Spatiotemporal Multi-Graph Convolution Network for Ride-Hailing Demand Forecasting

Xu Geng,[*1] **Yaguang Li,**[*2] **Leye Wang,**[1,3] **Lingyu Zhang,**[4] **Qiang Yang,**[1] **Jieping Ye,**[4] **Yan Liu**[2,4]
[1]Hong Kong University of Science and Technology, [2]University of Southern California, [3]Peking University, [4]Didi AI Labs, Didi Chuxing
xgeng@connect.ust.hk, yaguang@usc.edu, wly@cse.ust.hk, zhanglingyu@didichuxing.com,
qyang@cse.ust.hk, yejieping@didichuxing.com, yanliu.cs@usc.edu

Abstract

Region-level demand forecasting is an essential task in ride-hailing services. Accurate ride-hailing demand forecasting can guide vehicle dispatching, improve vehicle utilization, reduce the wait-time, and mitigate traffic congestion. This task is challenging due to the complicated spatiotemporal dependencies among regions. Existing approaches mainly focus on modeling the Euclidean correlations among spatially adjacent regions while we observe that non-Euclidean pair-wise correlations among possibly distant regions are also critical for accurate forecasting. In this paper, we propose the *spatiotemporal multi-graph convolution network* (ST-MGCN), a novel deep learning model for ride-hailing demand forecasting. We first encode the non-Euclidean pair-wise correlations among regions into multiple graphs and then explicitly model these correlations using multi-graph convolution. To utilize the global contextual information in modeling the temporal correlation, we further propose *contextual gated recurrent neural network* which augments recurrent neural network with a contextual-aware gating mechanism to re-weights different historical observations. We evaluate the proposed model on two real-world large scale ride-hailing demand datasets and observe consistent improvement of more than 10% over state-of-the-art baselines.

Introduction

Spatiotemporal forecasting is a crucial task in urban computing. It has a wide range of applications from autonomous vehicles operations, to energy and smart grid optimization, to logistics and supply chain management. In this paper, we study one important task: region-level ride-hailing demand forecasting, which is one of the essential components of the intelligent transportation systems. The goal of region-level ride-hailing demand forecasting is to predict the future demand of regions in a city given historical observations. Accurate ride-hailing demand forecasting can help organize vehicle fleet, improve vehicle utilization, reduce the wait-time, and mitigate traffic congestion (Yao et al. 2018b). This task is challenging mainly due to the complex spatial and temporal correlations. On the one hand, complicated dependencies are observed among different regions. For example, the demand of a region is usually affected by its spatially adjacent neighbors and at the same time correlated with distant regions with the similar contextual environment. On the other hand, non-linear dependencies also exist among different temporal observations. The prediction of a certain time is usually correlated with various historical observations, e.g., an hour ago, a day ago or even a week ago.

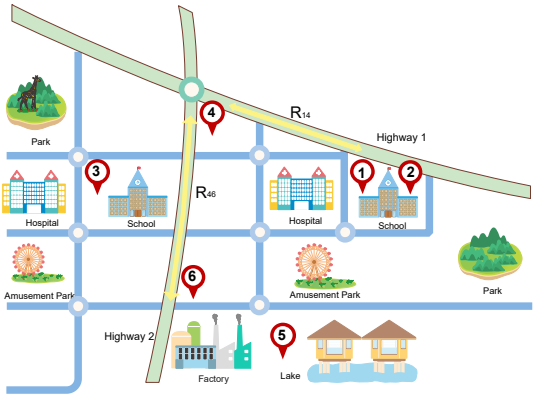

Figure 1: An example of different correlations among regions. To predict the demand in region **1**, spatially adjacent region **2**, functionality similar region **3** and transportation connected region **4** are considered more important, while distant and irrelevant regions **5** are less relevant.

Recent advances in deep learning enable promising results in modeling the complex spatiotemporal relationship in region-based spatiotemporal forecasting. With convolutional neural network and recurrent neural network, state-of-the-art results are achieved in (Shi et al. 2015; Yu et al. 2017; Shi et al. 2017; Zhang, Zheng, and Qi 2017; Zhang et al. 2018a; Ma et al. 2017; Yao et al. 2018b; 2018a). Despite promising results, we argue that two important aspects are largely overlooked in modeling the spatiotemporal correlations. First, these methods mainly focus on modeling the Euclidean correlations among different regions, however, we observe that non-Euclidean pair-wise correlations are also critical for accurate forecasting. Figure 1 shows an example. For region **1**, in addition to neighborhood region **2**, it may also correlate to a distant region **3** that shares similar functionality, i.e., they are both near

[*]Equal contribution. Work done primarily while authors were interns at Didi AI Labs, Didi Chuxing.

schools and hospitals. Besides, region **1** may also be affected by region **4**, which is directly connected to region **1** via a highway. Second, in these methods, when modeling temporal correlation with RNN, each region is processed independently or only based on local information. However, we argue that global and contextual information are also important. For example, a global increase/decrease in ride-hailing demand usually indicates the occurrence of some events that will affect future demand.

To address these challenges, we propose a novel deep learning model called spatiotemporal multi-graph convolution network (ST-MGCN). In ST-MGCN, we propose to encode the non-Euclidean correlations among regions into multiple graphs. Different from (Yao et al. 2018b), which uses the graph embedding as extra constant features for each region, we leverage the graph convolution to explicitly model the pair-wise relationship among regions. Graph convolution is able to aggregate neighborhood information when performing the prediction which is hard to achieve through traditional graph embedding. Furthermore, to incorporate global contextual information when modeling the temporal correlation, we propose contextual gated recurrent neural network (CGRNN). It augments RNN by learning a gating mechanism, which is calculated based on the summarized global information, to re-weight observations in different timestamps. When evaluated on two real-world large scale ride-hailing demand datasets, ST-MGCN consistently outperforms state-of-the-art baselines by a large margin. In summary, this paper makes the following contributions:

- We identify non-Euclidean correlations among regions in ride-hailing demand forecasting and propose to encode them using multiple graphs. Then we further leverage the proposed multi-graph convolution to explicitly model these correlations.
- We propose the Contextual Gated RNN (CGRNN) to incorporate the global contextual information when modeling the temporal dependencies.
- We conduct extensive experiments on two large-scale real-world datasets, and the proposed approach achieves more than 10% relative error reduction over state-of-the-art baseline methods for ride-hailing demand forecasting.

Related work

Spatiotemporal prediction in urban computing

Spatiotemporal prediction is a fundamental problem for data-driven urban management. There are rich amount of works on this topic, including predicting bike flows (Zhang, Zheng, and Qi 2017), the taxi demand (Ke et al. 2017b; Yao et al. 2018b), the arrival time (Li et al. 2018b), and the precipitation (Shi et al. 2015; 2017), where the prediction is aggregated in rectangular regions, and region-wise relationship is modeled by geographical distance. More specifically, the spatial structure of urban data is formulated as a matrix whose entries represent rectangular regions. In previous works, regions and their pair-wise relationships naturally formulate an Euclidean structure, and consequently convolution neural networks are leveraged for effective prediction.

Non-Euclidean structured data also exists in urban computing. Usually, station or point based prediction tasks, like traffic prediction (Li et al. 2018c; Yu, Yin, and Zhu 2018; Yao et al. 2018a), point-based taxi demand prediction (Tong et al. 2017) and station-based bike flow prediction (Chai, Wang, and Yang 2018) are naturally non-Euclidean as the data format is no longer a matrix and convolution neural networks becomes less helpful. Manual feature engineering or graph convolution networks are state-of-the-art techniques for handling non-Euclidean structure data. Different from previous works, ST-MGCN encodes pair-wise relationships among regions into semantic graphs. Though ST-MGCN is designed for region based prediction, the irregularity of region-wise relationship makes it a prediction problem for non-Euclidean data.

In (Yao et al. 2018b), the authors propose DMVST-Net which encodes the region-wise relationship as graph for taxi demand prediction. DMVST-Net mainly uses graph embedding as an external features for spatiotemporal prediction, and consequently fails to use the demand values from related regions. In (Yao et al. 2018a), the authors further improves (Yao et al. 2018b) by modeling the periodically shift problem with the attention mechanism. However, none of these approaches explicitly models the non-Euclidean pair-wise relationships among regions. In this work, ST-MGCN uses the proposed multi-graph convolution to incorporate features from related regions, which is able to make predictions from demand values of regions that are related in different perspective.

Recent research in neuroimage analysis for Parkinson's disease (Zhang et al. 2018b) shows the effectiveness of graph convolution network in spatial feature extraction. It uses GCN to learn features from most similar regions and proposed a multi-view structure to fuse different MRI acquisitions. However, temporal dependency is not considered in above work. ST-GCN is used in spatiotemporal prediction for skeleton based action recognition (Li et al. 2018a; Yan, Xiong, and Lin 2018).The transformation of ST-GCN is a combination of spatial dependency and local temporal recurrence. However, we argue in these models, the contextual information or the global information is largely overlooked in the temporal dependency modeling.

Graph convolution network

Graph convolution network (GCN) is defined over a graph $\mathcal{G} = (V, \boldsymbol{A})$, where V is the set of all vertices and $\boldsymbol{A} \in \mathbb{R}^{|V|\times|V|}$ is the adjacency matrix whose entries represent the connections between vertices. GCN is able to extract local features with different reception fields from translation variant non-Euclidean structures (Hammond et al. 2011). Let $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ denotes the graph Laplacian matrix, where $\boldsymbol{D}$ is the degree matrix, a graph convolution operation (Defferrard, Bresson, and Vandergheynst 2016) is defined as

$$\boldsymbol{X}_{l+1} = \sigma\Big(\sum_{k=0}^{K-1} \alpha_k \boldsymbol{L}^k \boldsymbol{X}_l\Big)\ ^{1} \tag{1}$$

[1] In a graph convolution layer with P inputs and Q outputs, there will be PQ convolution operations. Here, we only show one

where $\boldsymbol{X}_l$ denotes the features in the l-th layer, α_k is the trainable coefficient, $\boldsymbol{L}^k$ is the k-th power of the graph Laplacian matrix, σ is the activation function.

Channel-wise attention

Channel-wise attention (Hu, Shen, and Sun 2018; Chen et al. 2017) is proposed in the computer vision literature. The intuition behind channel-wise attention is to learn a weight for each channel, in order to find the most important frames and emphasize them by giving higher weights. Let $\boldsymbol{X} \in \mathbb{R}^{W \times H \times C}$ denotes the input, where W and H are the dimensions of the input image, and C denotes the number of channels, then the pipeline of channel-wise attention is defined as follows:

$$z_c = F_{pool}(\boldsymbol{X}_{:,:,c}) = \frac{1}{WH} \sum_{i=0}^{W} \sum_{j=0}^{H} X_{i,j,c} \text{ for } c = 1, 2, \cdots C$$
$$\boldsymbol{s} = \sigma(\boldsymbol{W}_2\, \delta(\boldsymbol{W}_1 \boldsymbol{z})) \tag{2}$$
$$\tilde{\boldsymbol{X}}_{:,:,c} = \boldsymbol{X}_{:,:,c} \circ s_c \quad \text{for } c = 1, 2, \cdots C$$

F_{pool} is a global average pooling operation, which summarizes each channel into a scalar $\boldsymbol{z}_c$ where c is the channel index. Then an attention operation is applied to generate adaptive channel weights $\boldsymbol{s}$ by applying non-linear transformations on the summarized vector $\boldsymbol{z}$, where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ is the corresponding weights, δ and σ is the ReLU and sigmoid function respectively. After that, $\boldsymbol{s}$ is applied to the input via channel-wise dot product. Finally, the input channels are scaled based learned weights. In this work, we adopt the idea of channel-wise attention, and generalize it for temporal dependency modeling among a sequence of graphs.

Methodology

We formalize the learning problem of spatiotemporal ride-hailing demand forecasting and describe how to model the spatial and temporal dependencies using the proposed *spatiotemporal multi-graph convolution network* (ST-MGCN).

Region-level ride-hailing demand forecasting

We divide a city into equal-size grids, and each grid is defined as a *region* $v \in V$, where V denotes the set of all disjoint regions in the city. Let $\boldsymbol{X}^{(t)}$ represent the number of orders in all regions at the t-th interval. Then the *region-level ride-hailing demand forecasting* problem is formulated as a single step spatiotemporal prediction given input with a fixed temporal length, i.e., learning a function $f : \mathbb{R}^{|V| \times T} \rightarrow \mathbb{R}^{|V|}$ that maps historical demands of all regions to the demand in the next timestep.

$$[\boldsymbol{X}^{(t-T+1)}, \cdots, \boldsymbol{X}^{(t)}] \xrightarrow{f(\cdot)} \boldsymbol{X}^{(t+1)}$$

Framework overview The system architecture of the proposed model ST-MGCN is shown in Figure 2. We represent different aspects of correlations between regions as multiple graphs, whose vertices represent regions and edges encode the pair-wise relationship among regions. First, we use the proposed *Contextual Gated Recurrent Neural Network* (CGRNN) to aggregate observations in different times considering the global contextual information. After that, multi-graph convolution is applied to capture different types of correlations between regions. Finally, a fully connected neural network is used to transform features into the prediction.

operation for simplicity.

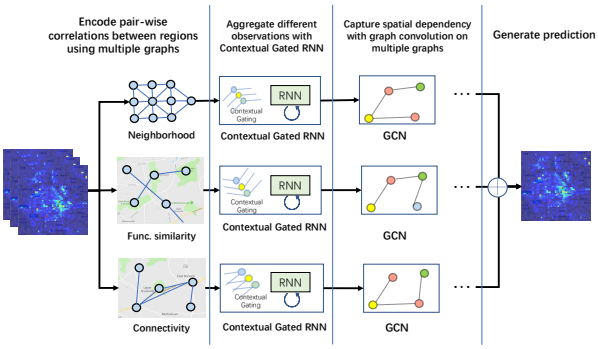

Figure 2: System architecture of the proposed *spatiotemporal multi-graph convolution network* (ST-MGCN). We encode different aspects of relationships among regions, including neighborhood, functional similarity and transportation connectivity, using multiple graphs. First, the proposed *contextual gated recurrent neural network* (CGRNN) is used to aggregate observations in different times considering the global contextual information. After that, multi-graph convolution is used to model the non-Euclidean correlations among regions.

Spatial dependency modeling

In this section, we show how to encode different types of correlations among regions using multiple graphs and how to model these relationships using the proposed multi-graph convolution.

We model three types of correlations among regions with graphs, including (1) the neighborhood graph $\mathcal{G}_N = (V, \boldsymbol{A}_N)$, which encode the spatial proximity, (2) functional similarity graph $\mathcal{G}_F = (V, \boldsymbol{A}_F)$, which encodes the similarity of surrounding Point of Interests (POIs) of regions, and (3) the transportation connectivity graph $\mathcal{G}_T = (V, \boldsymbol{A}_T)$, which encodes the connectivity between distant regions. Note that, our approach can be easily extended to model new types of correlations by constructing related graphs.

Neighborhood Neighborhood of a region is defined based on the spatial proximity. We construct the graph by connecting a region to its 8 adjacent regions in a 3×3 grid.

$$A_{N,ij} = \begin{cases} 1, & v_i \text{ and } v_j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Functional similarity When making prediction for a region, it is intuitive to refer to other regions that are similar to

this one in terms of functionality. Region functionality could be characterized using its surrounding POIs for each category, and the edge between two vertices (regions) is defined as the POI similarity:

$$A_{S,i,j} = \text{sim}(P_{v_i}, P_{v_j}) \in [0, 1] \quad (4)$$

where P_{v_i}, P_{v_j} are the POI vectors of regions v_i and v_j respectively, whose dimension equals to the number of POI categories and each entry represents the number of a specific POI category in the region.

Transportation connectivity The transportation system is also an important factor when performing spatiotemporal predictions. Intuitively, those geographically distant but conveniently reachable regions can be correlated. These kinds of connectivity are induced by roads like motorway, highway or public transportation like subway. Here, we define regions that are directly connected by these roads as "connected" and the corresponding edge is defined as:

$$A_{C,i,j} = max(0, \text{conn}(v_i, v_j) - A_{N,i,j}) \in \{0, 1\} \quad (5)$$

where $\text{conn}(u, v)$ is the indicator function of the connectivity between v_i and v_j. Note that, the neighborhood edges are removed from connectivity graph to avoid redundant correlations and also results in a sparser graph.

Multi-graph convolution for spatial dependency modeling With these graphs constructed, we propose the multi-graph convolution to model the spatial dependency as defined in Equation 6.

$$\boldsymbol{X}_{l+1} = \sigma\left(\bigsqcup_{\mathbf{A}\in\mathbb{A}} f(\mathbf{A}; \theta_i)\boldsymbol{X}_l\mathbf{W}_l\right) \quad (6)$$

where $\boldsymbol{X}_l \in \mathbb{R}^{|V|\times P_l}, \boldsymbol{X}_{l+1} \in \mathbb{R}^{|V|\times P_{l+1}}$ are the feature vectors of $|V|$ regions in layer l and $l+1$ respectively. σ denotes the activation function, and $\bigsqcup$ denotes the aggregation function, e.g., sum, max, average etc. $\mathbb{A}$ denotes the set of graphs, and $f(\mathbf{A}; \theta_i) \in \mathbb{R}^{|V|\times|V|}$ represents the aggregation matrix of different samples based on graph $\mathbf{A} \in \mathbb{A}$ parameterized by θ_i, while $\mathbf{W}_l \in \mathbb{R}^{P_l\times P_{l+1}}$ denotes the feature transformation matrix, For example, if $f(\mathbf{A}; \theta_i)$ is the polynomial function of the Laplacian matrix $\mathbf{L}$, then this will become ChebNet (Defferrard, Bresson, and Vandergheynst 2016) on multiple graphs. If $f(\mathbf{A}; \theta_i) = \mathbf{I}$, i.e., the identity matrix, then this will fall back to the fully connected network.

In the implementation, $f(\mathbf{A}; \theta_i)$ is chosen to be the K order polynomial function of the graph Laplacian $\mathbf{L}$, and Figure 3 shows an example of the value transformation for a centralized region through the graph convolution layer. Suppose all the entries in the adjacency matrix are 0 or 1, entry $L^k_{ij} \neq 0$ means v_i is able to reach v_j in k-hop. In terms of convolution operation, k defines the size of reception field during spatial feature extraction. Using road connectivity graph $\mathcal{G}_C = (V, \boldsymbol{A}_C)$ in Figure 1 to illustrate. In the adjacency matrix $\boldsymbol{A}_C$, we have:

$$A_{C,1,4} = 1; A_{C,1,6} = 0; A_{C,4,6} = 1,$$

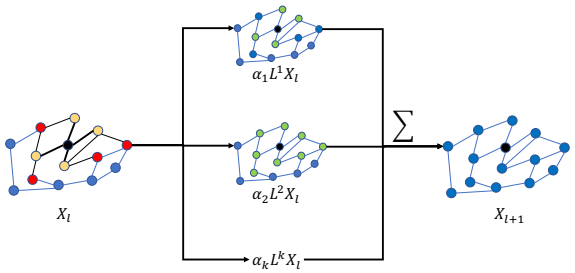

Figure 3: An example of the ChebNet graph convolution centralized at the black vertex. Left: The centralized region is marked black. The one-hop neighbors are marked yellow, while the two-hop neighbors are marked red. Middle: with the increase of degree of the graph Laplacian, the reception field grows (marks green). Right: The output of this layer is a sum among graph transformations with degree value from 1 to K.

and the corresponding entries of the 1-degree graph Laplacian are:

$$L^1_{C,1,4} \neq 0; L^1_{C,1,6} = 0; L^1_{C,4,6} \neq 0$$

If the maximum degree of graph Laplacian K is set to 1, the transformed feature vector of region 1, i.e., $\boldsymbol{X}_{l+1,1,:}$ will not contain the feature vector of region 6: $\boldsymbol{X}_{l,6,:}$, since $L^1_{C,1,6} = 0$. When increasing K to 2, the corresponding entry $L^2_{C,1,6}$ becomes non-zero, and consequently $\boldsymbol{X}_{l+1,1,:}$ can utilize information from $\boldsymbol{X}_{l,6,:}$.

The multi-graph convolution based spatial dependency modeling is not restricted to these three types of region-wise relationships mentioned above, and it can be easily extended to model other region-wise relationships as well as other spatiotemporal forecasting problems. It models spatial dependencies by feature extraction through region-wise relationship. With small reception field, the feature extraction will focus on close regions, i.e., neighbors that can be reaches with small number of hops. Increasing the max degree of graph Laplacian or stacking multiple convolution layers will increase the reception field and consequently encourage the model to capture more global dependencies.

Graph embedding is an alternative technique for modeling the region-wise correlation. In DMVST-Net (Yao et al. 2018b), the authors use graph embedding[2] to represent region-wise relationship, and then add these embeddings as extra features to each region. We argue that spatial dependency modeling approach in ST-MGCN is preferred for the following reasons: ST-MGCN encodes region-wise relationships into graphs and aggregate demand values from related regions by graph convolution. While in DMVST-Net, the region-wise relationship was embedded to a temporal invariant region-based feature as input to the model.

[2]The graph embedding is pre-computed. It produces a temporal-invariant feature vector for each region.

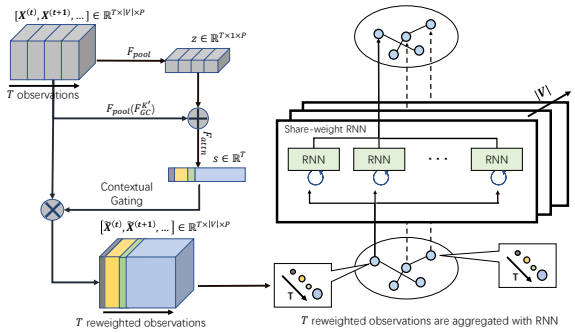

Figure 4: Temporal correlation modeling with contextual gated recurrent neural network (CGRNN). It first produces region descriptions using the global average pooling over the input and its graph convolution output for each observation. Then it transfer the summarized vector $\boldsymbol{z}$ into weights which are used to scale each observation. Finally, a shared RNN layer across all regions is applied to aggregate the gated input sequence of each region into a single vector.

Though DMVST-Net also captures the topological information, it is hard to aggregate demand values from related regions through the region-wise relationship. Also, invariant features have limited contribution to the model training.

Temporal correlation modeling

We propose the *Contextual Gated Recurrent Neural Network* (CGRNN) to model the correlations between observations in different timestamps. CGRNN incorporates contextual information into the temporal modeling by augmenting RNN with a context aware gating mechanism whose architecture is shown in Figure 4. Suppose, we have T temporal observations and $\boldsymbol{X}^{(t)} \in \mathbb{R}^{|V| \times P}$ denotes the t-th observation, where P is the feature dimensions, P will be 1 if the feature only contains the number of orders. Then the workflow of contextual gating mechanism is as follows.

$$\hat{\boldsymbol{X}}^{(t)} = [\boldsymbol{X}^{(t)}, F_{\mathcal{G}}^{K'}(\boldsymbol{X}^{(t)})] \text{ for } t = 1, 2, \cdots T \quad (7)$$

First, the contextual gating mechanism produces region descriptions by concatenating the historical data of a certain region with information from related regions. The information from related regions is regarded as contextual information, and is extracted by a graph convolution operation $F_{\mathcal{G}}^{K'}$ with max degree K' (Equation 7) using the corresponding graph Laplacian matrix. The contextual gating mechanism is designed to involve information from related regions by performing graph convolution operation before the pooling step.

$$z^{(t)} = F_{pool}(\hat{\boldsymbol{X}}^{(t)}) = \frac{1}{|V|} \sum_{i=1}^{|V|} \hat{X}_{i,:}^{(t)} \text{ for } t = 1, 2, \cdots T \quad (8)$$

Secondly, we use the global average pooling F_{pool} over all regions to produce the summary of each temporal observation (Equation 8).

$$\boldsymbol{s} = \sigma(\boldsymbol{W}_2\ \delta(\boldsymbol{W}_1 \boldsymbol{z})) \quad (9)$$

Then an attention operation (Equation 9) is applied to the summarized vector $\boldsymbol{z}$, where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ is the corresponding weights, δ and σ is the ReLU and sigmoid function respectively.

$$\tilde{\boldsymbol{X}}^{(t)} = \boldsymbol{X}^{(t)} \circ s^{(t)} \quad \text{for } t = 1, 2, \cdots T \quad (10)$$

Finally, $\boldsymbol{s}$ is applied to the scale each temporal observation (Equation 10).

$$\boldsymbol{H}_{i,:} = \text{RNN}(\tilde{\boldsymbol{X}}_{i,:}^{(1)}, \cdots, \tilde{\boldsymbol{X}}_{i,:}^{(T)}; \boldsymbol{W}_3) \text{ for } i = 1, \cdots, |V| \quad (11)$$

After the contextual gating, a shared RNN layer with weight $\boldsymbol{W}_3$ across all regions is applied to aggregate the gated input sequence of a region into a single vector $\boldsymbol{H}_{i,:}$ (Equation 11). The intuition of sharing RNN among regions is to find a universal aggregation rule for all regions, which encourages model generalization and reduces model complexity.

Experiments

In this section, we compare the proposed model ST-MGCN with other state-of-the-art baselines for region-level ride-hailing demand forecasting.

Dataset We conduct experiments on two real-world large scale ride-hailing datasets collected in cities: **Beijing** and **Shanghai**. Both of these datasets are collected in the main city zone of ride-hailing orders within the time period from Mar 1st, 2017 to Dec 31st, 2017. For data split, we use the data from Mar 1st 2017 to Jul 31st 2017 for training, data from Aug 1st 2017 to Sep 30th 2017 as validation, and the data from Oct 1st 2017 to Dec 31st 2017 is used for testing. The POI data is collected in 2017, and contains 13 primary POI categories. Each region is associated with a POI vector, whose entry is the number of instances of a certain POI category. The road network data used for transportation connectivity evaluation is provided by OpenStreetMap (Haklay and Weber 2008).

Experimental Settings

Recall that the learning task is formulated as learning a function $f : \mathbb{R}^{|V| \times T} \to \mathbb{R}^{|V|}$. In the experiment, we generate the region set V by partitioning city map into grids with size equals to $1km \times 1km$ [3]. There are totally 1296 regions in Beijing, and 896 regions in Shanghai. Following the practice in (Zhang, Zheng, and Qi 2017), the input of the network consists of 5 historical observations, including 3 latest **closeness** components, 1 **period** component and 1 latest **trend** component. In building the transportation connectivity graph, we consider the following high-speed roads, including motorway, highway and subway. Two regions are regarded as "connected" as long as there is a high-speed road directly connecting them.

In the experiment, $f(\mathbf{A}; \theta_i)$ in Equation 6 is chosen to be the Chebyshev polynomial functionl (Defferrard, Bresson, and Vandergheynst 2016) of the graph Laplacian with the degree K equals to 2, and $\bigsqcup$ is chosen to be the sum aggregation function. The number of hidden layers is 3, with

[3] Referred to industrial practice.

Table 1: Performance comparison of different approaches for ride-hailing demand forecasting. ST-MGCN achieves the best performance with all metrics on both datasets.

Method	Beijing		Shanghai	
	RMSE	MAPE(%)	RMSE	MAPE(%)
HA	16.14	23.9	17.15	34.8
LASSO	14.24±0.14	23.8±0.8	10.62±0.06	22.9±0.8
Ridge	14.24±0.11	23.8±0.9	10.61±0.04	23.1±0.8
VAR	13.32±0.17	22.4±1.6	10.54±0.18	23.7±1.4
STAR	13.16±0.22	22.2±1.9	10.52±0.21	23.2±1.4
GBM	13.66±0.16	23.1±1.5	10.25±0.11	23.4±1.2
STResNet	11.77±0.95	14.8±6.0	9.87±0.94	14.9±6.0
DMVST-Net	11.62±0.48	12.3±5.5	9.61±0.44	13.8±1.2
ST-GCN	11.62±0.36	10.1±5.1	9.29±0.31	11.2±1.3
ST-MGCN	**10.78±0.25**	**8.8±3.5**	**8.30±0.16**	**9.3±0.9**

64 hidden units each and an L2 regularization with a weight decay equal to 1e-4 is applied to each layer. Specially, the graph convolution degree K' in CGRNN equals to 1.

We use ReLU as the activation in the graph convolution network. The learning rate of ST-MGCN is set to 2e-3, and early stopping on the validation dataset is used. All neural network based approaches are implemented using Tensorflow (Abadi and others 2016), and trained using the Adam optimizer (Kingma and Ba 2015) for minimizing RMSE. The training of ST-MGCN takes 10GB RAM and 9GB GPU memory. The training process takes about 1.5 hour on a single Tesla P40.

Methods for evaluation We compare the proposed model (ST-MGCN) with the following methods for ride-hailing demand forecasting:

- **Historical Average** (HA): which models the ride-hailing demand as a seasonal process, and uses the average of previous seasons as the prediction. The period used is 1 week, and the prediction is based on aggregated data from the same time in previous weeks.
- **LASSO, Ridge**: which takes historical data from different timestamps as input for linear regression with L1 and L2 regularization respectively.
- **Auto-regressive model** (VAR,STAR): VAR is the multivariate extension of auto-regressive model which is able to model the correlation between regions. STAR (Pace et al. 1998) is a an AR extension specifically for spatiotemporal modeling problems. In the experiment, the number of lags used is 5.
- **Gradient boosted machine** (GBM): gradient boosting decision tree based regression implemented using LightGBM (Ke et al. 2017a). The following setting is used in the experiment: the number of trees is 50, the maximum depth is 4 and the learning rate is 2e-3.
- **ST-ResNet** (Zhang, Zheng, and Qi 2017): ST-ResNet is a CNN-based framework for traffic flow prediction. The model uses CNN with residual connections to capture the trend, the periodicity, and the closeness information.
- **DMVST-Net** (Yao et al. 2018b): DMVST-Net is a multi-view based deep learning approach for taxi demand prediction. It consists of three different views: the temporal view, the spatial view, and the semantic view modeled with LSTM, CNN and graph embedding respectively.
- **DCRNN, ST-GCN**: Both DCRNN (Li et al. 2018c) and ST-GCN (Yu, Yin, and Zhu 2018) are graph convolution based models for traffic forecasting. Both models use road network for building non-euclidean region-wise relationship. DCRNN models the spatiotemporal dependency by integrating graph convolution into the gated recurrent unit, while ST-GCN models the both the spatial and temporal dependencies with convolution structures and achieves better efficiency.

Performance comparison

For all approaches, we tune the model parameters using grid search based on the performance on the validation dataset, and report the performance on the testing dataset over multiple runs. We evaluate the performance of based on two popular metrics, i.e., Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) [4]. Table 1 shows the test error comparison of different approaches for ride-hailing demand forecasting over of ten runs.

We observe the following phenomena in both datasets: (1) deep learning based methods, including ST-ResNet, DMVST-Net, ST-GCN and the proposed ST-MGCN, which are able to model the non-linear spatiotemporal dependencies, generally outperform other baselines; (2) ST-MGCN achieves the best performance regarding all the metrics on both datasets, outperforming the second best baseline by at least 10% in terms of relative error reduction, which suggests the effectiveness of proposed approaches for spatiotemporal correlations modeling; (3) compared with other deep learning models, ST-MGCN also shows lower variance.

Effect of spatial dependency modeling

To investigate the effect of spatial and temporal dependency modeling, we evaluate the following variants of ST-MGCN by removing different components from the model, including: (1) the neighborhood graph, (2) the functional similarity graph, (3) the transportation connectivity graph. The result is shown in Table 2. Removing any graph component causes a significant error increase which justifies the importance of each type of relationship. These graphs encode the important prior knowledge, i.e., region-wise correlation, which is leveraged for more accurate forecasting.

To evaluate the effect of incorporating multiple region-wise relationships, we extend existing single graph-based models, including DCRNN (Li et al. 2018c) and ST-GCN (Yu, Yin, and Zhu 2018) with the multi-graph convolution framework and the resulted models are DCRNN+

[4]Following the practice in (Yao et al. 2018b), we filter the samples with demand values less than 10 when computing MAPE.

and ST-GCN+. As shown in table 3, both DCRNN+ and ST-GCN+ achieve improved performance which shows the effectiveness of incorporating multiple region-wise relationships.

Table 2: Effect of spatial correlation modeling on the Beijing dataset. Removing any component will result in a statistically significant error increase.

Removed component	RMSE
Neighborhood	11.47
Functional	11.42
Transportation	11.69
ST-MGCN	**10.78**

Table 3: Effect of adding multi-graph design to existing methodologies on the Beijing dataset. Adding extra graph to original model will result in a statistically significant error decrease.

Model	RMSE
ST-GCN	11.62
ST-GCN+	11.20
DCRNN	12.02
DCRNN+	11.55
ST-MGCN	**10.78**

Effect of temporal dependency modeling

To further investigate the effect of temporal dependency modeling, we evaluate the following variants of ST-MGCN using different methods for temporal modeling, including (1) **Average pooling**: which aggregates different temporal observations using the average pooling, (2) **RNN**: which aggregates temporal observations using the recurrent neural network (RNN) (3) **CG**: which uses contextual gating to re-weight different temporal observations but without RNN (4) **GRNN**: CGRNN without the graph convolution (Equation 7). The results are shown in Table 4. We observe the following phenomena:

- Average pooling which blindly averages different observations has the worst performance, while RNN which is able to do content dependent non-linear temporal aggregation achieves clearly improved results.
- CGRNN which augments RNN with contextual gating mechanism achieves further improved result than RNN. Besides, removing either the RNN (CG) or the graph convolution operation (GRNN) results in clear worse performance which justify the effectiveness of each component.

Effect of model parameters

To study the effects of different hyperparameters of the proposed model, we evaluate models on the Beijing by varying two of the most important hyperparameters, i.e., the degree

Table 4: Effect of temporal correlation modeling on the Beijing dataset

Temporal modeling approach	RMSE
Average pooling	12.74
RNN	11.05
CG	11.82
GRNN	10.91
CGRNN	**10.78**

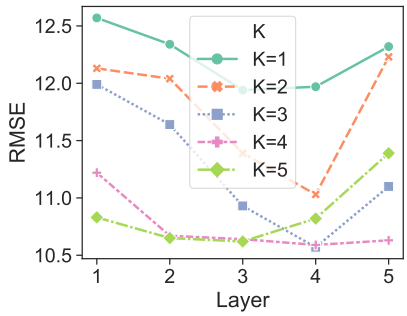

Figure 5: Effect of number of layers and the polynomial order K of the graph convolution on the Beijing dataset.

K and number of layers in the graph convolution. Figure 5 shows the performance on test set. We observe that with the increase of number of layers, the error first decreases and then increases. While the error first decreases and then plateaus with the increase of K. Larger K or the number of layers will enable the model capture more global correlation at the cost of increased model complexity and more prune to overfitting.

Conclusion and Future work

In this paper, we investigated the region-level ride-hailing demand forecasting problem and identified its unique spatiotemporal correlations. We proposed a novel deep learning based model which encoded the non-Euclidean correlations among regions using multiple graphs and explicitly captured them using multi-graph convolution. We further augmented the recurrent neural network with contextual gating mechanism to incorporate global contextual information in the temporal modeling procedure. When evaluated on two large scale real-world ride-hailing demand datasets, the proposed approach achieved significantly better results than state-of-the-art baselines. For future work, we plan to investigate the following aspects (1) evaluate the proposed model on other spatiotemporal forecasting tasks; (2) extend the proposed approach for multiple step sequence forecasting.

Acknowledgement

This research has been funded in part by NSF grants IIS-1254206, IIS-1539608, ITSP project No. ITS/391/15FX and

Hong Kong CERG grants 16209715, 16244616. The research is supported by Didi Chuxing. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the sponsors such as NSF.

References

Abadi, M., et al. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*.

Chai, D.; Wang, L.; and Yang, Q. 2018. Bike flow prediction with multi-graph convolutional networks. *SIGSPATIAL*.

Chen, L.; Zhang, H.; Xiao, J.; Nie, L.; Shao, J.; Liu, W.; and Chua, T.-S. 2017. SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 6298–6306. IEEE.

Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, 3844–3852.

Haklay, M., and Weber, P. 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7(4):12–18.

Hammond, D. K.; Vandergheynst, P.; Gribonval, R.; Hammond, D. K.; Vandergheynst, P.; and Gribonval, R. 2011. Wavelets on graphs via spectral graph theory. 30(2):129–150.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on*. IEEE.

Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; and Liu, T.-Y. 2017a. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, 3149–3157.

Ke, J.; Zheng, H.; Yang, H.; and Chen, X. M. 2017b. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies* 85:591–608.

Kingma, D. P., and Ba, J. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR '14)*.

Li, C.; Cui, Z.; Zheng, W.; Xu, C.; and Yang, J. 2018a. Spatio-temporal graph convolution for skeleton based action recognition. In *2018 AAAI Conference on Artificial Intelligence (AAAI'18)*.

Li, Y.; Fu, K.; Wang, Z.; Shahabi, C.; Ye, J.; and Liu, Y. 2018b. Multi-task representation learning for travel time estimation. In *International Conference on Knowledge Discovery and Data Mining (KDD '18)*.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2018c. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR '18)*.

Ma, X.; Dai, Z.; He, Z.; Ma, J.; Wang, Y.; and Wang, Y. 2017. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors* 17(4):818.

Pace, R. K.; Barry, R.; Clapp, J. M.; and Rodriquez, M. 1998. Spatiotemporal autoregressive models of neighborhood effects. *The Journal of Real Estate Finance and Economics* 17(1):15–33.

Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, 802–810.

Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.-Y.; Wong, W.-k.; and Woo, W.-c. 2017. Deep learning for precipitation nowcasting: A benchmark and a new model. In *Advances in Neural Information Processing Systems*, 5617–5627.

Tong, Y.; Chen, Y.; Zhou, Z.; Chen, L.; Wang, J.; Yang, Q.; Ye, J.; and Lv, W. 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. ACM.

Yan, S.; Xiong, Y.; and Lin, D. 2018. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *2018 AAAI Conference on Artificial Intelligence (AAAI'18)*.

Yao, H.; Tang, X.; Wei, H.; Zheng, G.; Yu, Y.; and Li, Z. 2018a. Modeling spatial-temporal dynamics for traffic prediction. *arXiv preprint arXiv:1803.01254*.

Yao, H.; Wu, F.; Ke, J.; Tang, X.; Jia, Y.; Lu, S.; Gong, P.; Ye, J.; and Li, Z. 2018b. Deep multi-view spatial-temporal network for taxi demand prediction. In *2018 AAAI Conference on Artificial Intelligence (AAAI'18)*.

Yu, R.; Li, Y.; Shahabi, C.; Demiryurek, U.; and Liu, Y. 2017. Deep learning: A generic approach for extreme condition traffic forecasting. In *SIAM International Conference on Data Mining (SDM)*.

Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *IJCAI'18)*.

Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X.; and Li, T. 2018a. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence* 259:147–166.

Zhang, X.; He, L.; Chen, K.; Luo, Y.; Zhou, J.; and Wang, F. 2018b. Multi-view graph convolutional network and its applications on neuroimage analysis for parkinson's disease. *arXiv preprint arXiv:1805.08801*.

Zhang, J.; Zheng, Y.; and Qi, D. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, 1655–1661.