Exploring the Side Information Fusion Method with Spatial-temporal Model for Taxi Demand Prediction

Jia Mou
School of Artificial Intelligence, University of Chinese Academy of Sciences
Beijing 100049, China
Institute of Automation, Chinese Academy of Sciences
Beijing 100049, China
mujia2018@ia.ac.cn

Yu Liu
Institute of Automation, Chinese Academy of Sciences
Beijing 100049, China
yu.liu@ia.ac.cn

Dongchang Liu†
School of Artificial Intelligence, University of Chinese Academy of Sciences
Beijing 100049, China
Institute of Automation, Chinese Academy of Sciences
Beijing 100049, China
dongchang.liu@ia.ac.cn

ABSTRACT

Taxi is one of the most common public transport, predicting taxi demand precisely within an area is of great significance for improving efficiency of traffic. Taxi demand is spatial-temporal data, and highly influenced by many external factors, such as time, weather. So there are two main problems on taxi demand prediction: the one is that modeling both spatial and temporal non-linear correlations is not easy, the other is that real scenarios exist temporal but non-spatial side information, which is hard to be fused with spatial-temporal taxi demand. To handle the two problems, in this paper,we propose a novel Side information fused Spatial-Temporal Network(SideInfo-STNet) framework to model correlations of time, space and side information. The framework has three main components: Spatial-temporal Taxi Demand (transforming raw taxi demand to taxi demand image sequences); Side Information (transforming side information to time series vectors); SideInfo-STLSTM (extending LSTM to has convolution structures and fusing side information into LSTM gate units). By using SideInfo-STNet to conduct extensive experiments on large-scale TLC trips of New York City, we validate that our model outperforms traditional and deep learning based models on taxi demand prediction.

CCS CONCEPTS

• **Information systems** → *Data stream mining*; *Location based services*.

KEYWORDS

SideInfo-STNet, Spatial-temporal data, Taxi demand prediction, Deep learning

†Corresponding author: dongchang.liu@ia.ac.cn

DSIT 2020, July 24–26, 2020, Xiamen, China

ACM ISBN 978-1-4503-7604-4/20/07…$15.00
https://doi.org/10.1145/3414274.3414510

ACM Reference format:
Jia Mou, Yu Liu and Dongchang liu. 2020. Exploring the Side Information Fusion Method with Spatial-temporal Model for Taxi Demand Prediction. In Proceedings of 2020 3rd International Conference on Data Science and Information Technology (DSIT'20). Xiamen, China, 7 pages. https:// doi.org/10.1145/3414274.3414510

1 Introduction

Traffic has great impact on the life of modern urban people, and the smooth traffic condition is also a major embodiment of the degree of civilization of a city.

Taxi is an important means of transportation, but because of the uncertainty of demand distribution, it often leads to a large number of vehicles and people can not effectively match. In order to effectively solve this problem, we can predict the demand of taxis in a region in advance, and then preallocate taxis to reduce unnecessary energy loss and reduce traffic congestion.

Taxi demand is spatial-temporal data, and highly influenced by many side information (e.g. time, weather) [1, 2]. So the two important issues of taxi demand prediction are: (1) capturing the temporal and spatial correlations of demand data adequately; (2) taking side information into consideration efficiently.

Traditional methods such as autoregressive integrated moving average (ARIMA) and its variants [3–5], geographically weighted regression(GWR) and generalized linear model(GLM) [6] are widely applied for traffic prediction, but they fail to adequately model complex spatial-temporal correlations, and those handmade feature designs are tedious.

With the development of deep learning, convolutional neural network (CNN) can effectively deal with spatial correlations, it has made great achievements in the field of computer vision [7]. Recent studies consider the traffic of the city as an image and the traffic flow as pixel value, then use CNN to capture the spatial correlations of traffic data [8, 9]. Recurrent neural network (RNN) is designed to deal with temporal correlations, a number of RNN-based models have been used to capture the temporal correlations of traffic data, in study [10] a RNN-based model is apply to predict passenger wait time. LSTM [11] is a special form of RNN, which introduces gate units to control the flow of information, so it is more suitable for handling long-term dependencies of taxi demand [12, 13].

Recently, a number of studies try to model both spatial and temporal correlations. A line of studies use CNN on spatial taxi demand to get feature vectors, then feed the vectors to a sequential model such as LSTM. Representatively, DMVST-Net [14] uses Local CNN and LSTM to capture the complex nonlinear correlations of space and time, the studies [9, 15] divide demand to many stages according to the time, and use CNN to extract features on each stage data, then a fusion layer used to concatenate those stage features. another line of studies incorporate CNN into the LSTM cells. The study [16] proposes ConvLSTM, which replaces linear layers with convolutional structures in LSTM, the D-GAN [17] has apply ConvLSTM to simultaneously extract spatial and temporal features of taxi demand data.

Travel plans are often affected b y e xternal f actors (e.g. time and weather), therefore the factors have substantial impact on the taxi demand in city, we regard those factors as side information of taxi demand prediction problem. The studies [14, 15, 17] use spatial-temporal models to encode demand data to vectors, and use another models to extract side information feature vectors, then fuse the demand data with side information by concatenating the two vectors, the fusion method is straightforward. However, multimodal researchs [18–20] show that explicitly modeling the correlation between different i nformation (sharing w eights a cross time steps and sharing weights across modalities) can significantly improve the prediction accuracy and robustness of model. Similarly, we think the correlations between side information and spatial-temporal taxi demand are subtle, the predict accuracy must be influenced b y t he f usion method.

To discover temporal and spatial correlations of taxi demand adequately and fuse side information into model efficiently, we propose a novel Side information fused Spatial-Temporal Network (SideInfo-STNet) framework, and conduct extensive experiments to compare our proposed model with competing methods, the result show SideInfo-STNet outperforms those methods. Our main contributions are as follows.

- We propose an end-to-end deep-learning-based framework (SideInfo-STNet), in which a noval LSTM structure called SideInfo-STLSTM can model spatial-temporal correlations of taxi demand and fuse side information with taxi demand in a subtle way.
- We compare the taxi demand patterns under different circumstances (e.g. time,weather), and prove that different side information has substantial impact on taxi demand.
- We validate the proposed framework by conducting extensive experiments on large-scale TLC trips of New York City, the results show the advantages of SideInfo-STNet compared with competing methods.

2 Preliminaries

The purpose of taxi demand prediction is to use the previously observed demand sequences and side information in the area to predict the taxi demand of next timestamp. In this section, we present some formal definitions of taxi demand prediction problem.

2.1 Spatial-Temporal Taxi Demand

The taxi demand is defined as the number of taxi requests in certain area per time point, we split the area into $N \times N$ grids by latitude and longitude, and split one day into M time intervals $T = [t_1, t_2, ..., t_M]$. We define the taxi demand in certain area and certain time interval t as $img^t, t \in T$, whose size is $N \times N$ and its pixel values are the sum of total demand in each grid within interval t. We arrange the spatial taxi demand in the order of time, then we get the spatial-temporal taxi demand $X = [img^1, img^2, ...img^L]$, where L is the length of dataset.

2.2 Side Information

Side Information can make the encoding length of source information shorter, which means the fusion of side information can improve the model a lot. External factors that affect travel plans must have substantial impact on taxi demand, so we define external factors (e.g. time has periodicity impact on human activities; weather has impact on travel plans) as the side information for taxi demand prediction. The side information in time interval t is defined as s^t, we also arrange the side information along the timeline, then we get $S = [s^1, s^2, ...s^L]$.

2.3 Taxi Demand Prediction Problem

The taxi demand prediction problem aims to predict the demand at time interval t, given the l frame historical spatial-temporal taxi demand $X^t = \{img^{t-l}, img^{t-l+1}, ..., img^{t-1}\}$ and corresponding side information $S^t = \{s^{t-l}, s^{t-l+1}, ..., s^{t-1}\}$. The formula can be written as $img^t = \Phi\left(X^t, S^t\right)$.

3 Methodology

In this section, we formally introduce the overall framework of SideInfo-STNet as shown in Fig. 1, including its components and algorithms.

Table 1: Side Information Structure

side information	range	levels	divide points
dow	[1,7]	7	1,2,3,4,5,6
hod	[0,23]	24	0,1,2,...,22
ws(m/s)	[0,25]	6	2,4,6,8,10
temperature$(^\circ F)$	[-1,94]	5	20,40,60,80

3.1 SideInfo-STNet components

SideInfo-STNet has three main components: Spatial-temporal Taxi Demand; Side Information; SideInfo-STLSTM.

Spatial-temporal Taxi Demand: The size of city is

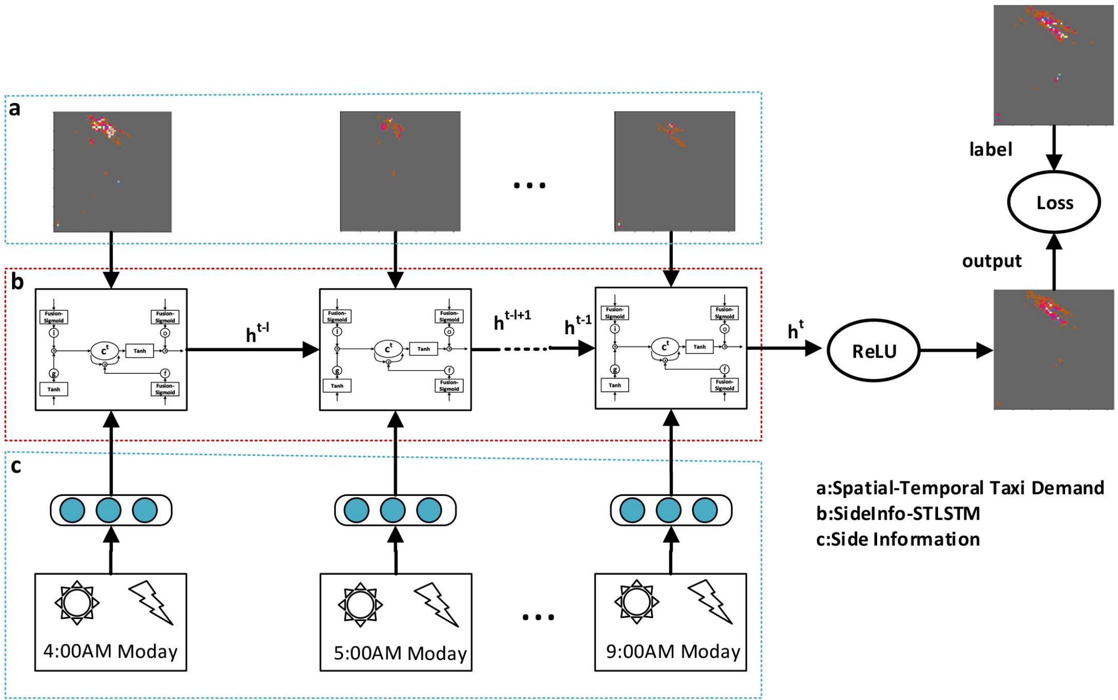
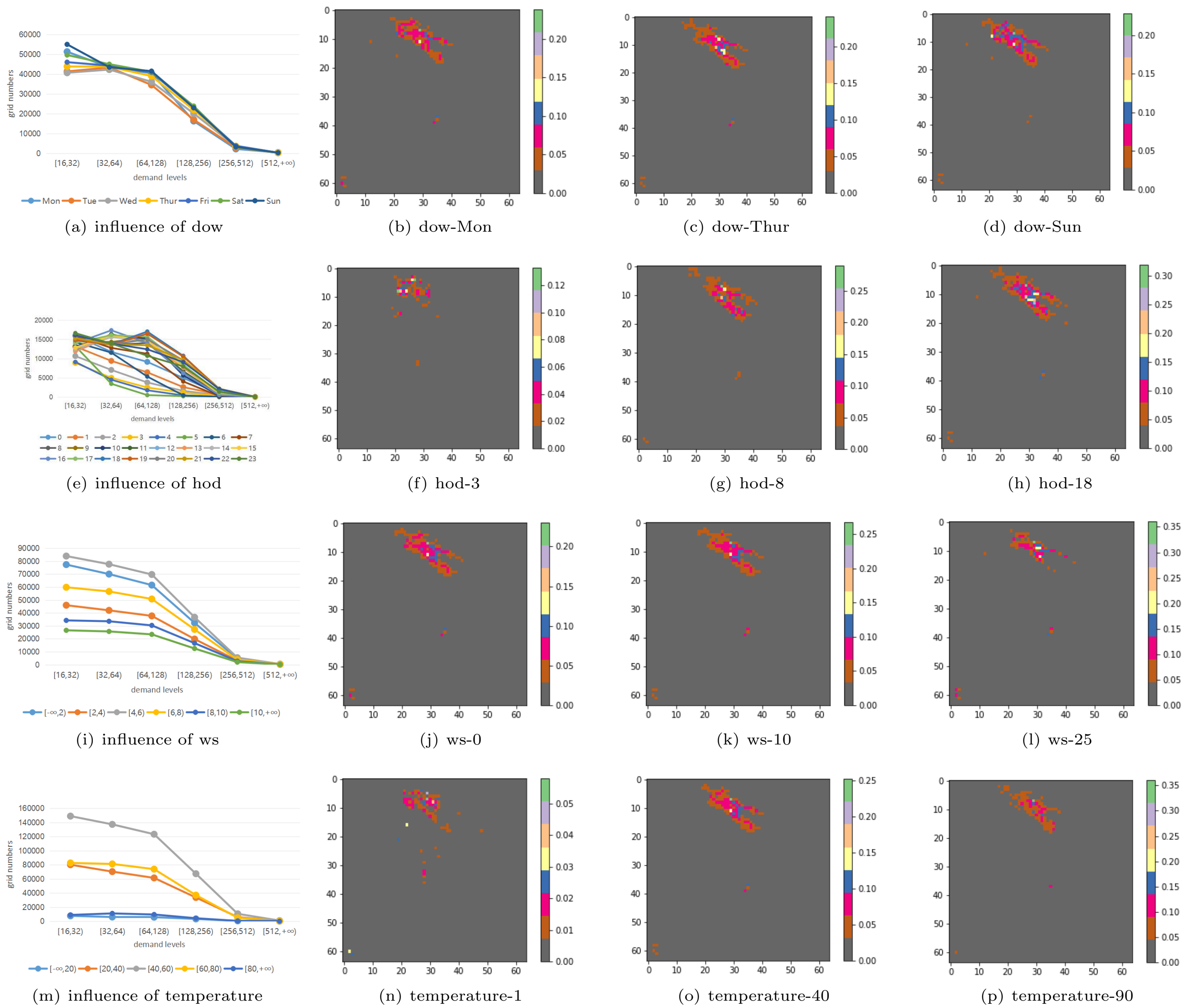


Figure 1: The framework of SideInfo-STNet. part.a transforms raw taxi demand to taxi demand image sequences by definition of spatial-temporal taxi demand; part.b extends LSTM to has convolution structures and fuses side information into LSTM gate units; part.c transforms side information to time series vectors by definition of side information; Finally, rectified linear unit(ReLU) is used after the last cell of SideInfo-STLSTM to scale the taxi demand image for prediction.

usually very large, and the taxi demand can vary drastically in different regions, so predicting the city-scale taxi demand as a whole is difficult. We need to divide the city into grids and predicting taxi demand for each of them, smaller granularity means more effective prediction. As we all know, taxi-dispatching is a slow process, there are obvious correlations between the taxi demands of adjacent time. To capture correlations of space and time, we construct spatial-temporal taxi demand in the form of definition of spatial-temporal taxi demand.

Side Information: According to definition of side information, time and weather have great influence on taxi demand. We choose day-of-week(dow), hour-of-day(hod) as important side information of time, and wind-speed(ws), temperature as important side information of weather. We divide those side information into different levels as shown in Table. 1. To illustrate the influence of those four factors on taxi demand, we count the grid numbers belong to different demand levels under all circumstances of those four side information, and select the levels from 16 to $+\infty$ (the reasons will be explained later) to show the influence of side information. For a more intuitive understanding, average taxi demand images under particular circumstances are calculated: dow(Mon, Thur, Sun); hod(3, 8, 18); ws(0, 10, 25); temperature(1, 40, 90) as shown in Fig. 2. By comparing dow-Mon with dow-Sun and hod-3 with hod-8, we note that those taxi demand patterns have notable differences, mainly because time factors (dow, hod) have periodicity impact on human activities, weekend and commuting time will lead to huger taxi demand. By comparing ws-10 with ws-25 and temperature-1 with temperature-40, we note that weather also has impact on taxi demand patterns. Because extreme weathers frustrate people's hopes of going out, taxi demand is smaller under extreme weathers. Beyond that, the side information mainly acts on the grids of high demand levels, we think it is because some areas of the city have very few people live in, no matter how side information changes, the taxi demand will still in very low value (approximately equal to 0).

SideInfo-STLSTM: The SideInfo fused Spatial-temporal LSTM (SideInfo-STLSTM) structure is shown as Fig. 3, and can be illustrated by Formula. 1. The memory cell c^t acts as an accumulator for states; the input gate i^t determines which information should be retained to update c^t; the forget gate f^t determines which information should be discard or be weakened; the output gate o^t determines which feature should be flow to hidden state h^t. The input taxi demand img^t, memory cell c^t, hidden state h^t are all 2D spatial tensors, and the side information s^t is 1D ont-hot tensor.



(a) influence of dow (b) dow-Mon (c) dow-Thur (d) dow-Sun

(e) influence of hod (f) hod-3 (g) hod-8 (h) hod-18

(i) influence of ws (j) ws-0 (k) ws-10 (l) ws-25

(m) influence of temperature (n) temperature-1 (o) temperature-40 (p) temperature-90

Figure 2: influence of side information on taxi demand: (a), (e), (i), (m) illustrate the influence of side information on grid numbers belong to different demand levels; the other pictures show the average taxi demand under different circumstances.

We fuse the side information with taxi demand by Fusion-Sigmoid in i^t, f^t, o^t shown in Formula. 1. In Fusion-Sigmoid, taxi demand img^t, hidden state h^t and side information s^t are all connected with 2D gate units, where the $*$ denotes convolution operator, the $\circ$ denotes Hadamard product and $\oplus$ denotes that the element after $\oplus$ will be expanded to the same shape as the element before $\oplus$, then return the sum of the two elements.

By SideInfo-STLSTM, we can capture the spatial and temporal correlations simultaneously, and fuse side information with taxi demand in a subtle way.

$$
\begin{aligned}
i^t &= \sigma\left(W_{xi} * img^t + W_{hi} * h^{t-1} \oplus W_{si} \circ s^t + b_i\right) \\
f^t &= \sigma\left(W_{xf} * img^t + W_{hf} * h^{t-1} \oplus W_{sf} \circ s^t + b_f\right) \\
o^t &= \sigma\left(W_{xo} * img^t + W_{ho} * h^{t-1} \oplus W_{so} \circ s^t + b_o\right) \\
g^t &= tanh\left(W_{xg} * img^t + W_{hg} * h^{t-1} + b_g\right) \\
c^t &= f^t \circ c^{t-1} + i^t \circ g^t \\
h^t &= o^t \circ tanh(c^t)
\end{aligned}
\tag{1}
$$

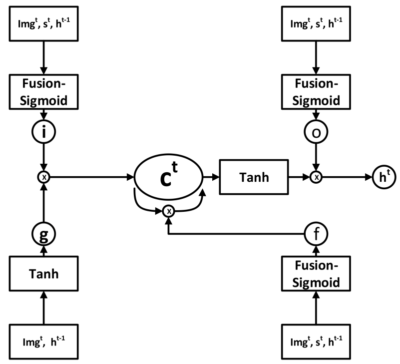


Figure 3: Structure of SideInfo-STLSTM

3.2 Prediction Component

The output of SideInfo-STLSTM is a 2D image, whose pixel values range from -1 to 1, we use rectified linear unit(ReLU) [21] after the last cell of SideInfo-STLSTM to scale the output to $[0, 1]$ as the taxi demand values will be normalized. Finally, we get the prediction (taxi demand image).

3.3 Trainning Algorithm

Algorithm. 1 shows the training process of SideInfo-STNet. Firstly, we construct the spatial-temporal taxi demand sequence and side information sequence according to the length of historical observations and the frame we need to predict. Then, the parameters of our model is optimized by minimizing the loss function in the form of mini-batch. The loss function used in this paper is Mean Squared Error (MSE):

$$MSE = \frac{1}{\wedge} \sum_{p \in \wedge} (\hat{img}_p - img_p)^2 \tag{2}$$

where $\hat{img}_p$, img_p are the true value and predicted value of taxi demand in grid p, $\wedge$ is grids of spatial-temporal taxi demand.

Algorithm 1 Training of SideInfo-STNet

Input: length of historical observations: l; spatial-temporal taxi demand: $\{img^1, ..., img^L\}$; side information: $\{s^1, ..., s^L\}$;
Output: taxi demand prediction img^t

$D \leftarrow \emptyset$
for $t = l \quad to \quad L$ **do**
  $X^t = \{img^{t-l}, img^{t-l+1}, ..., img^{t-1}\}$
  $S^t = \{s^{t-l}, s^{t-l+1}, ..., s^{t-1}\}$
  add $(\{X^t, S^t\}, \{img^t\})$ into D
end for
repeat
  Randomly select a batch D_b from D;
  Optimize parameters of model by minimizing the loss function
until stopping criteria is met;

4 Experiments

4.1 Preprocessing and Parameters

The taxi demand dataset used in this paper is collected and provided by the NYC Taxi and Limousine Commission (TLC) [22], which contains taxi demand data from 01/01/2009 to 06/30/2019. Due to the lack of accurate position information, we truncate the data from 08/01/2013 to 06/30/2016. According to the definition of spatial-temporal taxi demand, the raw data is transformed into 64×64 2D images, the size of each grid of image is approximately $1km \times 1km$, the pixel values range from 0 to 1549, we use the Min-Max normalization method to scale the pixel values into the range $[0, 1]$. And we divide one day into 24 time levels, which means one hour is a level. The shape of all spatial-temporal taxi demand is $25560 \times 64 \times 64$. Then we divide the data into train set, validation set and test set in the ratio of 5:3:2.

For side information we extract hour-of-day (hod), day-of-week (dow) from raw taxi requests as side information of time, then we find the local climatological data for NYC, extract wind speed(ws) and temperature as side information of weather. According to definition of side information, we process those information into tensors, whose shapes are $25560 \times (24, 7, 6, 5)$. similarly, the side information is divided in the ratio of 5:3:2.

In experiment,we use PyTorch [23] to build our model, the length of historical observations is 23, the batch size is 8, *Adam* is selected as optimizer, whose learning rate is initialized as 0.001, and the learning rate will be multiplied by 0.9 per epoch. we select *xavier* as initial method for all trainable parameters of our model. In SideInfo-STLSTM, the hidden dimension is 3, and corresponding sizes are (64, 64, 1), the convolution kernel size is 3×3. For the sake of fairness, all models mentioned in this paper use the same epoch number 50.

4.2 Evaluation Metric

In spatial-temporal taxi demand, most grid values in our spatial-temporal taxi demand are 0 or approximately equal to 0, so the taxi demand can be described as hyperimbalanced data. To illustrate the result of models on different demand levels, we divide taxi demand into 11 levels, then group grids in the same demand level, MAE is used in each level to measure the prediction accuracy, we call the indicator as Sliced MAE, which can be written as Formula. 3. Finally, we use MAE together with Sliced MAE to compare performance of different models.

$$MAE_i = \frac{1}{\wedge_i} \sum_{p \in \wedge_i} |\hat{img}_p - img_p| \tag{3}$$

where $\hat{img}_p$, img_p are the true value and predicted value of taxi demand in grid p, $\wedge_i$ is grids of spatial-temporal taxi demand in each demand level.

Table 2: Comparison of Different Models

Models\levels	[-∞,1)	[1,2)	[2,4)	[4,8)	[8,16)	[16,32)	[32,64)	[64,128)	[128,256)	[256,512)	[512,+∞)	MAE
HA	0.108	0.832	1.556	2.810	4.810	9.006	15.207	22.413	31.190	48.609	62.130	1.207
FC-LSTM	**0.006**	1.042	2.354	4.342	6.455	8.939	12.592	19.873	30.111	50.805	108.687	1.131
ConvLSTM	0.065	0.957	1.631	2.479	3.591	5.615	8.478	12.202	18.565	34.554	58.306	0.764
ST-ResNet	0.078	0.881	1.519	2.440	3.649	5.680	8.418	12.294	19.243	35.554	62.687	0.779
DMVST	0.107	**0.803**	**1.416**	**2.336**	3.615	5.953	9.264	13.303	20.302	38.589	62.544	0.837
SideInfo-STNet	0.064	0.940	1.604	2.417	**3.541**	**5.544**	**8.277**	**12.061**	**18.126**	**33.614**	**57.501**	**0.750**

Table 3: Comparison of Variants of SideInfo-STNet

Models\levels	[-∞,1)	[1,2)	[2,4)	[4,8)	[8,16)	[16,32)	[32,64)	[64,128)	[128,256)	[256,512)	[512,+∞)	MAE
ConvLSTM	0.065	0.957	1.631	2.479	3.591	5.615	8.478	12.202	18.565	34.554	58.306	0.764
SideInfo-STNet-with-dow	0.064	**0.940**	**1.604**	2.417	3.541	5.544	8.277	12.061	18.126	33.614	57.501	0.750
SideInfo-STNet-with-hod	**0.053**	0.958	1.639	2.460	3.540	5.447	8.233	11.971	18.362	34.359	57.062	0.743
SideInfo-STNet-with-ws	0.056	0.955	1.631	2.467	3.573	5.633	8.393	12.183	18.696	34.428	56.889	0.755
SideInfo-STNet-with-temperature	0.059	0.962	1.616	**2.414**	3.557	5.570	8.351	12.027	18.471	34.645	58.118	0.750
SideInfo-STNet-with-all	0.058	0.950	1.632	2.449	**3.516**	**5.407**	**8.156**	**11.954**	**18.121**	**32.844**	**56.241**	**0.741**

4.3 Performance Comparison

To illustrate the advantages of SideInfo-STNet, we compare it with competing methods, including the state of the art. The settings of those models are as follows:

- **Historical Average(HA):** HA predicts the taxi demand by computing the average value of historical taxi demand in the same timeslot and position, for example, the taxi demand at 6:00am-7:00am on Monday in grid $[0, 0]$ is predicted as the average value of all taxi demand at 6:00am-7:00am on Monday in grid $[0, 0]$ in train dataset.
- **Fully Connected LSTM(FC-LSTM):** FC-LSTM is a well-known model for time series prediction, we set the hidden state dimensions as 1024,1024,4096, then use linear layer to generate the output whose shape is 64×64.
- **Convolutional LSTM(ConvLSTM):** ConvLSTM [16] extends the FC-LSTM to have convolutional structures, it is widely used in spatial-temporal data prediction. we set the hidden state dimensions as 64, 64, 1.
- **ST-ResNet:** ST-ResNet [15] uses the residual neural network framework to model the different spatial-temporal correlations of traffic. We set $l_c = l_p = lq = 3$, p and q mean one-day and one-week, the residual-unit layer number is 4.
- **DMVST:** DMVST [14] apply CNN, LSTM to model spatial-temporal correlation, and graph embedding is used in semantic view. We set number of local CNN layer as 3.
- **SideInfo-STNet:** SideInfo-STNet is proposed model in this paper.

For fair comparisons, the common settings of all other models (e.g. taxi demand dimensions, epoch, optimizer) are the same as our proposed model and dow is selected as side information used in ST-ResNet, DMVST, SideInfo-STNet.

Table. 2 shows the result of comparison. HA and FC-LSTM have no ability to capture spatial-temporal correlations, so they perform poorly. Especially FC-LSTM can not distinguish different regions, so it suffers from the hyperimbalance of taxi demand, the most grids of taxi demand prediction is approximately equal to 0, then FC-LSTM achieves the meaningless lowest Sliced MAE in $[-\infty, 1]$. The ConvLSTM models spatial and temporal correlations simultaneously, it achieves better performance. ST-ResNet and DMVST model spatial and temporal correlations and fuse dow with taxi demand, they also obtain good results, and DMVST achieves the lowest Sliced MAE in $[1, 2), [2, 4), [4, 8)$ levels. But in high demand levels it gets worse result than ConvLSTM. As we illustrated before, side information has substantial impact on taxi demand, Especially in regions with huge taxi demand, our proposed SideInfo-STNet models correlations of space, time and side information simultaneously. Consequently it achieves the lowest MAE in most levels, especially those high demand levels. In MAE column, SideInfo-STNet also achieves the lowest value. The result shows that our proposed SideInfo-STNet outperforms other competing methods.

If we remove side information, our proposed SideInfo-STNet will degenerate to ConvLSTM (i.e. SideInfo-STNet-with-none). To show the improvement of accuracy caused by fusion of side information, we compare following variants of SideInfo-STNet with ConvLSTM, The result of comparison is shown in Table. 3. The result show that dow, hod, ws and temperature all have positive impact on taxi demand prediction, and the best model can be constructed by integrating the four factors into SideInfo-STNet called **SideInfo-STNet-with-all**.

- **SideInfo-STNet-with-dow:** SideInfo-STNet + spatial-temporal taxi demand + dow.
- **SideInfo-STNet-with-hod:** SideInfo-STNet + spatial-temporal taxi demand + hod.
- **SideInfo-STNet-with-ws:** SideInfo-STNet + spatial-temporal taxi demand + ws.
- **SideInfo-STNet-with-temperature:** SideInfo-STNet + spatial-temporal taxi demand + temperature.
- **SideInfo-STNet-with-all:** SideInfo-STNet + spatial-temporal taxi demand + dow + hod + ws + temperature.

For the purpose of keeping the results clear and tidy, we do not show all the experiments we have done here, but all the comparisons with SideInfo-STNet show that SideInfo-STNet outperforms other competing methods in MAE and Sliced MAE in all high demand levels, despite minor shortcomings in low demand levels. As overlooking the samples with demand values less than 10 is a common practice used in taxi industry [14]. The effect of this minor imperfection in low demand levels is quiet small. SideInfo-STNet is verified to have strong power on taxi demand prediction task.

5 Conclusions

In this paper, we propose a noval deep-learning model named SideInfo-STNet, which can adequately learn the complex spatial-temporal correlations simultaneously, and effectively fuse side information into the model. Then we validate our model on large-scale TLC trips of New York City. By carrying out experiments on large scale TLC trips with 5 models, we prove that our model is more applicable for taxi demand prediction with side information, especially in regions with huge taxi demand. In the future, we will take more side information (e.g. POI, traffic accident) into consideration, and apply SideInfo-STNet to more complex spatial-temporal prediction problems such as game state evaluation, regional rainfall prediction, visual tracker.

REFERENCES

[1] B. Hu, X. Xia, X. Shen, and X. Dong, "Analyses of the imbalance of urban taxis high-quality customers based on didi trajectory data," *Journal of Advanced Transportation*, vol. 2019, 2019.

[2] P. Bei, U. Demiryurek, and C. Shahabi, "Utilizing real-world transportation data for accurate traffic prediction," in *Data Mining (ICDM), 2012 IEEE 12th International Conference on*, Conference Proceedings.

[3] L. I. Xiaolong, G. Pan, W. U. Zhaohui, Q. I. Guande, L. I. Shijian, D. Zhang, W. Zhang, and Z. Wang, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers of Computer Science in China*, vol. 6, no. 1, pp. 111–121, 2012.

[4] Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, and J., "Predicting taxipassenger demand using streaming data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1393–1402, 2013.

[5] L. Moreira-Matias, R. Fernandes, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "On recommending urban hotspots to find our next passenger," p. 1723, 2013.

[6] J. Tang, F. Gao, F. Liu, W. Zhang, and Y. Qi, "Understanding spatio-temporal characteristics of urban travel demand based on the combination of gwr and glm," *Sustainability*, vol. 11, no. 19, p. 5525, 2019.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[8] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[9] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "Dnn-based prediction model for spatio-temporal data," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Conference Proceedings, pp. 1–4.

[10] Y. Lu, Z. Zeng, H. Wu, G. G. Chua, and J. Zhang, "An intelligent system for taxi service: Analysis, prediction and visualization," *AI Communications*, vol. 31, no. 1, pp. 33–46, 2018.

[11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[12] N. Davis, G. Raina, and K. Jagannathan, "A framework for end-to-end deep learning-based anomaly detection in transportation networks," *arXiv preprint arXiv:1911.08793*, 2019.

[13] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, "Deep learning: A generic approach for extreme condition traffic forecasting," in *Proceedings of the 2017 SIAM international Conference on Data Mining*. SIAM, Conference Proceedings, pp. 777–785.

[14] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Thirty-Second AAAI Conference on Artificial Intelligence*, Conference Proceedings.

[15] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Thirty-First AAAI Conference on Artificial Intelligence*, Conference Proceedings.

[16] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, Conference Proceedings, pp. 802–810.

[17] D. Saxena and J. Cao, "D-gan: Deep generative adversarial nets for spatio-temporal prediction," *arXiv preprint arXiv:1907.08556*, 2019.

[18] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," 2011.

[19] J. Ren, Y. Hu, Y.-W. Tai, C. Wang, L. Xu, W. Sun, and Q. Yan, "Look, listen and learna multimodal lstm for speaker identification," in *Thirtieth AAAI Conference on Artificial Intelligence*, Conference Proceedings.

[20] S. S. Rajagopalan, L.-P. Morency, T. Baltrusaitis, and R. Goecke, "Extending long short-term memory for multi-view structured learning," in *European Conference on Computer Vision*. Springer, Conference Proceedings, pp. 338–353.

[21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines vinod nair," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel*, Conference Proceedings.

[22] NYCtaxidataset:https://www1.nyc.gov/site/tlc/about/tlc-trip-recorddata.page.

[23] PyTorch:https://github.com/pytorch/pytorch.