

Research Article

Critical Segments Identification for Link Travel Speed Prediction in Urban Road Network

Xiaolei Ru,¹ Xiangdong Xu,^{1,2} Yang Zhou,¹ and Chao Yang ^{1,2}

¹Key Laboratory of Road and Traffic Engineering of the Ministry of Education, School of Transportation Engineering, Tongji University, Shanghai 201804, China

²Urban Mobility Institute, Tongji University, Shanghai 200092, China

Correspondence should be addressed to Chao Yang; tongjiyc@tongji.edu.cn

Received 5 April 2020; Revised 6 August 2020; Accepted 28 August 2020; Published 8 September 2020

Academic Editor: Francesco Galante

Copyright © 2020 Xiaolei Ru et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Predicting traffic operational condition is crucial to urban transportation planning and management. A large variety of algorithms were proposed to improve the prediction accuracy. However, these studies were mainly based on complete data and did not discuss the vulnerability of massive data missing. And applications of these algorithms were in high-cost under the constraints of high quality of traffic data collecting in real-time on the large-scale road networks. This paper aims to deduce the traffic operational conditions of the road network with a small number of critical segments based on taxi GPS data in Xi'an city of China. To identify these critical segments, we assume that the states of floating cars within different road segments are correlative and mutually representative and design a heuristic algorithm utilizing the attention mechanism embedding in the graph neural network (GNN). The results show that the designed model achieves a high accuracy compared to the conventional method using only two critical segments which account for 2.7% in the road networks. The proposed method is cost-efficient which generates the critical segments scheme that reduces the cost of traffic information collection greatly and is more sensible without the demand for extremely high prediction accuracy. Our research has a guiding significance on cost saving of various information acquisition techniques such as route planning of floating car or sensors layout.

1. Introduction

Traffic operational condition, measuring with traffic flow, travel time, and vehicle speed, is an important indicator to reflect the level of service of urban roads network. Travelers design efficient travel plans, including departure time, travel mode, and route, while traffic managers develop strategies to respond to various traffic situations in advance by predicting traffic operational conditions of the road network [1–3]. Thus, traffic state prediction is always a research hotspot. With various available high-quality datasets of traffic information, the vast, elaborate machine learning algorithms [4–8] were applied to deal with this problem and the prediction accuracy was being pushed to a fairly high level.

Although the principles of algorithms are different, their implementations are in the same procedures or scenes, shown in Figures 1(a) and 1(b). In Figure 1(a), traffic state

predictions of each segment are independent [9]. The historical information has an advantage in prediction; for example, the correlation analysis is used to analyze the relevance between the historical traffic flow and the traffic flow within the current interval [10]. In Figure 1(b), traffic state predictions of segments in road network are cooperative [11–13]. The entropy-based grey relation analysis is implemented to choose lane segments that are strongly correlated with the lane segments to be predicted [14]. The convolutional neural network was also used to extract the spatial features [15]. Both of them are based on the hypothesis that the traffic state at different locations of road network has correlation which is related to the spatial distribution [16]. In general, the predictive effect in the second scene is superior to the first. Then, we consider the third scene: predicting the vehicle speed of segments all over the network using available historical information of parts of segments (Figure 1(c)).

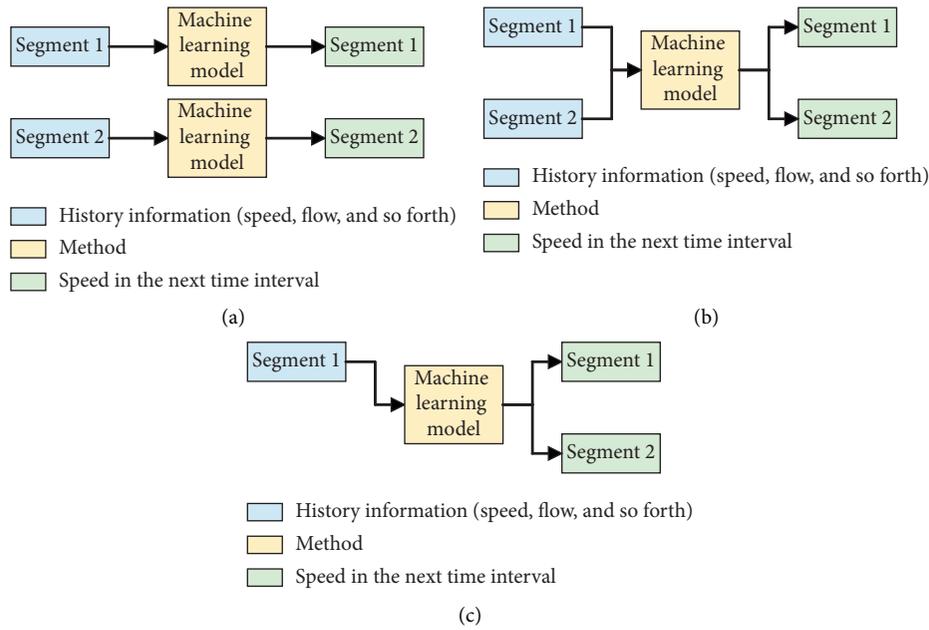


FIGURE 1: The implementation scene of the algorithms for vehicle speed prediction.

We discuss a practical application of the third scene here. We take an experimental road network consisting of seventy-five segments (directional) in Xi'an city of China (see Figure 2(a)). The mean and standard deviation of travel speed of each segment are shown in Figure 2(b). We suppose there are installed sensors in each segment to record the dynamics of the vehicle flow. Theoretically, the algorithms suited for the above first or second scene could accomplish the prediction task well. However, once sensors malfunction causes data missing or errors, those algorithms would not work. They require maintaining all the sensors frequently to guarantee the quality of real-time data. Indeed, the maintenance resource is usually limited because it is hard to ensure that plenty of sensors are fault-free simultaneously and chronically. A sensible alternative scheme is to maintain a small fraction of segments' sensors and establish a prediction algorithm based on the incomplete data only recorded on them. Even in the worst case, only this fraction of sensors is on operation; the prediction accuracy for the whole road network still meets the demand. Then, a challenge different from the pure issues of speed prediction raises: how to identify the critical segments that determine understanding the traffic state of the whole road network?

In order to solve this question, we introduced the graph neural network (GNN), an extended deep learning model to deal with graph data [17]. It is well-performed in finding the complex relation information among elements [18–20]. Combined with attention mechanism [21], we construct an GNN-based machine learning model that takes the historical traffic information of critical segments as input and predicts the link travel speed for each segment in the next time interval. The attention mechanism quantifies the contribution of each segment's traffic information to travel speed prediction of each link in the road network when the model achieves the downstream objective of minimizing the

prediction error. We take advantage of the quantitative contribution to design a heuristic algorithm, which removes the segment with minimal self-attention coefficient as the most trivial one iteratively. The remaining segments are finally identified as the critical segments. The results show that the model performance using the traffic data of only two critical segments is beyond the conventional method using historical average in the experimental road network. The application of the proposed method can reduce the amount of traffic information that needs to be collected significantly at the expense of a slight loss in prediction accuracy. Here, we introduce some related research [22–24]. An approach was proposed to exploit the spatial-temporal causality among travel speeds of road segments by a time-lagged correlation coefficient function and utilize the local stationarity of correlation coefficient to estimate the travel speeds of road segments to handle the problem of missing travel speed values of vehicles on some road segments, due to the coarseness of vehicular crowdsensing data [22]. However, the objective of our research is to reduce the data demand initiatively by identifying critical segments while these previous researches aimed to reduce the negative impacts of data missing passively.

In summary, our main contributions are as follows:

- (1) According to the application restriction, traffic information is expected to collect on segments as less as possible in order to reduce acquisition cost; we put forward a new research issue: how to identify the critical segments that contribute to guaranteeing traffic state prediction accuracy for all segments in road network in the most effective way.
- (2) We propose a heuristic algorithm to select segments of which the missing traffic information is hard to be remedied from others as the critical segments by attention mechanism.

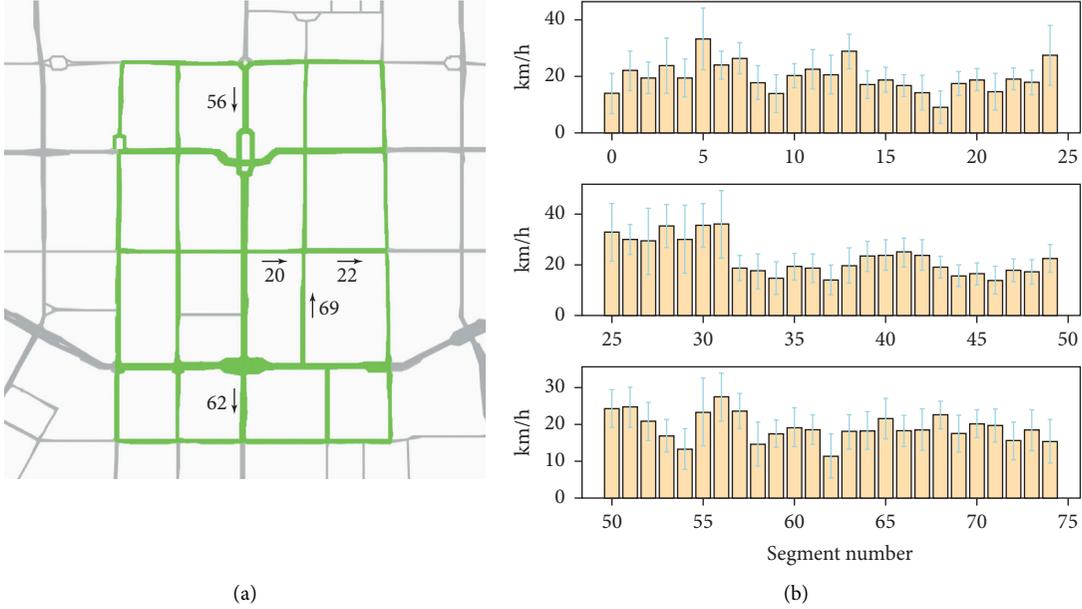


FIGURE 2: Study area and speed characteristics of each segment.

- (3) We make an experimental study that prediction with the data of 2.7% of segments can meet the accuracy demands. The critical segments schemes are highly cost-efficient and provide a cost saving thought for various information acquisition techniques.

1.1. Problem Formation. We describe the question concisely as follows:

$$\min l = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (V_{ij} - \hat{V}_{ij})^2, \quad (1)$$

$$\sum_{j=1}^N x_j \leq p, \quad (2)$$

$$x_j \in \{0, 1\} \quad \forall j = 1, \dots, N, \quad (3)$$

where V_{ij} is true vehicle speed of segment j in training sample i , $\hat{\mathbf{V}}$ is a vector of the prediction speed, x_j is a decision variable of j -th segment and $x_j = 1$ if it is selected, and p represents the cost limit. Objective function (1) minimizes the mean square error between true speed and prediction speed on all the segments:

$$\hat{V}_{ij} = f(\mathbf{d}_i), \quad (4)$$

$$\mathbf{d}_i = \left\{ d_{ij} \mid x_j = 1 \right\}, \quad (5)$$

where \mathbf{d} is the history information of vehicle dynamic and $f(\cdot)$ represents the complicated prediction relation from history information to the future vehicle speed. Equation (5)

indicates the history information collected only from the selected critical segments.

Then, the question could divide into two subproblems. (a) Decision variables assignment: the contribution of each single segment to vehicle speed prediction for road network is heterogeneous and influenced by road network topology and traffic assignment. Critical segments identification is a combinatorial optimization problem. We have to design a heuristics algorithm for this NP-hard problem. (b) Prediction relation $f(\cdot)$ establishment by machine learning model: it is a typical nonlinear regression problem.

2. Method

2.1. GNN-Based Machine Learning Model for Vehicle Speed Prediction. Among various GNN variations, the graph attention network [25] causes our attention. We exploit the self-attention mechanism to explore the contribution of each segment to predict vehicle speed for the whole network. Combined with our problem, a single graph attentional layer is described as follows:

$$\mathbf{h}_v^{l+1} = \sigma \left(\sum_{u \in \mathbf{N}(v)} a_{uv}^l \cdot \mathbf{W}^l \mathbf{h}_u^l + b_v^l \right), \quad (6)$$

where \mathbf{h}_v^l is the hidden feature vector of segment v in l -th layer. Initially, \mathbf{h}_v^0 is the history information recorded on segment v while $\mathbf{h}_v^{\text{last}}$ is the ultimate predicted speed of segment v . $\mathbf{N}(v)$ is the set of neighbor segments of segment v . $\mathbf{W} \in R^{d' \times d}$ is a learning weight matrix sharing by each \mathbf{h}_u . d'/d equals dimensionality of feature vector \mathbf{h}_v in the next layer and the current layer, respectively. σ is the activated function and $a_{uv} \in [0, 1]$ is the attention coefficient indicating the contribution of history information on segment u to predict vehicle speed on segment v . a_{uv} is calculated by attention mechanism as follows:

$$e_{uv} = g([\mathbf{U}\mathbf{h}_u, \mathbf{U}\mathbf{h}_v]), \quad (7)$$

$$a_{uv} = \text{softmax}_u(a_{uv}) = \frac{\exp(e_{uv})}{\sum_{u \in N(v)} \exp(e_{kv})}, \quad (8)$$

where $g(\cdot)$ is independent feedforward network sharing by any pair of \mathbf{h}_u and \mathbf{h}_v and quantifies the importance of \mathbf{h}_u to \mathbf{h}_v . \mathbf{U} is another learnable weight matrix, which transforms the hidden feature vectors into higher-level features before feeding to $g(\cdot)$. And $[\cdot]$ is the concatenation operation.

Here, we consider the neighbor segments $N(v)$. In graph theory, neighbor means the node linked to the current node directly. On the road network, segments' layout is constrained with geographic location and no explicit links, so the spatially close segments look like neighbors, such as segments 20/22/69 in Figure 2. However, can we assert that segment 56 has no correlation with segment 62 even when they are on opposite sides of the network? The answer is certainly not. The drastically increased vehicle flow on segment 56 may give rise to congestion on segment 62 in the next time interval. The road network is a complex system, which not only is an underlay topology structure but also carries the traffic dynamic. Thus, we take the road network as a full-connected network in which any pair of segments has a link and can be put into the machine learning model. The correlation strength on links is quantified by the attention mechanism intelligently.

2.2. Attention-Based Greedy Algorithm for Critical Segments Identification. The previous section solves the second sub-problem in Section "Problem Formation," which is the establishment of the prediction relation $f(\cdot)$ by machine learning model. For critical segments identification, we design a heuristics algorithm based on the accessory in $f(\cdot)$ -attention coefficient a_{uv} . After the model completes training, a_{uv} for each pair of segments is calculated on the test set. An indicator needs to be designed to heuristically decide which segment's data is abandoned or retained in each step.

Many studies about centrality of nodes in complex systems field [26, 27] indicated that the effect of node set was not simple sort combination of each node's effect. The effects for vehicle speed prediction among history information of segments were redundant and replaceable. Then, what is irreplaceable? A segment v , with higher self-attention a_{vv} , means the speed prediction on it is mainly dependent on history information itself. The vehicle dynamic on this segment is relatively independent of the road network more than the segments with lower self-attention. If the data of this segment is missing, it is hard to extract useful features from other segments' data for prediction.

According to this clue, we give out a greedy algorithm to generate critical segments' scheme iteratively. In each step, segment v with the lowest self-attention a_{vv} (see the red symbols in Table 1) is removed in a greedy way:

TABLE 1: Attention coefficients between the remaining segments.

a_{uv}	Segment 1	Segment 2	...	Segment N	c_v
Segment 1	a_{11}	a_{12}	...	a_{1N}	$\sum_{u=1}^N a_{1u}$
Segment 2	a_{21}	a_{22}	...	a_{2N}	$\sum_{u=1}^N a_{2u}$
...
Segment N	a_{N1}	a_{N2}	...	a_{NN}	$\sum_{u=1}^N a_{Nu}$

$$v = \arg \min_v \{a_{vv} \mid v \in N_r\}, \quad (9)$$

where N_r is the remaining segments with history information. And the machine learning model is retrained to renew $\{a_{vv}\}$ per iteration. The iteration will stop until the number of the remaining segments drops to cost limit p , as illustrated in Algorithm 1.

Equation (6) is specialized for the remaining segments, as GNN blocks in Figure 3. By removing segments not having history information, the predicted value of the vehicle speed is utilized in the hidden features of the remaining segments generated in GNN blocks, calculated as follows (linear blocks in Figure 3):

$$\mathbf{h}_v = \sum_{u \in N_r} \theta_{uv} \cdot \mathbf{h}_u + b_v, \quad (10)$$

where θ_{uv} is a learning weight (regression coefficient). The difference between θ_{uv} and a_{uv} is that θ_{uv} is a constant while a_{uv} is changed per sample in different time intervals.

3. Experiment and Result

3.1. Data and Machine Learning Model Configuration. The data we use is vehicle trajectories within the periods of ride-hailing orders in the second ring road area of Xi'an city. The data is from DiDi platform and spans from 10/01/2016 to 11/29/2016. The GPS points in the dataset cover the whole road network in Figure 2 and are processed by routing to ensure that the data can correspond to the actual road information. The collecting interval of GPS points is 2–4s. The main fields in the dataset contain the driver ID, order ID, timestamp, longitude, and latitude. After data preprocessing, the average vehicle speed is obtained on each segment per 5 minutes between 6:00 AM and 10:00 PM. We take the vehicle speed and flow volume in the previous two hours as historical traffic information to predict the vehicle speed in the next 5 minutes. The training set is the data of the former 48 days (10/01/2016–11/17/2016) and the test set for evaluation is the data of the latter 12 days (11/18/2016–11/29/2016).

The detailed structure of the proposed model is shown in Figure 4. A single GNN block consists of two graph attentional layers using Leaky Rule activation and one Conv1D layer using linear activation. The number of neurons is 32,16,1 in each layer, respectively. The attention coefficients calculated in the first graph attentional layer are adopted in the greedy algorithm. A feedforward neural network consisting of three Conv1D layers using Leaky Rule activation and one SoftMax layer is incorporated into a single graph attentional layer to compute the attention coefficient. The number of neurons is 16,16,1 in each

- (1) **Input:** whole segments set \mathbf{N} , history information on all segments \mathbf{d} , true vehicle speed on all segments \mathbf{V} , remaining segments set \mathbf{N}_r , history information on remaining segments \mathbf{d}_r .
- (2) Initialize $\mathbf{N}_r \leftarrow \mathbf{N}$, $\mathbf{d}_r \leftarrow \mathbf{d}$
- (3) while $|\mathbf{N}_r| > p$ do
- (4) train the machine learning model $f(\cdot)$ taking \mathbf{d}_r as samples and \mathbf{V} as labels.
- (5) get the self-attention coefficient $\{a_{vv} \mid v \in \mathbf{N}_r\}$ computed in $f(\cdot)$
- (6) remove history information on segment v with lowest a_{vv}
- (7) $\mathbf{N}_r \leftarrow \mathbf{N}_r - v$
- (8) $\mathbf{d}_r \leftarrow \mathbf{d}_r - \mathbf{d}_v$
- (9) End while
- (10) return \mathbf{N}_r

ALGORITHM 1: Critical segments identification for vehicle prediction.

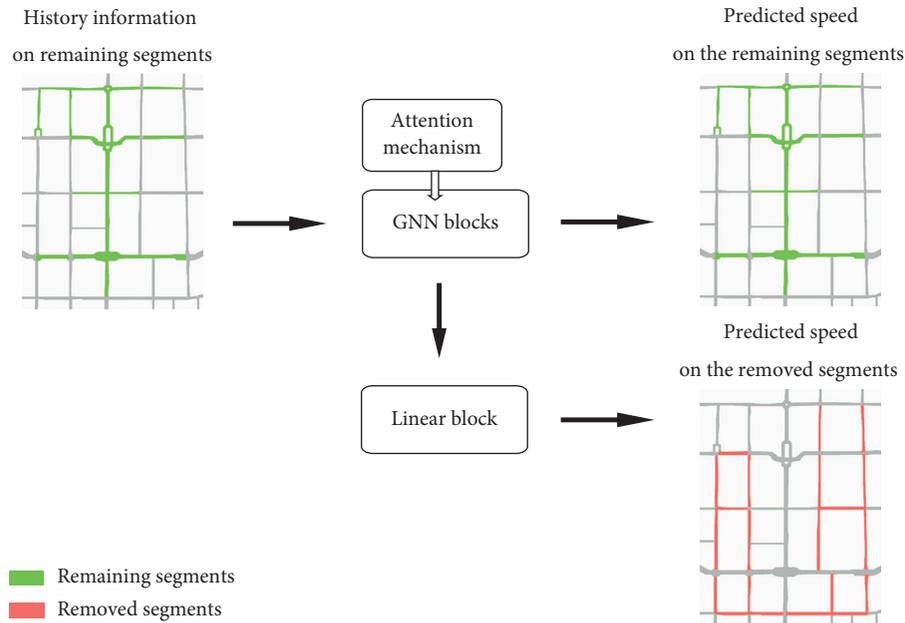


FIGURE 3: A graphical illustration of the proposed prediction model.

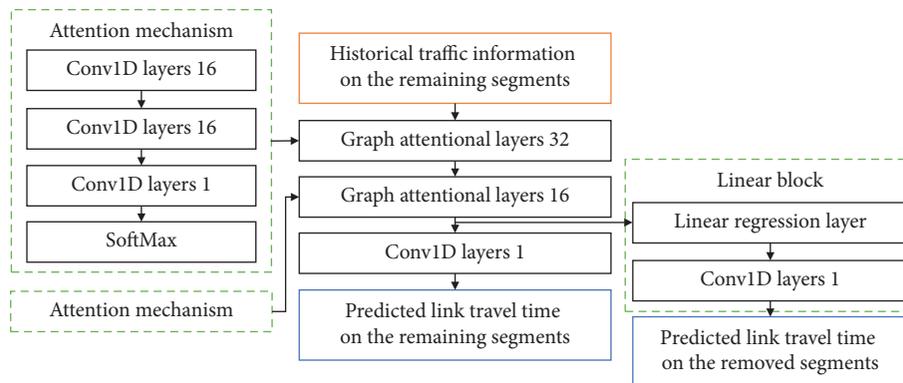


FIGURE 4: Details of the model configuration.

Conv1D layer, respectively. The hidden features outputting from the second graph attentional layer are taken as the inputs of the Linear Block. The Linear Block consists of one linear regression layer and one Conv1D layer using the linear activation. The number of regression coefficients in

the linear regression layer equals the number of the remaining segments multiplying the number of the removed segments. The neural network is trained by the method of Adam optimizer with the batch size of 256 and the learning rate of $10e-3$.

3.2. Result of Critical Segments Identification for Vehicle Speed Prediction. The results are shown in Figure 5(a). The green circles represent the average prediction accuracy of the road network by giving a certain number of critical segments selected by the greedy algorithm I. By contrast, we design the greedy algorithm II which removes the data of the segment with minimal contribution C_v per iteration:

$$v = \arg \min_v \{C_v \mid v \in N_r\}, \quad (11)$$

where the contribution C_v of the segment v is the sum of its contribution to other segments (see the green symbols in Table 1):

$$C_v = \sum_{u=1}^N a_{vu}^i. \quad (12)$$

Greedy algorithm II retains those suffering from the most attention from other segments as critical segments and is a more intuitive solution. The results are shown by yellow circles in Figure 5(a). Besides, we subjectively set a lower limit (red horizontal line in Figure 5(a) as a reference by the conventional method using historical average, which takes the average vehicle speed in the same time interval of history days (weekdays and weekends are distinguished) as the prediction values for each segment.

We observe that (a) the accuracy of our GNN-based machine learning model with complete data of all 75 segments increases by nearly 12% compared to the conventional methods (the leftmost green circle in Figure 5(a)); (b) the greedy algorithm I generates a scheme (called scheme I) containing only two critical segments in which the prediction accuracy by incomplete data is still beyond the low limit (the second green point on the right in Figure 5(a)); (c) the scheme (called scheme II) to meet this demand needs six critical segments generated by the greedy algorithm II (the sixth yellow circle on the right in Figure 5(a)); and (d) with the number of the selected critical segments reducing, the prediction accuracy descends. In each iteration, greedy algorithm I is superior to greedy algorithm II.

The prediction errors of the proposed model on all the segments are shown in Figure 5(b). The value on x -axis is the number of segments in road network. The blue bars are generated using incomplete data only containing history information on critical segments of scheme I. The orange bars are generated using complete data of the whole road network.

3.3. Interpretability of Critical Segments from Traffic Perspective. In order to give an insight into the characteristic of the critical segments, we visualize scheme I and scheme II on road network, as shown in Figure 6. We find both scheme I and scheme II prefer selecting the segments on the margin of road network. These particular segments perceive the external vehicle flow entering the network and the internal vehicle flow leaving network sensitively. Supervising the flow information on them is convenient to estimate the total volume of vehicle flow as well as the level of congestion in road network.

Segments 24 and 29 in east-west direction roads are the unique express way in this road network. Since the import and export of express way are controlled by flyover crossing, the dynamic of vehicle speed on it has a strong continuity at time sequence and is disturbed less by flow afflux from other segments. This result accords well with the logic of greedy algorithm I that selects critical segments with maximal self-attention. The right-turn lane on segment 29 with heavy traffic is a main link between the express way and road network. Scheme I also indicates that the current vehicle dynamic on these two critical segments is the worthiest to be paid attention to if we want to foresee the traffic situation of the whole road network in advance.

3.4. Interpretability of Critical Segments from Machine Learning Perspective. In order to analyze why scheme I is an efficient design (because scheme I needs less number of critical segments than scheme II to meet the prediction demand), we examine the representation of each segment learned by machine learning model, using a technique developed for visualization of high-dimensional features called t-Distributed Stochastic Neighbor Embedding (t-SNE) (see Figure 7). Specifically, two-dimension embedding is generated from the hidden features outputting from the first layer of GNN blocks by running t-SNE algorithm, which tends to map the representation of perceptually similar states to nearby points [28]. In other words, the nodes located closely in subplots of Figure 7 mean that the hidden features extracted from the history information recorded on these segments are highly similar. Two nodes representing two critical segments selected in scheme I are far apart in both morning peak hours (Figure 7(a)) and evening peak hours (Figure 7(c)). Their features are uncorrelated so that they have adequate ability to express cooperatively other nodes' features. Contrastively, the part of the six nodes selected in scheme II is distributed intensively especially in the evening peak, as shown in Figure 7(d), where seventy-five nodes are clustered into six categories by k-means and five critical segments selected in scheme II fall into the same category (green) indicating the configuration of scheme II is highly redundant. That is the reason why the number of the selected segments in scheme II is triple that of scheme I but the performance of scheme II is not better than scheme I. We conclude that it must be an excellent scheme where the hidden features embedding of critical segments belong to different categories in different time intervals.

3.5. Relation between Prediction Accuracy Improvement and the Number of Critical Segments. Apart from meeting the needs of prediction accuracy, we also consider the efficiency meaning the degree to which equivalent history information is converted into prediction accuracy. A representative case is demonstrated in Figure 8. We use greedy algorithm I to generate three schemes containing one/two/three critical segments, respectively. Obviously, the second scheme containing two critical segments is the same as the aforementioned scheme I. Then, the first and second ones are taken as a control group and the second

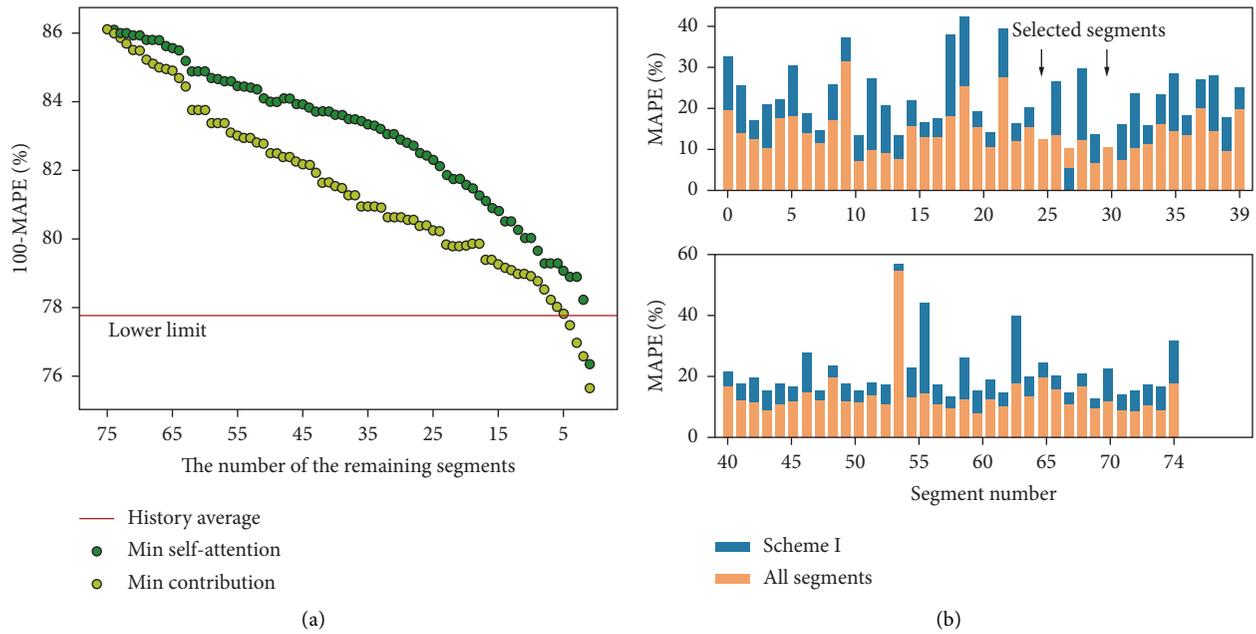


FIGURE 5: Results of critical segments identification.

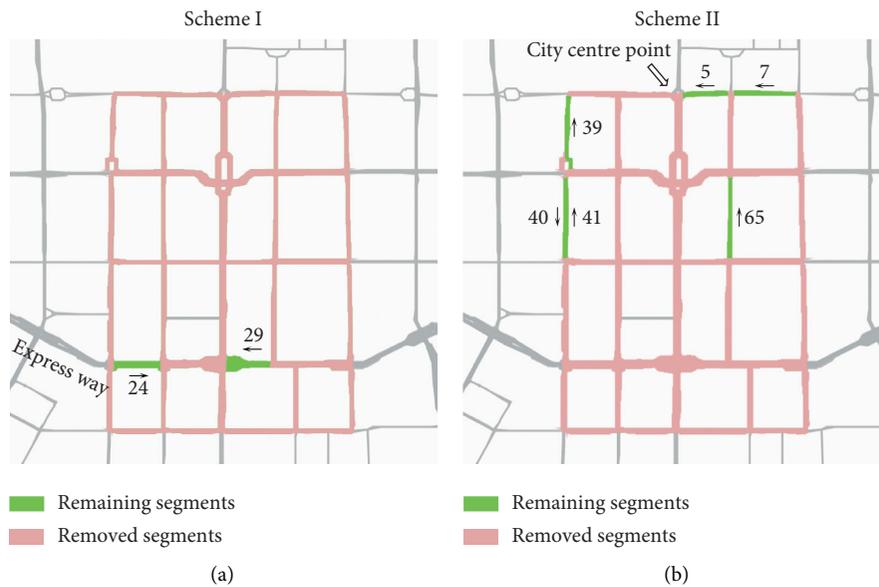


FIGURE 6: Geographic visualization of critical segments identification schemes.

and third ones are taken as another control group. With the number of critical segments rising from one to two, prediction accuracy on eighteen segments markedly improves (see Figure 8(a), highlighted by color orange) while the improvement of prediction accuracy on only one segment is beyond 2% (Figure 8(b)) when the number continues to increase to three. The digit beside segments in the figure is the reduction of prediction error (MAPE). Comparing Figure 8(a) with Figure 8(b), we find the growth of prediction accuracy is significantly different though the increment of the numbers of critical segments is the same. In other words, the average benefit for vehicle

speed prediction bringing by unit amount of history information, regarded as one segment, is different in various schemes.

3.6. Relation between the Prediction Accuracy Improvement and the Number of Critical Segments. Collecting vehicle flow information of segments needs cost, no matter by means of floating cars and sensors. In the practical application we discuss in the introduction, the cost may be maintenance cost or production cost of sensors. While the prediction accuracy improves along with the number of critical

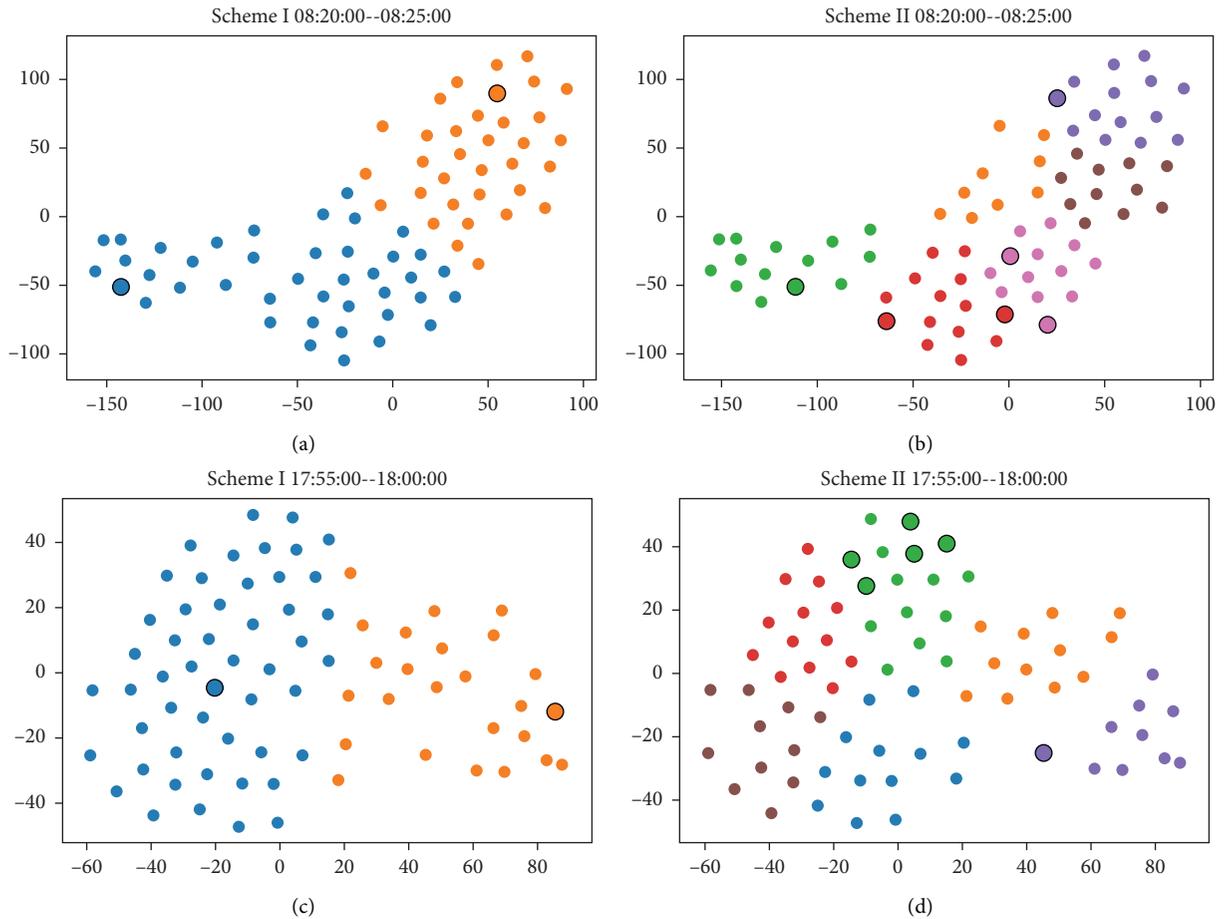


FIGURE 7: Feature visualization of schemes (critical segments are highlighted by a black circle).

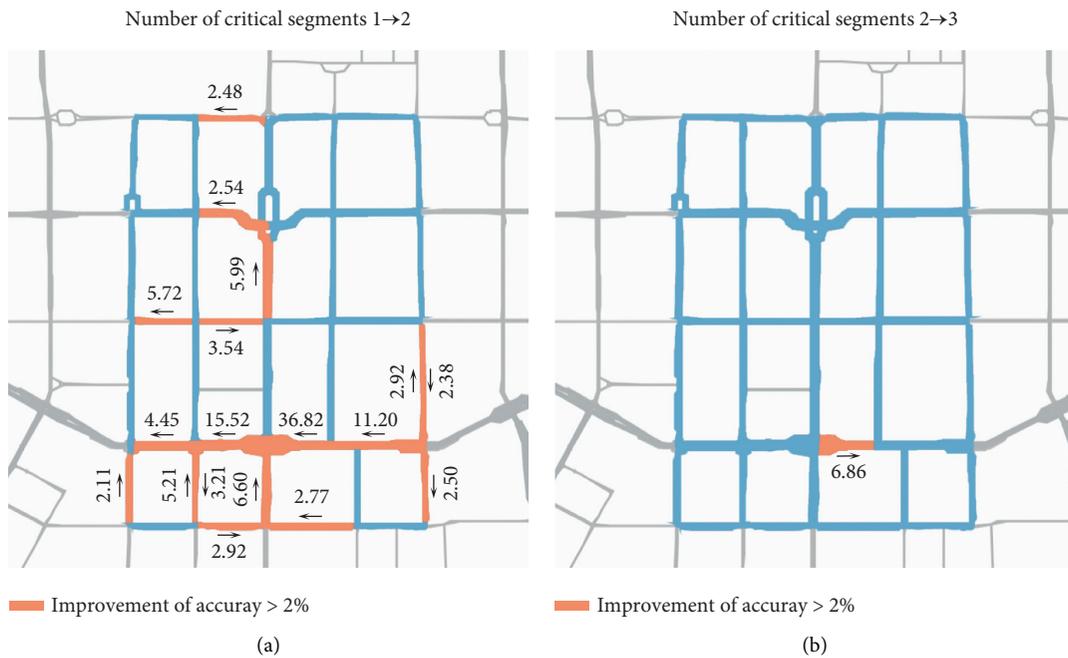


FIGURE 8: Accuracy improvement by progressively increasing the number of critical segments.

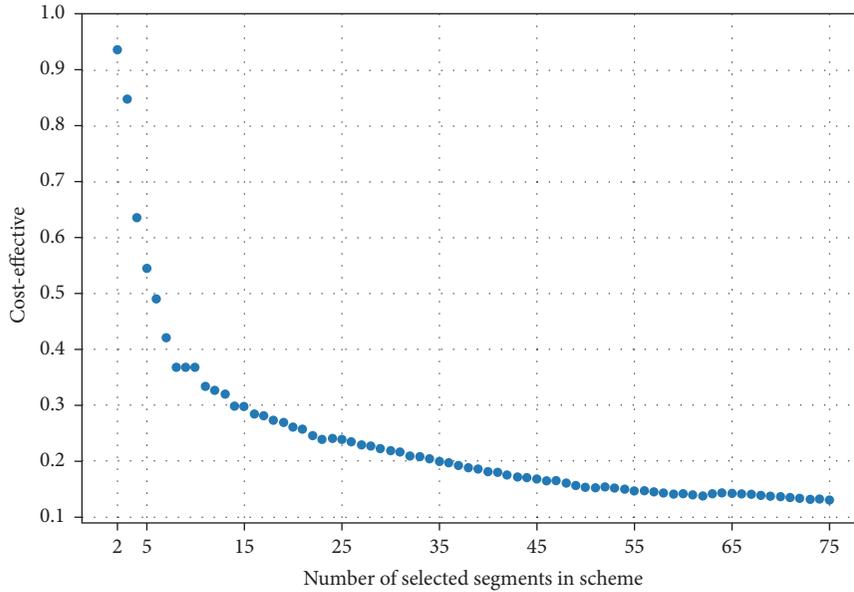


FIGURE 9: Cost-efficiency of the scheme.

segments, the scheme cost also increases. Assuming information collection for one segment is a unit cost, the cost-efficiency measuring the balance of performance and cost for schemes is quantified as follows:

$$c = \frac{P - P_b}{S}, \quad (13)$$

where P and P_b are prediction accuracy of the current scheme and the benchmark scheme containing only one critical segment, respectively, and S is the number of selected critical segments in the current scheme. The cost-efficiency of the schemes generated by greedy algorithm I present the downward trend as a whole with the increase of S . As shown in Figure 9, the curve descends rapidly before S reaches 8 and then gradually flattens. Comparing Figure 5(a) with Figure 9, we consider that the schemes containing a smaller number of critical segments are more advisable if no requirement of extremely high prediction accuracy exists. The cost-efficiency could be a reference index to aid in decision-making besides maximizing prediction accuracy as much as possible.

4. Discussions and Conclusions

The aim of our research is to identify a small number of the critical segments to reduce the collection amount of the traffic information significantly with the permission of slight loss in prediction accuracy, rather than blindly pursuing the extremely high prediction accuracy. We draw the following conclusions:

- (1) In experimental road network, the average prediction accuracy of travel speed for all the segments by the prediction model using the historical traffic information collected from only 2.7% of critical

segments is superior to the conventional method using historical average. The proposed greedy method could identify the critical segments efficiently to understand the traffic state of the whole road network.

- (2) Using the visualization technique of high-dimensional features t-SNE, we know that a scheme of critical segments is optimal if the distribution of the two-dimension embeddings generated from the information features of the critical segments is dispersive, indicating that the traffic information of the critical segments is not redundant.
- (3) The cost-efficiency of information acquisition meaning the efficiency of the equivalent traffic information for the improvement of prediction accuracy is continuously declining with richer and richer information. The traffic information acquisition should consider the acquisition cost and prediction accuracy requirements of traffic state comprehensively.
- (4) The results provide a thought for cost saving of information acquisition techniques. For example, since the traffic flow information of only a small number of critical segments needs to be recorded, the trip distance of the floating cars and the installation or maintenance number of sensors can be cut down dramatically.

Our research can be improved from two directions in the future. Firstly, the elaborate prediction models are designed further to establish a more precise relation between the history traffic information and the predicted vehicle speed, such as the existence of research of hybrids model where GNN and recurrent neural network combine [29] and dynamically modeling spatial dependencies of traffic flows

[30]. Secondly, the critical segments identification methods are further designed to find a more optimal combination of critical segments. Both of them attempt to make the curve in Figure 5(a) decline slowly.

Data Availability

The data supporting the results of our study can be found at <https://outreach.didichuxing.com/research/opendata/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] T. Litman, "Measuring transportation: traffic, mobility and accessibility," *Institute of Transportation Engineers Journal*, vol. 73, no. 10, pp. 28–32, 2003.
- [2] C. Yang, X. Ru, and B. Hu, "Route temporal-spatial information based residual neural networks for bus arrival time prediction," *Journal of Harbin Institute of Technology*, pp. 1–10, 2019.
- [3] Y. Zhang, W. Li, H. Deng, and Y. Li, "Evaluation of public transport-based accessibility to health facilities considering spatial heterogeneity," *Journal of Advanced Transportation*, vol. 2020, Article ID 7645153, 2020.
- [4] X. Yang, Y. Zou, and J. Tang, "Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models," *Journal of Advanced Transportation*, vol. 2020, Article ID 9628957, 2020.
- [5] T. Ma and Y. Pigné, "Bayesian dynamic linear model with adaptive parameter estimation for short-term travel speed prediction," *Journal of Advanced Transportation*, vol. 2019, Article ID 5314520, 2019.
- [6] X. Luo, D. Li, Y. Yang, and S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *Journal of Advanced Transportation*, vol. 2019, Article ID 4145353, 2019.
- [7] Y. Chen, Y. Chen, and B. Yu, "Speed distribution prediction of freight vehicles on mountainous freeway using deep learning methods," *Journal of Advanced Transportation*, vol. 2020, Article ID 8953182, 2020.
- [8] L. Li, L. Qin, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowledge-Based Systems*, vol. 172, pp. 1–14, 2019.
- [9] X. Ma, Y. Z. Tao, and Y. H. WangYu, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [10] Y. Gu, W. Lu, X. Xu, L. Qin, Z. Shao, and H. Zhang, "An improved bayesian combination model for short-term traffic prediction with deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1332–1342.
- [11] Y. Li, R. Yu, and C. Shahabi, "Diffusion convolutional recurrent neural network: data-driven traffic forecasting," 2017, <https://arxiv.org/abs/1707.01926>.
- [12] G. Boquet, A. Morell, and J. Serrano, "A variational autoencoder solution for road traffic forecasting systems: missing data imputation, dimension reduction, model selection and anomaly detection," *Transportation Research Part C: Emerging Technologies*, vol. 115, Article ID 102622, 2020.
- [13] Z. Cui, K. Henrickson, R. Ke, Z. Pu, and Y. Wang, "High-order graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting," 2018, <https://arxiv.org/abs/1802.07007>.
- [14] Y. Gu, W. Lu, L. Qin, M. Li, and Z. Shao, "Short-term prediction of lane-level traffic speeds: a fusion deep learning model," *Transportation Research Part C: Emerging Technologies*, vol. 106, pp. 1–16, 2019.
- [15] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 166–180, 2018.
- [16] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceeding of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 1655–1661, San Fransico, CA, USA, February 2017.
- [17] A. Santoro, D. Raposo, and D. G. T. Barrett, "A simple neural network module for relational reasoning," in *Proceedings of the Conference and Workshop on Neural Information Processing Systems*, Long Beach, CA, USA, December 2017.
- [18] P. W. Battaglia, J. B. Hamrick, and V. Bapst, "Relational inductive biases, deep learning, and graph networks," 2018, <https://arxiv.org/abs/1806.01261>.
- [19] T. Kipf, E. Fetaya, and C. Wang, "Neural relational inference for interacting systems," 2018, <https://arxiv.org/abs/1802.04687>.
- [20] P. W. Battaglia, R. Pascanu, and M. Lai, "Interaction networks for learning about objects, relations and physics," in *Proceedings of the Conference and Workshop on Neural Information Processing Systems*, Barcelona, Spain, December 2016.
- [21] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 5998–6008, Long Beach, CA, USA, December 2017.
- [22] C. Wang, Z. Xie, L. Shao, Z. Zhang, and M. Zhou, "Estimating travel speed of a road section through sparse crowdsensing data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3486–3495, 2019.
- [23] M. Zhou, H. Dong, P. A. Ioannou, Y. Zhao, and F.-Y. Wang, "Guided crowd evacuation: approaches and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 5, pp. 1081–1094, 2019.
- [24] J. Zeng, S. Yu, Y. Qian, and X. Feng, "Expressway traffic flow model study based on different traffic rules," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 6, pp. 1099–1103, 2018.
- [25] P. Veličković, G. Cucurull, and A. Casanova, "Graph attention networks," 2017, <https://arxiv.org/abs/1710.10903>.
- [26] A. F. Siegenfeld and Y. Bar-Yam, "An introduction to complex systems science and its applications," 2019, <https://arxiv.org/abs/1912.05088>.
- [27] M. Kitsak, S. Gallos, F. L. Liljeros, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.
- [28] L. Muchnik and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [29] C. Hyun-Ho and Y. Byoung-Jo, "High-speed data-driven methodology for real-time traffic flow predictions: practical applications of ITS," *Journal of Advanced Transportation*, vol. 2018, Article ID 5728042, 2018.
- [30] M. Xu, W. Dai, and C. Liu, "Spatial-temporal transformer networks for traffic flow forecasting," 2020, <https://arxiv.org/abs/2001.02908>.