TA-STAN: A Deep Spatial-Temporal Attention Learning Framework for Regional Traffic Accident Risk Prediction

Lei Zhu, Tianrui Li, Shengdong Du
School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China
Institute of Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China
zhulei@my.swjtu.edu.cn, {trli, sddu}@swjtu.edu.cn

***Abstract*—Accurate and effective prediction of future traffic accident risk is critical to reducing the number of traffic accidents, which is also of great help to personal safe travel. In our paper, we choose the real traffic administrative area as the way of regional division rather than grid map, so that our prediction results can be applied to true traffic scenarios directly. Instead of considering traffic flow as a single factor affecting traffic accidents, we divide traffic flow into multiple traffic volumes based on vehicle type. In order to better model the dynamic impact of different traffic flow data and traffic accident data in the local region and global regions for future traffic accident risk prediction, we design a deep learning framework to predict regional Traffic Accident risk that utilizes a Spatial-Temporal Attention Network (named TA-STAN). We also integrate many external environmental factors to further improve the accuracy. We evaluate our TA-STAN model on the real traffic accident dataset in New York City. The experimental results show that TA-STAN outperforms 6 baseline models in 3 evaluation metrics. More importantly, by visualizing the weight of attention, we can reasonably interpret the actual meaning of attention weights, which plays a crucial role in our model.**

Keywords—traffic accidents risk; spatial-temporal attention network; deep learning framework

I. INTRODUCTION

With the rapid development of urbanization and the realization of road motorization, people's living has been more and more convenient. At the same time, the explosive growth of motor vehicles has caused a series of social security issues, such as traffic jam, air pollution and traffic accident, which have brought great pressure on the government traffic control [1]. According to the Report of Global Road Safety published by the World Health Organization in 2015, road traffic accidents caused about 1.3 million deaths and 20 to 50 million non-fatal injuries worldwide annually [2]. Therefore, accurate and effective prediction of the number of traffic accidents in various parts of the city over a period of time is crucial to reduction of the number of traffic accidents. On the one hand, governments and city managers can artificially regulate the police force [3]. Moreover, by getting the future risk of traffic accidents, we can choose a safer way for our trip [4].

In the latest research of traffic accident risk prediction as well as other traffic prediction tasks, the main way of regional division is grid map, which divides the whole city into squares, and the models are based on statistic data by the grid [5-10]. This way destroys the inherent attributes of geography and breaks the inherent geographical information on space. And the predicted results are difficult to match to the original traffic regions. If the traffic accident risk can be studied on the basis of the division of the existing traffic administrative districts of the city, the modeling results will be more accurate and meaningful. Therefore, our data are counted on the basis of existing division of the New York taxi zones, as shown in Fig. 1.

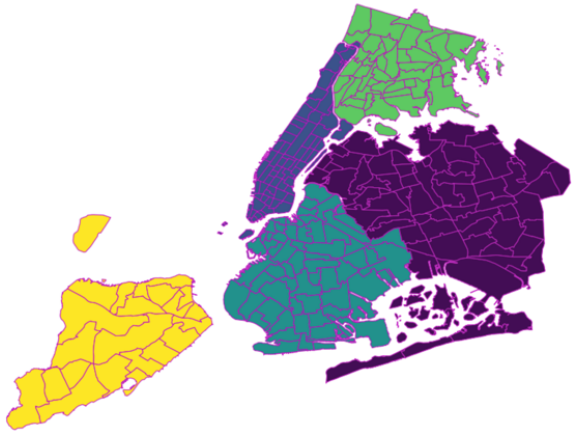

Fig. 1 Taxi zone map of New York City

Besides traffic accident data, the traffic flow data are key factors in the research of traffic accident risk [8]. In this paper, traffic accident data and various traffic flow data are all called "traffic indicators", which will be defined in Section III. We want to uniformly model traffic indicators from different regions to predict future traffic accident risk, which is actually very complex. The main difficulties are as follows:

- *The influence of traffic indicators in local region.* In the same area, different traffic indicators have different effects on traffic accident risk, which will change dynamically based on time. For example, in the Manhattan area of New York, people travel by multiple means of transportation (including yellow taxis and green taxis) during commuting hours, while the rest is mainly by yellow taxis. Obviously, the influence of yellow taxi and green taxi on traffic accident risk changes over time. We need a component to model the effects of local temporal and spatial features.
- *The influence of traffic indicators in global region.* Traffic indicators in surrounding regions also have a dynamic impact on traffic accidents of current region.

978-1-7281-2009-6/$31.00 ©2019 IEEE paper N-19880.pdf

Traffic flow in other regions will shift to the current region and change the traffic accident risk in the current region. The accident of other areas will affect the speed of the traffic flow transferring, so it will also be the import factor of the local traffic accident risk [11]. We need to have a component to model the impact of traffic indicators in all regions of a city on the current regional traffic accidents, which is also changed over time.

- *Influence of environmental factors*. Although traffic accident risk is mainly related to traffic volume, the weather conditions and temperature are also important. For example, high temperature in summer will easily lead to a tire burst, which will cause traffic accidents; rainy days will have reductive number of traffic accidents and so on [12]. Therefore, we need a component to integrate the impact of external factors on the traffic accidents risk.

In order to solve the above problems, we propose a deep learning framework to predict regional Traffic Accident risk that includes a Spatial-Temporal Attention Network (denoted as TA-STAN). Our main contributions are as follows:

1. In order to better capture the dynamic impact of different traffic indicators in local region and global regions, we design a spatial-temporal attention mechanism, which includes local spatial attention mechanism, global spatial attention mechanism and temporal attention mechanism. It is the first time that a deep learning framework with spatial-temporal attention mechanism is applied to the research of traffic accident risk prediction.

2. In order to get the impact of environmental factors, we extract many external features, and a simple component is designed to assist spatial-temporal attention mechanism with external features for final prediction.

3. We use the real New York traffic accident data set to evaluate our TA-STAN model. It is shown that the TA-STAN model performs the best under different evaluation metrics. Finally, by visualizing the weight of attention, we can reasonably interpret the actual meaning of attention weights.

The rest of the paper is organized as follows. Section II is the related work introduction. Section III defines our problem uniformly. Section IV describes the proposed traffic accident risk prediction model in detail. Section V presents the experimental design and evaluation results. Section VI summarizes the conclusions.

II. Related Work

A. *Analysis of influencing factors for Traffic Accidents*

Previous studies have focused on the important factors that affect traffic accidents on specific sections of the road [13]. In 1999, Karlaftis et al. [14] developed logistic regression models in their research. In 2000, Ivan et al. [15] explored the effects of traffic density and land use, as well as ambient light conditions and time of day on traffic accidents. In 2005, Chang et al. [16] put forward some constructive suggestions on how to reduce traffic accidents effectively by building tree-based models. In 2008, Ma et al. [17] offer a multivariate Poisson-lognormal (MVPLN) specification that simultaneously models crash counts by injury severity.

B. *Frequency Prediction of Traffic Accidents based on Certain Road*

With the development of statistics and the wide application of machine learning, many researchers pay more and more attention to predicting future traffic accidents on the certain road. Based on the assumption that the distribution of traffic accidents on a specific road is deterministic, Miaou et al. [18] used Poisson and negative binomial (NB) regression models to simulate the frequency of traffic accidents in 1994. In 2001, Oh et al. [19] found that the turbulence of traffic flow has a great influence on traffic accidents, so the real-time traffic accident frequency model based on Bayes network is built by using traffic flow as a feature. In 2016, Huang et al. [20] used an improved radial basis function neural network and extracted 15 characteristics that affect traffic accidents to model the frequency of traffic accidents. In 2017, Egilmez et al. [21] developed an optimized radial basis function neural network (RBFNN) to model seven factors that affect traffic accidents.

C. *The prediction of regional traffic accident risk*

With the large-scale application of sensors, people can obtain more and more multi-source and heterogeneous data related to traffic, such as real-time traffic flow and weather data. At the same time, deep learning was increasingly used to build large-scale traffic accident models with the ability to deal with high dimensional complex data. In 2016, Chen et al. [6] redefined the number of traffic accidents as traffic risks, and constructed a deep model of Stack denoise Autoencoder using GPS records from people in Japan to explore the impact of urban passenger flow changes on traffic accident risk. In 2018, Ren et al. [7] considered the temporal patterns of traffic accident risk, and used traffic flow, weather and air data, and applying LSTM model to study the temporal periodicity and trend of regional traffic analysis, which is a significant improvement compared with SAE and SVM models. In 2018, Yuan et al. [8] used a deep learning framework to model a large number of heterogeneous data related to traffic accidents, including traffic volume, road conditions, weather, etc. They extracted many temporal and spatial features from these data, and used Conv-LSTM to combine the features to find out the spatial-temporal patterns behind traffic accidents. Although the above researches took into account the temporal and spatial patterns of traffic accident data and achieved good prediction results, the artificial separation of the administrative area will break the whole pattern of the region and cause the deviation of the prediction.

III. Preliminary

This section introduces the definition of our problem. We propose some necessary definitions and then show the problem definition of traffic accident risk prediction. Suppose there are n_z zones in the city.

Definition 1. Range of Zone. We divide the city into irregular areas based on the map of traffic administrative districts. The range of i^{th} zone z_i is:

$$R(z_i) = \{p_1, p_2, \dots, p_m\} \in \mathbb{R}^m,$$

where p_k is the k^{th} 2-dimension point of geographic polygon, which is composed of boundary latitude and longitude coordinates of zone z_i.

As Ren has documented, we can't predict accurately whether traffic accidents will occur, but the traffic accidents counts at different timestamps of a day in a region are in a stable mode [7]. This mode is largely affected by changes of traffic flow in 24 hours every day. So we classify each traffic accident records into the corresponding zone according to latitude and longitude.

Definition 2. Traffic accident risk. The traffic accidents risk Similar to [6] for i^{th} zone z_i in timestamp t is:

$$TR_{\mathrm{i},t} = \sum_{j}^{Ins_t \in R(z_i)} (v_j + p_j),$$

where $Ins_t \in R(z_i)$ are the instances of traffic accident of time t belong to zone z_i; For each traffic accident j of $Ins_t \in R(z_i)$, v_j is the number of vehicles in the accident and p_j is amount of the injured people in the accident.

Definition 3. Traffic indicators. We call the different traffic data, including traffic accident data and various traffic flow data, are all traffic indicators. The traffic indicators for the zone z_i in time t are defined as:

$$TI_{i,t} = \left(TI_{i,t}^1, TI_{i,t}^2, \dots, TI_{i,t}^{n_I}\right) \in \mathbb{R}^{n_I},$$

where n_I is the number of traffic indicators; $TI_{i,t}^j$ represents the j^{th} traffic indicators of the zone z_i at time t. And $TI_{i,t}^j$ is obtained by counting the number of j^{th} traffic data belonging to the range of zone z_i.

Problem Definition. Given all the historical traffic indicators TI and external environmental features, predict the future traffic accidents risk $(TR_{T+1}, TR_{T+2}, \dots, TR_{T+t'})$ for a period of time t' in each zone of the city.

IV. Spatial-temporal Attention Network

We introduce the deep learning framework, TA-STAN, to predict regional traffic accident risk as shown in Fig. 2. This framework consists of 3 main parts: *1) Spatial attention mechanism*. In the phase of Encoder, we use two spatial attentions mechanisms, local spatial attention and global spatial attention, to capture the dynamic influence of historical regional traffic indicators. *2) Temporal attention mechanism*. We use temporal attention mechanism to associate each timestamp in Decoder with each timestamp in Encoder. *3) External feature extraction*. In the phase of Decoder, our model incorporates external features such as weather features, time features, and road design features as part of the final forecast to predict traffic accident risk over time more accurately.

A. Attention mechanism

Before introducing our spatial-temporal attention mechanism, we first introduce the attention mechanism, which was proposed by Bahdanau et al. in the neural machine translation task [22]. By calculating the attention coefficient between different hidden states of Encoder and Decoder, the Encoder hidden states and Decoder hidden states are aligned so that model can capture temporal relationship and contextual information are different for various timestamps in Decoder. The basic form of the attention mechanism is as follows:

$$e_i = score(u_j, v_i), \tag{1}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_k \exp(e_k)}, \tag{2}$$

$$c = \sum_i \alpha_i v_i . \tag{3}$$

For the set of source vectors $V = \{v_1, v_2, \dots, v_n\}$ and the set of target vectors $U = \{u_1, u_2, \dots, u_m\}$, we need to calculate the importance score of vectors in V to vectors in U, as shown in Equation (1). Normalization of the importance score is usually achieved using Softmax function, as shown in Equation (2). Finally, by combining V and attention weight α, we can get the set of vectors with attention weight relative to U.

The most important thing in calculating attention is the score of Equation (1), which is the attention function of calculating attention. The two commonly used attention functions are as follows [23]:

$$u^T W v \text{ (General)} \tag{4}$$

$$w_3^T \tanh(w_1 u + w_2 v) \text{ (Concat)} \tag{5}$$

We use both attention functions in various attention mechanisms.

B. Spatial attention mechanism

In this part we apply spatial attentions to capture the dynamic influence of historical regional traffic indicators. This section consists of three parts: *1) Local spatial attention mechanism.* We find the importance between local input of traffic indicators and target series with local spatial attention. *2) Global spatial attention mechanism.* We find the importance between global input, including global indicative input and global regional input, of traffic indicators and targets series with global spatial attention. *3) Spatial fusion module.* In order to fuse the above three kinds of spatial input and the attention weight, a spatial fusion module is used.

1) Local spatial attention mechanism

Local spatial attention mechanism is the weighting mechanism for model the input in Encoder, which pays more attention to the impact of current regional traffic indicators. For a region, there is a complex relationship between traffic flow data and traffic accident data, which will change over time. For example, in Lower Manhattan, yellow taxis are the main mode of driving, which has a great impact on traffic accidents. However, when the traffic capacity is insufficient at morning and evening rush hours, the use of green taxis and for-hired vehicles will be greatly increased, and the importance of these traffic indicators affecting the risk of traffic accidents will change. To describe the impact of this local change, we design an attention mechanism to capture the dynamic relationship between future local traffic accidents and historic local traffic indicators. Given j^{th} traffic indicator $TI_{i,t}^j$ of zone z_i in timestamp t, the attention mechanism is as follows:

$$a_t^j = w_1 \tanh\left(w_2 h_{t-1} + w_3 TI_{i,t}^j + b_0\right), \tag{6}$$

$$\alpha_t^j = \frac{\exp(a_t^j)}{\sum_q^{n_z} \exp(a_t^q)}, \tag{7}$$

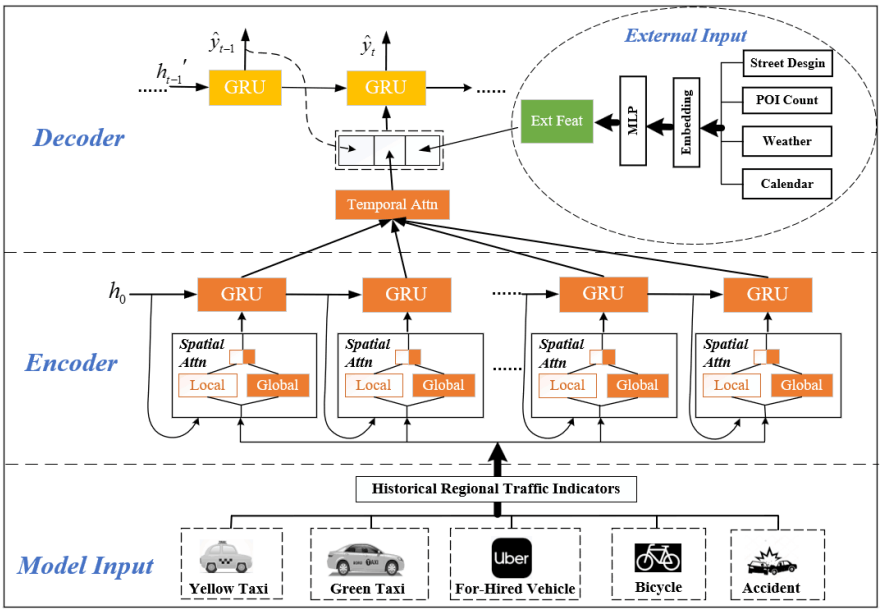

Fig. 2 Overall Framework of TA-STAN Model

where h_{t-1} denotes the hidden state of the previous timestamp in Encoder, and a_t^j describes the importance of the previous hidden state and local traffic indicators to the hidden state at the current timestamp.

2) Global spatial attention mechanism

For a regional traffic accident, the impact of the other areas is crucial as well and will change over time. If we simply input traffic indicators of all regions that affect the target series, the computational cost is very high and the experimental effect is bad. We note that the overall spatial impact can be divided into two perspectives: 1) Indicative: All areas of a traffic indicator produce an overall impact of a common indicator. 2) Regional: All traffic indicators in a region have an overall regional impact. For example, for the airport area, its fixed flights will export high-flow risk to Manhattan office area in the daytime and Brooklyn residential area in the evening, so the comprehensive influence of this area is relatively large, which is a regional impact. On the other hand, in the early rush hour, the traffic flow in all areas will increase, so a certain traffic indicator to play a leading role without distinguishing between regions, which is called an indicative effect. So we design two global spatial attention mechanisms, e.g. global indicative attention and global regional attention.

$$e_t^k = w_4 \tanh(w_5 h_{t-1} + w_6 TI_t^k + b_1), \quad (8)$$

$$\beta_t^k = \frac{\exp(e_t^k)}{\sum_q^{n_I} \exp(e_t^q)}, \quad (9)$$

$$r_t^l = w_7 \tanh\left(w_8 h_{t-1} + w_9 TI_{l,t} + b_2\right), \quad (10)$$

$$\gamma_t^l = \frac{\exp(r_t^l)}{\sum_q^{n_z} \exp(r_t^q)}, \quad (11)$$

where h_{t-1} still denotes the hidden state of the previous time in Encoder, e_t^k and β_t^k represent the impact factors and normalized impact factors of the k^{th} global traffic indicator in timestamp t, respectively, as shown in Equations (8) and (9). r_t^l and γ_t^l represent the influence factors and the normalized influence factors of the l^{th} global region in timestamp t, respectively, as shown in Equations (10) and (11). $TI_t^k \in \mathbb{R}^{n_z}$ and $TI_{l,t} \in \mathbb{R}^{n_I}$ are global indicative input and global regional input respectively.

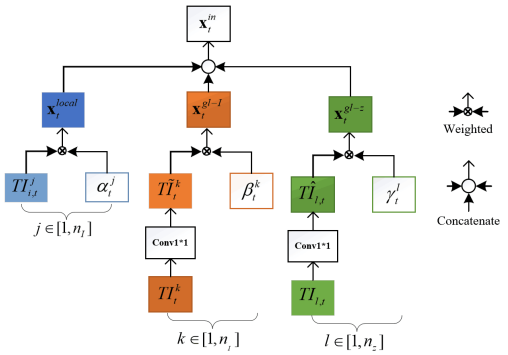

Fig. 3 Spatial fusion module

3) Spatial fusion module

After using the spatial attention mechanism, we extract three spatial inputs, and the next step is to combine these three inputs as a fused input to Encoder.

As shown in Fig. 3, after the operation of conv1*1, we get indicative feature $T\tilde{I}_t^k$. Then we combine indicative feature $T\tilde{I}_t^k$ and attention weight α_t^j, generated indicative vector $x_{gl-I}^t \in \mathbb{R}^{n_I}$. In the same way, we can get regional vector $x_{gl-z}^t \in \mathbb{R}^{n_z}$. Finally, we get the fused input of Encoder at time t as $x_{input}^t = \{x_{local}^t, x_{gl-I}^t, x_{gl-z}^t\} \in \mathbb{R}^{2\times n_I + n_z}$.

C. Temporal attention mechanism

Similar to neural machine translation, in the traffic accident risk prediction framework, the temporal attention mechanism is needed to associate each timestamp t' in Decoder with

timestamp t in Encoder. Temporal attention can add a connection to help the model better learn the impact of historical timestamps on future timestamps [24]. In the following equations, we calculate the attention score of hidden state in each time t' in Decoder and all hidden states in Encoder:

$$e_{t'}^{k} = \mathrm{d}_{t'-1} W h_k + b, \tag{12}$$

$$\delta_{t'}^{k} = \frac{\exp(e_{t'}^{k})}{\sum_{j=1}^{T} \exp(e_{t'}^{j})}, \tag{13}$$

$$c_{t'} = \sum_{p=1}^{T} \delta_{t'}^{k} h_k, \tag{14}$$

where $\mathrm{d}_{t'-1}$ is the previous hidden state in Decoder. Then we get the context information of Decoder every timestamp $c_{t'}$.

D. External feature extraction

This section is inspired by the influence of external factors in previous research of traffic accidents risk prediction [8]. Spatial-temporal traffic indicators have been combined with external environmental features such as weather, time and street design to better predict the future traffic accident risk. We design a simple and effective component to join the Decoder to assist in prediction.

As shown in Fig. 2, we first extract a series of features from street speed limit, POI, weather, zone number, and time data. For example, for street speed limit data, we get the average speed limit in an area, and count whether the number of streets in the area is limited or not. For POI data, it is mainly to count whether there are schools, commercial districts, government offices, etc. The time features are mainly the day of the week, the hour of the day, whether it is a working day and whether it is a holiday or not. We use one-hot coding for category data such as zone numbers. For all of the above extracted external features, we use a layer of embedding to encode these features and get our encoded external input $x_{ext}^{t'}$, which is used as a part of the timestamp t' in Decoder to help predict the traffic accident risk at timestamp t'.

V. Experimental results

A. Data Source

We chose New York City as the location for our research. New York City's road network is spread across the five major administrative districts. The complicated streets and dense population have brought traffic congestion and traffic accidents to the city. All the data we collect are about New York City for traffic accident risk. Its details are as follows：

Vehicle Collisions. Vehicle collision data we obtained from the New York of Police Department (NYPD). This data contains all the recorded traffic accidents for the period from 2012 to 2018. In addition to basic information such as time, place, and street, this data set also includes the reason of crash.

Motor Vehicle Trip Record Data. We obtain motor vehicle trip data from the NYC Taxi and Limousine Commission (TLC), which includes travel records for motor vehicles from 2009 to the present, each record containing the exact geographic coordinates and time of picking up and off. The motor vehicle contains various types of vehicles, e.g. yellow taxi, green taxi and for-hired vehicle.

Citi Bike Trip Data. The bicycle travel data includes every ride data from July 2013 to the present in New York. The data includes the starting point coordinates, the starting point time, the ending point coordinates and the time spent.

Taxi Zone Information. We use the divided area of the taxi as the division standard for all data. Since the traffic accident data is mainly a motor vehicle accident, the default bicycle division area here is similar to the taxi division area. The entire city of New York is divided into 263 areas.

Weather Data. Weather data comes from the National Weather Data Center and data dimensions contains date, time, site latitude and longitude, temperature, humidity, visibility, wind direction and weather.

Street Design Data. The road design data comes from the NYC open data, which includes the speed limit of the street, whether it includes the Bike Priority Area, and whether it includes the Left Turn Traffic Calming.

POI Data. POI data is also from the New York Public Data Center, which contains POI coordinates for schools, shopping malls, food, entertainment and sports.

B. Experimental setup

Data set partition. Based on historical 12-hour traffic accident data, traffic flow data and other external data, we predict the risk of traffic accidents for the next 12 hours. Our entire data set (1 year and a half, 13104 hours, 263 areas) is converted to 3,472,548 data. All data is divided into three parts. In the first part, the data in 2017 is used for training; in the second part, data from January to February in 2018 is used for validation; in the third part, data from March to June in 2018 is for test.

Evaluation task. Through experiments we hope to answer the following questions: 1) Q1: Are the results of our framework prediction better than those of previous baseline models? 2) Q2: Is it effective to split the traffic flow into different flow by type of vehicles? 3) Q3: Are the spatial attention mechanisms and temporal attention mechanisms effective? 4) Q4: What are the effect of attention functions on different attention mechanisms? 5) Q5: Are our predictions consistent with the reality?

Evaluation metric. We use MSE (mean square error) to minimize the squared loss error between the predicted value $\hat{y}$ and label y_{real}. We also use RMSE (root mean square error) and MAE (mean absolute error) to measure our predictions.

Baseline models. In order to compare the effects of our proposed model, we use the following models as the baseline models: 1) HA: Historical average. 2) LR: Linear regression. The input of LR comes from local historical traffic accident data. Affected by changes of crowd flows every day, the traffic data contains 3 time characteristics, time closeness, time periodicity and time trend [25]. Therefore, we take the traffic accidents risk in the last 10 hours of this area, the traffic accidents risk in the same day of the previous 4 weeks, and the traffic accidents risk of the same hour in the previous 7 days. 3) LSTM. The LSTM is trained using a two-layer stacked LSTM network with 512 neurons and Dropout = 0.2. 4) SdAE (Denoising auto-encoder). Here is the method used in [6] to predict urban traffic risk. 5) Xgboost. A gradient boosting algorithm of machine learning function library [26]. We extract the traffic accident, traffic flow

and external features. 6) Seq2seq. We generate a context feature using a LSTM layer encoding input, and then decode the target using this context feature and another LSTM layer.

TABLE I A COMPARISON OF MODELS

Model	MSE	RMSE	MAE
SdAE	4.225E-3	0.06512	0.0398
HA	3.481E-3	0.05902	0.0409
LR	2.304E-3	0.04801	0.0311
LSTM	1.953E-3	0.04423	0.0259
Seq2seq	1.376E-3	0.03712	0.0220
Xgboost	0.850E-3	0.02916	0.0186
TA-STAN	**0.172E-3**	**0.01312**	**0.0082**

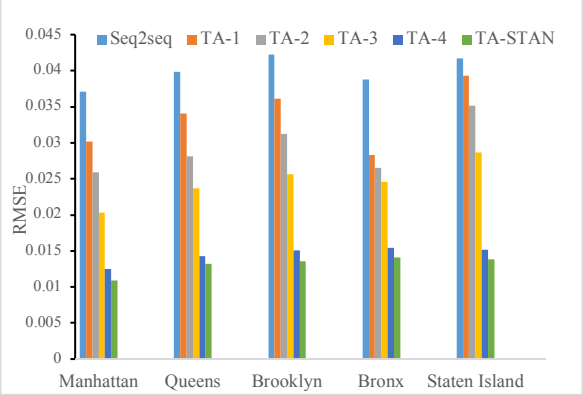

Fig. 4 Variant models

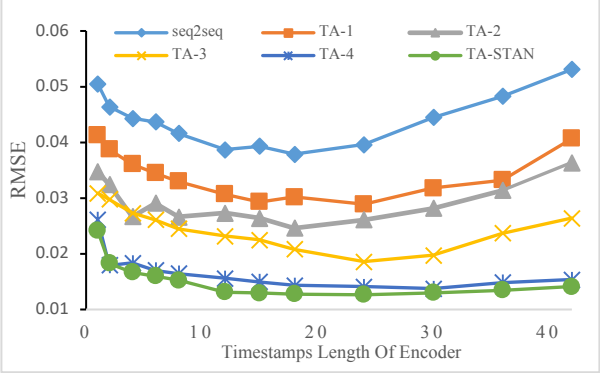

Fig. 5 Variant models with different timestamps length of Encoder

C. Model comparison (for question Q1)

In this section, we compare our model to the baseline models. To be fair, we show the best results for each model. The experimental results are shown in TABLE I. We can see that HA is better than SdAE, indicating that traffic accidents in each area have a certain periodicity and seasonality. The results of LR are better than HA and SdAE, indicating that traffic accidents in historical close time are more able to reflect the future situation. LSTM is better than LR, which indicates that the RNN structure with memory unit can capture the relationship between different timestamps. Seq2seq is better than LSTM because of adding the Decoder structure, and Seq2seq structure has better performance in long term prediction. The Xgboost model is only second to our proposed model for that the traffic accident risk is not only affected by its own historical data, but also related to the traffic flow data and external factors. Our proposed TA-STAN model has achieved great improvement in the three evaluation metrics of MSE, RMSE and MAE, showing our model can well capture the dynamic impact of three aspects of local traffic indicators, other regional traffic indicators and external factors on future traffic accident risk.

D. The role of each component (for question Q2 and Q3)

In this section we explore the role of each of our TA-STAN components through experiments. There are five components including local spatial attention (L), global indicative spatial attention (GI), global regional spatial attention (GZ), temporal attention (T), and external environmental feature (E). We add one component at a time to see the different effects of the model in the five borough of New York. There are several variant models as follows.

- TA-1: L
- TA-2: L+GI
- TA-3: L+GI+GZ
- TA-4: L+GI+GZ+T
- TA-STAN: L+GI+GZ+T+E

1) Evaluation of Local spatial attention

As shown in Fig. 4, in five boroughs of New York, the TA-1 model is better than seq2seq, especially Brooks, in which training error drops the most, because Brooks has less interaction with other regions and the traffic situation is more affected by local traffic flow and traffic accidents. From the above analysis, we can see that the TA-1 model has a much lower error than Seq2seq. This result shows that the method of subdividing traffic flow into multiple traffic flows is more effective for predicting future traffic accident risk. On the one hand, the local spatial attention mechanism does capture the dynamic changes of the impact of different traffic flows on future traffic accident risks.

2) Evaluation of global spatial attention

We continue to analyze Fig. 4. After adding the component GI, the error of model TA-3 has a certain decline compared with the TA-2; TA-4 adds the group GZ on the basis of TA-3, and also has a large loss decline in all boroughs. The results of this experiment fit the fact that the traffic states between the five boroughs interact with each other. Especially in the Manhattan area, the model upgrade is particularly obvious after joining the GI and GZ. This may be because the daily commute between Manhattan and the surrounding area is particularly frequent, so the other areas' dynamic impact on traffic accident risk in Manhattan can be well captured through two spatial attention mechanisms.

3) Evaluation of Temporal Attention

Compared with the TA-4, temporal attention mechanism of TA-5 has improved in all regions, which indicates that the temporal attention mechanism can obtain the correlation between historical timestamps with the future timestamps. Since the temporal attention mechanism is used to align the hidden states of Decoder and Encoder, the length of Encoder's timestamps will affect the alignment effect. We also explore changes in the effect of the temporal attention mechanism under different Encoder lengths. The result is shown in Fig. 5, where the length of the Decoder is 12, and it can be observed that the minimum RMSE value of the six models are reached when the Encoder length is 24, indicating that there is a significant daily

periodic for traffic accident. In addition, we observe that after the Encoder length is greater than 30, as the length continues to increase, the RMSE error value gradually increases, but the errors of our TA-5 and TA-STAN models are relatively stable and will not continue to increase because these two models contain the temporal attention mechanism.

4) Evaluation of External feature extraction

As an empirical feature component, this section provides a number of additional features to make our experimental results more accurate. We add external features based on the TA-4. The model drops in the case of the original low loss, and the decline in each region is relatively close, indicating that the external factors are indispensable.

TABLE II EFFECTIVENESS OF ATTENTION FUNCTION

Model	Attention Mechanism	RMSE	MAE
TA-1	L (general)	0.0314	0.0223
	L (concat)	0.0307	0.0217
TA-2	GI (general)	0.281	0.0192
	GI (concat)	0.0273	0.0183
TA-3	GZ (general)	0.0239	0.0158
	GZ (concat)	0.0232	0.0149
TA-4	T (general)	0.0156	0.0117
	T (concat)	0.0170	0.0125

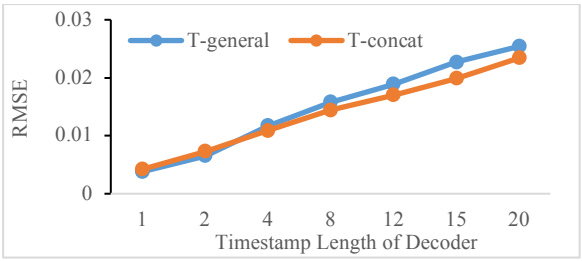

Fig. 6 Influence of attention function

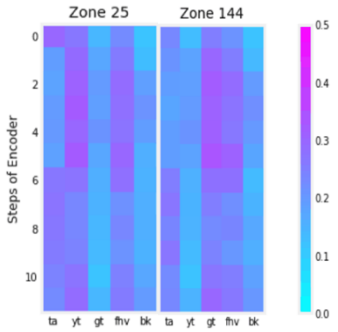

Fig. 7 Weight visualization of local spatial attention mechanism

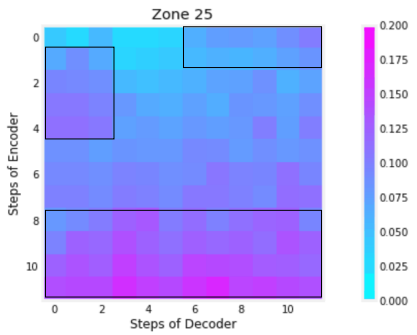

Fig. 8 Weight visualization of temporal attention mechanism

E. The Effectiveness of Attention Function (Q4)

In this section, we discuss the effectiveness of different attention mechanisms (L, GI, GZ, and T) with different attention functions (General, Concat). When we test each attention function, other attention mechanisms use optimal parameters and structures. The experimental results are shown in TABLE II.

Among the three attention mechanisms of L, GI and GZ, the Concat method is better than the General method. It may be that the General method itself is a single-layer neural network, which can better establish the nonlinear importance between Encoder's hidden states and input data. For the *T* attention mechanism, the General method's RMSE loss is smaller in totally and we also test the model RMSE of the two attention functions at different Decoder lengths, as shown in Fig. 6. We can see that when the length of the Decoder is relatively short, the Concat method is better than the General method, but as the length grows, the General method is significantly better than the Concat method, which shows that the General approach is more accurate in capturing the importance between history and future timestamps. This is because the General method adds a weight between the two time hidden units. This weight is similar to the similarity calculation, and the importance value obtained by the similarity calculation method is more generalized in practice.

F. Case study (for question Q5)

We use the zone numbers 25 and 144 as the focus, and show the case analysis of the time period of 6-18 pm on May 17, 2018. The analysis uses the local spatial attention mechanism and the temporal attention mechanism as examples to explain the practical meaning of attention weight. Zone 25 is the core financial district of Midtown Manhattan, and is also the area where people gather during the day. Zone 144 is the residential area of Brooklyn, so a large number of people commute between Zone 25 and Zone 144 every day.

We first look at the visualization of the local spatial attention weights, as shown in Fig. 7. We can see that the traffic accident itself is more important for the two regions in the two periods 0-2 and 6-10. These two time periods correspond to the early morning and noon respectively, and the population has no obvious trend, so the predicted value is more relevant to the traffic accident itself. And the timestamps 2-6 and 10-12 correspond to the morning peak and the evening peak of the day, and the importance of the four modes of transportation has increased, indicating that the traffic volume has a greater impact on traffic accidents. In addition, in the zone 25, the importance of green taxis is generally low, and the overall importance of yellow taxis and for-hired vehicle is higher. In contrast, in zone 144, yellow taxis are less important, green taxi is of high importance. This is because the green taxi has a smaller service in downtown Manhattan, and the yellow taxi has a smaller service in Brooklyn. From the above analysis, we can see that local spatial attention mechanism can indeed grasp the importance of the local time series.

In the following we explore the visual map of time attention mechanism, as shown in Fig. 8. Overall, the closer the historical time is, the greater the importance of the historical time is. But we can also see that the historical 1-4 timestamps in Encoder is also important to the timestamps of 0-2 in Decoder, because the historical interval is the morning rush hour, and the forecast interval is the evening rush hour. The traffic volume during the

morning rush hours can also relatively reflect the traffic volume and traffic accident risk during the evening rush hours. The 8-10 time period in Decoder is nightlife, so it pays attention to two parts: 1) 10-12 timestamps in Encoder, which is the neighboring point, reflecting the traffic volume of the day. 2) The 0-1 timestamp in the Encoder. This is the situation of last night, indicating that the traffic pattern in the zone has a bit periodicity. Through the above analysis, we can see that temporal attention accurately captures the dynamic time correlation of different timestamps between Encoder and Decoder.

VI. Conclusion

This paper explored an urban regional traffic accident risk prediction model based on multi-source heterogeneous data using deep learning. In order to model the spatial-temporal characteristics of traffic, we applied the Encoder-Decoder framework that includes the spatial-temporal attention mechanism. In our multi-source heterogeneous data, we paid more attention to the traffic volume data due to its higher correlation with traffic accident, and subdivided the traffic volume into multiple traffic flow based on different vehicles. In order to better capture the dynamic impact of different traffic indicators on future traffic risks, we have designed three attention mechanisms, which are local spatial attention mechanism, global spatial attention mechanism and temporal attention mechanism. Finally, in Decoder's prediction stage, we integrated the impact of external environmental factors on traffic risk, making the prediction more accurate. We employed real traffic data from New York City as experimental data. The experimental results showed that our training error is much less than six basic models on all the three evaluation metrics. We have done a series of experiments to explore the effectiveness of each component in this deep learning framework and the effects of different functions on attention components. Finally, we visualized the attention weight to analyze its practical significance. The results indicated that our attention weights have strong interpretability.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 61773324, 61876158).

References

[1] M. Peden, R. Scurfield, D. Sleet, et al., World report on road traffic injury prevention, 2004.

[2] D. Shinar, Traffic safety and human behavior. *Emerald Group Publishing*, 2007.

[3] F. Mannering, C. Bhat, "Analytic methods in accident research: Methodological frontier and future directions", *Analytic Methods In Accident Research*, vol. 1, pp. 1-22, 2014.

[4] C. Chen, X. Fan, C. Zheng, et al., "SDCAE: Stack denoising convolutional autoencoder model for accident risk prediction via traffic big data", *Proceedings of the 6th International Conference on Advanced Cloud and Big Data (CBD). IEEE*, pp. 328-333, 2018.

[5] Z. Yuan, X. Zhou, T. Yang, et al., "Predicting traffic accidents through heterogeneous urban data: A case study", *Proceedings of the 6th International Workshop on Urban Computing*, vol. 14, pp. 1-9, 2017.

[6] Q. Chen, X. Song, H. Yamada, et al., "Learning deep representation from big and heterogeneous data for traffic accident inference", *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI),* pp. 338-344, 2016.

[7] H. Ren, Y. Song, J. Wang, et al., "A deep learning approach to the citywide traffic accident risk prediction", *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITS). IEEE*, pp. 3346-3351, 2018.

[8] Z. Yuan, X. Zhou, T. Yang, "Hetero-ConvLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data", *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). ACM*, pp. 984-992, 2018.

[9] S. Du, T. Li, X. Gong, et al., "A hybrid method for traffic flow forecasting using multimodal deep learning", *arXiv preprint arXiv*: 1803.02099, 2018.

[10] D. Wang, Y. Yang, S. Ning, "DeepSTCL: A deep spatio-temporal ConvLSTM for travel demand prediction", *2018 International Joint Conference on Neural Networks (IJCNN). IEEE*, pp. 1-8, 2018.

[11] Y.C. Chiou, C. Fu, H. Chih-Wei, "Incorporating spatial dependence in simultaneously modeling crash frequency and severity", *Analytic Methods Accident Research*, vol. 2, pp. 1-11, 2014.

[12] F. L. Mannering, V. Shankar, C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data", *Analytic Methods Accident Research*, vol. 11, pp. 1-16, 2016.

[13] Z. Zhang, Q. He, J. Gao, et al., "A deep learning approach for detecting traffic accidents from social media data", *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 580-596, 2018.

[14] M. Karlaftis, P. Latoski, N. Richards, et al., "ITS impacts on safety and traffic management: An investigation of secondary crash causes", *Journal Intelligent Transportation Systems*, vol. 5, no. 1, pp. 39-52, 1999.

[15] J.N. Ivan, C. Wang, N.R. Bernardo, "Explaining two-lane highway crash rates using land use and hourly exposure", *Accident Analysis and Prevention*, vol. 32, no. 6, pp. 787-795, 2000.

[16] L. Chang, W. Chen, "Data mining of tree-based models to analyze freeway accident frequency", *Journal of Safety Research*, vol. 36, pp. 365-375, 2005.

[17] J. Ma, K. Kockelman, P. Damien, "A multivariate Poisson-lognormal regression model for prediction of crash counts by severity using Bayesian methods", *Accident Analysis and Prevention*, vol. 40, no. 3, pp. 964-975, 2008.

[18] S.P. Miaou, "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regression", *Accident Analysis and Prevention*, vol. 26, no. 4, pp. 471-482, 1994.

[19] C. Oh, J.S. Oh, S. Ritchie, et al., "Real-time estimation of freeway accident likelihood", *Presented at the 80th Annual Meeting of the Transportation Research Board (TRB)*, Washington, DC, 2001.

[20] H. Huang, Q. Zeng, X. Pei, et al., "Predicting crash frequency using an optimised radial basis function neural network model", *Transportmetrica A: Transport Science*, vol. 12, no. 4, pp. 330-345, 2016.

[21] G. Egilmez, D. McAvoy, "Predicting nationwide road fatalities in the US: a neural network approach", *International Journal of Metaheuristics*, vol. 6, no. 4, pp. 257-278, 2017.

[22] D. Bahdanau, K. Cho, Y. Bengio, "Neural machine translation by jointly learning to align and translate", *arXiv preprint arXiv*:1409.0473, 2015.

[23] T. Luong, H. Pham, C. D. Manning, "Effective approaches to attention-based neural machine translation", *Proceedings of Empirical Methods Natural Language Processing (EMNLP)*, pp. 1412-1421, 2015.

[24] D. Britz, A. Goldie, M.T. Luong, et al., "Massive exploration of neural machine translation architectures", *arXiv preprint arXiv*:1703.03906, 2017.

[25] J. Zhang, Y. Zheng, D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction", *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1655-1661, 2017.

[26] T. Chen, C. Guestrin, "Xgboost: A scalable tree boosting system", *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). ACM*, pp. 785-794, 2016.