Traffic flow prediction based on graph wave adaptive spatiotemporal graph convolution network

Mengxiao Diao, Yansong Qu, Lin Wu and Zhenlong Li*
College of Metropolitan Transportation, Beijing University of Technology, Beijing, 100124, China
*Corresponding author's e-mail: lzl@bjut.edu.cn

ABSTRACT

Graph-based traffic flow predictioning is widely applied in traffic systems, where constructing intricate spatiotemporal correlation models from relevant time series data is imperative for comprehending the dynamics of the traffic system. The extraction of features from graphical data, coupled with the integration of time series data, serves to enhance the accuracy of traffic flow predictions. Additionally, the challenge arises when real traffic flow data often contains missing values. Predicting traffic in scenarios with missing data proves to be challenging, as existing traffic flow predictioning methods frequently lack the capacity to model dynamic spatiotemporal correlations in the presence of such gaps, leading to unsatisfactory prediction outcomes. This paper introduces the Adaptive Spatiotemporal Graph WaveNet-based Graph Convolutional Network (AST-GW-GCN) to tackle traffic flow predictioning. AST-GW-GCN comprises three independent components, each modeling short-term, daily, and weekly dependencies of traffic flow. Within each component, Gated Temporal Convolutional Network (TCN) and Graph Convolutional Network (GCN) serve as encoders, conducting spatial convolution to extract spatial correlations and temporal convolution to capture temporal correlations, thereby generating hidden features. The Gated Recurrent Unit (GRU) is employed to decode these hidden features and weigh the outputs of the three components, producing the final prediction result. The spatial convolution module establishes an adaptive adjacency matrix to overcome the physical constraints of the graph structure, facilitating improved extraction of hidden spatial dependencies within the data. Furthermore, experiments are conducted under various data missing patterns and missing ratios. The experimental findings, based on the PeMS08 real dataset, demonstrate that the proposed AST-GW-GCN comprehensively captures spatiotemporal correlations in the data, outperforming baseline models in terms of performance.

Key words: deep learning, Intelligent Transportation System, Graph convolutional neural network

1. INTRODUCTION

The significant rise in the number of automobiles has contributed to an escalating gap between road supply and demand, consequently leading to congested traffic conditions. Confronting this challenge, many nations have taken a proactive stance in the advancement of Intelligent Transportation Systems (ITS) as a strategic initiative for the effective management and handling of traffic. The development and application of Intelligent Transportation Systems (ITS) have proven to be effective in improving traffic conditions, reducing congestion, and enhancing travel efficiency. Real-time, accurate, and reliable traffic information serves as a fundamental basis for ITS, and this information is often embedded within a vast amount of traffic data. Among these data, traffic flow serves as a fundamental indicator that reflects the state of the road. Accurate prediction of traffic flow allows traffic management authorities to make more informed decisions regarding vehicle dispatch and management, ultimately enhancing the operational efficiency of the road infrastructure.

Road traffic flow prediction is a general prediction problem that relies on real-time spatiotemporal data. With the deployment of data collection equipment such as cameras and sensors on highways, rich traffic time series data sets have been accumulated, providing a solid foundation for traffic flow prediction. Traffic flow prediction is complex, influenced by factors such as temporal variations (time of day, seasonality), spatial considerations (geographical location, intersections), historical data, and external events (accidents, weather). Social aspects, including special events and holidays, along with traffic management measures, public transportation, and economic factors, collectively shape traffic dynamics.Early simple linear time series predicted models were used for traffic flow predictioning. However, these methods require the data to satisfy certain assumptions, which are often not met by the complicatedity and nonlinear nature of traffic data, causing them to perform poorly in practice. Subsequently, in response to the evolving landscape, machine

International Conference on Smart Transportation and City Engineering (STCE 2023),
edited by Miroslava Mikusova, Proc. of SPIE Vol. 13018, 1301817

learning methods were leveraged to model increasingly intricate and complex datasets. This strategic adoption of machine learning techniques reflects a dynamic effort to capture and analyze the nuances present in the growing complexity of data structures. However, their limitations included, on one hand, they could not effectively consider the spatiotemporal relationship of high-dimensional traffic data, and on the other hand, the prior knowledge provided by expert experience greatly affected the model. accuracy. In the contemporary era, the utilization of deep learning methods has not only gained widespread recognition but has also become a focal point of interest among scholars. This increasing attention underscores the significance and potential of deep learning approaches in addressing the complexities inherent in traffic flow prediction. Convolutional neural networks (CNN) can extract spatial features from grid data, combined with graph convolutional networks (GCN) to capture spatiotemporal correlations. Nonetheless, these methods still have room for improvement in capturing spatiotemporal dependencies and dynamic correlations in traffic flow data.

In practical scenarios, real-world traffic data frequently display imperfections, and the task of accurate traffic flow predictioning becomes challenging when faced with data loss or missing information. This inherent data imperfection poses a notable obstacle in achieving precise predictions of traffic flow. Encoder-decoder based traffic prediction models have received little research attention.

In summary, we propose a new model: the Spatiotemporal Graph Convolutional Network based on Graph Wavelet Adaptation. This model facilitates the modeling of spatiotemporal graphs under various data missing patterns. The paper introduces a dynamic dependency matrix, acquired through the learning process of node embeddings, effectively capturing latent spatial dependencies within the data. The contributions can be delineated as follows:

- The use of an adaptive adjacency matrix that does not rely on the actual graph structure is employed to capture spatiotemporal correlations within the data. Specifically, it breaks free from the physical constraints of the traditional real road network graph structure, allowing for a better exploration of hidden spatiotemporal dependencies and dynamic correlations among nodes in real highway traffic scenarios.
- In consideration of the inevitability and ubiquity of data missing in real traffic scenarios, the study investigates prediction results under various data missing patterns. Given the richness of real traffic data, predicting traffic under missing data scenarios presents a more challenging and practical aspect.
- Modeling is carried out for three traffic cycles: recent, daily, and weekly cycles, incorporating temporal dimension convolution within the dependency relationships in different time cycle dimensions.
- The model utilizes an encoder-decoder architecture to effectively capture long-term dependencies in input sequences of vast and complex nonlinear traffic data.

2. RELATED WORK

2.1. Traffic prediction

Model-driven methods primarily involve simulating the instantaneous or steady-state relationships between traffic parameters in specific scenarios. They offer the advantage of modeling complex traffic scenarios, capturing the movement characteristics of individual traffic entities, and understanding the overall operational rules of the traffic system. However, they also lack precision in scenarios with large data volumes, and perform suboptimally in irregular traffic conditions.

In the realm of data-driven approaches, several early methodologies [1-2] have been explored. Though effective, they are still inferior to data-driven methods.

Recurrent neural networks (RNN) including long short-term memory (LSTM)[3] and gated recurrent unit (GRU)[4] can extract temporal correlations effectively. Models that combine RNN and 1D CNN[5] usually achieve better performance. However, considering the complicacy of the traffic system structure, models based on RNN or one-dimensional CNN still have limitations in modeling spatial dependence.

2.2. Graph Neural Network for traffic prediction

The transportation network is a typical graph structure. In order to simultaneously capture spatiotemporal dependencies, by performing convolution on graph-structured data, graph convolution can directly learn the representation information hidden in the graph structure, and can also consider the internal features of the traffic network to extract semantic dependencies.

Diffusion-convolutional neural network[6] takes graph convolution for a process of passing from one node to another node until convergence is stable. Graph WaveNet[7] put forwards an adaptive adjacency matrix as a supplement to the priori adjacency matrices. ST-A-PGCL[8] extracts spatiotemporal correlation with missing values through periodical adaptive graph contrastive learning.

3. METHODOLOGY

3.1. Problem description

Given a traffic network $G=(V,E,W)$, where V denotes road nodes, E denotes edges connecting these nodes, and W represents a weighted adjacency matrix derived representing edge weight information. If $v_i, v_j \in V$ and $(v_i, v_j) \in E$,then W_{ij} is 1, otherwise it is 0. Given X, expressed as

$$X=\{X_{:,0}, X_{:,1}, \dots, X_{:,t}, \dots\} \tag{1}$$

where$X_{:,t}=\{x_{1,t}, x_{2,t}, \dots, x_{i,t,} \dots, x_{N,t}\}^T \in R^{N\times 1}$ is the traffic value of N nodes on the transportation network in the past τ time slice Historical measurement values, the problem becomes predicting t steps of future traffic flow data based on the passing T steps, that is:

$$\{X_{:,t+1}, X_{:,t+2}, \dots, X_{:,t+\tau}\} = \varphi(X_{:,t}, X_{:,t-1}, \dots, X_{:,t-T+1}; \mathrm{G}) \tag{2}$$

Consider the scenario where traffic flow data is gathered q times each day, and the window size is denoted as T_p. Three time series segments, with lengths T_h, T_d and T_w, are respectively selected as inputs for the recent, daily, and weekly periodicities. Here, T_h, T_d and T_w are integer multiples of T_p.

Through the learning mapping function f of the neural network, the missing value $\hat{Y}=f(X_h, X_d, X_w)$ is calculated to make $\hat{Y}$ as close to Y.

By employing the neural network's learning mapping function, denoted as f, the missing value $\hat{Y}$ is computed as $f(X_h, X_d, X_w)$ with the objective of minimizing the difference between $\hat{Y}$ and the true value Y.

3.2. Model framework

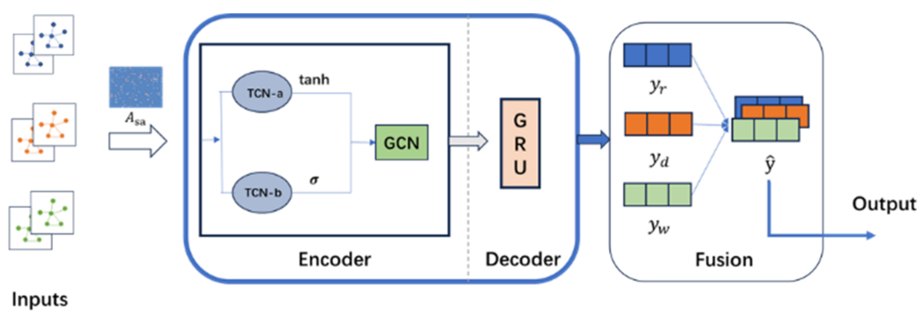

Fig 1. The framework of AST-GW-GCN.

Figure 1 shows the architecture of the AST-GW-GCN with four parts.

3.3. Inputs

3.3.1 Periodic components

The recent period refers to the input that is contiguous to the prediction period. Intuitively, the development and dissipation of traffic congestion occur gradually. Hence, traffic flows in the recent past are bound to influence future traffic flows.

The daily period is the historical time series data calculated in daily units in the past few days that is similar to the prediction period. Traffic data may have repetitive patterns due to people's daily routines, such as the daily morning rush hour.

The weekly period comprises segments from recent weeks that share the same weekly features and time intervals as the prediction period. Hence, the cycle component is devised to capture the cyclic characteristics inherent in traffic data.

3.3.2 Missing patterns

Random missing is the random occurrence of missing values in the data. We randomly replace a certain proportion of the data in the data set with null values to simulate the missing situation.Block missing is a large area of missing values in the data. We randomly replace a certain area of data in the data set with null values to simulate the missing situation.

3.4. Encoders

3.4.1 Spatial Convolution

GCN is essentially a matrix neural network based on a graph structure. Its matrix layer structure contains predefined adjacency matrices and their related feature matrices as input signals, and its equation is expressed as[9]:

$$y_{l+1} = \sigma(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}y_l W_l) \tag{3}$$

where σ is denotes activation function, $\tilde{A} = A + I$ denotes self-loop adjacency matrix, $\widetilde{D}$ is the correspondence matrix, W_l is the weight matrix of the lth convolutional layer, y_{l+1} represents the output and $y_0 = X$.

In Graph WaveNet[7], the adaptive adjacency matrix $\tilde{A}_{ad}$ can discover hidden spatial dependencies on its own. The adaptive adjacency matrix randomly initializes two embedding dictionaries, in which the parameters $E_1, E_2 \in R^{N\times c}$can be automatically learned, and the adaptive matrix is expressed as[7]

$$\tilde{A}_{ad} = SoftMax(ReLU(E_1 E_2^T)) \tag{4}$$

Among them, E_1 is the source node, E_2 is the target node, ReLU is the activation function, and SoftMax function normalizes the adaptive matrix. We utilize the adaptive adjacency matrix to substitute the adjacency matrix in equation (3), and finally The convolutional layer is expressed as

$$y_{l+1} = \sigma(\widetilde{D}^{-\frac{1}{2}}\tilde{A}_{ad}\widetilde{D}^{-\frac{1}{2}}y_l W_l) \tag{5}$$

3.4.2 Temporal Convolution

Gating in recurrent neural networks can effectively get command of the message passing between temporal network layers. Given the input $X \in R^{N\times D\times S}$, its form is[7]

$$h = g(\theta_1 * X + b)\odot\sigma(\theta_2 * X + c) \tag{6}$$

Among them, $\theta_1, \theta_2, b, and\ c$ are model parameters, $\odot$ denotes element product, $g(\cdot)$ and $\sigma(\cdot)$ are activation functions, which determine the information output ratio.

3.5. Decoders

The model employs a GRU[4] as the decoder to unravel the hidden representation from the encoder and subsequently conduct traffic predictioning on the time series. The GRU is illustrated in Figure 2.

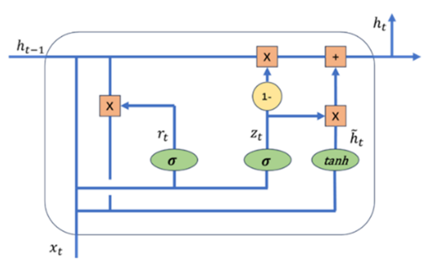

Fig 2. Overall structure of gated recurrent unit

The following is the formula representation of GRU，

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{7}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{8}$$

$$\tilde{h}_t = tanh\ (W \cdot [r_t * h_{t-1}, x_t]) \tag{9}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{10}$$

Among them, r_t is the reset gate, x_t is the current input, h_{t-1} is the previous hidden layer output, z_t is the update gate, $\tilde{h}_t$ is the update value, h_t is the final output.GRU combines the local input and the upper layer output with Sigmoid to obtain the reset gate and update gate.

3.6. Outputs

The three periodic components are dynamically fused to obtain the final output result.The equation is expressed as:

$$H_f = \sum_{i=0}^{N} W_i \times H_i \tag{11}$$

Where H_f is the fused hidden feature, H_i is the candidate feature, N is the number of candidate features, and W_i is the weight parameter.

4. EXPERIMENTS

4.1. Experimental setups

We use the dataset PeMS08 published by Song[10]. The data set is aggregated at 5-minute intervals, with all data collected every 30 seconds. The PeMS08 dataset spans 62 days and contains 170 Nodes and 275 Edges. This research was carried out within the Python 3.7 environment. The dataset was assigned to training, validation, and testing data as a ratio of 7:1:2. All core models and the AST-GW-GCN model were trained under random missing and block missing modes, yielding missing rates of 20%, 40%, and 60%, respectively. The average performance index of the models was then computed.

4.2. Baselines

Evaluate the efficacy of four fundamental models using the PeMS08 dataset under the 20% random missing mode. Utilize a 12-time step historical data to prediction future traffic flow at intervals of 9, 12, 15, 18, 21, 24, 27, and 30 time steps, respectively.. LSTM-NN[6], GRU-NN[7] and GWNet[12] were selected as baseline models.

5. RESULTS

Table 1. In the case of 20% random missing, the comparison results of prediction indicators based on the PeMS08 data set and the baseline model

Model	MAE	RMSE
LSTM-NN	49.3089	69.0101
GRU-NN	48.3451	67.6521
GWNET	31.5187	45.9158
AST-GW-GCN	**19.5140**	**29.9073**

As depicted in Table 1, our model has demonstrated superior performance across diverse predictioning metrics. With the incorporation of the transformer architecture, our model exhibits advancements in comparison to time series predictioning models like LSTM and GRU,The long-term dependency problem can be better solved, and our model performance far exceeds these two baseline models. Compared with the GWNET baseline spatiotemporal prediction model, our model architecture is able to detect the spatiotemporal correlation at each time stage. The Transformer structure introduces a self-attention mechanism to also make the input data more interpretable.

Table 2. Model prediction index results under different data missing modes and proportions

Missing Patterns	MAE	RMSE
BLOCK0.2	21.0241	31.5903
BLOCK0.4	22.2518	33.2242
BLOCK0.6	22.2083	33.9239
RANDOM0.2	**19.5140**	**29.9073**
RANDOM0.4	20.7546	31.7451
RANDOM0.6	23.1397	34.9427

Table 2 contrasts the predictive metrics of the model in the block missing mode and the random missing mode at missing proportions of 20%, 40%, and 60%, respectively. The model achieves optimal predictioning performance under the 20% random missing mode.

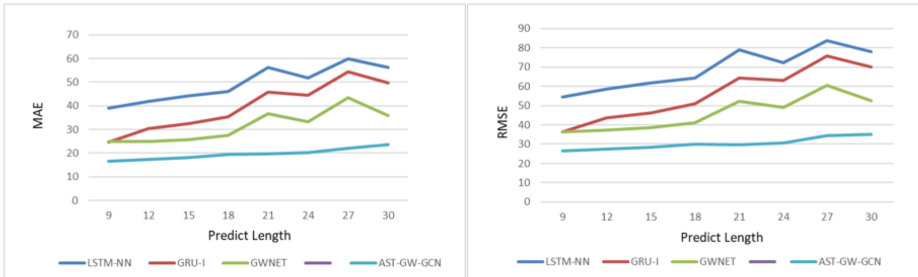

Fig 3. The predictive efficacy of AST-GW-GCN and baseline models varies with the prediction length.

With the extension of the prediction horizon, the baseline model exhibits a gradual decline in predictive performance. Notably, the prediction performance of the AST-GW-GCN model experiences a more pronounced decrease compared to the other baseline models.. This is because the input of the model contains periodic branches that express long-term traffic, and because the use of an encoder-decoder structure allows the model to effectively capture the long-range dependencies of time series data, our model is more efficient than the baseline model. Suitable for long-term series predictioning tasks.

6. CONCLUSION

We introduce an adaptive spatiotemporal graph convolutional network (AST-GW-GCN) founded on the GraphNet[12] framework for the prediction of traffic flow. This model can predict long-term series of traffic flow data without pre-set graph structure. Our model uses an encoder-decoder architecture and captures spatiotemporal dependencies simultaneously. Within the input module, we incorporate near-period, daily period, and weekly period. In the encoder module, the adaptive adjacency matrix enables the autonomous learning of concealed spatial dependencies from the data, all without prior knowledge. The model further conducts comparative experiments with the baseline model across various data missing patterns and proportions, revealing superior prediction performance. In future endeavors, we aim to consider external factors like weather and traffic accidents to enhance performance.

REFERENCES

[1] Williams,B.,Hoel,L.(2003)Modeling and predictioning vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results.J. Transp.Eng., 6:664–672.

[2] Guo,J., Huang,W., Williams,B.(2014) Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification.Transp. Res. Part C Emerg. Technol.,43 :50–64.

[3] Hochreiter,S.,Schmidhuber,J.(1997)Long short-term memory.Neural Comput.,8:1735–1780.

[4] Cho,K.(2014)Learning phrase representations using RNN encoder-decoder for statistical machine translation.Comput.Sci.

[5] Wang,K.,Ma,C.,Qiao,Y.,Lu,X.,Hao,W.,Dong,S.(2021)A hybrid deep learning model with 1DCNN-LSTM-attention networks for short-term traffic flow prediction.Phys. Stat. Mech. Its Appl.,583: 126293.

[6] Atwood,J.,Towsley,D.(2016)Diffusion-convolutional neural networks. The 30th International Conference on Neural Information Processing Systems. USA.pp.2001-2009.

[7] Wu,Z.,Pan,S.,Long,G.,Jiang,J.,Zhang,C.(2019)Graph wavenet for deep spatial-temporal graph modeling. In: The 28th International Joint Conference on Artificial Intelligence.pp.1907–1913.

[8] Qu,Y.,Rong,J.,Li,Z.,Chen,K.(2023) ST-A-PGCL: Spatiotemporal adaptive periodical graph contrastive learning for traffic prediction under real scenarios. Knowledge-Based Systems,272.

[9] Kipf,T.,Welling,M.(2017)Semi-supervised classification with graph convolutional networks.In: ICLR.

[10] Song,C.,Lin,Y.,Guo,S.,Wan,H.(2020)Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data predictioning.In:Proc. AAAI Conf. Artif. Intell.pp.914–921.