GSNet: Learning Spatial-Temporal Correlations from Geographical and Semantic Aspects for Traffic Accident Risk Forecasting

Beibei Wang,[1,2] **Youfang Lin,**[1,2,3,4] **Shengnan Guo,**[1,2,4] **Huaiyu Wan**[1,2,3*]

[1]School of Computer and Information Technology, Beijing Jiaotong University, Beijing, China
[2]Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing, China
[3]CAAC Key Laboratory of Intelligent Passenger Service of Civil Aviation, Beijing, China
[4]Key Laboratory of Transport Industry of Big Data Appalication Technologies for Comprehensive Transport, Beijing, China
{wangbb, yflin, guoshn, hywan}@bjtu.edu.cn

Abstract

Traffic accident forecasting is of great importance to urban public safety, emergency treatment, and construction planning. However, it is very challenging since traffic accidents are affected by multiple factors, and have multi-scale dependencies on both spatial and temporal dimensional features. Meanwhile, traffic accidents are rare events, which leads to the zero-inflated issue. Existing traffic accident forecasting methods cannot deal with all above problems simultaneously. In this paper, we propose a novel model, named GSNet, to learn the spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting. In the model, a Spatial-Temporal Geographical Module is designed to capture the geographical spatial-temporal correlations among regions, while a Spatial-Temporal Semantic Module is proposed to model the semantic spatial-temporal correlations among regions. In addition, a weighted loss function is designed to solve the zero-inflated issue. Extensive experiments on two real-world datasets demonstrate the superiority of GSNet against the state-of-the-art baseline methods.

Introduction

Traffic accidents have caused heavy loss of life and property every year. According to the WHO[1], the number of annual road traffic deaths on the earth reaches 1.35 million in 2018. Hence, traffic accident forecasting is very significant for public safety and city construction. If the traffic accident risk can be predicted accurately in advance, governments can make better traffic planning to reduce traffic accidents, administrators can issue traffic accident risk warnings and drivers can choose safer routes to avoid traffic hazards.

Traffic accidents occur in certain spatial and temporal scenarios, so they are affected by both spatial and temporal dimensional factors. In recent years, researchers have already proposed many deep learning-based models to forecast traffic accidents. Existing deep learning-based models can be broadly categorized into two classes. One class only captures spatial or temporal features and the other class captures both spatial and temporal features. More specifically, the first class mainly employs recurrent neural networks (RNN) to model traffic accidents' underlying temporal correlations, such as TARPML (Ren et al. 2018), or convolutional neural networks (CNN) to capture adjacent regions' spatial correlations, such as SDCAE (Chen et al. 2018). The other class combines RNN and CNN to model both the temporal periodicity and spatial correlations of traffic accidents, such as Hetero-ConvLSTM (Yuan, Zhou, and Yang 2018).

However, these existing traffic accident forecasting methods cannot effectively solve the following three problems:

1) The causes of traffic accidents are complex. In reality, lots of factors such as weather, time, traffic flow, etc., can affect the occurrence of traffic accidents. How to take all these factors into account when forecasting traffic accident risk is challenging.

2) Traffic accidents usually exhibit multi-scale dependencies in both the spatial and temporal dimensions, which can be called geographical spatial-temporal correlations and semantic spatial-temporal correlations. Firstly, taking the geographical spatial-temporal correlations shown in Figure 1 as an example, in the spatial dimension, traffic accidents of neighboring regions (e.g., region ① and ②) are usually affected by each other due to the road connections and traffic flows. In the temporal dimension, traffic accidents often have short-term proximity and long-term periodicity. Besides, there are semantic spatial-temporal correlations among regions. For example, region ① and ③ in Figure 1 share similar features, i.e., road structure, POI (point of interest) distribution and traffic accident pattern. Therefore, they might have a similar trend in traffic accident risk. In a word, it is essential to model the multi-scale correlations of regions for accurate traffic accident risk forecasting.

3) The zero-inflated issue. Generally, traffic accidents are rare events, which means too excessive number of zeros exist in traffic accident risk values. In the model training phase, if the zeros are not properly handled, they will further cause the prediction toward zero. We call this phenomenon the zero-inflated issue. (Bao, Liu, and Ukkusuri 2019) mentioned that when the spatial-temporal resolution of the prediction tasks increases, the zero-inflated issue will occur.

To address the above problems, we propose a Geographical and Semantic spatial-temporal Network (GSNet) for

*Corresponding author: hywan@bjtu.edu.cn

[1]https://www.who.int/

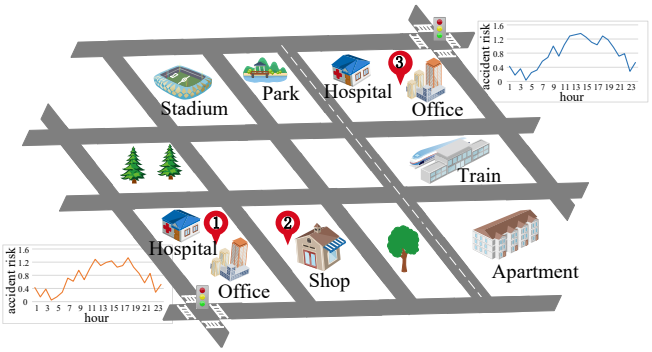

Figure 1: An example of geographical and semantic spatial-temporal correlations. Region ① and ② are road connected, so they have geographical spatial-temporal correlations. Region ① and ③ share similar road structure, POI distribution and traffic accident pattern, so they have semantic spatial-temporal correlations.

traffic accident risk forecasting. With multi-source spatial-temporal data as input, we respectively design a Spatial-Temporal Geographical Module which uses convolution, GRU and attention mechanism to capture the geographical spatial-temporal correlations among regions, and a Spatial-Temporal Semantic Module which employs multiple graph convolutions, GRU and attention mechanism to capture the semantic spatial-temporal correlations among regions.

Our contributions can be summarized as follows:

- A Geographical and Semantic spatial-temporal Network (GSNet) model is proposed for traffic accident risk forecasting, which takes multi-source spatial-temporal factors into account and is able to model the spatial-temporal correlations of the traffic accident data from geographical and semantic aspects.
- We design a weighted loss function to address the zero-inflated issue, which pays more attention to the samples with high traffic accident risk to adjust the model to predict unbiased results.
- Extensive experiments are conducted on two real-world traffic accident datasets, which demonstrate our model surpasses the state-of-the-art methods.

Related Work

Traffic Accident Forecasting

Traffic accident forecasting has attracted many attentions and some effective works have been developed for years. These works can be classified into statistic learning-based methods and deep learning-based methods. Statistic learning-based methods mainly include negative binomial regression, decision tree, and k-nearest neighbor. (Caliendo, Guida, and Parisi 2007) utilized negative binomial regression to investigate the impact of multilane roads conditions on traffic accidents. In (Olutayo and Eludire 2014), the authors used decision tree to model the severity of injury resulting from traffic accidents. (Lv, Tang, and Zhao 2009) adopted the k-nearest neighbor method to investigate how to identify the traffic accident potential. However, these approaches assume traffic accidents are independent, which is hard to accord with reality.

Recently, researchers attempted to utilize deep learning-based methods to predict traffic accidents. (Chen et al. 2016) employed stack denoise autoencoder (SDAE) and for the first time to estimate traffic accident risk on a city scale. To further improve the performance of prediction, (Chen et al. 2018) proposed stack denoise convolutional autoencoder (SDCAE), which stacks CNN to extract spatial correlations among regions. However, these two models ignore the temporal dependencies of traffic accidents. Later, (Yuan, Zhou, and Yang 2018) proposed the Hetero-ConvLSTM model to capture spatial heterogeneity by moving window. Besides, it captures temporal auto-correlation from a geographical aspect. (Zhou et al. 2020) proposed the RiskOracle model, which captures spatial-temporal correlations from different periodic urban graphs and then uses multi-task learning to predict traffic accident risk and traffic flow. However, all of the above works still have some shortcomings in simultaneously considering the geographical and semantic spatial-temporal features of traffic accidents.

Therefore, in this paper we propose a novel model to capture the geographical spatial-temporal correlations and semantic spatial-temporal correlations simultaneously, which is believed to improve the performance of traffic accident risk forecasting.

Graph Convolution Networks in Spatial-Temporal Forecasting

As mentioned in (Guo et al. 2019), although traditional convolutions can effectively extract the geographical patterns of data, they can only be applied for the standard grid data. Nowadays, graph convolutional networks (GCN) have achieved great success in spatial-temporal tasks, such as traffic flow forecasting, ride-hailing demand forecasting, etc. (Guo et al. 2019) proposed the ASTGCN model, which uses attention based spatial-temporal graph convolutions to model dynamic spatial-temporal features of traffic flows. ST-MetaNet (Pan et al. 2019) employs a sequence-to-sequence architecture and utilizes a meta graph attention network to model the diverse spatial correlations. ST-MGCN (Geng et al. 2019) deploys multiple graph convolutions to explicitly model the pairwise spatial correlations among regions. In addition, to model the temporal dependencies, it designs Contextual Gated RNN to incorporate the global contextual information. (Song et al. 2020) proposed the STSGCN model, which introduces a spatial-temporal synchronous graph convolution mechanism and achieves the state-of-the-art performance in traffic flow forecasting.

Motivated by the above works, we construct multi-view graphs to represent the similarities among regions from different semantic perspectives, and then we employ multiple graph convolutions to capture the semantic spatial-temporal correlations of traffic accident patterns.

Preliminaries

Definition 1: *Region*. We partition a city into $I \times J$ grids based on the longitude and latitude, where a grid i represents a region and all regions have the same size. It is noted that, the shape of a city is usually irregular, so only $N (N \leq I \times J)$ regions have road segments. In these N regions we can collect their actual features and traffic accident data, while in other regions we set zero values for their features.

Definition 2: *Traffic Accident Type*. According to the number of casualties in traffic accidents, we define three traffic accident types, i.e., minor accidents, injured accidents and fatal accidents, and corresponding risk values are set to be 1, 2 and 3 respectively.

Definition 3: *Traffic Accident Risk*. Let $\mathbf{Y}_t^i$ be the sum of traffic accident risk values in region i at time interval t. For example, $\mathbf{Y}_t^i$ is 5 if two minor accidents and one fatal accident have happened in region i at time interval t.

Definition 4: *Similarity Graph*. To describe the similarities among regions in a city from different semantic aspects, We define three undirected graphs: (1) Risk Similarity Graph $\mathcal{G}_K = (V, E_K, \mathbf{A}_K)$, which represents the similarity of traffic accident risk patterns among regions; (2) Road Similarity Graph $\mathcal{G}_D = (V, E_D, \mathbf{A}_D)$, which represents the similarity of road characteristics (e.g., the number of roads, road types, etc.) among regions; and (3) POI Similarity Graph $\mathcal{G}_P = (V, E_P, \mathbf{A}_P)$, which represents the POI distribution similarity among regions. Here V is the set of nodes of the three graphs, and each node $i \in V$ represents a region with some road segments. According to Definition 1, $|V| = N$. $E_\star$ denotes the set of edges and $\mathbf{A}_\star \in \mathbb{R}^{N \times N}$ denotes the adjacency matrix of graph $\mathcal{G}_\star$, where $\star \in \{K, D, P\}$.

To determine $\mathbf{A}_\star$, we first need to calculate the similarity score $\mathrm{Sim}_\star(i, j) \in [0, 1]$ between node i and j. Motivated by (Zhou et al. 2020), we utilize the Jensen-Shannon divergence to measure the similarity, and the details will be introduced later in Eq. 6. After obtaining the pairwise similarities of all the nodes, we select top-L most similar nodes for each node as its first-order neighbors. That is, let $e_\star^{i,j} \in E_\star$ if node i and j are first-order neighbors with each other. Then we construct the adjacent matrix $\mathbf{A}_\star$ as follow:

$$\mathbf{A}_\star^{i,j} = \begin{cases} \mathrm{Sim}_\star(i, j), & e_\star^{i,j} \in E_\star, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Problem Statement: *Traffic Accident Risk Forecasting*. Let $\mathbf{X}_t \in \mathbb{R}^{I \times J \times d_r}$ denote the grid features of all the regions at time interval t, including the information of weather, POI, traffic flow and traffic accident risk, where d_r is the dimension of region features. Let $\mathbf{S}_t \in \mathbb{R}^{N \times d_g}$ denote the signal matrix of the three graphs at time interval t. Each row represents a node's features, including the values of traffic flow and traffic accident risk, where d_g is the dimension of node features. Let $\mathbf{z}_t \in \mathbb{R}^{d_t}$ be the time information of time interval t, including hour of day, day of week and if it is a holiday, where d_t is the dimension of time features. Given the historical observations of region features $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_T)$, graph signal matrices $(\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_T)$ and $\mathbf{z}_{T+1}$, our goal is to predict the traffic accident risk at the next time interval, i.e., $\mathbf{Y}_{T+1} \in \mathbb{R}^{I \times J}$.

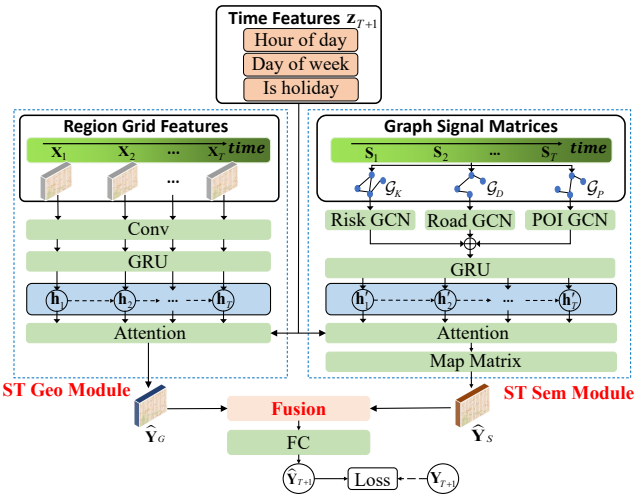

Figure 2: Architecture of GSNet.

The GSNet Model

The main idea of our proposed model is to fuse the geographical and semantic spatial-temporal features to improve the accuracy of traffic accident risk forecasting. Figure 2 presents the architecture of our GSNet model, which mainly consists of two modules, i.e., Spatial-Temporal Geographical Module and Spatial-Temporal Semantic Module. Specifically, the Spatial-Temporal Geographical Module takes spatial-temporal grid features and time features as input, and it uses convolution, GRU and temporal attention to model the geographical spatial-temporal correlations among regions. The Spatial-Temporal Semantic Module takes the graph signal matrices and time features as input, and it employs multiple GCN, GRU and temporal attention to capture semantic spatial-temporal correlations among regions. Finally, the outputs of the two modules are fused dynamically to make the final prediction.

Spatial-Temporal (ST) Geographical (Geo) Module

The ST Geo Module, as shown in the left part of Figure 2, intends to capture the geographical spatial-temporal correlations among regions. It first utilizes convolutions to model the geographical spatial correlations, and then uses GRU and temporal attention mechanism to dynamically capture the short-term and long-term temporal correlations.

Geographical Spatial Convolutions Traffic accidents usually exhibit complex spatial correlations among regions. For example, the accidents of two nearby regions tend to be strongly correlated in peak hours due to tidal flows. We utilize convolutions to capture such geographical spatial correlations. The convolution at time interval t is described as:

$$\mathbf{X}_t^k = f(\mathbf{W}_t^k * \mathbf{X}_t^{k-1} + \mathbf{b}_t^k), \tag{2}$$

where * represents convolution operation, and $\mathbf{W}^k$, $\mathbf{b}^k$ are learnable parameters of the k-th convolutional layer. Note

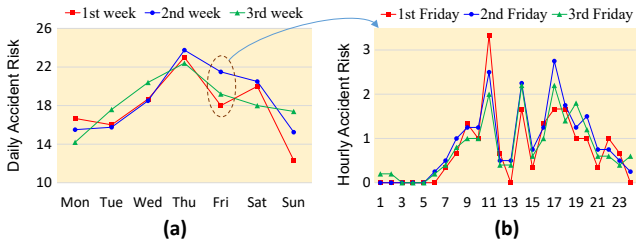

Figure 3: (a) Daily traffic accident risk curves of a region over three consecutive weeks; (b) Hourly traffic accident risk curves of the region on three consecutive Fridays.

that the convolutions for all time intervals share the same parameters. $f(\cdot)$ is the ReLU activation function. $\mathbf{X}_t^k$ represents the output of the k-th convolutional layer at time interval t, and $\mathbf{X}_t^0 = \mathbf{X}_t$. After K convolutional layers, the output is denoted as $\mathbf{X}_t^K \in \mathbb{R}^{I \times J \times d_K}$, where d_K is the number of convolution kernels in the K-th convolutional layer. We use $\mathbf{X}'_t$ to represent $\mathbf{X}_t^K$ for short in the following sections.

Temporal Representation of Geographical Spatial Features Besides the geographical spatial correlations, traffic accidents usually have short-term proximity and long-term periodicity in the temporal dimension. For example, traffic accidents are usually affected by the traffic flows and road conditions of a region in the recent time periods. And for observing the long-term periodicity, we chart the daily traffic accident risk curves of a region over several consecutive weeks, as shown in Figure 3(a), and the hourly risk curves of the region on several consecutive Fridays, as shown in Figure 3(b). We can find that the traffic accidents in this region have a strong weekly periodicity. Therefore, it is essential to model the short-term and long-term temporal correlations for accurate forecasting.

To capture the short-term proximity and long-term periodicity of the geographical spatial features, we fetch temporal data from the recent p time intervals and the same time interval in the previous q weeks to form a sequence of data $\mathbf{X}_1 \ldots, \mathbf{X}_q, \ldots, \mathbf{X}_T$ $(T = p+q)$, which is taken as the input of the ST Geo Module. In our model, we use GRU to capture the underlying temporal correlations of traffic accidents:

$$\mathbf{h}_t^i = \mathrm{GRU}(\mathbf{x}_t'^i, \mathbf{h}_{t-1}^i), \tag{3}$$

where $\mathbf{x}_t'^i \in \mathbb{R}^{d_K}$ denotes the output of the geographical spatial convolutions of region i at time interval t, $\mathbf{h}_t^i \in \mathbb{R}^{d_h}$ denotes the hidden states of region i at t, and d_h is the number of hidden units. Note that the GRUs for all the regions share the same parameters. Let $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_T]$, in which $\mathbf{h}_t \in \mathbb{R}^{I \times J \times d_h}$ is the hidden states of all the regions.

In reality, different historical data have different influences on the target time interval, and the influences vary over time. Motivated by STG2Seq (Bai et al. 2019), we introduce a temporal attention mechanism to adaptively capture the dynamic correlations in the temporal dimension by computing the attention scores between $\mathbf{H}$ and the time features $\mathbf{z}_{T+1}$ at the target time interval $T+1$:

$$\alpha = \mathrm{softmax}(\mathrm{ReLU}(\mathbf{H}\mathbf{W}_H + \mathbf{z}_{T+1}\mathbf{W}_\mathbf{z} + \mathbf{b}_\alpha)), \tag{4}$$

where $\mathbf{W}_H \in \mathbb{R}^{d_h \times 1}$, $\mathbf{W}_\mathbf{z} \in \mathbb{R}^{d_t \times T}$ and $\mathbf{b}_\alpha \in \mathbb{R}^T$ are learnable parameters. $\alpha \in \mathbb{R}^T$ is the temporal attention score vector, which indicates the importance distribution of different historical time intervals on the target time interval. Finally we get the output of the ST Geo Module by dynamically merging the hidden temporal information:

$$\hat{\mathbf{Y}}_G = \sum_{i=1}^{T} \alpha_i \cdot \mathbf{h}_i, \tag{5}$$

where $\hat{\mathbf{Y}}_G \in \mathbb{R}^{I \times J \times d_h}$.

Spatial-Temporal (ST) Semantic (Sem) Module

The ST Sem Module, as shown in the right part of Figure 2, is used to capture spatial-temporal correlations from different semantic aspects. It employs three graph convolution networks to model three kinds of spatial correlations respectively, i.e, risk similarity, road similarity and POI similarity. Like the ST Geo Module, it then utilizes GRU and temporal attention to capture short-term proximity and long-term periodicity of semantic spatial features. Finally, the map matrix is used to map the graph data into grid data.

Semantic Spatial Multi-Graph Convolutions To capture the three kinds of spatial correlations from different semantic aspects, we construct three types of similarity graphs, including risk similarity graph $\mathcal{G}_K$, road similarity graph $\mathcal{G}_D$, and POI similarity graph $\mathcal{G}_P$.

Here we introduce how to construct these similarity graphs. Firstly, we calculate the risk, road and POI similarity scores between any two nodes. In this study, we use the Jensen-Shannon divergence (Zhou et al. 2020) to measure the similarity. Taking the POI similarity as an example, the calculation method is as follows:

$$\begin{aligned} &\mathrm{Sim}_P(i,j) = 1 - \mathrm{JS}(\mathbf{R}_P^i, \mathbf{R}_P^j), \\ &\mathrm{JS}\left(\mathbf{R}_P^i, \mathbf{R}_P^j\right) = \frac{1}{2} \sum_{1 \le l \le q} \left(\begin{array}{l} \mathbf{R}_P^i(l) \log \frac{2\mathbf{R}_P^i(l)}{\mathbf{R}_P^i(l)+\mathbf{R}_P^j(l)} + \\ \mathbf{R}_P^j(l) \log \frac{2\mathbf{R}_P^j(l)}{\mathbf{R}_P^i(l)+\mathbf{R}_P^j(l)} \end{array} \right), \end{aligned} \tag{6}$$

where $\mathbf{R}_P^i, \mathbf{R}_P^j \in \mathbb{R}^q$ denote the POI distribution of region i and j, which the sum is 1. $\mathbf{R}_P^i(l)$ means the l-th dimension of $\mathbf{R}_P^i$. Similarly, we calculate $\mathrm{Sim}_K(i,j)$ through the risk vectors of regions in all time intervals, and $\mathrm{Sim}_D(i,j)$ through the road properties vectors, which include the length and width of roads, road types, snow removal priorities, etc. Secondly, we select top-L most similar regions for each region to construct the adjacency matrices $\mathcal{A} = [\mathbf{A}_K, \mathbf{A}_D, \mathbf{A}_P]$.

After constructing the three graphs, multi-graph convolutions are used to model the semantic spatial correlations among regions. We stack two graph convolutional layers. At time interval t, the graph convolution can be described as:

$$\mathbf{S}'_t = \sum_{\star \in \{K,D,P\}} \mathrm{ReLU}(\mathbf{A}_\star \mathrm{ReLU}(\mathbf{A}_\star \mathbf{S}_t \mathbf{W}_\star^{(0)} + \mathbf{b}_\star^{(0)}) \mathbf{W}_\star^{(1)} + \mathbf{b}_\star^{(1)}), \tag{7}$$

where $\mathbf{W}_\star^{(0)} \in \mathbb{R}^{d_g \times d_c}$, $\mathbf{W}_\star^{(1)} \in \mathbb{R}^{d_c \times d_c}$, $\mathbf{b}_\star^{(0)}, \mathbf{b}_\star^{(1)} \in \mathbb{R}^{d_c}$ are learnable parameters. d_c denotes the number of ker-

nels in the graph convolutional operations. d_g is the dimension of node features in $\mathbf{S}_t$. Notice that the graph convolutions for all time intervals share the same parameters. Then, $\mathbf{S}'_t \in \mathbb{R}^{N \times d_c}$ represents the output of the multi-graph convolutions at time interval t.

Temporal Representation of Semantic Spatial Features Similar to the ST Geo Module, we utilize GRU and temporal attention to capture temporal correlations of the semantic spatial features. Likewise, we select the recent p time intervals and the same time interval in the previous q weeks to form the input $\mathbf{S}_1 \ldots, \mathbf{S}_q, \ldots, \mathbf{S}_T$ of the ST Sem Module. Then GRU is used to model temporal correlations:

$$\mathbf{h}'^i_t = \mathrm{GRU}(\mathbf{s}'^i_t, \mathbf{h}'^i_{t-1}), \tag{8}$$

where $\mathbf{s}'^i_t \in \mathbb{R}^{d_c}$ is the output of the multi-graph convolutions of node i at time interval t, $\mathbf{h}'^i_t \in \mathbb{R}^{d'_h}$ denotes the hidden states of i at t, and d'_h is the number of hidden units. The GRUs for all the nodes share the same parameters. Let $\mathbf{H}' = [\mathbf{h}'_1, \mathbf{h}'_2, \ldots, \mathbf{h}'_T]$, in which $\mathbf{h}'_t \in \mathbb{R}^{N \times d'_h}$ is the hidden states of all the nodes.

Then we utilize a temporal attention to dynamically capture the temporal correlations by computing the attention scores between $\mathbf{H}'$ and the time features $\mathbf{z}_{T+1}$:

$$\alpha' = \mathrm{softmax}(\mathrm{ReLU}(\mathbf{H}'\mathbf{W}'_H + \mathbf{z}_{T+1}\mathbf{W}'_{\mathbf{z}} + \mathbf{b}'_\alpha)), \tag{9}$$

where $\mathbf{W}'_H \in \mathbb{R}^{d'_h \times 1}$, $\mathbf{W}'_{\mathbf{z}} \in \mathbb{R}^{d_t \times T}$ and $\mathbf{b}'_\alpha \in \mathbb{R}^T$ are learnable paparemters. $\alpha' \in \mathbb{R}^T$ is the temporal attention score vector, which indicates the importance distribution of different historical time intervals on the target time interval. Then, we get the output of the temporal attention:

$$\mathbf{F} = \sum\nolimits_{i=1}^{T} \alpha'_i \cdot \mathbf{h}'_i, \tag{10}$$

where $\mathbf{F} \in \mathbb{R}^{N \times d'_h}$.

For better fusing the graph signals with the grid features later, we construct a pre-computed map matrix $\mathbf{M} \in \mathbb{R}^{(I \times J) \times N}$ to transform $\mathbf{F}$ into grid format, in which $\mathbf{M}_{i,n} = 1$ if node n corresponds to region i, otherwise $\mathbf{M}_{i,n} = 0$. Finally, we get the output of the ST Sem Module:

$$\hat{\mathbf{Y}}_S = \mathbf{M}\mathbf{F}, \tag{11}$$

where $\hat{\mathbf{Y}}_S \in \mathbb{R}^{I \times J \times d'_h}$.

Feature Fusion

Generally, geographical spatial-temporal correlations and semantic spatial-temporal correlations have different degrees of influence on the target region. Therefore, instead of concatenation, we use two weight matrices and a fully connected layer to dynamically fuse the outputs of the ST Geo Module and ST Sem Module:

$$\hat{\mathbf{Y}} = \mathrm{FC}(\mathbf{W}_1 * \hat{\mathbf{Y}}_G + \mathbf{W}_2 * \hat{\mathbf{Y}}_S), \tag{12}$$

where $*$ denotes convolution operation, and 1×1 convolution kernels are used here. $\mathbf{W}_1$ and $\mathbf{W}_2$ are parameters of convolution kernels. $\mathrm{FC}(\cdot)$ denotes fully-connected layer.

Dataset	NYC	Chicago
Time range	1/1/2013 - 12/31/2013	2/1/2016 - 9/30/2016
#Traffic accidents	147k	44k
#Taxi orders	173,179k	1,744k
#POIs	15,625	None
Hours of weather	8,760	5,832
#Road segments	103k	56k

Table 1: Statistics of Datasets.

After the dynamic fusion, it generates the final prediction $\hat{\mathbf{Y}} \in \mathbb{R}^{I \times J}$, which represents the traffic accident risks at the next time interval.

Loss Function

To address the zero-inflated issue, we design a weighted loss function. In the training phase, the loss function assigns higher weights to the samples with high traffic accident risks to avoid the final prediction toward zero values. Specifically, we classify all samples into four levels by their accident risks, i.e., $\mathbf{I} = \{0, 1, 2, \geq 3\}$. We use $\mathbf{Y}(i)$ to represent the samples whose traffic accident risk level is i. The final loss function is as follow:

$$\mathrm{Loss}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{2} \sum\nolimits_{i \in \mathbf{I}} \lambda_i (\mathbf{Y}(i) - \hat{\mathbf{Y}}(i))^2, \tag{13}$$

where $\mathbf{Y}$ denotes the ground truth, and $\hat{\mathbf{Y}}$ is the prediction of the model. λ_i is the weight of the samples with traffic accident risk level i, which is a hyperparameter.

Experiments

Datasets

We use two public real-world datasets collected from NYC[2] and Chicago[3]. The statistics of the datasets are shown in Table 1. The traffic accident data includes location, time and the number of casualties. The taxi order data includes pick-up time, longitude and latitude, and drop-off time, longitude and latitude. The POI data includes seven categories: residence, school, culture facility, recreation, social service, transportation and commercial. The weather data includes temperature and sky condition. The road segment data includes road length, width, type and snow removal priority.

For the NYC dataset, we respectively construct the risk, road and POI similarity graphs based on its traffic accident, road segment and POI data. For the Chicago dataset, due to the lack of POI data, we only construct the risk and road similarity graphs.

Settings

We partition all data on the time axis with ratio $6 : 2 : 2$ into training, validation and test set. The whole city is split into rectangle regions, and the size of a region is about $2km \times 2km$. The length of time interval is set as 1 hour.

We implement our GSNet model in PyTorch. All data is normalized into the range $[0, 1]$ by Max-Min normalization.

[2]https://opendata.cityofnewyork.us/

[3]https://data.cityofchicago.org/

When evaluating, we denormalize the prediction into the normal values. The hyperparameters are determined based on the model's performance on the validation data. For the length of short-term and long-term historical data, we set $p = 3$ and $q = 4$. When constructing similarity graphs, L is set to 10. In geographical spatial convolutions, $K = 2$ and the convolution kernel size is 3×3. In semantic spatial multi-graph convolutions, we stack 2 graph convolutional layers with 64 filters for each. In GRU, d_h and d'_h are both set to 256. Additionally, the batch size is 32 and the learning rate is $1e^{-5}$. In the loss function, the sample weights $\lambda_{i\in\mathbf{I}}$ are respectively set to 0.05, 0.2, 0.25 and 0.5.

Metrics

We evaluate the performance of GSNet from two perspectives, including regression and ranking. In the regression perspective, we use RMSE to evaluate the predicted risks of all regions. In the ranking perspective, inspired by (Ma et al. 2018), we use Recall and MAP to evaluate the percentage of accurate predictions of the regions with high traffic accident risks. For time interval t ($1 \leq t \leq T$), if there are traffic accidents in k_t regions, Recall indicates the percentage of the intersection of true k_t regions and top k_t regions with highest predicted risks. MAP indicates the mean average precision, and the rank of predicted k_t regions are more relevant to that of true k_t regions. Lower RMSE values indicate the model predicts more accurate risks in all regions, while higher Recall and MAP values indicate the model performs better in high-risk regions, which means the model can figure out more high-risk regions.

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum\nolimits_{t=1}^{T}\left(Y_t - \hat{Y}_t\right)^2}, \tag{14}$$

$$\text{Recall} = \frac{1}{T}\sum\nolimits_{t=1}^{T}\frac{S_t \cap R_t}{|R_t|}, \tag{15}$$

$$\text{MAP} = \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{j=1}^{|R_t|} pre(j) \times rel(j)}{|R_t|}, \tag{16}$$

where Y_t is the ground truth and $\hat{Y}_t$ is the predicted values of all regions at time interval t. R_t is the set of regions where traffic accidents have really occurred at t. S_t is a set of regions with top $|R_t|$ highest predicted risks. $pre(j)$ denotes the precision of a cut-off rank list from 1 to j. $rel(j)$ is the recall of region j, where $rel(j) = 1$ if there are traffic accidents in this region, otherwise $rel(j) = 0$.

To evaluate the model's performance more comprehensively, we use RMSE, Recall and MAP to indicate the performance on all time intervals, i.e., 0:00-24:00. Additionally, we use RMSE*, Recall* and MAP* to indicate the performance on the time intervals with high frequency of traffic accidents, i.e., 7:00-9:00 and 16:00-19:00.

Baseline Methods

We compare our model with the following 8 baselines:

- **HA**: Historical Average. The average value of the traffic accident risks at the same time interval in short-term and long-term historical data is taken as prediction.
- **XGBoost** (Chen and Guestrin 2016): XGBoost is a powerful model based on the boosting tree.
- **MLP** : Multiple Layer Perceptron. We select ReLU as its activation function.
- **GRU** (Chung et al. 2014): Gated Recurrent Unit. It is a time series prediction model, which can capture historical temporal dependencies.
- **SDCAE** (Chen et al. 2018): Stack Denoise Convolutional Autoencoder. It captures the geographical spatial features by stacking multiple denoise convolution layers to predict traffic accident risks in the city scale.
- **ConvLSTM** (Shi et al. 2015): It combines CNN and LSTM to model both spatial and temporal dependencies.
- **Hetero-ConvLSTM** (Yuan, Zhou, and Yang 2018): Heterogeneous ConvLSTM. It uses a moving window to get subsets of spatial regions to capture the heterogeneity of rural and urban regions.
- **Graph WaveNet** (Wu et al. 2019): A deep learning model that stacks graph convolution layers to predict spatial-temporal graph data with long-range temporal sequences.

Experiment Results

Table 2 shows the prediction performance of different methods on the NYC and Chicago datasets. It can be seen that our GSNet model achieves the best performance on the both datasets in terms of all metrics. Specifically, we can observe that HA and XGBoost don't perform well, due to their limited ability of modeling complex dependencies of spatial-temporal data. Compared with those conventional methods, deep learning-based models achieve better performances. GRU can capture short-term and long-term temporal features, while SDCAE can model spatial correlations by stacking convolutional layers. But they cannot model the spatial and temporal dependencies simultaneously. Compared with them, ConvLSTM and Hetero-ConvLSTM combine CNN and LSTM to model the geographical spatial-temporal correlations and make further improvements. But they overlook the semantic similarities among regions. Graph WaveNet considers multiple similarity graphs and employs TCN and GCN to model semantic similarities. However, it is not good at capturing geographical spatial-temporal correlations.

Overall, our GSNet model simultaneously considers the geographical spatial-temporal correlations and the semantic spatial-temporal correlations among regions. Consequently, GSNet achieves the best performance among all the methods. In addition, our model has much better performance in high-risk time intervals against other methods, which further demonstrates the superiority of GSNet in modeling multi-scale spatial-temporal correlations of traffic accident data.

Effects of Different Components

To further illustrate the effectiveness of different components, we design four variants to conduct ablation experiments on the NYC dataset. (i) Concat: Instead of dynamic feature fusion, we simply concatenate the geographical features $\hat{\mathbf{Y}}_G$ and the semantic features $\hat{\mathbf{Y}}_S$; (ii) -Attention: We remove the temporal attention from both the ST Geo and Sem Modules; (iii) -Geo: We remove the ST Geo Module; and (iv) -Sem: We remove the ST Sem Module.

Dataset	Metric	HA	XGBoost	MLP	GRU	SDCAE	ConvLSTM	Hetero-ConvLSTM	Graph WaveNet	**GSNet (ours)**
NYC	RMSE	10.3243	11.0165	8.4289	8.3375	7.9774	7.9505	7.9731	7.7358	**7.5158**
	Recall	24.42%	23.14%	27.28%	28.09%	30.81%	30.99%	30.42%	31.78%	**34.20%**
	MAP	0.1049	0.1008	0.1196	0.1228	0.1594	0.1526	0.1454	0.1623	**0.1861**
	RMSE*	9.4994	10.1730	7.6379	7.3546	7.2806	7.2554	7.2750	7.0958	**6.7206**
	Recall*	26.94%	25.22%	29.51%	30.76%	31.22%	32.61%	31.43%	33.04%	**35.30%**
	MAP*	0.1258	0.1119	0.1338	0.1301	0.1536	0.1557	0.1498	0.1647	**0.1808**
Chicago	RMSE	14.9581	15.6946	12.5116	12.6482	11.3382	11.1309	11.3033	11.0835	**10.5989**
	Recall	13.80%	12.58%	17.53%	17.83%	18.78%	18.84%	18.43%	18.95%	**20.69%**
	MAP	0.0572	0.0545	0.0631	0.0664	0.0753	0.0789	0.0716	0.0805	**0.0903**
	RMSE*	10.2564	10.3685	8.9500	9.0421	8.7543	8.5254	8.5437	8.4484	**8.1496**
	Recall*	15.89%	15.22%	18.93%	18.66%	20.58%	20.30%	18.93%	20.42%	**22.77%**
	MAP*	0.0644	0.0614	0.0748	0.0758	0.1002	0.0925	0.0770	0.0933	**0.1266**

* represents the performance on time interval with high frequency of accidents.

Table 2: Performance comparison of different approaches.

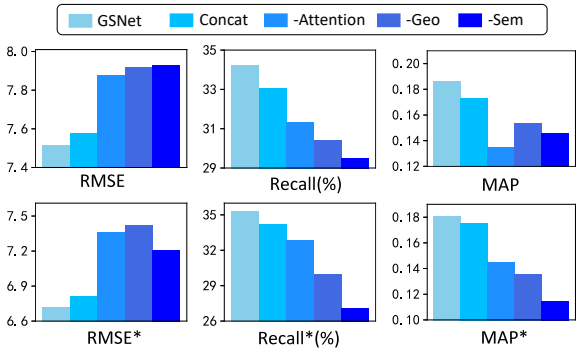

Figure 4: Component analysis of GSNet.

The results are shown in Figure 4. We observe that simple concatenation of features causes the performance worse a little, which demonstrates the dynamic feature fusion is useful. Removing the temporal attention greatly hurts the performance, which proves the effectiveness of dynamically modeling the importance of historical information. The results of the -Geo and -Sem are much worse than the GSNet model, which demonstrates that simultaneously capturing the geographical and semantic spatial-temporal correlations is crucially important for traffic accident risk forecasting.

Effects of Different Similarity Graphs

We also investigate the effects of different similarity graphs on the NYC dataset. We construct three variants (i.e., GSNet-Risk, GSNet-Road, GSNet-POI), each of which only introduces one of the risk, road and POI similarity graphs. And the other settings are the same as GSNet. The results are shown in Table 3. Compared with GSNet-Road and GSNet-POI, GSNet-Risk achieves better performance. It demonstrates that the traffic accident risk of a certain region is more relevant to the regions with similar risk patterns. GSNet-Road outperforms GSNet-POI a little, which reveals regions with similar road features are more likely to share similar accident risk patterns. It also shows roads have more important influence on traffic accidents than POIs which represent urban functionality. Overall, the GSNet achieves the best performance. It demonstrates the necessity of modeling spatial-temporal correlations from multiple semantic aspects.

Model	RMSE/RMSE*	Recall/Recall*(%)	MAP/MAP*
GSNet-Risk	7.7135/6.8858	32.42/33.41	0.1769/0.1710
GSNet-Road	7.7537/6.9000	32.24/32.54	0.1753/0.1695
GSNet-POI	7.8317/6.9748	31.86/32.75	0.1695/0.1648
GSNet	**7.5158/6.7206**	**34.20/35.30**	**0.1861/0.1808**

Table 3: Comparison of different similarity graphs.

Effects of Weighted Loss Function

We further conduct the experiments of the weighted loss function and unweighted loss function on the NYC dataset, with the other setting remaining the same. Table 4 shows that those two loss functions are at the same level in terms of Recall and MAP, but RMSE and RMSE* of the weighted loss function is 26.4% and 28.2% lower than the unweighted loss function. The results further illustrate that the proposed loss function works better in predicting all regions' traffic accident risk and address the zero-inflated issue to some degree.

Model	RMSE/RMSE*	Recall/Recall*(%)	MAP/MAP*
unweighted	10.2113/9.3556	**34.22**/34.46	**0.1872**/0.1771
weighted	**7.5158/6.7206**	34.20/**35.30**	0.1861/**0.1808**

Table 4: Comparison of loss functions.

Conclusion

In this paper, we propose a novel GSNet model for traffic accident risk forecasting. To capture multi-scale spatial-temporal dependencies, we respectively design a Spatial-Temporal Geographical Module to capture the geographical spatial-temporal correlations among regions, and a Spatial-Temporal Semantic Module to describe the semantic spatial-temporal correlations among regions. Besides, we design a weighted loss function to address the zero-inflated issue. The experiments on two real-world datasets show that our model outperforms the state-of-the-art methods. Our proposed model might be able to address other sparse spatial-temporal prediction problems, such as crime prediction.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. 2019JBM024).

References

Bai, L.; Yao, L.; Kanhere, S.; Wang, X.; and Sheng, Q. 2019. STG2Seq: Spatial-Temporal Graph to Sequence Model for Multi-step Passenger Demand Forecasting. In *IJCAI*.

Bao, J.; Liu, P.; and Ukkusuri, S. V. 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. *Accident Analysis & Prevention* 122: 239–254.

Caliendo, C.; Guida, M.; and Parisi, A. 2007. A crash-prediction model for multilane roads. *Accident Analysis & Prevention* 39(4): 657–670.

Chen, C.; Fan, X.; Zheng, C.; Xiao, L.; Cheng, M.; and Wang, C. 2018. Sdcae: Stack denoising convolutional autoencoder model for accident risk prediction via traffic big data. In *2018 Sixth International Conference on Advanced Cloud and Big Data (CBD)*, 328–333. IEEE.

Chen, Q.; Song, X.; Yamada, H.; and Shibasaki, R. 2016. Learning deep representation from big and heterogeneous data for traffic accident inference. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chung, J.; Gulcehre, C.; Cho, K.; and Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning*.

Geng, X.; Li, Y.; Wang, L.; Zhang, L.; Yang, Q.; Ye, J.; and Liu, Y. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 3656–3663.

Guo, S.; Lin, Y.; Feng, N.; Song, C.; and Wan, H. 2019. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 922–929.

Lv, Y.; Tang, S.; and Zhao, H. 2009. Real-time highway traffic accident prediction based on the k-nearest neighbor method. In *2009 international conference on measuring technology and mechatronics automation*, volume 3, 547–550. IEEE.

Ma, C.; Zhang, Y.; Wang, Q.; and Liu, X. 2018. Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 697–706.

Olutayo, V.; and Eludire, A. 2014. Traffic accident analysis using decision trees and neural networks. *International Journal of Information Technology and Computer Science* 2: 22–28.

Pan, Z.; Liang, Y.; Wang, W.; Yu, Y.; Zheng, Y.; and Zhang, J. 2019. Urban traffic prediction from spatio-temporal data using deep meta learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1720–1730.

Ren, H.; Song, Y.; Wang, J.; Hu, Y.; and Lei, J. 2018. A deep learning approach to the citywide traffic accident risk prediction. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 3346–3351. IEEE.

Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, 802–810.

Song, C.; Lin, Y.; Guo, S.; and Wan, H. 2020. Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatial-Temporal Network Data Forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 914–921.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In *IJCAI*, 1907–1913.

Yuan, Z.; Zhou, X.; and Yang, T. 2018. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 984–992.

Zhou, Z.; Wang, Y.; Xie, X.; Chen, L.; and Liu, H. 2020. RiskOracle: A Minute-Level Citywide Traffic Accident Forecasting Framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 1258–1265.