Dual-Stage Attention Based Spatio-Temporal Sequence Learning for Multi-Step Traffic Prediction

Ziqiang Cui. Chunhui Zhao*

State Key Laboratory of Industrial Control Technology,
College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, China
(Corresponding author, email: chhzhao@zju.edu.cn, 21732060@zju.edu.cn)*

Abstract: Traffic prediction has great significance including but not limited to mitigating traffic congestion, reducing traffic accidents, and reducing waiting time. At the same time, traffic prediction, especially multi-step prediction, faces many difficulties including temporal correlations and spatial correlations. We propose a dual-stage attention based spatio-temporal sequence learning for multi-step traffic prediction which can not only express temporal correlation and spatial correlation, but also can adaptively learn the contribution weights of different related roads and historical moments. More specifically, for spatial dependencies, we first generate the input vector for each historical moment considering the information of relevant road segments by the method of spatial region of support and further add the first-stage attention termed spatial attention to automatically determine the weight of each relevant road segment for each historical moment. For temporal dependencies, we use LSTM based encoder-decoder networks to fully learn the temporal characteristic and make multi-step prediction considering temporal correlation between multi steps. We further add the second-stage attention termed temporal attention in the decoder part to automatically learn the contribution of different historical moments to each prediction moment. In addition, we consider external factors including weather and holidays and characterize their impacts using fully connected networks. Finally, the effectiveness of the proposed method is evaluated using traffic data in Hangzhou, China.

Keywords: multi-step traffic prediction, attention mechanism, sequence learning, spatial correlation, temporal correlation.

1. INTRODUCTION

With the increment of global population and the acceleration of urbanization, the number of road networks and the number of motor vehicles have continued to increase, and traffic congestion has become a more and more serious problem. At the same time, traffic congestion directly leads to an increase in people's travel time, and it also leads to an increase in the time cost on work and a decline in the productivity of the whole society.

In order to effectively alleviate the above problems, research on traffic prediction has gained increasing attention. As early as the 1980s, many scholars began to explore the use of statistical learning methods for short-term traffic prediction. Typical examples include the History Average Model, the Time Series Analysis Model, etc. Smith and Demetsky (1994) compared several traffic flow prediction algorithms including the historical average model. Williams used an exponential smoothing algorithm to optimize the historical average model when predicting traffic flow on urban highways. The time series data prediction methods can be used for traffic prediction. G. Box et al. (2010) analyzed the application of Auto-Regressive Integrated Moving Average (ARIMA) in time series forecasting and control. The model are relatively simple, which generally have low precision and cannot fit complex traffic flow relationships. An empirical models and frequency-band separation based method (Zhao et al., 2014) is proposed to make online time series prediction which does not need to adjust model parameters when predicting future data for other subjects. Another prediction model which allows incremental learning and can be migrated to new subjects is proposed for time series prediction (Luo and Zhao, 2019), which, however, can not well represent complex characteristics of traffic data since the base model is linear. Recently, more and more research has applied machine learning including deep learning in the field of traffic prediction. A traffic flow forecasting method based on Support Vector Regression and Particle Swarm Optimization (PSO) is proposed (Hu et al., 2015). Lv et al. (2015) proposed a deep learning based method which is the first time that a deep architecture model is applied using auto encoders as building blocks to represent traffic flow features for prediction. However, the above methods only consider the values of adjacent moments before the predicted time, ignoring the influence of longer time scales and the spatial correlation. The deep learning model represented by the recurrent neural network (RNN) and the convolutional neural network (CNN) has been used by more and more scholars in traffic prediction in recent years. Zhang et al.(2016) proposed CNN based prediction model and its improved version ST-ResNet (Zhang et al., 2017) which used the advantages of CNN network to fully express the spatial correlation of traffic flow, but the temporal characteristics are not fully expressed and the grid region has no actual physical meaning in real world. The LC-RNN model (Lv et al., 2018) effectively integrates RNN and CNN, and has the advantage that RNN is good at capturing timing characteristics and CNN is good at capturing spatial characteristics. An encoder-decoder networks for traffic speed prediction is proposed which considered temporal correlation and spatial correlation and can make multi-step prediction (Liao et al. 2018). In addition to deep learning, a spatio-temporal broad learning networks for traffic speed prediction (ST-BLN) (Cui and Zhao, 2019) is proposed which can make one-step traffic prediction faster and more accuracy. And the method of spatial region of support and temporal region of support is also proposed in this paper to effectively select related road segments and historical moments. But all of the above work did not consider the contribution of different historical moments and the contribution of different related sections.

The traffic prediction problem is a very complicated problem, and there are still areas for improvement. In this paper, we propose dual-stage attention based spatio-temporal sequence learning for multi-step traffic prediction. Specifically, we utilize Long Short Term Memory (LSTM) based encoder-decoder networks to express the temporal dependencies and timing characteristics. To deal with the spatial correlations, we generate the input vector for each historical moment considering the information of relevant road segments using the method of spatial region of support. And we further add the first-step attention to the input vector to automatically determine the weight of each relevant road segment for each moment. Then the information of multiple historical moments is sequentially input to the encoder part for encoding. In the decoding part, we also add the attention mechanism to to adaptively select relevant encoder hidden states across all time steps, i.e., the second-stage attention in the proposed method. Thus, when the decoder part decodes at each prediction moment, it will comprehensively consider the information of all the moments of the encoder part and automatically learn to determine the weight of the information of different historical moments. The outputs of the decoder at each moment would be further combined with external factors including weather and holidays into a fully connected network and finally outputs the prediction results of multiple time steps in the future. Our proposed method can not only express temporal dependencies and spatial dependencies, but also automatically determine the importance of different relevant road segments and different historical moments. In addition, it fully considers external factors and can perform multi-step prediction.

Our contributions are three-fold:

- Considering different contribution of different related road segments to the target road segment, we add the first-stage attention, termed spatial attention, to the input vectors. It can automatically learn the weights of various relevant road segments to obtain a new weighted input vector for each moment.
- Considering the different impacts of different historical moments on the forecasting moment, we add the second-stage attention, termed temporal attention, in the decoder part. It can automatically learn the contribution weights of different historical moments to each prediction moment.
- We utilize the architecture of encoder decoder to make multi-step prediction and utilize fully connected networks to express the influence of external factors, including weather and special events etc.

The rest of this paper is structured as follows. We briefly introduce the attention mechanism in Section 2. The architecture of the proposal method is introduced in Section 3. The effectiveness of the proposed method is evaluated in Section 4. Finally, the conclusion is summarized in Section 5.

2. PRELIMINARY

In this section, we briefly introduce LSTM which has been widely used in time series prediction. LSTM is an improved method based on RNN which can solve the problem of gradient disappearance in long-term dependencies by introducing three sigmoid gates, forget gate, input gate and output gate.

Given the input sequence $\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_1,...,\mathbf{x}_T)$ with $\mathbf{x}_t \in \mathbb{R}^n$, where n is the dimension of exogenous series. The forget gate $\mathbf{f}_t$, input gate $\mathbf{i}_t$ and output gate $\mathbf{o}_t$ are defined and updated as follows.

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1};\mathbf{x}_t]+\mathbf{b}_f) \tag{1}$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1};\mathbf{x}_t]+\mathbf{b}_i) \tag{2}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1};\mathbf{x}_t]+\mathbf{b}_o) \tag{3}$$

where $[\mathbf{h}_{t-1};\mathbf{x}_t]$ is a concatenation of the previous hidden state $\mathbf{h}_{t-1}$ and the current input $\mathbf{x}_t$. $\mathbf{W}_f,\mathbf{W}_i,\mathbf{W}_o$ and $\mathbf{b}_f,\mathbf{b}_i,\mathbf{b}_o$ are parameters to learn. σ is a logistic sigmoid function.
Then the current cell state and hidden state are calculated as follows

$$\mathbf{s}_t = \mathbf{f}_t \odot \mathbf{s}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_s[\mathbf{h}_{t-1};\mathbf{x}_t]+\mathbf{b}_s) \tag{4}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{s}_t) \tag{5}$$

where $\mathbf{W}_s$, $\mathbf{b}_s$ are parameters to learn, $\odot$ is an elementwise multiplication.

The core concept of LSTM is the state of the cell and the gate structure. Cell state can be seen as the memory of the network. The state of the cell is able to pass on important information in the sequence. Therefore, even information of an earlier time step can be carried into cells of a later time step, which overcomes the effects of short-term memory. The addition and removal of information is achieved through the gate structure, which learns what information to save or forget during the training process.

3. METHODOLOGY

In this section, the details of the proposed dual-stage attention based spatio-temporal sequence learning for multi-step traffic prediction will be introduced. Traffic flow has strong temporal dependences and spatial dependences. More specifically, traffic at historical moments will have an impact on current prediction moment, and the impact of different historical moments is different. At the same time, Traffic on a single road segment is affected by some relevant road segments in the road network, the contribution of each relevant road segment is also different. In addition, some external factors such as weather, holidays also affect the traffic. In response to the above points, we design a model to solve these problems well. The architecture of our method is shown as Fig. 1.

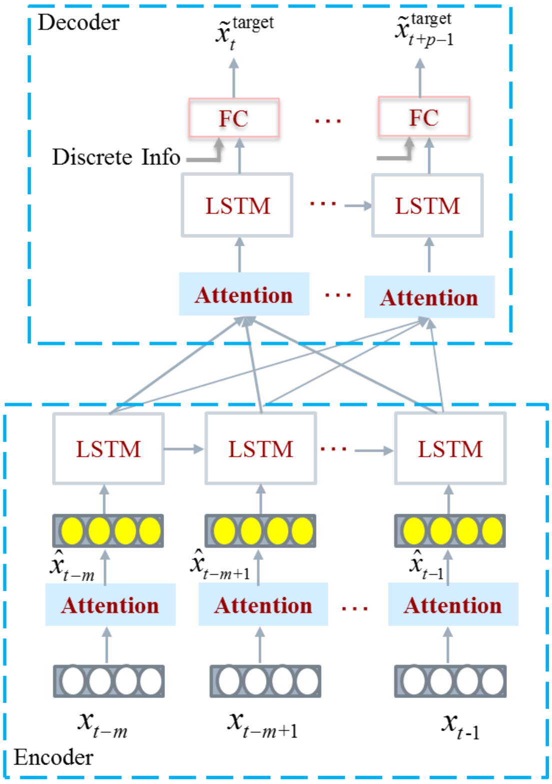

Fig. 1. Architecture of dual-stage attention based spatio-temporal sequence learning for multi-step traffic prediction

The encoder-decoder model is utilized as the base in which both of the encoder part and decoder part are composed of some LSTM units. LSTM can fully extract the timing characteristics of time series data and avoid gradient disappearance in long-term dependency problems. First, the method of spatial region of support is used to filter out the top k relevant road segments of the target road segment. Thus we get a k-dimensional input vector at each moment of m historical intervals. And we add the first-stage attention to learn and **a**utomatically determine the weight of each relevant road segment at each historical moment**.** After the first-stage attention, a new input vector is generated and sent to the LSTM unit in encoder part. In decoder part, for each time interval in multiple prediction intervals, we add the second-stage attention to learn the weights of all the encoder information of hidden states and get a weighted sum of these hidden states as the information of hidden layer for the current prediction moment. Then the output of the current LSTM unit will be combined with external discrete information and sent to a fully connected network. $\tilde{x}_t^{\text{target}}$ is the final prediction result at t moment of the target road segment. Next, we will introduce the encoder part and the first-stage attention and the decoder part and the second-stage attention separately.

3.1 The Encoder Part and First-stage Attention

In encoder part, there are m LSTM units each of which represents one historical interval where m is the size of the historical time window. Define t_p as the first prediction time interval, for the ith historical interval ($i = t_p - m, t_p - m + 1, ..., t_p - 1$), there is a raw input vector $\mathbf{x}_i = [\mathbf{x}_i^1, \mathbf{x}_i^2, ..., \mathbf{x}_i^k]$ generated by the method of spatial region of support. Where k is the dimension which represents the number of related road segments used in our model and each dimension represents the traffic flow or speed of a related road segment at ith historical interval. We do not send the raw input vector to the encoder networks because there are k related road segments which are not equally important to the target road. Instead, we add the first-stage attention called spatial attention to the input vector in order to learn the weights of these related road segments. The spatial attention mechanism computes the attention weights for input series conditioned on the previous hidden state and cell state in the encoder and then feeds the newly weighted input vector into the encoder LSTM unit.

The computation process of ith input moment is as follows. First, we generate the score, which can be represented as follows

$$e_i^j = \mathbf{v}_e^T \tanh(\mathbf{W}_e[\mathbf{h}_{i-1}; \mathbf{s}_{i-1}] + \mathbf{U}_e \mathbf{x}^j) \tag{6}$$

where $j = 1, 2, ..., k$, $\mathbf{h}_{i-1} \in \mathbb{R}^f$ and $\mathbf{s}_{i-1} \in \mathbb{R}^f$ are the previous hidden state and the cell state in the encoder LSTM unit, $\mathbf{x}^j$ is the jth dimension of input series, $\mathbf{v}_e \in \mathbb{R}^m, \mathbf{W}_e \in \mathbb{R}^{m\times 2f}, \mathbf{U}_e \in \mathbb{R}^{m\times m}$ are parameters to learn.

Next we convert the score into a probability distribution by normalization as follows.

$$\alpha_i^j = \exp(e_i^j) / \sum_{g=1}^{k} \exp(e_i^g) \tag{7}$$

Finally we get a newly computed input as follows

$$\hat{\mathbf{x}}_i = [\alpha_i^1 \hat{\mathbf{x}}_i^1, \alpha_i^2 \hat{\mathbf{x}}_i^2, ..., \alpha_i^k \hat{\mathbf{x}}_i^k] \tag{8}$$

In this way, we emphasize the weight of different road segments at different times. Thus our method can express the spatial correlation closer to the reality then others. The LSTM units can be updated as (4) and (5) above and all of the hidden states of encoder LSTM units will be used in decoder part.

3.2 The Decoder Part and Second-stage Attention

Another LSTM based networks is used to predict the result in decoder part. Temporal dependences are significant in traffic prediction because historical traffic information is related to the current moment. Traditional decoder with LSTM units only uses the hidden layer information at the last moment of the encoder, so the information of early moments deteriorates rapidly as the length of the input sequence increases. At the same time, the information of historical moments play a important role to the prediction and different historical moments have different effects on the forecasting moment which should be reflected in our model. To solve the problems, we add the second-stage attention to the decoder part to adaptively select relevant encoder hidden states across all time steps.

The implementation process is as follows. Define t as the current prediction moment, first we generate the score conditioned on the previous hidden state and cell state in the decoder and each hidden state in encoder which can be represented as follows

$$l_t^i = \mathbf{v}_d^T \tanh(\mathbf{W}_d[h'_{t-1}; s'_{t-1}] + \mathbf{U}_d \mathbf{h}_i),\ 1 \le i \le m \tag{9}$$

where h'_{t-1} and s'_{t-1} are respectively the previous hidden state and cell state of the prediction LSTM unit, $\mathbf{h}_i$ is the hidden state of the ith LSTM unit of encoder, m is the number of LSTM units in encoder part, $\mathbf{v}_d, \mathbf{W}_d, \mathbf{U}_d$ are parameters to learn.

The second step is to convert the score into a probability distribution by normalization as follows

$$\beta_t^i = \exp(l_t^i) / \sum_{j=1}^{m} \exp(l_t^j) \tag{10}$$

Then the weighted sum of m encoder hidden states is calculated as follows

$$\mathbf{c}_t = \sum_{i=1}^{m} \beta_t^i \mathbf{h}_i \tag{11}$$

We concatenate $\mathbf{c}_t$ with the previous prediction result y_{t-1} to generate a new input vector for moment t in decoder.

$$\mathbf{x}_t = [\mathbf{c}_t; y_{t-1}] \tag{12}$$

Then we can update the LSTM units in decoder as (4) and (5) mentioned above. In this way, when making prediction at the current time, we not only make full use of the information of historical moments, but also take into account the differences in the influence of different historical moments on the current moment. Thus we fully express the temporal dependences in traffic prediction. At the same time, the network structure can naturally perform multi-step prediction and maintain timing correlation for each prediction step.

3.3 External Factors

In addition to the spatio-temporal dependences, we also take the external factors including weather condition, holidays and weekends into consideration. When predicting the traffic at *t* moment, we concatenate the output of decoder and the discrete vector composed of external factors and feed them to a fully-connected network as follows

$$\hat{y}_t = \sigma(\mathbf{W}_t[y_t; \mathbf{d}_t] + b_t) \tag{13}$$

where y_t is the output of decoder, $\mathbf{d}_t$ is external discrete vector, $\mathbf{W}_t$ and b_t are parameters to learn.

4. EXPERIMENTS AND RESULTS

In this section, experiment is conducted on a real world traffic speed prediction case to evaluate the effectiveness of the proposed method.

4.1 Dataset

The experimental data used in this paper is derived from Hangzhou taxi GPS data. Hangzhou taxis are equipped with GPS, so we can get their location information, including longitude and latitude and instantaneous speed. At the same time, some external factors, such as weather, holidays and other events also affect traffic conditions, so we have also collected them. The details of the data set and processing method are as follows.

Data of 8000 taxis in Hangzhou, China from October 1, 2013 to January 31, 2014 is consisted in this traffic dataset. Each record contains the instantaneous speed, the corresponding latitude and longitude coordinate. The road network in Hangzhou is roughly divided into many road segments. Without loss of generality, we randomly choose 200 road sections for analysis.

The external factors that we consider in this paper include weather and holidays. The weather dataset is obtained from Chinese Weather Report Net, which contains information on precipitation, temperature, and wind speed from October 1, 2013 to January 31, 2014. The weather information is updated once per hour. The weather states are divided into 3 kinds, depending on the precipitation because the pavement humidity has a direct influence on people's driving behaviour. For special days, dates are divided to three categories including working days, weekends and festivals because people have different travel rules on dates of different categories.

4.2 Preprocessing

In order to estimate the speed of the traffic flow, we define the average speed of a road segment over a certain time interval, that is, for each road segment, calculate the average speed (*km/h*) of all GPS records of the road segment during a specific time interval. Due to the large size of the road network but the limited number of taxis, we set the time interval to one hour to ensure at least one record in each time period. For example, for a road segment, the average speed at 9 am on November 1 is the expected speed value for all GPS records from 9 am to 10 am. At the same time we will smooth out the outliers.

4.3 Settings

The data from September 1, 2013 to January 31, 2014 is used to build and evaluate the ST-BLN methodology. We randomly selected 10 road segments to model their traffic speed and analyze the results. We utilize 3000 sets of training data to predict the future traffic speed, while the other 500 sets of data are adopted to test the prediction performance. In our experiment, the prediction step is set to 3, the size of encoder m is set to 24. The dimension of LSTM hidden state is set as 96.

The proposed method is compared with the previous method ST-BLN (Cui and Zhao, 2019), Deep Neural Network (DNN) and Support Vector Regression (SVR) (C.Cortes and V.Vapnik, 1995). For ST-BLN, the number of enhancement nodes is set to 100. For DNN, the depth is set to 3, and every layer has 100 neurons. For SVR, we use RBF kernel. Because the latter three methods can only predict the one-step traffic speed, whose goals are slightly different from our model. So on the testing stage, we treat prior prediction as observations and use them for next prediction.

4.4 Evaluation Metric

We measure our method by Root Mean Square Error (RMSE) as

$$RMSE = \sqrt{(1/z) \times \sum_i (x_i - \hat{x}_i)^2} \tag{14}$$

where $\hat{x}$ and x are the predicted value and ground truth, respectively, z is the number of all predicted values.

4.5 Results and Analysis

Table 1 shows the RMSE of our proposed method and other three methods on 10 randomly selected road segments. It can be observed that the proposed method in this paper outperforms the other three methods on each road segments for each prediction step. The RMSE value of our method is 1.36% to 2.90% lower than other methods, which is a significant improvement of accuracy. Besides, to observe the stability of the proposed method, we calculate the standard deviation (std) of 10 road segments for each prediction step, the results indicate the proposed method is the most stable of all the four methods.

To show results more intuitively, we plot the mean value and standard deviation of RMSE of our method and other three methods on 10 road segments which is shown in Fig. 2. From the figure, we can further verify the validity of the results. In addition, we can see that as the prediction step size increases, the accuracy of our method drops significantly slower than other methods. Our method makes full use of historical time information for each prediction step, adaptively adjusts the weight of each historical moment, and consider the previous prediction value. However, other methods rely too much on the previous prediction value and can not adjust the weight of information at each historical moment which amplifies the

Table 1. RMSE comparing between our method and other methods

Road Seg-ment	Our Method			ST-BLN			DNN			SVR		
	60 (min)	120 (min)	180 (min)	60 (min)	120 (min)	180 (min)	60 (min)	120 (min)	180 (min)	60 (min)	120 (min)	180 (min)
1	2.754	3.154	3.715	2.773	3.451	4.126	3.042	3.478	4.253	3.014	3.501	4.261
2	2.613	3.122	3.601	2.649	3.272	4.028	2.837	3.302	4.029	3.191	3.256	4.004
3	1.729	2.244	2.803	1.753	2.503	3.219	1.763	2.597	3.322	1.774	2.518	3.327
4	2.816	3.412	4.017	2.842	3.615	4.315	2.888	3.665	4.396	2.893	3.671	4.385
5	2.350	2.965	3.456	2.357	3.102	3.902	2.352	3.213	3.917	2.427	3.217	3.900
6	2.083	2.602	3.204	2.109	2.794	3.621	2.251	2.681	3.713	2.532	2.803	3.681
7	2.362	2.919	3.412	2.378	3.067	3.812	2.520	3.124	3.808	2.490	3.121	3.826
8	1.815	2.475	2.982	1.902	2.591	3.205	2.001	2.672	3.212	2.107	2.659	3.204
9	2.798	3.125	3.573	2.903	3.262	3.904	2.924	3.311	3.952	2.910	3.278	3.868
10	1.812	2.361	2.808	1.825	2.552	3.225	1.887	2.578	3.238	1.854	2.564	3.215
Mean	2.313	2.838	3.357	2.349	3.021	3.736	2.447	3.062	3.784	2.519	3.059	3.767
Std	0.431	0.392	0.402	0.435	0.393	0.402	0.467	0.399	0.414	0.491	0.401	0.413

error as the time step increases.

To verify the significant difference of our method comparing with others, we do the pair-t test on the RMSE on 10 road segments of our method and the other five methods for each prediction step with setting the significance level to 0.05. The result fully proves that there is a significant difference between the RMSE of the proposed method and the other three methods.

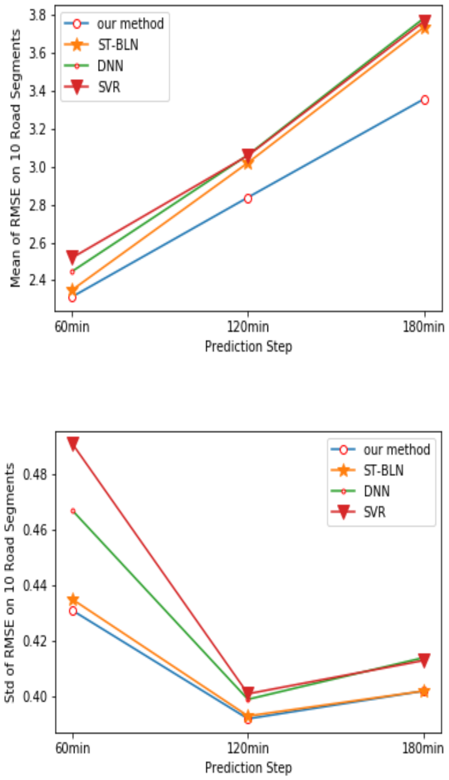

Fig. 2. Mean and std value of RMSE on 10 roads for four methods

5. CONCLUSIONS

In this paper, a dual-stage attention based spatio-temporal sequence learning method is proposed for multi-step traffic prediction. In the proposed method, the differences in contributions to predictions at different historical moments and different road segments are considered, revealing more granular spatial and temporal dependences in traffic prediction. On the basis of encoder decoder architecture, we propose the first attention termed spatial attention to adaptively learn the contribution weights of related road segments. Further, we add the second attention termed temporal attention to the decoder part to adaptively select relevant encoder hidden states across all time steps. In addition, external factors are also taken into consideration. The experiment verifies the effectiveness of the proposed method. Besides traffic speed, the method is also applicable to other scenes such like traffic flow prediction.

ACKNOWLEDGEMENTS

This work was supported by NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (No. U1709211), and Zhejiang Key Research and Development Project (2019C03100 and 2019C01048).

REFERENCES

Box G.E.P, Jenkins G.M., Reinsel G.C. (2010). Time series analysis: forecasting and control [J]. *Journal of Time*, 31(4):303-303.

Bahdanau D, Cho K, Bengio Y. (2014). Neural machine translation by jointly learning to align and translate [J]. *arXiv preprint*, arXiv:1409.0473.

Chunhui Zhao*, Eyal Dassau, Lois Jovanovič, Howard C. Zisser, Francis J. Doyle III, Dale E. Seborg*. (2014). Online Prediction of Subcutaneous Glucose Concentration for Type 1 Diabetes Using Empirical Models and Frequency-Band Separation, *AIChE Journal*. 60(2), 574-584.

Ziqiang Cui, Chunhui Zhao. (2019). Spatio-Temporal Broad Learning Networks for Traffic Speed Prediction, *12th Asian Control Conference (ASCC)*, IEEE, 2019: 1536-1541.

C. Cortes and V. Vapnik. (1995). "Support-vector networks" *Machine Learning*. 20(3):273-297.

J. Zhang, Y. Zheng, et al. (2016). DNN-Based Prediction Model for Spatial-Temporal Data, *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,* ACM, 2016: 92.

J. Zhang, Y. Zheng, et al. (2017). Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction, *Thirty-First AAAI Conference on Artificial Intelligence.*.

Lv Y, Duan Y , Kang W , et al. (2015). Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):865-873.

Lv Z, Xu J, Zheng K, et al. (2018). LC-RNN: A Deep Learning Model for Traffic Speed Prediction[C], *IJCAI*. 2018: 3470-3476.

Liao, Binbing, Jingqing Zhang, et al. (2018). Deep Sequence Learning with Auxiliary Information for Traffic Prediction, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* ACM, 2018: 537-546.

Smith B L, Demetsky M J. (1994). Short-term traffic flow prediction: neural network approach [J]. *Transportation Research Record Journal of the Transportation Research Board*, 1453(1453):98-104.

Shengwei Luo, Chunhui Zhao. (2019). Transfer and Incremental Learning Method for Blood Glucose Prediction of New Subjects with Type 1 Diabetes. *ASCC* 2019: 73-78

W. Hu, L. Yan, K. Liu and H. Wang. (2015). PSO-SVR: A Hybrid Short-term Traffic Flow Forecasting Method, *IEEE International Conference on Parallel and Distributed Systems (ICPADS)*, Melbourne, VIC, pp. 553-561.