CITY-SCALE TAXI DEMAND PREDICTION USING MULTISOURCE URBAN GEOSPATIAL DATA

Jialin Yan, Longgang Xiang[*], Chenhao Wu, Huayi Wu

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China - (jialinyan, geoxlg, ngch, wuhuayi)@whu.edu.cn

Commission IV, WG IV/3

KEY WORDS: Taxi Demand Prediction, Pixel-Adaptive Convolution, Multisource Data, Data Fusion, Deep Learning

ABSTRACT:

Real-time, accurate taxi demand prediction plays an important role in intelligent traffic system. It can help manage taxi patching and minimize the time and energy waste caused by waiting. In the era of big data, a diversity of urban data and increasingly complex traffic data have been collected and published. Traditional forecasting methods have been unable to cope with the heterogeneous massive traffic data, whereas deep learning, as a new data-oriented technique, has been widely used in the field of traffic prediction. This paper aims to utilize multisource data and deep learning techniques to improve the accuracy of taxi demand prediction. In this paper, a joint guidance residual network (JG-Net) is proposed for city-scale taxi demand prediction. Taxi order data and multiple urban geospatial data (POI, road network and population distribution data) are integrated into the JG-Net. Regional features are considered in the prediction process by three guidance branches composed of pixel-adaptive convolutional networks, each of which applies one type of urban data. JG-Net assigns learnable weights to different branches and regions to combine the output of the branches, then further aggregates weather and time information to forecast the taxi demand. Extensive experiments and analyses are conducted, which show our method outperforms traditional methods. The mean square error of the prediction result on the testing set is 1.868, which is 12.3% lower than related models. The positive influence of combining multiple geospatial data is also validated by ablation experiments.

1. INTRODUCTION

Transportation is the lifeblood of urban vitality. The development of smart city is inseparable from effective intelligent transportation system. Taxi is an important part of urban public transportation. Thousands of people choose to take taxis every day because of convenience and rapidity. If we are able to accurately predict the taxi demand in various areas of the whole city, reasonable vehicle pre-allocation will be achieved, and then the time waste of passengers and drivers and the waste of fuel caused by empty vehicles can be greatly reduced. With the advent of the big data era, massive multisource urban data, such as GPS data, statistical data, and satellite imagery, have been collected through different sensors. Using these heterogeneous data to assist better city-scale taxi demand prediction is the problem we aim to solve in this paper.

Vehicle demand prediction is a branch of generic traffic prediction which also includes road speed prediction (Yao et al., 2017), travel time prediction (Wu et al., 2003), vehicle flow prediction (Abadi et al., 2015), etc. The formulas of the prediction problem for different spatio-temporal traffic data are very similar and the main purposes are almost identical—calculate the future traffic-relevant value (Yao et al., 2018). Scholars have put forward a series of models for traffic prediction.

The conventional prediction models are time series models, whose structure is predetermined based on certain theoretical assumptions, and the model parameters can be calculated using data. Autoregressive Integrated Moving Average (ARIMA) is a time series parametric model which assumes that traffic is a stationary process with constant mean, variance and autocorrelation. In the past few decades, scholars have proposed many traffic prediction models based on ARIMA (Ahmed and Cook, 1979; Lee and Fambro, 1999; Van Der Voort et al., 1996). ARIMA was also extended with Poisson (Moreira-Matias et al., 2013a) or perceptron (Moreira-Matias et al., 2013b) to predict taxi demand of a given region.

Machine learning prediction methods include clustering algorithms, support vector machine (SVM) prediction models and neural network prediction models. Clustering algorithm can aggregate the discrete taxi points into regions to estimate the demand of each region in the city (Chang et al., 2010; Davis et al., 2016). The essence of SVM is to map original data dimension to high-dimensional feature space through nonlinear transformation and then perform linear regression in this space. SVM-based traffic prediction methods (Sun et al., 2015; Wu et al., 2003; Yao et al., 2017) were proven to be superior to time series and regression-based methods. Deep learning doesn't require pre-determined features, reducing incompleteness caused by handcrafted features. Deep neural networks establish complex nonlinear relationship through distributed and hierarchical feature representations, providing a deeper representation of the data, which has been successfully applied in many fields, such as natural language processing (Graves et al., 2013), image recognition (Kemker et al., 2018; Marmanis et al., 2016), etc. Convolutional neural network (CNN) (Zhang et al., 2018, 2016) and recurrent neural network (RNN) (Ma et al., 2015a, 2015b; Xu et al., 2018) have been the two deep learning models mainly applied in traffic prediction application because the former can capture spatial dependency and the latter for temporal dependency. Some integrated models that combine CNN and RNN were also studied (Yao et al., 2018).

[*] Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-213-2020 | © Authors 2020. CC BY 4.0 License.

Despite the progress made over the years, city-scale taxi demand prediction is still a challenge because of region diversity, people's travel pattern and external influence such as weather and big events. Also, predicting taxi demand for all parcels at one time is tricky. Previous researches all trained their models and applied the same parameters to all city regions regardless of region features. Moreover, the methods of multiple data fusion were quite rough, which were either concatenation or summation.

To address the mentioned limitations above, we proposed a joint guidance residual network using historical taxi order data and multisource urban geospatial data (POI, road network and population distribution) to forecast taxi demand in every region. The categories of POI in a region can infer the region's function (Yuan et al., 2012), and we assume that the taxi demand is relatively high in areas with large POIs and population, as well as dense road networks. Raw vector data and raster data were processed. Then, we carried out correlation analysis between taxi demand and the above data. JG-Net is composed of three branches, each of which deploys pixel-adaptive convolution in order to implement region adaptive prediction. By feeding POI/road/population data into the net, region features can be learned. The model dynamically aggregates the output of aforementioned three branches guided by different types of data, then further combines context information, such as weather and time, to obtain the prediction result. Using taxi order data of Chengdu city, we conducted extensive experiments to test the performance of the proposed method, and results show that our model can predict the taxi demand of the entire city with good performance.

2. DATA PROCESSING AND CORRELATION ANALYSIS

This section introduces the data, the processing methods, and the correlation analysis. The data include taxi order data, multisource urban geospatial data (POI data, road data, population data), weather data and time metadata.

2.1 Taxi order data

We used the online taxi order open dataset from Didi Chuxing (Didi Data Center, 2016), one of the largest Internet ride-sharing and ride-hailing Uber-like companies in China. Our study area is Chengdu city, the capital city of Sichuan Province and the technology, financial and transportation center in southwestern China. It is reported that Didi Chuxing has provided ride service to 8.5 million users in Chengdu until September 2016, which means that 6 out of every 10 Chengdu residents have used it once (CBNData and Didi Research Institute, 2016). The dataset covers the whole Chengdu city within G4201 ring road and was collected in November 2016 with an average daily order of 200,000. Taxi order dataset is complete and covers a large proportion of population in Chengdu. There is no missing value in time nor space.

Each record in taxi order data represents a ride, including seven attributes: order ID, timestamp of the origin, latitude and longitude of the origin, timestamp of the destination, latitude and longitude of the destination. Taxi order data were transformed into time-series images by following processing steps. Regular grids are usually applied for statistical study on this topic. First, according to previous work that was also based on taxi order data from Didi Chuxing (Yao et al., 2018), we divided the entire city into 40 by 40 grids and each of them denotes a 0.7km by 0.7km region. Order ID was used for data deduplication. Taxi demand is defined as the number of origin points at one region per time interval. And drop-off amount is the number of destination points. Then we calculated demand and drop-off amount in the regions every half hour. Min-Max normalization was used to scale all data into the range [0,1]. A sample image of taxi demand at one time interval is shown in Figure 1.

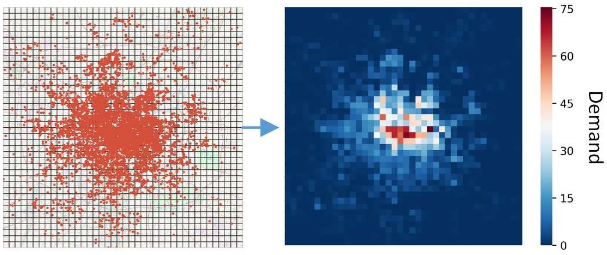


Figure 1. Taxi order data processing result

2.2 Multiple urban geospatial data

Urban geospatial data used in this paper include POI, road network, and the population distribution. POI data was collected from Baidu Map (https://map.baidu.com), which contains 19 categories and 162955 points in total. We counted the amount of each POI type in city grids and used it as the POI feature of the region. Raw road network data was obtained from Open Street Map (OSM), including primary roads, secondary roads, pedestrian roads, etc. For every grid, the length and number of roads were counted. The last kind of geospatial data, population distribution, was obtained from Global Human Settlement (GHS) website (https://ghsl.jrc.ec.europa.eu), which depicts the distribution of population, expressed as the number of people per cell. Since it is a raster dataset, we downsampled it from a resolution of 0.25 km to 0.7 km, the same as the city grid. The above three data can all cover the whole study area without missing value. They were not collected in the same year as taxi order data, but we believe POI, roads and population in a city won't dramatically change within a few years. Min-Max normalization was used to scale all data into the range [0,1]. Figure 2 shows the processing result of the above three types of urban data.

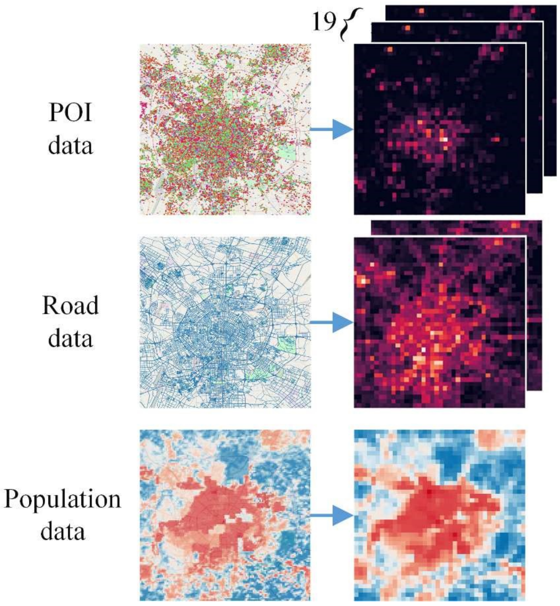


Figure 2. Multisource urban geospatial data processing results

2.3 Weather and time

The impact of weather and time on people's travel behavior is also considered in our study. Weather data came from the Weather Underground website (https://www.wunderground.com) and the raw data was collected per hour. No missing value is found and all values fall into reasonable range. First, the weather records were interpolated by an interval of half an hour. Then, one-hot encoding was applied to transform weather conditions and other time metadata (i.e., day_of_week, time_of_day) into binary vectors called context features. At last, we used Min-Max normalization to scale temperature and wind speed into the range [0,1].

2.4 Correlation analysis

In order to quantitatively analyze the relationship between taxi order data and multisource urban geospatial data, Pearson correlation coefficient, as defined as Eq. (1), was calculated between average daily demand, the amount of average daily drop-offs, summation of POIs, summation of road length, summation of road amount, and the population in each region.

$$CC = \frac{\sum (D_i - \bar{D})(M_i - \bar{M})}{\sqrt{\sum (D_i - \bar{D})^2 \sum (M_i - \bar{M})^2}} \qquad (1)$$

D_i and M_i are respectively the taxi demand and other data in region i, $\bar{D}$ and $\bar{M}$ are their average value. CC is correlation coefficient. All calculation results were proven to be significant through the hypothesis test. Table 1 shows that the drop-off amount and POI amount are both strongly related to taxi demand, while road and population are moderately related. Thus, combining the geospatial data properly can assist taxi demand prediction to a certain degree.

Data	CC	Description
Drop-off	0.98	Very strong
POI	0.76	Strong
Road	0.44	Moderate
Population	0.53	Moderate

Table 1. Correlation coefficient between demand and other data

3. JOINT GUIDANCE RESIDUAL NETWORK FOR TAXI DEMAND PREDICTION

In this section, we first present the general framework of the proposed prediction model—JG-Net, and then specifically introduce the structure of guidance branches that incorporate multisource geospatial data using pixel-adaptive convolution (PAC). At last, a fusion mechanism is designed to further aggregate all information together.

3.1 Framework Overview

We assume that the taxi demand in a certain region is greatly influenced by the surrounding area and remote areas can also influence the demand of current region through road connections. Besides, different regions serve as different function areas, their taxi demand patterns vary from one to another. JG-Net aims to capture the spatial-temporal dependency and region diversity in demand prediction problem. The purpose of taxi demand prediction problem is to forecast the demand value of the entire city at future time interval t, given historical data from past $t-1$ time intervals. At time interval t, we denote city-scale taxi demand in all $H \times W$ regions as a tensor $X_t \in \mathbb{R}^{H \times W}$, where H and W are height and width of one image. Thus, a time-series image $X \in \mathbb{R}^{T \times H \times W}$, where T represents many time intervals can express the dynamic changes of the demand.

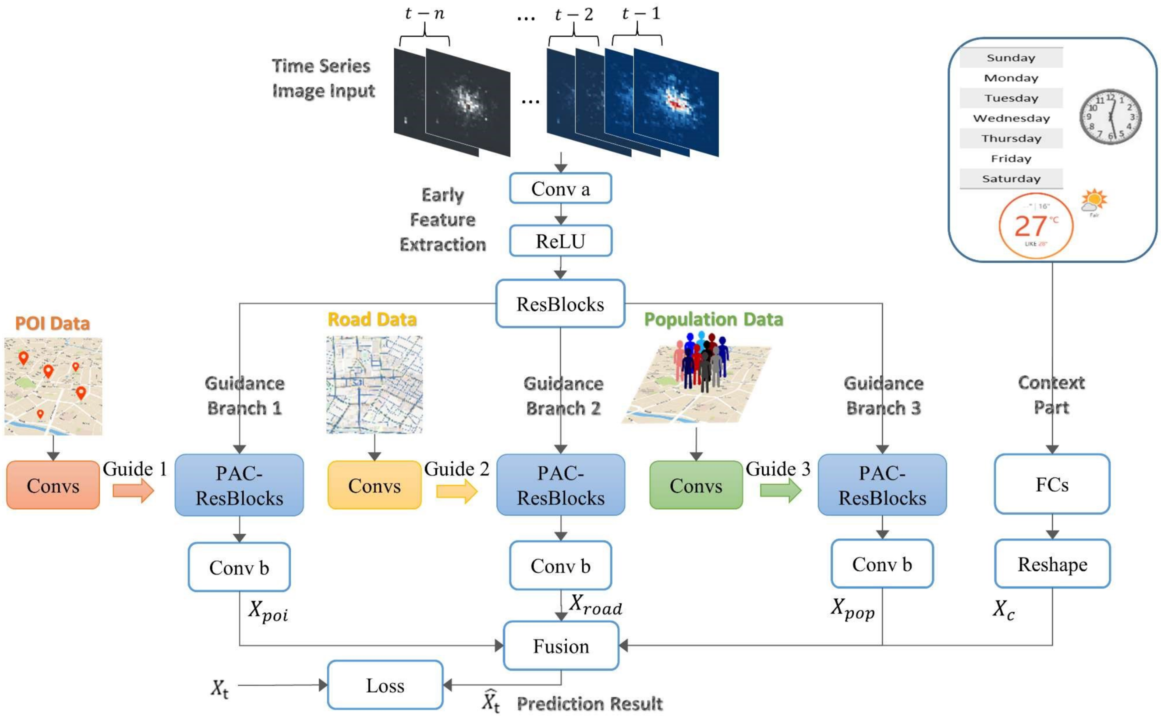


Figure 3. Joint Guidance Residual Network (JG-Net) architecture. Conv: Convolutional layer; ResBlock: Residual block; PAC: Pixel-adaptive convolution; FC: Fully connected layer.

Figure 3 depicts the architecture of JG-Net that is composed of early feature extraction, guidance branches, context part and feature fusion. The main model input is historical time-series images, and each image has two channels: demand and drop-off amount. We concatenate the images by the temporal axis (i.e. first image channel) and obtain a tensor that represents historical information. For early feature extraction, the original input is fed into a convolutional layer with an activation function, then into some residual blocks, each of which is composed of two convolutional layers and two activation layers. JG-Net applies pixel-adaptive convolution operation with the guidance of region features inferred from multisource urban geospatial data. POIs, population distribution, and road networks are used to assign features to pixels (i.e. regions) in each branch, respectively. The outputs of the three guidance branches are merged and further integrated with context features extracted from fully connected layers. At last, the future taxi demand prediction is the fusion result, which has the same height and width as the input images.

3.2 Structure of the guidance branch based on pixel-adaptive CNN

Three guidance branches share a similar structure that contains a guidance learning part and a residual pixel-adaptive CNN part followed by a convolutional layer. The taxi demand patterns for different functional areas may vary widely as shown in Figure 4. For example, Grid_1322 is a commercial area whose demand is the largest and increases on weekends when people are more likely to go shopping. In contrast, the demand in grid_1320 drops on weekends and is less than grid_1322 considering that grid_1320 is an educational and medical region. Since grid_1910 is near the airport and only travellers call for taxis there, the demand is more stationary and less than other areas.

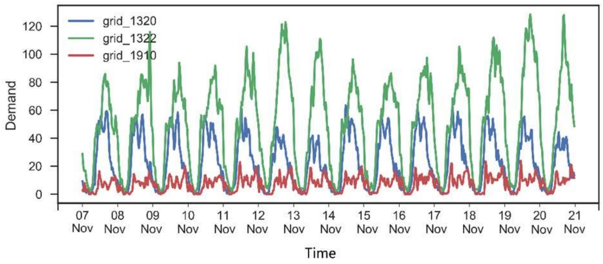


Figure 4. Taxi demand patterns of three different regions

Convolution is a fundamental operation in many traditional or deep learning computer vision applications (Krizhevsky et al., 2012; Lawrence et al., 1997; Simard et al., 2003). In standard CNN, the filter weights are shared spatially across the whole image, which is one of the main reasons for its popularity but also the major inherent limitation since it makes the convolution independent of content. For dense pixel prediction tasks, such as image segmentation and frame prediction, the optimal gradient for each pixel is different because each pixel should have a distinct output. However, CNN looks for global optimal parameters on the entire image that may not perfectly fit every pixel. Content-independent means that the trained CNN filter weights are used in all parts of the image, regardless of any local features. Similarly, in our demand prediction problem, the same parameters are applied in all regions of the city, regardless of whether the region is a commercial area, a cultural area or a residential area. To handle the limitations of traditional CNN, in this work, pixel-adaptive convolution operation (Su et al., 2019) is employed. In pixel-adaptive convolution, filter weights are not consistent but can change in different pixels. Each pixel in the image represents a region in the city. In this way, JG-Net is able to adaptively generate forecasting demand for the city in reference to individual features of each region.

Pixel-adaptive Convolution. The pixel-adaptive convolution (PAC) modifies traditional spatially invariant filter weights by a spatially varying kernel $K \in \mathbb{R}^{c' \times c \times s \times s}$ that depends on pixel features p . A pixel-adaptive convolution operation of input $x = (x_1, x_2, ..., x_n)$, $x_i \in \mathbb{R}^c$ over n pixels and c channels with c' -channel output is defined as

$$x_i' = \sum_{j \in M(i)} K(p_i, p_j) w_m x_j + b \tag{2}$$

where $M(i)$ is an $s \times s$ sliding convolution window, $w = (w_1, w_2, ..., w_{s \times s})$, $w_m \in \mathbb{R}^{c' \times c}$ are the filter weights, and $b \in \mathbb{R}^{c'}$ defines bias. The kernel function K has a fixed expression such as Gaussian in Eq. (3). By bringing pixel features p into K , the modifying parameters that can directly change original w are calculated. The parameters adapt the standard spatial filter w to each distinct pixel. The adapting pixel features p can be handcrafted or derived from end-to-end learning as our work in this paper.

$$K(p_{i,} p_j) = \exp\{-\frac{1}{2}(p_i - p_j)^\top (p_i - p_j)\} \tag{3}$$

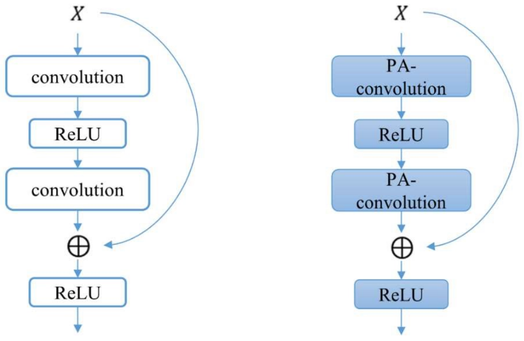
In the guidance learning part, we use several normal convolutional layers to extract adapting pixel features p , called guide, from multisource geospatial data (POI data, road data and population data). These features then serve as guidance in three model branches as shown in Figure 3. Different from traditional methods that train and apply the same parameters to all city regions, our method takes region characteristics into consideration, which is expected to help improve the accuracy of demand prediction.

Residual Block. Residual learning (He et al., 2016) is deployed in both the early feature extraction part and three guidance branches in JG-Net. Each convolution neuron in the feature map has a local reception field that is connected to the neurons in the upper layer by trainable weights (Lecun et al., 2015), which makes CNN fit for capturing dependencies between neighboring regions. All areas in a city are connected by complicated road networks, leading to the potential spatial dependency between distant areas, such as residential and commercial areas. Deep CNN structure can break the local limitations since it has a reception field that is wide enough to capture the spatial dependency of regions even though they are far apart. Much progress has been made in computer vision applications using deep networks as they integrate multilevel features (He et al., 2015; Jin et al., 2017). Residual learning has been proven to be very effective in deep structure because it can alleviate the problems of vanishing/exploding gradients and training accuracy degradation (He et al., 2016), knowing as the two main challenges in deep networks. Therefore, we design a deep model based on CNNs and PA-CNNs where residual learning is applied.

In the proposed JG-Net, a residual block is defined as:

$$X' = \mathcal{F}(X, W) + X \tag{4}$$

where X and X' are input and output tensors. $\mathcal{F}(X,W)$ denotes the residual function and W are all learnable parameters including weights and biases. Here we use a function $\mathcal{F}$ with two convolutional or PA-convolutional layers and a ReLU layer since a single layer is proven to be unhelpful (He et al., 2016). As shown in Figure 5, the only difference between a PAC-Residual Block and a Residual Block is that the former has pixel-adaptive operation in its convolutional layer. To generate shape-identical input and output and avoid resolution decline, we eliminate pooling operation in our model and apply the same padding in all convolutional layers. With L PAC-residual blocks and a final convolution, the output of three branches are X_{poi}, X_{road} and X_{pop} in Figure 3.



(a) Residual Block (b) PAC-Residual Block

Figure 5. (a) Structure of residual block and (b) PAC-residual block

3.3 Fusion

In this section, we introduce the fusion mechanism that combines three guidance branches and context information. First, the pixel-adaptive convolutional branches are merged using weight matrices, then the result is further merged with weather and time metadata by other weight matrices.

Taxi demand is related to POI, road and population distribution features as described in section 2.4. However, different regions are influenced by them to different degrees. For example, there may be only one POI in a transportation center (such as a train station), but its surrounding is densely distributed with roads. And there may be only one road through a commercial street while the number and variety of POIs are enormous. Additionally, in residential areas or schools, POIs and roads can be both sparse but the population density is large, leading to large taxi demand. A parametric-matrix-based fusion method was proposed for branch aggregation by assigning learnable weights to the branches and regions (Zhang et al., 2018). Inspired by the above phenomenon and parametric-matrix-based fusion method, we use three weight matrices learned from backpropagation to fuse POI guided branch, road guided branch and population guided branch as follows:

$$X_{branch} = W_p \circ X_{poi} + W_r \circ X_{road} + W_o \circ X_{pop} \tag{5}$$

where $\circ$ is Hadamard multiplication, W_p, W_r and W_o are learnable weight matrices that represent different influence degrees. Another two weight matrices W_c and W_b are learned to merge X_c and X_{branch}, then we get our city-scale prediction result $\hat{X} \in \mathbb{R}^{H\times W}$ in Eq. (6). Mean square error is the loss function in JG-Net.

$$\hat{X} = W_c \circ X_c + W_b \circ X_{branch} \tag{6}$$

4. EXPERIMENTS

To evaluate the proposed Joint Guidance Residual Network using multiple geospatial data, we carried out comprehensive demand prediction experiments on Didi Chuxing taxi order dataset as described in Section 2. Details about experiment settings and prediction results are provided in the subsequent subsections.

4.1 Experiment implementation

Experiment settings: We run the experiments using PyTorch library on a cluster of two NVIDIA GeForce 1080Ti GPUs that has 10GB memory each. The dataset contains taxi orders from 2016-11-01 to 2016-11-30 in Chengdu. Multisource data are Baidu Map POI data, OSM road data, GHS population data and web weather data. After data processing and normalization, we obtained 48 time intervals (half an hour) per day and 1440 in total. The historical data from the previous 8 time intervals is the input of all models. Data from the first 26 days are used for training and validation, and the last 4 days are for testing.

In training process, RMSprop optimization algorithm is deployed and the learning rate is set to 0.001. We trained all models for 200 epochs. In JG-Net, convolutional layers all have 64 filters except for the last layers in each branch. Two residual blocks are used for early feature extraction and two PAC residual blocks are included in each guidance branch. Table 2 introduces all parameters in JG-Net. Since original POI image has 19 channels, more filters are needed to extract features.

Layers	Filter Size	Filters
Guidance (POI) Learning CNN	3×3	32-16-16
Guidance (road) Learning CNN	3×3	8-4-4
Guidance (population) Learning CNN	3×3	8-4-4
Convolutional Layer a	3×3	64
Residual Block × 2	3×3	64
PAC-Residual Block × 2	3×3	64
Convolutional Layer b	3×3	1

Table 2. Model parameters for JG-Net

Metrics: We use Mean Square Error (MSE) for evaluation in our experiments. They are defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \hat{X}_i\right)^2 \tag{7}$$

where X_i and $\hat{X}_i$ are ground truth and forecasted value, N is the number of samples.

4.2 Overall comparison

We compared our JG-Net model with the following six related methods:

- Historical Average (HA): The demand of future time intervals is calculated by the average demand of historical time intervals.
- ARIMA: We used auto.arima function in R to fit the best ARIMA model.
- Multiple Layer Perceptron (MLP): Several fully connected layers are stacked and the numbers of hidden units are 4096, 2048,

https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-213-2020 | © Authors 2020. CC BY 4.0 License.

1024, 2048 and 1600. Original image input is flattened into a one-dimensional tensor.

- Long Short-Term Memory (LSTM): Three gates are included in LSTM to capture near and long dependencies in sequence learning. The numbers of hidden units are 2048, 1024, 1024 and 1600.
- Spatiotemporal Recurrent Convolutional Networks (SRCNs): SRCNs (Yu et al., 2017) combines LSTM and CNN to predict traffic speed of road network. We applied the same structures in their paper except that the last layer has 40×40 = 1600 neurons.
- ST-ResNet: ST-ResNet (Zhang et al., 2018) has a residual learning structure and considers three temporal dependencies. Due to the data limitation, we only included the closeness and period parts.

Model	MSE
HA	3.293
ARIMA	3.181
MLP	2.988
LSTM	2.651
SRCNs	2.119
ST-ResNet	2.13
JG-Net	
No Branch	2.21
POI Branch	1.945
Road Branch	1.897
Population Branch	1.931
3 Branches without Fusion	2.092
3 Branches with Fusion	**1.868**

Table 3. Comparison among different methods

Table 3 shows the prediction results. The proposed JG-Net reaches the lowest MSE (1.868), which reduces the error by 12.3%. Among these methods, HA and ARIMA only consider historical values of a single region, thus they perform poorly. SRCNs and ST-ResNet have similar prediction error because they both use CNN to capture spatial dependency of the whole city, which makes them superior to MLP to LSTM where spatial features are flattened.

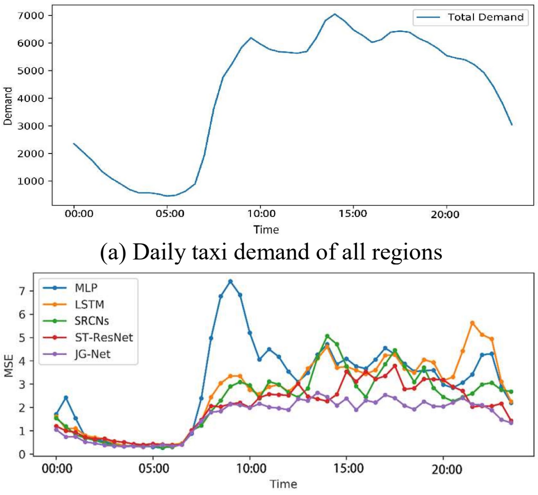
In order to verify the impact of multisource data on forecast results, we also compared the performance of several model variants:

- No branch: The proposed JG-Net degrades to a regular residual net. Only taxi order data is used in this model.
- POI Branch, Road Branch, Population Branch: The prediction is guided by only one type of the urban data, either POI data, road data or population data.
- 3 Branches without Fusion: This model variant simply sums the outputs of three branches.

Compared to the no-branch model, variants integrated with urban geospatial data perform better, verifying that region features learned from these data can improve demand prediction. Therefore, three kinds of data are used in our final JG-Net. When we simply sum the results of three branches instead of using weight matrices, the error increases, which demonstrates the effectiveness of our Hadamard product fusion mechanism.

4.3 Performance of different time and locations

We compared the forecasting performance of the JG-Net at different time. Figure 6 shows how prediction error changes in a day. Figure 6(a) is the daily demand of all regions. The taxi demand is the lowest at 5 am and so is MSE. Then the city comes alive, and in the meantime, the prediction error of these models begins to increase. The demand reaches its peak in the afternoon. Among the models, MLP is the most unstable and has the highest rising degree in MSE. The error of LSTM is smaller compared to MLP. The error curve of ST-ResNet is smoother than that of SRCNs because of residual learning. Our proposed JG-Net has the most stable curve, and relatively good prediction results are obtained at all time in a day.



(a) Daily taxi demand of all regions

(b) MSE over time

Figure 6. Prediction performance at different tim

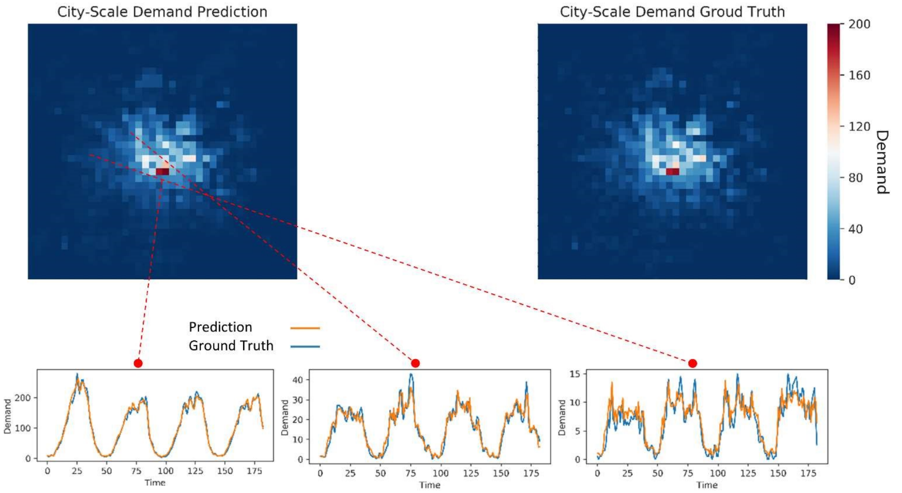


Figure 7. Prediction performance at different location

To examine the model performance in the entire city, we carried out some experiments at different locations. The x-axis of three subplots in Figure 7 represents predicted time intervals. Figure 7 shows great similarity between the city-scale prediction value and ground truth at one time interval, which is a solid proof of the superiority of our model at the city level. Among three typical locations, JG-Net fits the true values best at the commercial region where taxi demand is high and has a regular pattern. The fitting results of the model in other regions are also reasonable because we consider the individual characteristics of each region in the prediction process.

5. CONCLUSION AND FUTURE WORK

In this paper, inspired by region diversity in cities, we propose a joint guidance residual network named JG-Net for city-scale taxi demand prediction based on residual learning and pixel-adaptive convolution. The model has three guidance branches, and in each branch, multiple geospatial data, including POI data, OSM road data and population data, are applied to assign features to regions. This work is the first attempt to consider region characteristics in the city-level prediction process. Besides, weather information and time metadata are also considered. Using Didi Chuxing taxi order dataset of Chengdu, we carried out data processing and demand prediction experiments. By correlation analysis, we found that the above three types of geospatial data are related to taxi demand value, especially POI data. The experiment results show that the proposed model performs stably at different locations and time intervals in comparison with other related models. The overall prediction error on the testing set is reduced by 12.3%. The effectiveness of integrating multiple sources of geospatial data is also validated by ablation experiments.

To further improve the prediction accuracy, we will consider to collect more data, such as traffic density data and crowd mobility data. Taxi demand prediction for the entire city is a spatiotemporal problem. The present work has not sufficiently addressed the temporal dimension. In future studies, we will conduct more in-depth research on how to consider the temporal dimension, such as using 3D convolution or channel attention.

ACKNOWLEDGEMENTS

This research was supported by the National Key Research and Development Project (Grant No. 2018YFB2100603) and the National Science Foundation of China (Grant No. 41771474).

REFERENCES

Abadi, A., Rajabioun, T., Ioannou, P.A., 2015. Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data. *IEEE Transactions on Intelligent Transportation Systems*, 16(1), 653–662.

Ahmed, M.S., Cook, A.R., 1979. Analysis of Freeway Traffic Time-series Data by Using Box-jenkins Techniques. *Transportation Research Record*, 1–9.

CBNData, Didi Research Institute, 2016. Smart Travel Big Data Report [WWW Document]. URL https://www.cbndata.com/report/374/detail?isReading=report&page=7 (accessed 7.10.19).

Chang, H. wen, Tai, Y. chin, Hsu, Y. jen J., 2010. Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence and Data Mining*, 5(1), 3–18.

Davis, N., Raina, G., Jagannathan, K., 2016. A multi-level clustering approach for forecasting taxi travel demand. In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. Institute of Electrical and Electronics Engineers Inc., pp. 223–228.

Didi Data Center, 2016. Taxi Order Dataset of Chengdu [WWW Document]. URL https://gaia.didichuxing.com (accessed 8.1.19).

Graves, A., Mohamed, A.R., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. pp. 6645–6649.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, pp. 770–778.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1904–1916.

Jin, K.H., McCann, M.T., Froustey, E., Unser, M., 2017. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, 26(9), 4509–4522.

Kemker, R., Salvaggio, C., Kanan, C., 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145, 60–77.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*. pp. 1097–1105.

Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D., 1997. Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks*, 8(1), 98–113.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature*, 521(7553), 436-444

Lee, S., Fambro, D.B., 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. *Transportation Research Record*, 179–188.

Ma, X., Tao, Z., Wang, Yinhai, Yu, H., Wang, Yunpeng, 2015a. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54, 187–197.

Ma, X., Yu, H., Wang, Yunpeng, Wang, Yinhai, 2015b. Large-scale transportation network congestion evolution prediction using deep learning theory. *PLoS ONE*, 10(3).

Marmanis, D., Wegner, J.D., Galliani, S., Schindler, K., Datcu, M., Stilla, U., 2016. Semantic Segmentation of Aerial Images with an Ensemble of CNNs. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information*, CWG III/VII, 473–480.

Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., Damas, L., 2013a. Predicting taxi-passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, 14(3), 1393–1402.

Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., Damas, L., 2013b. On predicting the taxi-passenger demand: A real-time approach. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 54–65.

Simard, P.Y., Steinkraus, D., Platt, J.C., 2003. Best practices for convolutional neural networks applied to visual document analysis. In: *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. IEEE Computer Society, pp. 958–963.

Su, H., Jampani, V., Sun, D., Gallo, O., Learned-Miller, E., Kautz, J., 2019. Pixel-Adaptive Convolutional Neural Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 11166–11175.

Sun, Y., Leng, B., Guan, W., 2015. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing*, 166, 109–121.

Van Der Voort, M., Dougherty, M., Watson, S., 1996. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, 4(5), 307–318.

Wu, C.H., Wei, C.C., Su, D.C., Chang, M.H., Ho, J.M., 2003. Travel time prediction with support vector regression. In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. Institute of Electrical and Electronics Engineers Inc., pp. 1438–1442.

Xu, J., Rahmatizadeh, R., Boloni, L., Turgut, D., 2018. Real-Time prediction of taxi demand using recurrent neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 19(8), 2572–2581.

Yao, B., Chen, C., Cao, Q., Jin, L., Zhang, M., Zhu, H., Yu, B., 2017. Short-Term Traffic Speed Prediction for an Urban Corridor. *Computer-Aided Civil and Infrastructure Engineering*, 32(2), 154–169.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Li, Z., Ye, J., Chuxing, D., 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In: *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*. AAAI press, pp. 2588–2595.

Yu, H., Wu, Z., Wang, S., Wang, Y., Ma, X., 2017. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors (Switzerland)*, 17(7), 1501.

Yuan, J., Zheng, Y., Xie, X., 2012. Discovering regions of different functions in a city using human mobility and POIs. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 186–194.

Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., 2016. DNN-based prediction model for spatio-temporal data. In: *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*. Association for Computing Machinery, pp. 1-4.

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