TAXI ORIGIN-DESTINATION DEMAND PREDICTION WITH CONTEXTUALIZED SPATIAL-TEMPORAL NETWORK

Zhilin Qiu, Lingbo Liu, Guanbin Li, Qing Wang, Nong Xiao, Liang Lin[*]

School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China

ABSTRACT

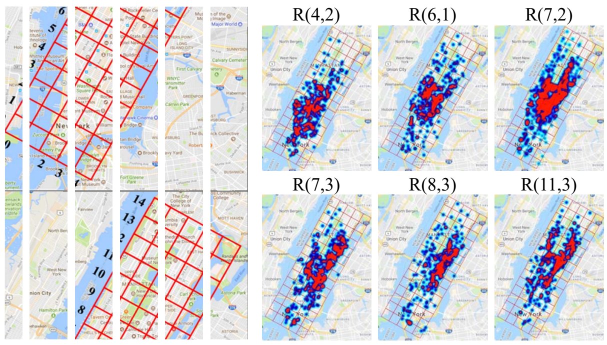
Taxi demand prediction has recently attracted increasing research interest due to its huge potential application in large-scale intelligent transportation systems. However, most of the previous methods only considered the taxi demand prediction in origin regions, while ignoring the modeling of the specific situation of the destination passengers. In this paper, we present a more challenging task, called taxi origin-destination demand prediction, which aims at predicting the taxi demand between all origin-destination (OD) pairs in a future time interval. Its main challenges lie in how to effectively capture the diverse contextual information to learn the demand patterns. We address this problem with a novel Contextualized Spatial-Temporal Network (CSTN), which can effectively capture various context of taxi demand into a unified framework. Specifically, the proposed network consists of three components for the modeling of local spatial context (LSC), temporal evolution context (TEC) and global correlation context (GCC) respectively. Extensive experiments and evaluations on a large-scale dataset well demonstrate the significant superiority of our CSTN over other compared methods of taxi origin-destination demand prediction.

***Index Terms*—** Mobility Data, Heat Map, Taxi Demand, Spatial-Temporal Modeling

1. INTRODUCTION

Taxi as one of the most common travel modes for urban residents, has greatly penetrated into people's daily life. With the increasing popularity of online taxi services platforms such as Uber and Didi Chuxing, we are able to collect the mobility data of taxi requests continuously. How to utilize such large-scale data to improve the demand prediction is an interesting and critical real-world problem.

As a crucial task in intelligent transportation system (ITS), taxi demand prediction [1–3] has attracted a wide range of research interest. The recent deep learning based methods [4, 5] divide a city into a grid map based on geographical coordinate and transform the taxi request data into heat maps to forecast the future demand. However, these methods only model the taxi demand at the departure place and estimate the requests for taxis in all regions or some specific regions, ignoring the influence of each passenger's destination. We believe the information of passengers' destinations is critical for the taxi allocation. Some taxi preallocation systems deploy the taxis in advance based solely on the predicted taxi origin demand, which may suffer from the following issues:



(a) (b)

Fig. 1. **(a)** Illustration of the region partition on a city. We partition a city into a grid map based on the longitude and latitude. Here is the example of the Manhattan in New York City. **(b)** Visualization of the heat maps of taxi demand by mapping the passengers' pick-up locations back to the geo-coordinates on Google map. The sub-figure with title "R(i,j)" is the taxi demand from all regions to the region R(i,j).

- Limited by the city management rules (such as the driving restriction policy in Beijing), some drivers are only allowed to drive in some specified regions. If a taxi driver is assigned to a region, where most passengers are to go to a restricted area of the driver, he/she cannot take orders, which may result in a waste of resources.
- Some drivers prefer to carry passengers in their familiar regions. Meanwhile, other drivers are unwilling to take the short trip orders for mild profit. If the destinations of most passengers in the driver's preallocated region are out of his/her operating regions or too close to the pickup locations, the driver may reject those requests.

Corresponding author is Liang Lin.

This work was supported in part by the National Key Research and Development Program of China under Grant No.2018YFC0830103, in part by the National Science Foundation of China under Grant No.U1811463, No.61602533 and No.61702565, in part by the Fundamental Research Funds for the Central Universities under Grant No.18lgpy63.

978-1-5386-9552-4/19/$31.00 ©2019 IEEE
DOI 10.1109/ICME.2019.00136

In this paper, we propose a more challenging task, taxi OD demand prediction, which aims at predicting the future taxi demand between any two regions. If this fine-grained taxi demand is well predicted, we can efficiently preallocate the taxis to meet the passengers' requests. The main challenges of this task lie in how to effectively capture the diverse spatial-temporal contextual information to learn the demand patterns. For example, some regions that are spatially adjacent usually have the similar demand patterns (e.g, the number of taxi requests and the demand trends), which is called the local spatial context (LSC) in our work. Moreover, even though two regions are spatially distant, their demand patterns may still have some relevance, if they share similar functionality (e.g., both of them are residential districts). We call this relationship between two far-apart regions as global correlation context (GCC). Finally, taxi demand is a time-varying process and its evolution is related to various factors, such as its current states and the ever-changing meteorology, which is formulated as temporal evolution context (TEC).

In this paper, we address this problem with a novel Contextualized Spatial-Temporal Network (CSTN), which well integrates the local spatial context, temporal evolution context and the global correlation context into a unified framework. Specifically, our proposed network consists of three modules respectively for the three types of context modeling. First, a LSC module utilizes two convolution neural networks to learn the local spatial dependencies of taxi demand respectively from the origin view and the destination view. The output of these two networks are combined to generate the local spatial feature, which involves the hybrid information of taxi demand patterns from different views. Second, a TEC module incorporates the local spatial features of taxi demand and the meteorological information to a ConvLSTM [6] for the analysis of taxi demand evolution. Third, to capture the correlation between the far-apart regions, a GCC module is proposed to compute the similarity between any two regions and to generate the global correlation feature of each region, by performing a weighted summation on the features of all regions. In this way, the representation of each region involves the information of all regions and is highly correlated with the regions with high similarity values. Finally, the local spatial-temporal feature generated by TEC module as well as the global correlation feature generated by GCC module are integrated to predict the future taxi OD demand. The main contributions of this work are two-fold:

- We extend the existing taxi demand prediction to the taxi OD demand prediction, which is more meaningful for ITS. To the best of our knowledge, we are the first to touch this interregional taxi demand prediction.
- We propose a novel Contextualized Spatial-Temporal Network, which can effectively capture the local spatial context, temporal evolution context and global correlation context into a unified framework.

2. RELATED WORK

Taxi Demand Prediction: Recently, deep neural network has been widely used in ITS [7–10] and some deep learning based methods have also been proposed for taxi demand prediction. For instance, Wang et al. [11] designed a neural network framework using context data from multiple sources to predict the gap between taxi supply and demand. Yao et al. [4] proposed a Deep Multi-View Spatial-Temporal Network framework to model both spatial and temporal relations of taxi demand. Zhou et al. [12] built an attention-based neural network to predict the passenger pickup/dropoff demand on each region, but they were still not applicable to taxi demand prediction among OD pairs. **However**, all the above methods only forecast the taxi request number in each region or at some specific locations. In contrast, our method attempts to predict interregional taxi demand, which can help taxi preallocation systems to allocate the taxis more efficiently.

Origin Destination Estimation: This task aims at estimating the flow between the endpoints (e.g., highway toll booths and bus stations) of the traffic network. Related works can be divided into static estimation [13] and dynamic estimation [14]. The former considers the traffic flow as time-independent and estimate the average demand, which is suitable for long-time transportation planning and design purpose. The latter estimates the time-variant flow between each origin and destination, which can be used for short time route guidance and dynamic traffic assignment. **However**, these methods are designed to estimate the traffic flow between some specific positions and are not effective for citywide taxi preallocation, as taxi passengers can be located in any area.

3. PRELIMINARIES

Region Partition: In this work, we focus on the taxi demand prediction between regions, rather than the specific positions. Following the previous works [4, 10], we partition a city into $H \times W$ non-overlapping grid map based on the longitude and latitude. Each rectangular grid represents a different geographical region in the city. The region on the i^{th} row and j^{th} column of the grid map is denoted as $R(i,j)$. Figure 1(a) illustrates the partitioned regions of the New York City.

Taxi Origin-Destination Demand: We denote the taxi origin-destination demand in time interval t as a 3D OD matrix $X_t \in R^{N \times H \times W}$, where N is the total number of the regions and it is equal to $H \cdot W$. Specially, $X_t(d, i_1, j_1)$ is the number of taxi requests from origin region $R(i_1, j_1)$ to region $R(i_2, j_2)$, where the index d is equal to $W \cdot i_2 + j_2$. In particular, the d^{th} channel of X_t, denoted as $X_t(d) \in R^{H \times W}$, is the taxi demand from all regions to region $R(i_2, j_2)$ with index d. Furthermore, the taxi origin demand, denoted as $O_t \in R^{H \times W}$, can be easily calculated by $\sum_{d=0}^{N-1} X_t(d)$ and most previous works only predict the demand O_t.

Taxi Origin-Destination Demand Prediction: In this

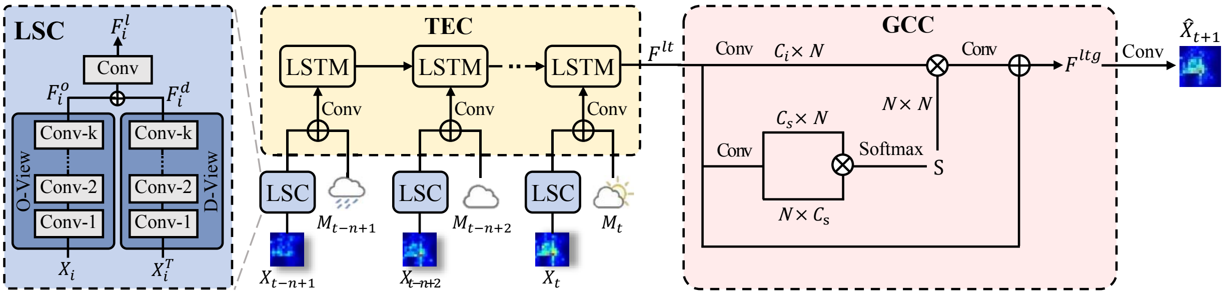


Fig. 2. The architecture of the Contextualized Spatial-Temporal Network. X_i is the OD matrix in time interval i, while X_i^T is the DO matrix called in our work. M_i is the meteorological data. C_{lt} and C_s are the channel numbers of feature F^{lt} and F_s respectively. N is the total number of regions. "$\bigoplus$" denotes feature concatenation and "$\bigotimes$" refers to the dot product operation.

work, we aim to predict the future taxi OD demand with historical data. Inspired by the previous work [4], we consider the impact of meteorological factors in the modeling process, as it is one of the core factors affecting taxi demand. We denote the meteorological data in time interval i as M_i. The preprocessing of taxi demand data and meteorological data are described in Section 5.1. Overall, our final goal is to predict X_{t+1} given the historical data $\{X_i, M_i | i = t-n+1, ..., t\}$, where n is the sequence length of time intervals.

4. METHOD

In this section, we present a novel Contextualized Spatial-Temporal Network (CSTN) for taxi OD demand prediction. As shown in Figure 2, our network consists of three components for three types of context modeling respectively. First, the LSC module which utilizes two CNN to learn the local spatial context of taxi demand from both the origin view and the destination view. Second, the TEC module incorporates both the local spatial features of taxi demand and the meteorological information to a ConvLSTM for the analysis of taxi demand evolution. Third, the GCC module generates the global correlation feature of each region by summing the features of all regions with the calculated similarity weights.

4.1. Local Spatial Context Modeling

Generally, the taxi demand is usually related to local spatial location, and the spatially adjacent regions may have similar demand patterns. For instance, people tend to depart from residence regions in city suburb and head to employment regions in city center in the morning rush hours. This local spatial context is for taxi demand pattern modeling.

In this work, our LSC module captures the local spatial context of taxi demand from both the origin view and destination view. As each channel of X_i is the taxi demand from all origin regions to the corresponding region, we define the convolution operations on X_i as origin view modeling. To model the local spatial context from destination view, we generate a DO matrix X_i^T from X_i with the transformation process described in Figure 3. Each channel of X_i^T is the taxi demand from the corresponding region to all destination regions.

Specifically, our LSC module consists of two CNNs, each of which contains K convolutional layers with 16 filters of kernel size 3×3. In time interval t, the first CNN, called origin view CNN, takes X_i as input and its output feature F_i^o only contains the local spatial context of origin view. Meanwhile, the second CNN, called destination view CNN, takes X_i^T as input and its output feature F_i^d only contains the local spatial context of destination view. Then, we concatenate these two features and fuse them using a convolutional layer with 32 filters to generate the final local spatial feature F_i^l , which simultaneously contains the local spatial context of taxi demand from both the origin view and the destination view.

4.2. Temporal Evolution Context Modeling

Taxi demand is time-varying and it is affected by diverse complicated factors. For instance, a sustained snowfall may weaken the travel willingness of residents and cause a decrease in taxi demand. Therefore we incorporate the historical demand and the ever-changing meteorology to grasp the evolving tendency of taxi demand along the temporal dimension.

In this paper, we model the temporal evolution context of taxi demand with ConvLSTM [6]. We first encode the meteorological data M_i with a Multiple Layer Perceptron, which is implemented by three fully-connected layers with 64, 16 and 8 neurons respectively. The output of the MLP is copied $H{\cdot}W$ times to construct a 3D feature $F_i^m \in R^{8 \times H \times W}$. We generate a feature F_i^{lm} by feeding the concatenate of F_i^l and F_i^m into a convolutional layer with 32 filters. F_i^{lm} is the local spatial feature that integrates the meteorological information.

Then we feed the features $F_{t-n+1}^{lm}, F_{t-n+2}^{lm}, ..., F_t^{lm}$ into the ConvLSTM sequentially. At iteration i, the ConvLSTM takes F_{t-n+i}^{lm} as input and accumulates the previous sequential information to the memory cell. After n iteration, the hidden state of ConvLSTM is denoted as h_n. Finally, we feed h_n into a convolutional layer with C_{lt} filters to generate a local spatial-temporal feature F^{lt}, which encodes the temporal evolution context of the taxi demand.

4.3. Global Correlation Context Modeling

The taxi demand spatial distribution is also related to the attribute of the regions, e.g., most of the residential regions in different areas may have high taxi demands in the morning. Therefore, even if the two regions are far apart in distance, they may still have similar taxi demand patterns as long as the attributes of the two are consistent. We call this relationship between far-apart regions as global correlation context. Inspired by work [15], we capture this context with a global feature fusion. Specifically, we generate the global correlation feature of each region as a weighted sum of all regional features, with the weights being calculated as the similarity between the corresponding region pairs. Thus, each region contains the information of all regions and it is mainly relevant to the regions of high similarities with it.

Firstly, we feed F^{lt} into a convolutional layer with C_s filters to generate an embedded feature F_s and then reshape it into a 2D matrix, which is expressed as:

$$F_s = \textbf{Conv}(F^{lt}, w_s), \quad F_s : R^{C_s \times H \times W} \rightarrow R^{C_s \times N}, \tag{1}$$

where w_s is the convolutional parameters. Each column of F_s stands for the feature of a region. We calculate a similarity matrix $S \in R^{N \times N}$ with a dot-product operation between F_s and its transposed matrix $F_s^T \in R^{N \times C_s}$, and perform a Softmax operation on each column of S, which is expressed as:

$$S = \textbf{Softmax}(F_s^T \otimes F_s), \tag{2}$$

where $\otimes$ denotes the dot-product. $S_{i,j}$ is the normalized similarity weight between the regions with index i and j.

After obtaining the similarity matrix S, we compute the global correlation feature of each region by summing the features of all region with the similarity weights. We implement this process with another dot-product operation. We reshape F^{lt} to dimension $C_{lt} \times N$ and then dot-product F^{lt} and S to compute the global feature F^g, which is expressed as:

$$F^{lt} : R^{C_{lt} \times H \times W} \rightarrow R^{C_{lt} \times N}, F^g = F^{lt} \otimes S, \tag{3}$$

F^g is further reshaped back to dimension $C_{lt} \times H \times W$. The feature F^g encodes the global correlation context, but lacks of structural locality. Therefore, we concatenate F^{lt} and F^g to generate a new feature F^{ltg}, which incorporates the hybrid information of the local spatial context, temporal evolution context and global correlation context.

Finally, we predict the OD demand $\hat{X}_{t+1} \in R^{N \times H \times W}$ by feeding F^{ltg} into a regression, which is formulated as:

$$\hat{X}_{t+1} = tanh(\mathcal{T}(F^{ltg})), \tag{4}$$

where $\mathcal{T}$ is the linear regression implemented by a convolutional layer with N filters and the hyperbolic tangent tan ensures the output values are between -1 and 1[1].

4.4. Implementation Details

We implement our proposed network with Tensorflow. In LSC module, the layer number K is set to 3. In TEC module, all convolutional layers in LSTM have 32 filters and the channel number C_{lt} is set to 75. In GCC module, the channel number C_s is set to 64. The sequence length n is set to 5. The parameters of all convolutional layers and the fully-connected layers in the model are initialized by Xavier. The minibatch size is 64. The learning rate is initially set to 10^{-4} and multiplied by 0.1 every 200 epochs. We optimize our network with Adam optimization [16] by minimizing the Euclidean loss between the ground truth and the predicted result.

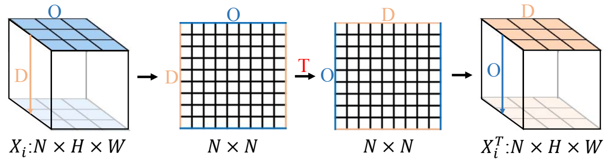


Fig. 3. The generation process of DO matrix from OD matrix. N is equal to $H{\cdot}W$ and T is a matrix transposition operation.

5. EXPERIMENTS

5.1. Dataset

We create a large-scale benchmark, denoted NYC-TOD, for the taxi OD demand prediction task. It contains the taxi OD demand data and the meteorological data of the New York City (NYC) in 2014. We choose the data of the last sixty days as the testing set, and all data before that as the training set.

Taxi Origin-Destination Demand Data: We choose the Manhattan borough in New York City as the study area and divide it into a 15×5 grid map based on the longitude and latitude. The detailed partitioned regions are shown in Figure 1(a). We use the NYC yellow taxi trip records collected by NYCTLC[2] in 2014 to construct the NYC-TOD dataset. Each raw trip record contains the timestamp and the geo-coordinates of origin/destination locations. After excluding the trips where the departure or destination is not in the Manhattan, we get 132 million taxi trip records. Finally, we can generate the taxi OD demand matrix in each time interval by calculating the number of taxi trips between all regions according to the timestamps and geo-coordinates of taxi trip records. The time interval is set to half an hour in this dataset.

Meteorological Data: We collect the meteorological data of NYC from Wunderground[3], which provides the temperature, windchill, humidity, visibility, wind speed, precipitation and 23 types of weather conditions (e.g. Sunny and Snowy). We encode the weather condition into a One-Hot vector and scale other six numeric indicators into the range [0,1] with Min-Max linear normalization. Finally, the meteorological data in time interval t is denoted as a vector $M_t \in R^{29}$.

[1]When training, we use Min-Max linear normalization method to scale the OD demand matrices into the range $[-1, 1]$. We rescale the predicted result back to the normal values when evaluating.

[2]`www.nyc.gov/html/tlc/html/about/trip_record_data.shtml`

[3]`https://www.wunderground.com/`

Table 1. Performance Comparison on the NYC-TOD.

Method	OD-MAPE	OD-RMSE	O-MAPE	O-RMSE
HA-All	37.71%	1.93	45.04%	52.44
HA-Rec	35.46%	1.89	47.59%	54.33
Lasso	33.85%	1.65	34.89%	33.00
OLSR	33.86%	1.65	33.09%	32.68
XGBoost	32.04%	1.54	37.78%	31.23
MLP	30.70%	1.49	25.24%	25.60
ST-ResNet	28.53%	1.38	24.16%	22.43
ConvLSTM	27.99%	1.36	19.89%	21.02
CSTN	**27.37%**	**1.32**	**18.48%**	**19.85**

5.2. Evaluation Metric

Following the previous work [10], we adopt the Mean Average Percentage Error (MAPE) and Rooted Mean Square Error (RMSE) as the evaluation metrics, which are defined as:

$$MAPE = \frac{1}{z}\sum_{t=1}^{z}\frac{\|\hat{X}_t - X_t\|}{X_t}, RMSE = \sqrt{\frac{1}{z}\sum_{t=1}^{z}\|\hat{X}_t - X_t\|^2}, \quad (5)$$

where z is the total number of testing samples, $\hat{X}_t$ and X_t are the predicted demand and the ground truth in time interval t.

In this work, we not only evaluate the task of taxi OD demand prediction, but also evaluate the taxi origin demand prediction. The predicted origin demand can be calculated by $\sum_{d=0}^{N-1}\hat{X}_t(d)$. For convenience, the MAPE and RMSE of the former task are denoted as OD-MAPE and OD-RMSE, while these metrics of the latter task are denoted as O-MAPE and O-RMSE. While performing the evaluation, we follow the previous work [4] to filter the OD pairs or the origin regions with ground truth less than 5 in each time interval, since such low taxi demand is always ignorable in real-world applications.

5.3. Performance Comparison

We compare the performance of our proposed method with the following basic and advanced methods:

- **Historical Average (HA)** predicts the future demand by averaging the historical demands, e.g, (1) the average of the historical demands in the same time intervals of every day, denoted as **HA-All**; (2) the average of the taxi demands of current n time intervals, denoted as **HA-Rec**.
- **Linear Regression** predicts the future demand with different versions of linear regression methods, e.g., Ordinary Least Squares Regression (**OLSR** [17]) and **Lasso** Regression [18] with ℓ_1-norm regularization.
- **XGBoost [19]** is an effective boosting-tree based method. XGBoost takes the concatenation of the demand matrices of previous n time intervals to predict the taxi demand.
- **Multiple Layer Perceptron (MLP)**: A neural network consists of four fully connected layers with 128, 128, 64 and 75 neurons respectively. The MLP forecasts the every channel of X_t by taking the corresponding channels of demand matrices of previous n time intervals as input.

Table 2. Performance Comparison of Various Contexts.

Method	LSC	LSC+TEC	LSC+TEC+GCC
OD-MAPE	28.54%	27.80%	27.27%
O-MAPE	20.80%	19.41%	18.48%

- **ST-ResNet [10]** is a deep learning method for traffic flow forecast and we apply it to predict the taxi demand.
- **ConvLSTM [6]** is our LSC module + TEC module. Specifically, the LSC module in this network only contains the origin view CNN and it takes X_i as input.

The performance of all methods are summarized in Table 1 and our method outperforms other methods by a significant margin on both the tasks of (1) taxi OD demand prediction and (2) taxi origin demand prediction. Notice that some competed methods, e.g., MLP, ST-ResNet and ConvLSTM, also adopt deep models to predict the taxi demand, but their performances are unsatisfactory. The main reasons are that MLP fails to capture the local spatial context and ST-ResNet does not explicitly learn temporal evolution context, while ConvLSTM does not model the global correlation context. In contrast, our method integrates the above various context into a unified framework to predict the demand.

5.4. Ablation Study

Influence of Various Context: To explore the influence of various context on taxi demand prediction, we implement the following variants of our model with different modules:

- **LSC Net**: contains the LSC module and directly concatenates the local spatial features of each time interval to predict taxi demand with a convolutional layer.
- **LSC+TEC Net**: composed of the LSC module and TEC module. The last hidden state of TEC module is used to predict the taxi demand with a convolutional layer.
- **LSC+TEC+GCC Net**: As the full version of CSTN, this network integrates the local spatial context, temporal evolution context as well as the global correlation context for taxi demand prediction.

As shown in Table 2, the LSC Net achieves an OD-MAPE of 28.54% and an O-MAPE of 20.80%. When explicitly modeling the temporal evolution context with LSTM, the LSC+TEC Net gets an OD-MAPE of 27.80% and an O-MAPE of 19.41%, achieving an obvious performance improvement compared to the LSC Net. After integrating the global correlation context with the GCC module, the LSC+TEC+GCC Net can further decrease the OD-MAPE to 27.27% and OD-MAPE to 18.48%, with 2.5% relative performance improvement on average. The experimental result shows that our network can achieve notable performance improvement by modeling these context, which also indicates the effectiveness of these context for taxi demand prediction.

Table 3. Effectiveness of the Two-View CNN in LSC module.

Method	OD-MAPE	O-MAPE
Origin View	28.94%	23.03%
Origin View + Destination view	28.54%	20.80%

Table 4. Influence of Local and Global Context.

Method	Global	Local	Local+Global
OD-MAPE	28.56%	27.80%	27.37%
O-MAPE	20.25%	19.41%	18.48%

Effectiveness of the Two-View ConvNets in the LSC module: To validate the effectiveness of the Two-View ConvNets for local spatial context modeling, we train a variant of LSC Net that only takes the OD matrix X_i as input to learn the local spatial context from origin view. As shown in Table 3, simply using origin view ConvNet, the LSC Net obtains very inferior results. After adding the destination view, the performance improves by 0.5% and 2.03% w.r.t metrics OD-MAPE and O-MAPE respectively, which evidences that the destination view context is also beneficial for taxi demand and our LSC module can effectively capture the local spatial context.

Influence of Local and Global Context: In CSTN, we forecast the taxi demand with both the local feature F^{lt} and global feature F^g. To analyze how these two features contribute to the performance, we train other two variants of CST-N which only take F^{lt} or F^g to predict the taxi demand with a convolutional layer. As shown in Table 4, the performance of the local feature F^{lt} is better than that of the global feature F^g, which indicates that the local feature is more efficient for this task. When combining the local and global feature for final prediction, our proposed model achieves the best performance, which evidences the complementary of both local context and global context in taxi demand prediction.

Influence of Meteorology: To explore the influence of meteorology, we train another LSC+TEC Net and CST-N without modeling the meteorology. As shown in table 5, without considering the meteorological condition, the performances of both LSC+TEC Net and CSTN degrades to some extent, which indicates the meteorological information can help to improve the performance of taxi demand prediction.

6. CONCLUSION

In this paper, we introduce a worth-exploring task, taxi origin-destination demand prediction, which aims at predicting the taxi demand between all OD pairs in the future time intervals. This problem can help taxi preallocation systems to allocate the taxi more efficiently. We address this problem with a Contextualized Spatial-Temporal Network, which integrates local spatial context, temporal evolution context and global correlation context in one united framework for future taxi demand prediction. Extensive experiments demonstrate that our proposed method achieves superior performance in comparison to the existing methods.

Table 5. Comparison of the Taxi Demand Prediction with or without Meteorology.

Method	OD-MAPE	O-MAPE
LSC+TEC Net W/O Meteorology	28.08%	20.03%
LSC+TEC Net W/- Meteorology	27.80%	19.41%
CSTN W/O Meteorology	27.69%	19.72%
CSTN W/- Meteorology	27.37%	18.48%

7. REFERENCES

[1] Florian Toqué, Etienne Côme, Mohamed Khalil El Mahrsi, and Latifa Oukhellou, "Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks," in *ITSC*, 2016.

[2] Chao Yang, Fenfan Yan, and Xiangdong Xu, "Daily metro origin-destination pattern recognition using dimensionality reduction and clustering methods," in *ITSC*, 2017.

[3] Yongxin Tong, Yuqiang Chen, Zimu Zhou, Lei Chen, Jie Wang, Qiang Yang, Jieping Ye, and Weifeng Lv, "The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms," in *KDD*, 2017.

[4] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *AAAI*, 2018.

[5] Lingbo Liu, Ruimao Zhang, Jiefeng Peng, Guanbin Li, Bowen Du, and Liang Lin, "Attentive crowd flow machines," in *ACM MM*. ACM, 2018.

[6] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *NIPS*, 2015.

[7] Zhongjian Lv, Jiajie Xu, Kai Zheng, Hongzhi Yin, Pengpeng Zhao, and Xiaofang Zhou, "Lc-rnn: A deep learning model for traffic speed prediction.," in *IJCAI*, 2018.

[8] Lingbo Liu, Hongjun Wang, Guanbin Li, Wanli Ouyang, and Liang Lin, "Crowd counting using deep recurrent spatial-aware network," in *IJCAI*, 2018.

[9] Zhilin Qiu, Lingbo Liu, Guanbin Li, Qing Wang, Nong Xiao, and Liang Lin, "Crowd counting via multi-view scale aggregation networks," in *ICME*, 2019.

[10] Junbo Zhang, Yu Zheng, and Dekang Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction.," in *AAAI*, 2017.

[11] Dong Wang, Wei Cao, Jian Li, and Jieping Ye, "Deepsd: supply-demand prediction for online car-hailing services using deep neural networks," in *ICDE*, 2017.

[12] Xian Zhou, Yanyan Shen, Yanmin Zhu, and Linpeng Huang, "Predicting multi-step citywide passenger demands using attention-based neural networks," in *ICWSDM*, 2018.

[13] Martin L Hazelton, "Some comments on origin–destination matrix estimation," *TRP A*, 2003.

[14] Martin L Hazelton, "Statistical inference for time varying origin-destination matrices," *TRP B*, 2008.

[15] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He, "Non-local neural networks," *arXiv preprint arXiv:1711.07971*, 2017.

[16] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[17] BD Craven and SM Islam, *Ordinary least-squares regression*, Sage Publications, 2011.

[18] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B*, pp. 267–288, 1996.

[19] Tianqi Chen and Carlos Guestrin, "Xgboost: A scalable tree boosting system," in *KDD*, 2016.