chi_{t-k\Delta}(v_i, v_j)]}{\sum_{v_l \in N_{t-k\Delta}(v_i)} \exp[\chi_{t-k\Delta}(v_i, v_l)]} \end{aligned} \tag{9}$$

where $N_{t-k\Delta}(v_j)$ indicates all the neighbors of region v_i in $\mathcal{G}_{t-k\Delta}(\mathcal{V},\mathcal{E})$. And the spatial representation of v_i can be denoted by

$$H^{v_i}_{t-k\Delta} = \sigma\left(\frac{1}{B}\sum_{b=1}^{B}\sum_{v_j \in N_{t-k\Delta}(v_i)} \mathrm{X}^{b}_{t-k\Delta}(v_i, v_j)\omega^{b}_{v_j}\mathcal{F}^{v_j}_{t-k\Delta}\right) \tag{10}$$

where σ corresponds to the LeakyReLU activation function [11], and B corresponds to the number of attention heads. Notice that we here employ the Multi-Head mechanism to obtain different types of spatial attention by calculating multiple attention coefficients in parallel. Hence, we can obtain the neighborhood-aware spatial embeddings for subsequent prediction tasks, where contextual and surrounding information is involved.

3.2.2 Transfer-LSTM based pick-up prediction. As discussed above, pick-up prediction can be divided into self-tendency learning and drop-to-pick transitions. Since LSTM has been demonstrated its powerful capabilities to handle time series problems, therefore, in this subsection, we are going to propose a novel Transfer-LSTM module to cooperatively learn these two critical elements for final predictions. To be detailed, for region v_i during time $t-(m-1)\Delta$ to t, we first take the historical pick-ups $\left[Hp^{v_i}_{t-(m-1)\Delta}, \cdots, Hp^{v_i}_{t}\right]$, as the input of our Transfer-LSTM, to learn the self-periodic and tendency-aware correlations (Line 1-4 in Eq.11). Then we investigate how the drop-offs influence future pick-ups by proposing a gated filter mechanism. Intuitively, drivers tend to look for passengers around the location of the last drop-offs, so the transition from drop-offs to pick-ups for a taxicab is mainly concentrated in the nearby areas of the last drop-offs. To this end, we also embed all drop-offs information as $\left[Hd^{v_i}_{t-(m-1)\Delta}, \cdots, Hd^{v_i}_{t}\right]$ can with previous proposed neighboring spatial embedding mechanism. Since the total pick-ups of one sub-region are composed of self-periodical tendencies and the additional transitions, here we especially design a transfer-gate z_t to explicitly capture the additional drop-off to pick-up transitions by taking previous drop-offs $Hd^{v_i}_{t-1}$ as an additional input to Transfer-LSTM (Line 1 in Eq.11). Then the $\tanh(Hd^{v_i}_{t-1})$ will multiply with hidden drop-pick transfer-state transfer-gate z_t to obtain the final transition quantification by joint optimization. The modified drop-pick transfer-gate can be rewritten as below,

$$\begin{aligned} z_t &= \sigma(U_z h_{t-1} + W_z Hd^{v_i}_{t-1} + b_z) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t + z_t \odot \tanh(Hd^{v_i}_{t-1}) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \tag{11}$$

where $\{f_t, i_t, z_t, o_t\}$ are forget-gate, input-gate, transfer-gate, and output-gate respectively, and other parameters are all learnable. The modified hidden state h_t will be transferred into the next time step. Meanwhile, the output of Transfer-LSTM indicates the future pick-ups, i.e., $\widehat{\mathcal{P}^{v_i}_{t+\Delta}}$. In summary, z_t can effectively capture the spatial transfer patterns in two ways. First, we incorporate historical drop-off information with an additional gate. Second, we also add the learned transfer information to the final prediction of future pick-ups in cooperative optimization, which takes full advantage of the spatial correlation from drop-offs to pick-ups.

3.2.3 Multi-granularity transition based drop-off prediction. Even though drop-offs are all transitioned from pick-ups, some drop-offs will experience a long-term running from their pick-up points. To this end, for achieving short-term drop-offs forecasting, we still disentangle the self-tendency and volumes transitioned from nearby pick-ups. We will elaborate the transition pattern extraction since most drop-offs are contributed by the nearby pick-ups. Given all historical regional pick-ups and drop-offs, regarding time interval $[t, t+\Delta)$, the transition matrix from pick-ups to drop-offs can be easily defined as

$$M_t = \begin{bmatrix} \beta^t_{00} & \cdots & \beta^t_{0(K-1)} \\ \vdots & \ddots & \vdots \\ \beta^t_{(K-1)0} & \cdots & \beta^t_{(K-1)(K-1)} \end{bmatrix} \tag{12}$$

where element β_{ij} represents the ratio of the business number from v_i to v_j to the total number of the business that start from v_i and can be calculated as

$$\beta^t_{ij} = \frac{\left|\left\{\tau | t_{p_\tau} \in [t, t+\Delta) \cap l_{p_\tau} \in v_i \cap l_{d_\tau} \in v_j\right\}\right|}{\sum_{k=0}^{(K-1)} \left|\left\{\tau | t_{p_\tau} \in [t, t+\Delta) \cap l_{p_\tau} \in v_i \cap l_{d_\tau} \in v_k\right\}\right|} \tag{13}$$

Obviously, matrix M_t represents the transition ratio from pick-ups to drop-offs of all clustered region pairs during interval $[t, t+\Delta)$. Given these matrices with regard to all historical time intervals, to predict the transition matrix from pick-ups to drop-offs during interval $[t+\Delta, t+2\Delta)$ from a multi-granularity perspective for a comprehensive prediction, we first extract three different matrix sequences from all historical matrices for exploiting the closeness, periodicity, and tendency [1] of the transition matrices from pick-ups to drop-offs. By taking the average value of each individual sequence, we then generate three matrices $\widehat{\mathcal{M}^c}$, $\widehat{\mathcal{M}^p}$, and $\widehat{\mathcal{M}^t}$, correspondingly. Finally, we predict the multi-granularity transition matrix from pick-ups to drop-offs during interval $[t+\Delta, t+2\Delta)$ by calculating a weighted sum of these three matrices, i.e.,

$$\widehat{M_{t+\Delta}} = \Pi^c \odot \widehat{\mathcal{M}^c} + \Pi^p \odot \widehat{\mathcal{M}^p} + \Pi^t \odot \widehat{\mathcal{M}^t} \tag{14}$$

where Π^c, Π^p, and Π^t are all learnable matrices, and $\odot$ corresponds to the Hadamard product.

With the predicted multi-granularity transition patterns $\widehat{M_{t+\Delta}}$, and the predicted future pick-ups of region v_i during $[t+\Delta, t+2\Delta)$ $\widehat{\mathcal{P}^{v_i}_{t+\Delta}}$, we are able to calculate the main contributor of drop-offs, transition-inferred future drop-offs $\widehat{\Lambda^{v_i}_{t+\Delta}}$ by

$$\widehat{\Lambda^{v_i}_{t+\Delta}} = \widehat{\mathcal{P}^{v_i}_{t+\Delta}} \cdot \widehat{M_{t+\Delta}} \tag{15}$$

Finally, the future drop-offs should introduce the previous drop-offs as the periodicity and tendency-aware components for fusion learning. Hence, we denote the spatiotemporal representations of historical drop-offs during the interval δ after time t as $\delta^{v_i}_t$, then the drop-offs can be predicted by

$$\widehat{\mathcal{D}^{v_i}_{t+\Delta}} = \mathrm{FC}\left(\mathrm{Concat}\left(\widehat{\Lambda^{v_i}_{t+\Delta}}, \delta^{v_i}_t\right)\right) \tag{16}$$

where function FC represents a fully connected layer and Concat corresponds to the concatenation operation.

3.2.4 Loss function. Finally, we present the joint losses to optimize the co-prediction tasks. The loss function for training our Co-prediction with Mutual Transition framework can be defined as

$$Loss(\Xi) = \frac{1}{K}\sum_{i=0}^{K-1}\left[\left(\widehat{\mathcal{P}^{v_i}_{t+\Delta}} - \mathcal{P}^{v_i}_{t+\Delta}\right)^2 + \left(\widehat{\mathcal{D}^{v_i}_{t+\Delta}} - \mathcal{D}^{v_i}_{t+\Delta}\right)^2\right] \tag{17}$$

[1] The sequence for exploiting the closeness is composed by the last λ_c time interval before t, the sequence for exploiting the periodicity consists of the same intervals as $[t, t+\Delta)$ during the previous λ_p days, and the sequence for exploiting the tendency is formed by the same intervals as $[t, t+\Delta)$ in the same day of a week during the previous λ_d weeks.

where Ξ is the combined learnable parameter set of the proposed network, and we use Adam optimizer to optimize our proposed multi-task learning network in this paper.

Until now, we can achieve the final results of joint predictions of both future pick-ups $\widehat{\mathcal{P}^{v_i}_{t+\Delta}}$ and drop-offs $\widehat{\mathcal{D}^{v_i}_{t+\Delta}}$ at $t + \Delta$ interval. Actually, we especially design a Transfer-LSTM and multi-granularity transition matrix to realize the intuitions of mutual transitions and jointly optimize with above dual learning objectives.

4 EXPERIMENTS

4.1 Datasets

In this section, we conduct extensive experiments on two real-world taxicab datasets collected from NYC and Suzhou. The detailed information about these two datasets is as follows.

Table 1: Data statistics and implementation details

Taxicab data	MI, NYC	SIP, Suzhou
Data type	Taxi business records	Taxi trajectories
Time span	01/01/2015-06/30/2015	01/01/2017-03/31/2017
Length of time interval	30 minutes	30 minutes
Number K of cluster regions	30	60
Grid division of city	3×10	12×5

(1) **NYC Taxicab business dataset.** This dataset includes all the taxicab business records of NYC from Jan. 1, 2015 to Jun. 30, 2015, and each business record contains taxi trip duration as well as the time points, longitudes, and latitudes information of both pick-ups and drop-offs.

(2) **Suzhou Taxicab trajectory dataset.** This dataset consists of the trajectory information of all taxicabs in Suzhou from Jan. 1, 2017 to Mar. 31, 2017. Each piece of the trajectory of a taxicab includes the detailed monitored GPS information of this taxicab with an interval of 20 seconds. And the time and location of pick-ups and drop-offs of this taxicab are also contained in the corresponding piece of record.

4.2 Implementation Details

As taxicabs are mostly concentrated on most developed sub-regions, we select Manhattan Island (MI) and Suzhou Industrial Park (SIP) in NYC and Suzhou as the experimental areas respectively. The data statistics and implementation details are demonstrated in Table 1. We grid the urban area with pick-up and drop-off point locations with squares r = 100 meters, where the tiny grid is small enough and has little influence on performance variations. We are only expected to adjust the number of clustering K to achieve our semantic graph. Here, unless specified, we set the K = 30, K = 60 in MI and SIP datasets, respectively according to the following experiments. For fairness, we carefully conduct experiments to achieve an equilibration regarding divisions in both grid-based methods and our clustering method, which maintains a comparative region size. To be specific, for CNN-based baselines, we divide MI and SIP into 3×10 and 12×5 uniform grids respectively to make sure each region in these two grids is about 1.3km $\times$ 1.3km approximately. And we set K of the clustering algorithm as 30 and 60 respectively for MI and SIP, which usually cover approximately 1.7 km^2. We set 4 attention heads in our GAT for both datasets and optimize it with experiments.

For temporal division and aggregation, for both datasets, we slice the temporal information into slots of 30 minutes, and λ_c, λ_p, and λ_d of the three different matrix sequences for exploiting the closeness, periodicity, and tendency are all set to 3, following the common practice in the field of traffic forecasting [5–7, 29]. Notice that such settings may be related to the results of inference accuracy but is orthogonal to the generalities of our proposals with fair comparisons. For each round of evaluations, we split all datasets with ratio 6:2:2 into training sets, validation sets and test sets. Regarding the efficiency of offline training and online inference, our method can be trained within half an hour and infer the new sample within several seconds, which can sufficiently support the real-time taxicab demand forecasting.

4.3 Baselines and Evaluation Metrics

4.3.1 Baselines. We evaluate the performance of our CMT-Net by comparing it with the following baseline models. To be fair, all baselines listed here are employed the proposed cluster-based methods for urban division. Moreover, all baselines are implemented in a joint-prediction manner. Noted that the quantitative co-prediction verification will be included in the ablation study part.

(1) Historical Average (HA). HA predicts both the pick-ups and drop-offs demands of each region during a specific time interval with the average pick-ups and drop-offs demands of the corresponding region during all the same intervals in history.

(2) Autoregressive Integrated Moving Average (ARIMA) model. It combines moving average with autoregression to predict future values in time series.

(3) Long Short-Term Memory (LSTM). LSTM is a specially designed RNN for additionally involving long-term dependence in future predictions.

(4) Deep Spatio-Temporal Residual Networks (ST-ResNet) [27]. It is a deep learning model for predicting traffic flows considering both spatial and temporal correlations of traffic flows.

(5) Convolutional LSTM (ConvLSTM) [20]. It combines CNN with LSTM into one integrated model, which can also capture both spatial and temporal correlations in predictions.

(6) Deep Multi-View Spatial-Temporal Network (DMVST-Net) [24]. It combines spatial, temporal and semantic views.

(7) Co-prediction method based on Spatio-Temporal neural Network (CoST-Net) [25]. It predicts the multiple demands of taxicab and sharing bike simultaneously. To be fair, we only conduct the task of forecasting taxi demand.

(8) Spatial-Temporal Synchronous Graph Convolutional Networks (STSGCN) [15]. It is a spatiotemporal synchronous model for traffic prediction based on the graph convolutions.

4.3.2 Evaluation metrics. We use two classic metrics, Rooted Mean Square Error (RMSE) and Mean Average Percentage Error (MAPE), to evaluate the performances of our spatiotemporal regression model and other baselines.

4.4 Evaluation Results and Analysis

4.4.1 Effect analysis of cluster-based partition. To investigate the effects of our multi-kernel clustering partition in CMT-Net, we

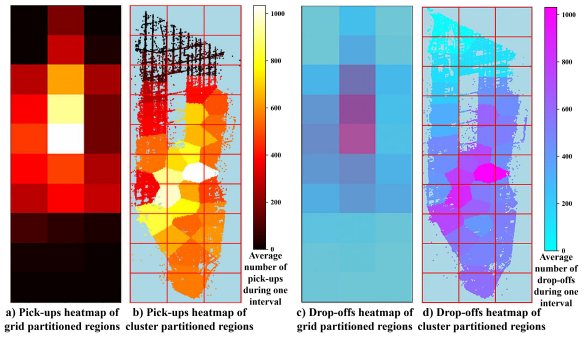


Figure 3: Heatmaps of average numbers of pick-ups and drop-offs events with different region divisions

first analyze the distributions of pick-up and drop-off points on MI dataset based on grid division and cluster-based partition, and demonstrate the results as heatmaps in Figure 3. Regarding the grid division, it exhibits a large dispersion on averaged regional numbers of both pick-ups and drop-offs during one interval, where the central Manhattan region attracts almost 5 times of the volumes as marginal regions, which are with barely few pick-ups and drop-offs events (Figure 3 (a) and (c)). The reasons may lie in that most marginal regions in uniform grid division are principally covered with surfaces of water, while the central regions are usually experiencing heavy traffics and frequent human mobility due to their specific functionalities. Thus, the imbalanced data distribution and sparsity issues impose great challenges to forecasting tasks. In contrast, our cluster-based partition shows that the numbers of both pick-ups and drop-offs among different regions are more equilibrated in Figure 3 (b) and (d). Moreover, the divided regions can reveal more rational shapes since those car-free areas like sea, river, lake, and park have been excluded.

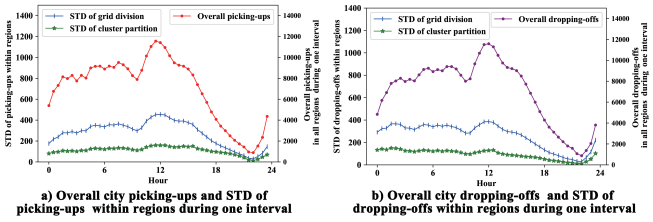


Figure 4: Overall city events and STD of events within regions during different intervals

We also evaluate the superiority of our multi-kernel clustering method in a quantitive manner. Here, we present the Standard Deviations (STD) of the numbers of interval-level pick-ups and drop-offs among different regions in Figure 4. As shown, the STD curves of both pick-ups and drop-offs in the cluster partition are more smooth than the corresponding curves in the uniform grid division. This indicates that the regional numbers of pick-ups and drop-offs are more equilibrated in the cluster partition which benefits the deep learning process and enhance the accuracy of prediction missions. In particular, the STDs of regional pick-ups and drop-offs in the grid division are 273.58 and 260.57 respectively, while those in the cluster partition are 108.58 and 93.72. Hence, our framework effectively decreases the STDs by 60.31% and 64.03% respectively, and facilitates the equilibration and learning mechanism. Further, the following results of the best baseline CoST-Net with pure grid division, which also has the equilibrated region size with ours, are inferior to ours (decreases by 14.11% and 10.92% on RMSE of MI), indicating the advantage in training and predictions of our cluster-based region partition considering both taxicab behavior and geographic proximity.

Based on the above, we can conclude that our cluster-based partition can gains benefits of the human mobility clustering and volumes balances in each cluster for easy training with this geographic and taxicab behavior awareness mechanism.

Table 2: Performance comparisons on MI and SIP datasets

	Methods	RMSE		MAPE	
		pick-ups	drop-offs	pick-ups	drop-offs
MI	HA	99.40	100.05	0.3595	0.3686
	ARIMA	58.05	62.15	0.2416	0.2805
	LSTM	47.24	48.19	0.1879	0.1951
	ConvLSTM	44.04	45.90	0.1740	0.1847
	ST-Resnet	43.90	44.83	0.1707	0.1759
	DMVST-Net	42.87	43.67	0.1621	0.1702
	STSGCN	42.73	42.36	0.1653	0.1742
	CoST-Net	39.34	40.31	0.1570	0.1639
	CMT-Net	**33.79**	**36.73**	**0.1343**	**0.1448**
	Methods	RMSE		MAPE	
		pick-ups	drop-offs	pick-ups	drop-offs
SIP	HA	29.30	30.59	0.3683	0.3732
	ARIMA	24.55	27.48	0.2567	0.2631
	LSTM	21.48	22.26	0.1901	0.1965
	ConvLSTM	17.84	19.38	0.1816	0.1839
	ST-Resnet	15.45	16.52	0.1724	0.1782
	DMVST-Net	14.78	15.67	0.1695	0.1715
	STSGCN	14.13	15.27	0.1679	0.1696
	CoST-Net	13.98	14.85	0.1629	0.1657
	CMT-Net	**13.08**	**13.62**	**0.1503**	**0.1598**

4.4.2 The performances of CMT-Net. The performances of our proposed approach and all alternative baselines on real-world datasets of both NYC and Suzhou are demonstrated in Table 2. As can be easily observed, in both NYC and Suzhou, the proposed CMT-Net outperforms all alternative solutions in terms of the RMSE and MAPE of both pick-ups and drop-offs demand predictions, and this directly and powerfully verifies the validity of our proposed method on pick-ups and drop-offs co-prediction. Specifically, in MI, compared with the state-of-the-art solution, our proposed framework can reduce the RMSE and MAPE by 14.11% and 19.58% respectively in predicting future pick-ups demands and by 13.19% and 14.92% respectively in predicting future drop-off events.

Furthermore, the performances of ARIMA and LSTM are relatively poor since these two models are both time series prediction models and cannot extract any spatial characteristics. Regarding ConvLSTM and ST-ResNet, these two methods extract limited

Table 3: Performances of ablative variants on MI dataset

Methods	RMSE		MAPE	
	pick-ups	drop-offs	pick-ups	drop-offs
CMT-Net-Single	37.21	39.46	0.1546	0.1613
CMT-Net-RMT	36.65	39.12	0.1499	0.1598
CMT-Net-RPDT	35.91	38.77	0.1472	0.1581
CMT-Net-RDPT	35.27	37.17	0.1454	0.1552
CMT-Net	**33.79**	**36.73**	**0.1343**	**0.1448**

spatial correlations with traditional CNN, consequently contributing to better performances when compared to time series-based models. Besides considering spatiotemporal correlations with traditional methods, both DMVST-Net and CoST-Net involve inter-region semantic similarities, and hence they achieve better performances than the baselines mentioned above. STSGCN employs the spatiotemporal synchronization mechanism to capture complex local spatiotemporal correlations, thus its performance is close to DMVST-Net and CoST-Net. However, these baselines all ignore the mutual transfer information among taxicab pick-ups and drop-offs. Regarding our proposed CMT-Net, and compared with other spatiotemporal models, the improvement can be attributed to the unique insight on explicitly reconstructing mutual transitions among pick-ups and drop-offs. We will further investigate our conjecture in the following parts.

4.4.3 Ablation Study. We perform ablation studies in this subsection, and the ablative variants are as follows, i.e.,

(1) CMT-Net-Single. This variant removes the joint optimization mechanism and respectively optimize the pick-ups and drop-offs predictions, which aims to verify the necessity of our joint co-prediction intuition.

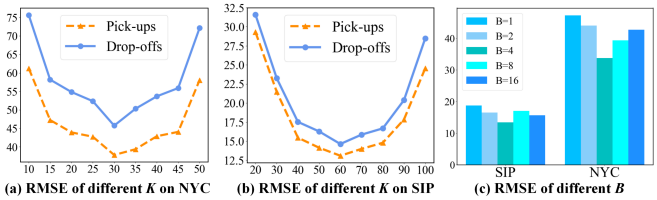
(2) CMT-Net-RMT. This variant doesn't include any transition information, and the transition patterns are replaced with the distance-aware geographical adjacent matrices.

(3) CMT-Net-RPDT. The multi-granularity transition matrix extracted from pick-ups to drop-offs is singly removed, but this variant considers the latent transition from drop-offs to pick-ups in Transfer-LSTM.

(4) CMT-Net-RDPT. The drop-pick transition in Transfer-LSTM is removed, but this variant contains the multi-granularity transition matrix from pick-ups to drop-offs.

As demonstrated in Table 3, we first verify the necessity of co-predictions by observing 10.12% and 7.43% increases of RMSEs in pick-ups and drop-offs single-task testing. After then, both two transition mechanisms can individually enhance the performances of CMT-Net. Specifically, in MI, the mutual transition aware mechanism can reduce the RMSE and MAPE by 7.80% and 10.41% respectively in predicting future pick-up demands, thanks to the drop-pick Transfer-gate. Analogously, RMSE and MAPE are reduced by 6.11% and 9.39% respectively in predicting future drop-offs events due to the well-learned multi-granularity pick-drop transition matrix.

4.4.4 Hyperparameter Study. There are two hyperparameters in our framework that need further careful analysis, i.e., the number of clustering K for graph construction, and the number of multiple heads B in spatial aggregation.



(a) RMSE of different K on NYC (b) RMSE of different K on SIP (c) RMSE of different B

Figure 5: Performance on different parameter settings

The performance variations on two datasets are characterized in Figure 5. According to our comprehensive results, we obtain the optimal settings of $K = 30$ and $K = 60$ on MI and SIP, and $B = 4$ on both datasets.

5 CONCLUSION

In this paper, we propose a mutual-transition-aware learning framework, CMT-Net, to address the challenge of taxicab pick-ups and drop-offs co-prediction. We first employ multi-kernel K-means clustering to realize the compact region division. Next, we construct the cluster partitioned sub-regions to a semantic graph based on the geographical proximity and taxicab behavior relevance among regions. We further extract mutual spatiotemporal transition patterns with a transfer-gated LSTM and multi-granularity transition matrix for drop-pick and pick-drop transitions, respectively. Performance evaluations on two real-world datasets and the ablation studies both powerfully demonstrate the effectiveness of our proposal. Therefore, our work provides a brand-new solution to tackle the taxicab pick-ups and drop-offs event co-prediction in a mutual transition learning perspective, which consequently enhances the service qualities of online taxicab calling platforms, and improves understanding of the interactions between human mobility and social economics. Regarding future work, we are going to optimize the vacant taxicab scheduling by co-predictions and the drop-pick transition-aware mechanism, and also we are intended to investigate more general flow transition quantification tasks such as inter-migration of people and animals for nature discovery.

ACKNOWLEDGMENTS

This work is partially supported by Project of Stable Support for Youth Team in Basic Research Field, CAS (No.YSBR-005), NSFC (No.62072427), Anhui Science Foundation for Distinguished Young Scholars (No.1908085J24), Zhejiang Lab's International Talent Fund for Young Professionals, and Jiangsu Natural Science Foundation (No.BK20191193).

REFERENCES

[1] Yi Ai, Zongping Li, Mi Gan, Yunpeng Zhang, Daben Yu, Wei Chen, and Yanni Ju. 2019. A deep learning approach on short-term spatiotemporal distribution forecasting of dockless bike-sharing system. *Neural Computing and Applications* 31, 5 (2019), 1665–1677.

[2] CAOCAO. 2020. https://www.caocaokeji.cn/.

[3] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 3656–3663.

[4] Yongshun Gong, Zhibin Li, Jian Zhang, Wei Liu, and Jinfeng Yi. 2020. Potential Passenger Flow Prediction: A Novel Study for Urban Transportation Development. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 04 (2020), 4020–4027.

[5] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (Jul. 2019), 922–929.

[6] Shengnan Guo, Youfang Lin, Shijie Li, Zhaoming Chen, and Huaiyu Wan. 2019. Deep Spatial–Temporal 3D Convolutional Neural Networks for Traffic Data Forecasting. *IEEE Transactions on Intelligent Transportation Systems* 20, 10 (2019), 3913–3926.

[7] Shengnan Guo, Youfang Lin, Huaiyu Wan, Xiucheng Li, and Gao Cong. 2021. Learning Dynamics and Heterogeneity of Spatial-Temporal Graph Data for Traffic Forecasting. *IEEE Transactions on Knowledge and Data Engineering* (2021), 1–1.

[8] Shan Jiang, Wentian Chen, Zhiheng Li, and Haiyang Yu. 2019. Short-Term Demand Prediction Method for Online Car-Hailing Services Based on a Least Squares Support Vector Machine. *IEEE Access* 7 (2019), 11882–11891.

[9] Jintao Ke, Xiaoran Qin, Hai Yang, Zhengfei Zheng, Zheng Zhu, and Jieping Ye. 2019. Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. *arXiv preprint arXiv:1910.09103* (2019).

[10] Li Kuang, Xuejin Yan, Xianhan Tan, Shuqi Li, and Xiaoxian Yang. 2019. Predicting taxi demand based on 3D convolutional neural network and multi-task learning. *Remote Sensing* 11, 11 (2019), 1265.

[11] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML '01)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 282–289.

[12] Lingbo Liu, Zhilin Qiu, Guanbin Li, Qing Wang, Wanli Ouyang, and Liang Lin. 2019. Contextualized spatial–temporal network for taxi origin-destination demand prediction. *IEEE Transactions on Intelligent Transportation Systems* 20, 10 (2019), 3875–3887.

[13] Xinwang Liu, Yong Dou, Jianping Yin, Lei Wang, and En Zhu. 2016. Multiple kernel k-means clustering with matrix-induced regularization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[14] Luis Moreira-Matias, Joao Gama, Michel Ferreira, Joao Mendes-Moreira, and Luis Damas. 2013. Predicting taxi–passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems* 14, 3 (2013), 1393–1402.

[15] Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. 2020. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 914–921.

[16] Yongxin Tong, Yuqiang Chen, Zimu Zhou, Lei Chen, Jie Wang, Qiang Yang, Jieping Ye, and Weifeng Lv. 2017. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 1653–1662.

[17] Uber. 2020. https://www.uber.com/.

[18] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[19] Chao Wang, Yi Hou, and Matthew Barth. 2019. Data-Driven Multi-step Demand Prediction for Ride-Hailing Services Using Convolutional Neural Network. In *Science and Information Conference*. Springer, 11–22.

[20] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*. 802–810.

[21] Jun Xu, Rouhollah Rahmatizadeh, Ladislau Bölöni, and Damla Turgut. 2017. Real-time prediction of taxi demand using recurrent neural networks. *IEEE Transactions on Intelligent Transportation Systems* 19, 8 (2017), 2572–2581.

[22] Ying Xu and Dongsheng Li. 2019. Incorporating graph attention and recurrent architectures for city-wide taxi demand prediction. *ISPRS International Journal of Geo-Information* 8, 9 (2019), 414.

[23] Hai Yang, Yan Wing Lau, Sze Chun Wong, and Hong Kam Lo. 2000. A macroscopic taxi model for passenger demand, taxi utilization and level of services. *Transportation* 27, 3 (2000), 317–340.

[24] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[25] Junchen Ye, Leilei Sun, Bowen Du, Yanjie Fu, Xinran Tong, and Hui Xiong. 2019. Co-prediction of multiple transportation demands based on deep spatio-temporal neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 305–313.

[26] Gustav Zander. 2017. Predicting taxi passenger demand using artificial neural networks.

[27] Junbo Zhang, Yu Zheng, and Dekang Qi. 2016. Deep spatio-temporal residual networks for citywide crowd flows prediction. *arXiv preprint arXiv:1610.00081* (2016).

[28] Lingyu Zhang, Tao Hu, Yue Min, Guobin Wu, Junying Zhang, Pengcheng Feng, Pinghua Gong, and Jieping Ye. 2017. A taxi order dispatch model based on combinatorial optimization. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 2151–2159.

[29] Xiyue Zhang, Chao Huang, Yong Xu, Lianghao Xia, Peng Dai, Liefeng Bo, Junbo Zhang, and Yu Zheng. 2021. Traffic Flow Forecasting with Spatial-Temporal Graph Diffusion Network. *Proceedings of the AAAI Conference on Artificial Intelligence* 35, 17 (May 2021), 15008–15015.

[30] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Hengchang Liu. 2020. RiskOracle: A Minute-Level Citywide Traffic Accident Forecasting Framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1258–1265.

[31] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Chaochao Zhu. 2020. Foresee Urban Sparse Traffic Accidents: A Spatiotemporal Multi-Granularity Perspective. *IEEE Transactions on Knowledge and Data Engineering* (2020), 1–1.