Urban traffic flow forecast based on dual path network

Hao Li, Xudong Liu, Yan Kang*, Yachuan Zhang and Rongjing Bu

1 Software College, Yunnan University, Kunming, Yunnan, 650000, China

*Corresponding author’s e-mail: kangyan@ynu.edu.cn

Abstract. Traffic flow prediction is very important for city construction. Time and space factors have a great impact on traffic flow. Traditional traffic prediction methods can only capture temporal correlations and not capture spatial and regional correlations. The use of convolutional neural networks can well capture the correlation between regions and the dependence between time and space, which can make traffic prediction more accurate. Therefore we introduced the dual-path network, and divided the traffic profile after convolution into two paths and trained at the same time. One path is ResNet and one path is DenseNet. The dual path network combines the advantages of these two networks for traffic prediction. The experimental results show that compared with the traditional traffic prediction model, the model not only improves the efficiency of the network but also improves the prediction accuracy of the network.

1. Introduction

This traffic flow data is obtained by collecting real-time trajectories of taxis. These data can well map the traffic congestion in the city. We divide Beijing into 32x32 grid traffic maps. The data on each grid represents the traffic in the area. The flow rate of each grid is divided into inflow and outflow, and the change of traffic is counted every half hour so that 48 inflow graphs and outflow graphs per grid per day are obtained, and the future flow is predicted based on historical traffic flow.

We have several contributions: Traditional traffic prediction methods such as sequentially models can only solve correlation of time, and cannot obtain the characteristics of space and area, so the convolution kernel is used to capture the correlation between time and space and region more accurately. i) Reuse the features extracted by the convolutional neural network using the ResNet network structure, which preserves the dependence between regional traffic. ii) Use the DenseNet structure to re-extract useful information from the output of all previous layers. It can effectively use the high-level information to re-examine the underlying new features and better capture the implicit correlation of some regions. iii) Converged ResNet and DenseNet, taking full advantage of the advantages of the two networks, so as to better predict traffic. There are three trends in traffic flow information, one is short-term, one is periodic, and one is long-term. Use this data as input to the DPN, and assign different weights, while further aggregating many other factors such as holidays and rain to make predictions more accurate.

2. Related work

Traffic flow prediction methods are constantly improving. With the development of neural networks and deep learning, the accuracy of traffic flow prediction has increased rapidly. Models such as RNN and CNN have achieved very good results in traffic flow prediction.
2.1. Model based on Knowledge

Traditional traffic prediction model parameters are not automatically updated with changes in observations, making it difficult to achieve real-time traffic forecasting. For example, ARIMA requires that the timing is stable or stable after differentiation. It essentially captures linear relationships and cannot capture nonlinear relationships. SARIMA adds seasonal adjustments based on ARIMA, but its effect on time series prediction is still not as good as LSTM. Although LSTM can make better predictions of time series, it is difficult to capture correlations between regions. So we used the dual path model for traffic flow prediction. By comparing the root mean square error our model is more effective than the above model.

3. Problem definition

**Trajectory data:** Trajectory data is information data of a time-space feature that can be collected by GPS, mobile phones, which can represent one or more moving objects. The trajectory data generally includes human activity trajectories, vehicle activity trajectories

**Area:** According to the real map latitude and longitude. We divide the Beijing city into 32*32 areas.

**Traffic image:** Each pixel on the picture represents the traffic in that area. We use tensorflow to convert these pictures into tensors. Because the traffic is converted into the grid changes with time, and statistics can be obtained at regular intervals. In this paper, the flow of the entire area is counted every half hour, and 48 flow images in which the size is 32*32 are obtained in one day.

**Traffic prediction:** We can find the trend of human traffic and the dependence of traffic changes in different grids and different periods by using the corresponding prediction model, the traffic flowy at time t in the region can be predicted based on the traffic flow I at time t-1

**Flow channel:** Like the RGB channel of the picture, the flow picture also has channels. The size of the experimental input flow picture is 32*32*2, 32*32 represents the size of the picture, and 2 represents the depth of the picture. In this paper, 2 indicates the inflow and outflow.

4. Dual path network structure

**DPN consists of DenseNet and ResNet together.** And we specifically introduced a dual-path network framework, and how do we use DPN for traffic prediction.

4.1. DenseNet

Based on forwarding propagation, each layer of the network can receive the feature maps of all the layers in front of it, and the data aggregation uses the splicing instead of the addition in Resnet. This connection has a big advantage. In the forward propagation, the deep network can obtain shallow information, while in the backward propagation, the shallow network can obtain deep gradient information. This maximizes the flow of data between networks. Also, this structure has a large number of feature reuse. Therefore, only a few parameters are needed, and very good results can be achieved. Mainly reflected in the number of channels of the feature map, DenseNet may only need about 12 or 24 compared to hundreds of channels of VGG and ResNet. DenseNet may only need about 12 or 24 or so. DenseNet can be expressed as Equation 1:

$$x_c = H_c(\left[x_0, x_1, ..., x_{c-1}\right])$$  

Where $x_c$ represents the current feature map, $x_0$, $x_1$, ..., $x_{c-1}$ represents the input map and the convolution map of the previous layer, and $H_c$ represents the merge operation, which combines $x_0$, $x_1$, ..., $x_{c-1}$ using concat mode. So that the characteristics of each layer are directly passed to the next layer, and the features are utilized more effectively. The network model is shown in Figure 1 below:
4.2. ResNet network
As the layers of neural networks continue to deepen, it becomes very difficult to train them. Some scholars have found that using $y = x$ as an identity map can reduce the problem of network gradient disappearance and explosion. After adding identity mapping on each layer of the network, the number of network layers can be continuously deepened, so that more features can be learned, its calculation formula is as follows:

$$y = H(x, \{W_i\}) + x$$

(2)

$X$ represents the input tensor, $Y$ represents the output tensor obtained after a layer of ResNet. When the number of network layers deepens, gradients may disappear or gradients may explode, which makes back propagation difficult. ResNet adds the information from the previous layer to the next layer, effectively saving the integrity of the information, so that the training results become more accurate. The structure of ResNet is shown in Figure 2 below:

ResNet has been reusing the features that have been extracted in the previous layer and has a very high reuse rate. However, except for those directly connected multiplexing features, it is a new feature that has not been extracted before, so the features extracted by ResNet. Medium redundancy is relatively low. The features extracted from the front layer of DenseNet are no longer simply reused by the latter layers, but create new features. The features extracted by the convolution in the back layer of this structure are likely to have been extracted from the previous layer. Therefore, the feature redundancy extracted by DenseNet is high. ST-DPN combines the core ideas of ResNet and DenseNet to achieve advantages and disadvantages [31]. The DPN slices the convolved feature map into two paths, one for ResNet and one for DenseNet. By combining the advantages of these two models, you can do better than the separate ResNet and DenseNet. The structure of the dual-path network is shown in Figure 3.
4.3. Additional Information Processing Component.
Traffic flow is likely to be affected by many external factors such as weather and holidays. For example, in the holiday season, the traffic volume of the work area greatly reduced compared with the normal day. So to predict the flow of people in the time interval, we map the extra data through the fully connected layer to the same dimension as the traffic matrix.

4.4. Treatment of additional factors
We used the data of the day, the data of the previous week and the three data of the previous month as the input of the network. Using these three types of data can better simulate the trend of traffic flow. We use the following formula to calculate the total flow:

\[ X = W_N \cdot X_N + W_p \cdot X_p + W_q \cdot X_q \]  

(3)

We have also added additional factors that affect traffic flow, such as weather and holidays, defining Xt as:

\[ X_t = \tanh(X) \]  

(4)

Dataset: We used TaxiBJ open source dataset. The specific content of the data set is shown in Table 1.

| Table 1. Formatting sections, subsections and subsubsections. |
|-----------------|------------------|-----------------|
| Data set        | TaxiBj           |                 |
| Data type       | Taxi Gps         |                 |
| Place           | Beijing          |                 |
| Time slot       | 2013/7/1~2013/10/30 |
|                 | 2014/3/1~2014/6/30 |
|                 | 2015/3/1~2015/6/30 |
|                 | 2015/11/1~2016/4/10 |
| Time interval   | 30 min           |                 |
| Network size    | (32,32)          |                 |
| Trajectory data |                  |                 |
| Taxi count      | 34000+           |                 |
| Available time interval | 22459    |                 |

There are six types of comparison models:
ARIMA: Differential integrated moving average autoregressive model, one of the time series predictive analysis methods.
SARIMA: Is an improved version of Arima, very effective in time series prediction.
RNN: RNN is very effective in dealing with time series problems. We convert the input tensor into a fully connected layer of 1024, and then use RNN processing. We used 4 different levels of RNNs as experiments.

LSTM: LSTM is an enhanced version of RNN, which effectively solves the problem that RNN cannot remember time state well. Similarly, we also used 4 different layers of LSTMs as experiments.

ResNet: This model uses ResNet to simulate the connections between the near and far regions so that it can accurately predict regional traffic.

DenseNet: Another convolutional neural network proposed after the milestone ResNet network successfully surpassed ResNet on the ImageNet dataset.

Dual path network: The simultaneous training of ResNet and DenseNet combines the advantages of both to enable a more accurate prediction of regional traffic.

Experiment procedure: We use tanh as our last output activation function, it keeps all our values in the [-1,1] range. During training, we use the standardization of dispersion to map the data to [-1,1], so that the impact of the data value on the result can be minimized during training, but we need to restore the normalized data during verification. Go to the real data and compare it with the original geogrid data. For the extra factor, we use one-hot encoding to convert the raw data into a binary vector and also use the dispersion normalization to map the weather and wind speed to [0,1].

Hyperparameters: All learned parameters are initialized with the average distribution of keras. All residual units use 64*3*3 convolution kernels. Using Adam as the optimizer, the batch size is 32. We chose 90% of the data as training data and 10% of the data as test data.

Evaluation Indicator: We tested our model by Root Mean Square Error (RMSE).

\[
\text{RMSE} = \frac{1}{z} \sqrt{\frac{1}{2} \sum (x_i - \hat{x}_i)^2}
\]

Where \(x\) represents the actual grid flow graph, \(\hat{x}_i\) indicates the corresponding predicted value, and \(z\) refers to the total number of data for the current batch.

4.5. Verification

We judge the quality of our model based on the results of RMSE. The experimental results in Table 2 demonstrate the superiority of our model. The results show that the RMSE mean of our model is 27.12% higher than ARIMA, 34.81% more than SARIMA, 34% to 71% higher than RNN model, and 25% to 57% more than LSTM, 3.62% more than ResNet. Because LSTM can better handle the state of time series, the result is much better than RNN.

| Model          | RMSE |
|----------------|------|
| ARIMA          | 22.78|
| SARIMA         | 26.88|
| RNN-3          | 30.41|
| RNN-12         | 47.25|
| RNN-24         | 52.53|
| RNN-48         | 43.81|
| LSTM-24        | 26.25|
| LSTM-48        | 28.87|
| ResNet         | 17.17|
| DenseNet       | 17.43|
| Dual path network | 16.80|

5. Summary and future work

To better handle the characteristics of time and space, we used dual path networks. The dual path network uses convolution kernels to capture the correlation between time and space and regions. Through DenseNet and ResNet, can better extract traffic features for training. Our models have
achieved significant results with additional factors that have significant impact. By comparing the RMSE, Compared with other models, our model has higher accuracy and more accurate predictions.

References
[1] Zhang N, Chen H, Chen X, et al. Forecasting Public Transit Use by Crowdsensing and Semantic Trajectory Mining: Case Studies[J]. ISPRS International Journal of Geo-Information, 2016, 5(10):180.
[2] Rasmussen T K, Ingvardson J B, Nielsen O A. Improved methods to deduct trip legs and mode from travel surveys using wearable GPS devices: A case study from the Greater Copenhagen area[J]. Computers Environment & Urban Systems, 2015, 54:301-313.
[3] Zhang J, Zheng Y, Qi D. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction[J]. 2016.
[4] Fan Z, Song X, Shibasaki R, et al. CityMomentum: an online approach for crowd behavior prediction at a citywide level[C]// ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2015:559-569.
[5] Ashbrook D, Starner T. Using GPS to learn significant locations and predict movement across multiple users[J]. Personal & Ubiquitous Computing, 2003, 7(5):275-286.
[6] Wen H, Zhongwei H U, Guo J, et al. Operational Analysis on Beijing Road Network during the Olympic Games[J]. Journal of Transportation Systems Engineering & Information Technology, 2008, 8(6):32-37.
[7] C. K. Moorthy, B. G. Ratcliffe. Short Term Traffic Forecasting Using Time Series Methods[C]// Transportation Planning Technology. 1988:45-56.
[8] Shekhar S, Williams B M. Adaptive Seasonal Time Series Models for Forecasting Short-Term Traffic Flow[M]. 2007.
[9] Guo J, Huang W, Williams B M. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification[J]. Transportation Research Part C, 2014, 43:50-64.
[10] Zhang L, Ma J, Zhu C. Theory modeling and application of an adaptive Kalman filter for short-term traffic flow prediction[J]. Journal of Information & Computational Science, 2012, 9(16):5101-5109.