A Deep Learning Approach to the Prediction of Short-term Traffic Accident Risk

Ren Honglei
Beihang University (BUAA)

Jingxin Liu
Beihang University (BUAA)

Yucheng Hu
Hunan University

Jinzhi Lei
Tsinghua University

A Deep Learning Approach to the Prediction of Short-term Traffic Accident Risk

Honglei Ren*, You Song*, JingXin Liu*, Yucheng Hu+, and Jinzhi Lei+

*School of Software, Beihang University, Beijing, China,
songyou@buaa.edu.cn, renhongleiz@buaa.edu.cn, jhljx@buaa.edu.cn
+Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University, Beijing, China,
huyc@tsinghua.edu.cn, jzlei@mail.tsinghua.edu.cn

***Abstract*—With the rapid development of urbanization, the boom of vehicle numbers has resulted in serious traffic accidents, which led to casualties and huge economic losses. The ability to predict the risk of traffic accident is important in the prevention of the occurrence of accidents and to reduce the damages caused by accidents in a proactive way. However, traffic accident risk prediction with high spatiotemporal resolution is difficult, mainly due to the complex traffic environment, human behavior, and lack of real-time traffic-related data. In this study, we collected heterogeneous traffic-related data, including traffic accident, traffic flow, weather condition and air pollution from the same city; proposed a deep learning model based on recurrent neural network toward a prediction of traffic accident risk. The predictive accident risk can be potential applied to the traffic accident warning system. We ranked the predictive power of various factors considered in our model through the method of Granger causality analysis, and established the order of predictive power as traffic flow $>$ traffic accident $>$ geographical position $\gg$ weather + air quality + holiday + time period, which indicate that traffic flow is the most essential factor for the occurrence of traffic accidents. The proposed method can be integrated into an intelligent traffic control system toward a more reasonable traffic prediction and command organization.**

I. Introduction

In modern society, the rapid development of urbanization has resulted in the boom of vehicles, causing a number of problems, such as traffic congestion, air pollution, and traffic accidents. These problems have caused huge economic loss as well as human casualties. According to *Global Status Report on Road Safety*, published by World Health Organization in 2015, about 1.25 million people were killed in traffic accidents every year. With the help of big traffic data and deep learning, real-time traffic flow prediction has enabled people to avoid traffic jam by choosing less congested routes. Big traffic data and deep learning may also provide a promising solution to predict or reduce the risk of traffic accidents.

One important task in traffic accident prevention is to build an effective traffic accident risk prediction system. If a traffic accident risk in a certain region can be predicted, we can disseminate this information to the nearby drivers to alert them or make them choose a less hazardous road. However, accurate prediction of traffic accident risk is very difficult because many related factors could affect traffic accident. For example, different regions have tremendous difference on traffic accident rate. In addition, poor weather condition such as snow or fog can reduce road visibility and traffic capacity, thus increase the change of traffic accidents. Traffic accident rate varies at different time of a day, possibly related to the physical condition of the drivers. Although many researchers have focused on the identification of key factors associated with traffic accident [1], [2], [3], effective prediction of the traffic accident risk dynamically remains to be a challenge problem.

With the development of deep learning, methods based on deep learning and big data have shown favorable results in traffic related problems, such as traffic flow prediction [4], arrival time estimation [5], origin-destination forecasting [6], etc. As for traffic accident risk prediction based on deep learning, to our best knowledge, the only work is done by Chen et. al., who use human mobility features extracted from Stack denoise Autoencoder to infer traffic accident risk in Japan [7]. However, they did not consider several important factors such as traffic flow and weather condition. For example, traffic accidents are more common in rush hours because of the elevated vehicle flow and human flow. Without the information of traffic flow, the predictive power may be weakened. Another important information that they missed is the periodical pattern of traffic accidents. In particular, traffic accidents may closely related to the day of week or the day of a year (such as holidays). Other important factors that may have significant impact on traffic accident include weather condition, air quality and regional characteristic, etc. To improve the power of traffic accident risk prediction, it is important to combine all these factors into a comprehensive prediction model.

In this paper, we collected big and heterogeneous data related with traffic accident and built a deep model for traffic accident risk prediction based on recurrent neural network. The model integrates historical short-term and periodic features of traffic accident, traffic flow, weather condition and air quality. Numerical results based on real traffic data shows that our method can learn the patterns of traffic accident more effectively than traditional machine learning methods and Stack denoise Autoencoder method[7].

In our study, we proposed a deep learning model for short-term traffic accident risk prediction. The model can learn deep connections between traffic accident and closely related factors. The proposed approach was aimed at ranking the

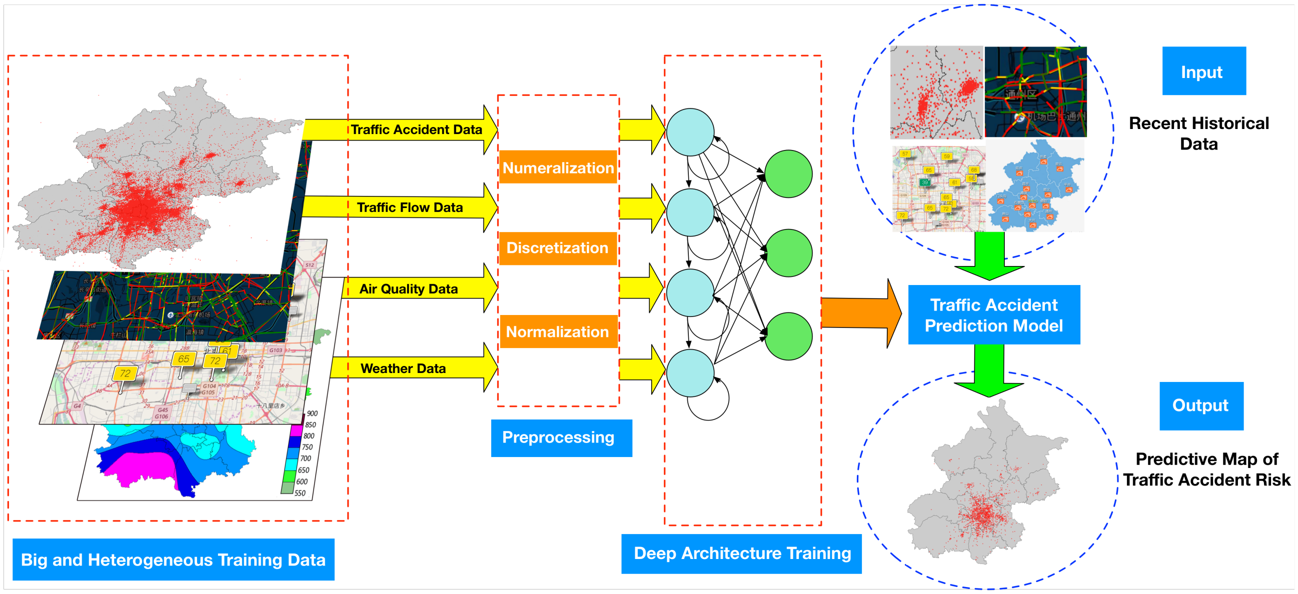

Fig. 1. Workflow of our traffic accident risk prediction method. First, the big and heterogeneous data related with traffic accident is collected. It contains the data of traffic accident, traffic flow, air quality and weather condition. Secondly, the data is preprocessed by numeralization, discretization, normalization, and then feed to the deep model for training. After training, we feed real-time data to the trained model and obtain traffic accident risk prediction.

predictive power of factors that associated with traffic accident according to Granger causality based on real data. As a potential application, the traffic accident prediction system based on our method can be used to help traffic enforcement department to allocate police forces in advance of traffic accidents.

The rest of this paper is organized as follows: Section 2 introduces some previous works that are related with the present one. Section 3 describes the big and heterogeneous data source we used in this paper. Section 4 introduces our deep learning model for traffic accident risk prediction. Section 5 explores the results of experiment and gives the predictability rank of factors that related with traffic accident on the basis of Granger causality. Section 6 gives the conclusions and future works.

II. Related Work

1) Identification of Traffic Accident Trigger: Tremendous efforts have been devoted to the identification of key conditions or particular traffic patterns that could lead to traffic accident. For instance, Oh proposed the assumption that disruptive traffic flow is a trigger to crash [2]. Based on the loop detector data and crash data, they found that 5-min standard deviation of speeds right before a traffic accident is an effective indicator of crash. Abdel-Aty took the spatial variation of traffic patterns into account and found that 5-min average occupancy from upstream and 5-min coefficient of variation in speed from downstream of the accident place can be indicators of accident [3]. Although different crash indicators have been proposed, they could not meet the requirement of accurate accident prediction because numerous factors have complex connections with traffic accidents.

2) Real-time Traffic Accident Prediction: With the development of machine learning, many researchers start to focus on real-time traffic accident prediction. Lv chose feature variables based on Euclidean metric and utilized k-nearest neighbor method to predict traffic accident [8]. Hossain proposed a method on the basis of Bayesian network to infer highway crash in real-time [9].Park collected big traffic accident data of highway in Seoul and build a prediction workflow based on k-means cluster analysis and logistic regression [10]. Recently, Chen used human mobility data in Japan and build a Stack denoise Autoencoder to infer the real-time traffic risk [7]. One limitation of these works is that, they did not incorporate several importance factors such as traffic flow, weather condition, air quality into their model. Without these information, the predictive power of the model could be weakened.

3) Deep Learning: The success of deep learning has proved its power in discovering intricate structures in high-dimensional data [11]. It has been widely used as the state-of-the-art technique in image recognition [12], speech recognition [13], natural language understanding [14], etc. Early researches of learning systems based on neural network can date back to 1966 [15]. Later, with the help of back propagation algorithm [16] and convolutional neural network (also known as CNNs), neural network successfully used in hand-written digit recognition [17]. However, the requirement of huge amount of training data restricted the application of neural network in limited and certain tasks. With the development of General Purpose Graphics Processing Unit (GPGPU) and available data, CNNs start to show its capability in image by making breakthrough in ImageNet competition in 2012. After then, it has become the dominant technique in image related fields. Like the success of CNNs, another kind of neural network –

Recurrent Neural Network (RNN for short) becomes popular in tasks associated with sequence inputs, such as speech and language. Nevertheless, the traditional RNN has the difficulty in learning and storing long information. Therefore, several improved neural network with explicit memory is proposed later, such as Long Short-Term Memory network (also known as LSTM) [18], neural Turing machine [19] and memory network [20]. As for researches on intelligent transportation system, a number of studies focus on traffic flow prediction based on deep learning [21], [4]. In a longer time scale, some studies try to predict the congestion evolution of large-scale transportation network [22]. Another interesting application utilized deep reinforcement learning to control the timing of traffic signal [23].

III. Big and Heterogeneous Data

In this study, to predict traffic accident risk, big and heterogeneous data related with traffic accident was collected. Details of the data are as follows:

- Traffic accident data: The traffic accident records of Beijing in the year 2016 was collected. Each record contains the time, severity level and GPS coordinate of the accident event. The severity level can be graded as slight injury (level 1), heavy injury (level 2) or fatal (level 3).
- Traffic flow data: GPS (Global Positioning System) record of positions and speed information of all taxis in Beijing throughout August 2016.
- Weather data: Verbal description of daily weather data of Beijing was collected from January 1, 2016 to December 31, 2016.
- Air quality data: Daily PM2.5 (Particulate Matter 2.5) index data of Beijing with the time-span the same as that of the weather data.

IV. Deep Model of Traffic Accident Risk Prediction

As Chen et. al. has documented, after some analysis of traffic accident data, we find it is difficult to predict whether traffic accident will happen or not directly, because complex factors can affect traffic accident, and some factors, such as the distraction of drivers, can not be observed and collected in advance [7]. Therefore, as Chen et. al. did, we try to predict the risk of traffic accident as well. Based on the consideration that the frequency and severity of traffic accident can reflect the risk in some degree, we define risk level, in the same way as Chen et. al, as the sum of severity level in each traffic accident record. For instance, risk level is 5 if 1 fatal accident and 2 slightly injury accidents happened in a region. Hence, hot spots of traffic accident can be detected if they show high risk levels.

To achieve the goal of traffic accident risk prediction, we collected big and heterogeneous data related with traffic accident and built a real-time prediction model based on deep learning. Our method is illustrated in Figure 1. First, we preprocessed the heterogeneous data to transform it into a processable format for machine learning. Then we constructed a deep model on the basis of recurrent neural network to infer traffic accident risk. After training the data, we input the recent historical data of traffic accident, traffic flow, weather and air quality into the trained model. The recent historical data means the data of last a few hours, yesterday and last week. By feeding the recent historical data into model, we can obtain a prediction of traffic accident risk from the output of model.

A. Preprocessing

Before we build deep learning model and start traffic accident risk prediction, a proper data structure is necessary. Therefore, we first preprocess our raw data by numeralization, discretization and normalization.

The traffic accident data was first discretized in space and time. The temporal resolutions were 30 or 60 minutes for different time horizon of prediction, and spatial resolution dimension was 1000m×1000m in uniform grids. After discretization, we obtained a matrix S whose element $S_{r,t}$ is all the severity levels of traffic accidents happened within region r and time slot t.

After we have obtained matrix S, we define risk level matrix R whose element $R_{r,t}$ is:

$$R_{r,t} = \sum_{i=1}^{n} S_{i,r,t} \tag{1}$$

where n is the number of traffic accidents happened in region r at time slot t, and $S_{i,r,t}$ is the severity of i-th traffic accident. For convenience, we denote $R_{r,t}$ as $R_r(t)$ hereafter.

For traffic flow data, from the taxi GPS data we obtained the average vehicle speed on each grid at every 5 minutes. This is done by first assigning each GPS record to its corresponding grid (the same uniform grids as the one used previously) and time slot, and then averaging the speeds of each records within one grid to get the average speed. The speed matrix, denoted by V, represents the mean speed of taxis in each region and time slot.

Because patterns of traffic flow and traffic accident in holidays are quite different from that in weekdays, besides matrixes of traffic accident and traffic flow, we also constructed a matrix H to label whether the time slot t belongs to a holiday or not. Holidays include Lunar New Year, Spring Festival, QingMing Festival (also known as Tomb Sweeping Day) , Labor Day, Dragon Boat Festival, Chinese National Day and Mid-Autumn Festival. In addition, traffic accident patterns also change drastically in the time of the day. For example, traffic accident is more frequent at rush hours than that at off-peaks. According to working time pattern and Chinese lifestyle, we divided the time of one day into 7 blocks based partially on [1]: 00:00–06:59 (mid-night to dawn), 07:00–08:59 (morning rush hours), 09:00–11:59 (morning working hours), 12:00–13:59(lunch break), 14:00–16:59 (afternoon working hours), 17:00–19:59 (afternoon rushing hours), and 20:00–23:59 (nighttime).

For weather data, the raw data we collected is daily weather description, such as cloudy, rainy, etc. Because literal data

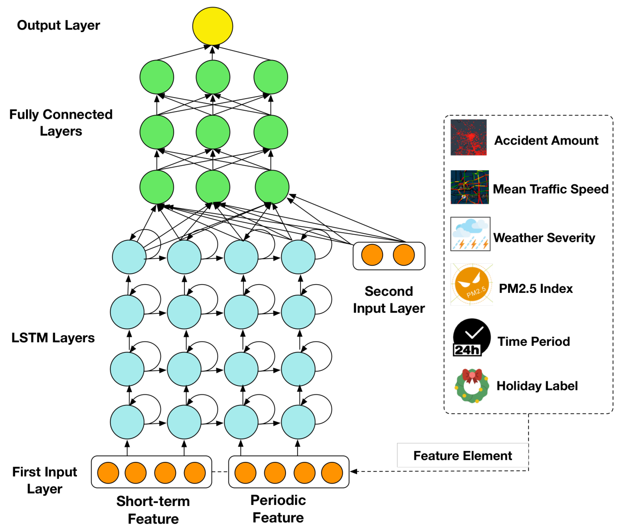

Fig. 2. The Deep Model of Traffic Accident Risk Prediction Method based on LSTM. The left part of figure shows the architecture of neural network, which is consisted of 2 separated input layers (the first input layer and the second input layer), 4 LSTM layers, 3 fully connected layers and 1 output layer. The first input layer is made up of short-term feature and periodic feature, and each feature contains several feature elements. As shown in right part, a feature element is a vector which includes the preprocessed data: traffic accident amount, mean traffic speed, weather severity, normalized PM2.5 index, time period, holiday label. The second input layer which shown in middle of figure is consisted of normalized longitude and latitude of the region center that we want to predict.

like this is difficult to process, we transformed them into quantitative values. The processing details can be found in supplemental material [24]. On the basis of date correspondence between traffic accident matrix S and weather data, we labeled each element of S with weather severity and aggregate the labels into a weather severity matrix W. For instance, element $W_r(t)$ means the weather severity in region r and time slot t.

For air quality data, we normalized it by dividing max PM2.5 index (500), i.e., $normalized\ PM2.5\ index = PM2.5\ index/500.0$. Similar to weather severity matrix, we utilized the normalized data and built an air quality matrix A. For instance, element $A_r(t)$ means normalized air quality index in region r and time slot t.

After preprocessing, we obtained 6 matrices, namely the risk level (R), traffic flow (V), holiday (H), time period (P), weather severity (W) and air quality (A). Aggregating these matrices together according to their temporal and spatial correspondence, a new integrated matrix I is constructed. For a particular region r and time slot t, $I_r(t) = [R_r(t), V_r(t), H_r(t), P_r(t), W_r(t), A_r(t)]^T$.

B. Model

In this subsection, we will introduce our Traffic Accident Risk Prediction Method based on LSTM (TARPML), and Figure 2 illustrates our deep model of TARPML. The input layers are consisted of two parts, namely, the first input layer and the second input layer . The first input layer is a vector T, and it aggregate the short-term and periodic feature elements, which includes traffic accident amount, mean traffic speed, weather severity, normalized PM2.5 index, time period and holiday label. The second input layer contains the normalized longitude and latitude of the region center that we expected to predict, and it directly input into fully connected layers rather than LSTM layers. The hidden layers of deep model is consisted of 4 LSTM layers and 3 fully connected layers sequentially. The last layer of model is output layer, which outputs the risk level for given input features.

The reason why we chose LSTM is that LSTM can capture the periodic feature of traffic accident, and traditional RNNs shows poor performance and intrinsic difficulties in training when it has long time period. These weaknesses have been proved in researches related with traffic flow prediction[22]. On another hand, the explicit memory cell in LSTM can avoid the problems of gradient vanish or gradient explosion existed in traditional RNNs. The structure of LSTM is similar to traditional RNNs, and it consisted of one input layer, one or several hidden layer and one output layer. The core concept of LSTM is its memory cell in hidden layer, it contains 4 major parts: an input gate, a neuron with a self-recurrent link, a forget gate and an output gate, and its inner structure is shown in Figure 3. More details about LSTM can be found in [18].

Later subsections will demonstrate short-term feature, periodic feature, geographical position information of inputs and the details of deep model.

1) Short-term Feature: Short-term features reflect the traffic related conditions that happened several hours right before the time to be predicted. According to the matrix introduced above, our short-term feature of input can be denoted as $T_s = [I_r(t-n), I_r(t-n+1), ..., I_r(t-1)]^T$. The symbol of n is the picked time period length, and $I_r(t)$ is the element of matrix I, which represents the traffic accident count, traffic flow, weather condition, air quality, time slot and whether its time slot is in holidays and festivals.

2) Periodic Feature: Traffic phenomenons, such as traffic flow or traffic accident, usually repeat some patterns at the same time of different periods. The length of this period may be a day or a week because people often commute in a fixed pattern every working day, and in weekends traffic patterns show a obvious periodicity, which is distinct from that in weekdays. This historical and periodic traffic patterns in a day or a week can be utilized to improve traffic accident risk prediction.

In this subsection, we proposed the concept of periodic feature as another input of LSTM beside short-term feature, and this periodic feature is consisted of daily periodic feature and weekly periodic feature. Suppose we need to predict whether traffic accident will happen in region r and time slot t, then the daily periodic feature can be denoted as $T_d = [I_r(t^d - n^d), I_r(t^d - n^d + 1), ..., I_r(t^d + n^d)]^T$. Similarly, the weekly periodic feature can be expressed as $T_w = [I_r(t^w - n^w), I_r(t^w - n^w + 1), ..., I_r(t^w + n^w)]^T$. In the expressions, t_d and t_w denote the same time slot of t

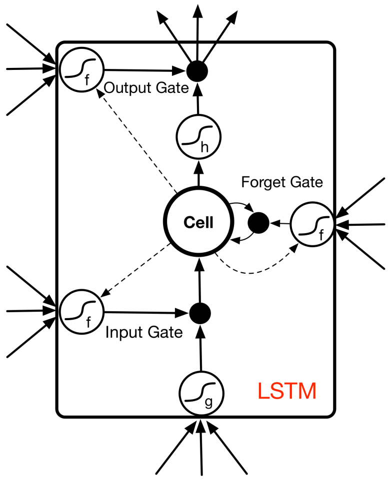

Fig. 3. The structure of LSTM Cell, which is consisted of an input gate, a neuron with a self-recurrent link, a forget gate and an output gate.

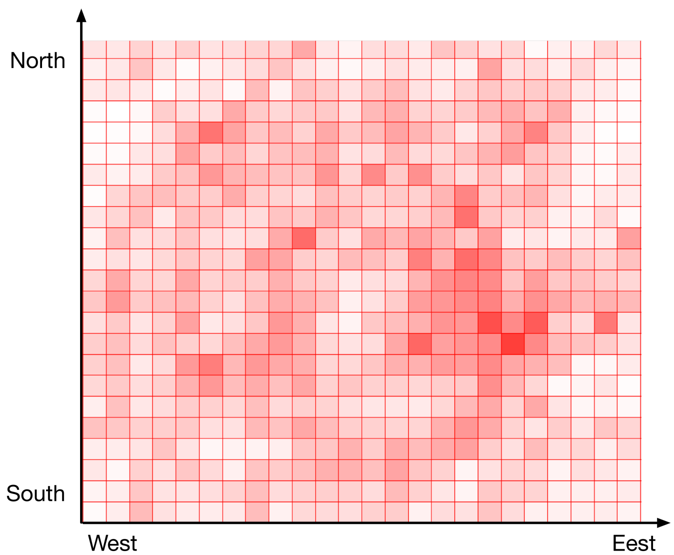

Fig. 4. The heatmap of traffic accident in Beijing in 2016 with 1000m*1000m spatial resolution. Deeper red indicates higher frequencies of traffic accident.

in yesterday and last week respectively, and n_d and n_w are selected time lags of daily periodicity and weekly periodicity respectively.

3) Geographical Position Information: To explore whether traffic accident is highly associated with region characteristics, we plot the heatmap of traffic accident in Beijing in 2016 (Figure 4). As shown in Figure 4, the traffic accident frequency is not uniform distributed, and it is highly related with geographical region. Therefore, we decided to add normalized latitude and longitude of the region center as a geographical position information into our model.

4) Deep Model: We sequentially concatenate short-term features and periodic features into a feature vector and input it into the first input layer, and the vector can be represented as $T = [T_s, T_d, T_w]^T$. The geographical position information was input into the second input layer, which directly connected with fully connected layer. Our deep model contains 4 LSTM layers and 3 fully connected layers orderly. To avoid overfitting, we add a dropout layer with 0.5 dropout rate between each two fully connected layers [25]. The activation function of fully connected layers and output layer is Rectified Linear Units (RELU), which can be denoted as $\max(0, x)$ mathematically.

V. Experiments and Results

In this section, we compare TARPML with several traditional machine learning methods, including Support Vector Machine, Logistic Regression, Decision Tree, and a Stack denoise AutoEncoder method based on deep learning [7]. In addition, we utilized Granger casualty to rank the predictive power of different factors and features. All of our experiments are performed by a PC (CPU: Intel Xeon(R) CPU E5-2609, 32GB memory, GPU: Tesla K20C).

A. Experimental Setup

Because the traffic flow data has the least time horizon (August, 2016), we chose the data of matrix I with the same time horizon as that of traffic flow data for experiment. In addition, since our model is temporal related, we arrange the data chronologically. After that, we chose top 80% data for training, and the rest is for testing. From the training data, we selected 20% of the last part as validation data. The sample size of training, validation, testing is 333,152, 65,888 and 134,560 respectively.

After hyper-parameters selection by using GridSearchCV method of scikit-learn [26], we chose 4 as time window size n, and it represents 4 hours historical data right before the time of prediction is used to infer next 1 hour traffic accident. The time lags of daily periodicity n_d and weekly periodicity n_w are set as 3, which represents the data related with traffic accident before and after 3 hours in last day and weekday are used to perform accident prediction. The architecture of TARPML are built upon Keras, which is a Python Deep Learning library [27]. According to the length of input features, the input shape of LSTM layers are set as $(n + 2n_d + 2n_w + 2, 6)$, whose first dimension means the length of input sequence, and the second dimension is the length of each feature. To keep the consistence with the total length of LSTM and second input layer, the number of neurons of each hidden layer is set as $n + 2n_d + 2n_w + 4$. We chose mean squared error as objective function of optimization, and selected RMSProp as the optimizer [28].

B. Performance Evaluation

1) Simulation Results: To evaluate the effectiveness of our model, here we selected different time slots of August 26th (Friday) in 2016 for testing. Figure 5 shows the visualization of our predicted traffic accident risk map (upper row), and real traffic accident risk map (lower row) at different time.

In figure 5, it can be seen three circles in each subfigure. Two red circles are the major commercial and business areas of

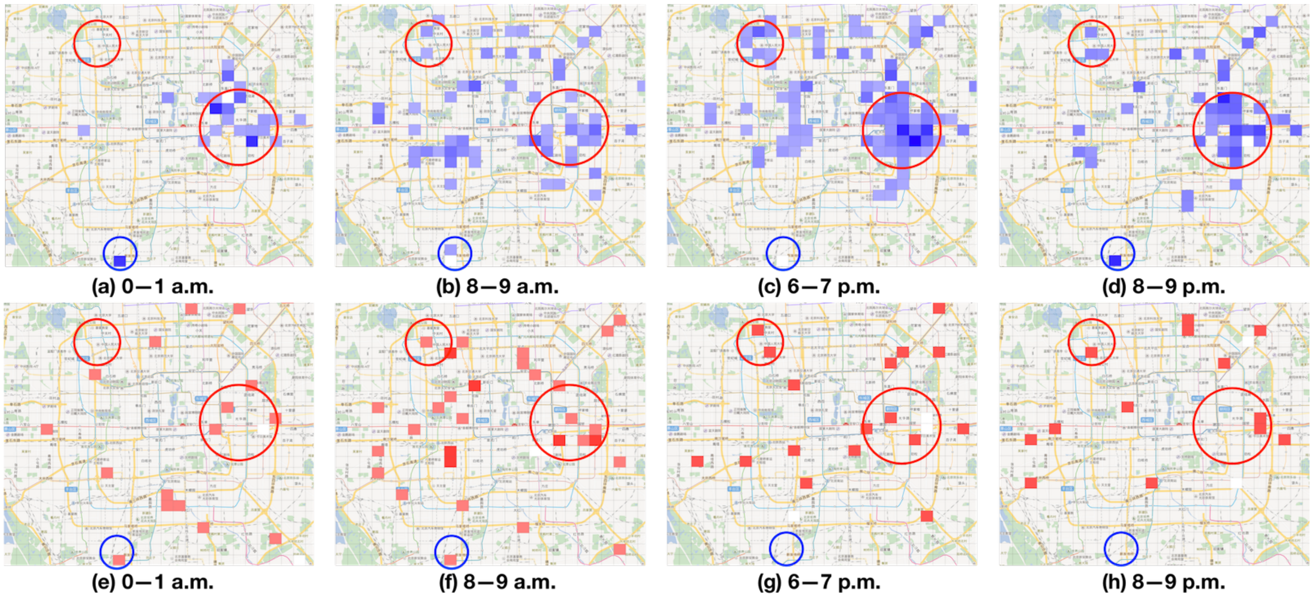

Fig. 5. Visualization of prediction results. These figures illustrate our predicted traffic accident risk map (figure (a)∼(d)), and real traffic accident risk map (figure (e)∼(h)). Deeper blue pixels in predicted risk map represent higher predicted risk level, and deeper red pixels in real risk map mean higher real risk calculated by Eq.(1). The two red circles in each subfigure are the major commercial and business areas of Beijing. The blue circle in each subfigure is an area of a huge farmer's market.

Beijing, the blue circle is a region of a huge farmer's market. The areas inside circles usually face higher traffic accident risk because of the huge vehicle traffic and human volume at rush hours. From figure 5(a) and (d), we can see that the model predicted circle regions will face higher risk. This is consistent with the risk of real data (figure 5(e) and (h)) and common sense, because regions in big red circle are major entertainment areas, where have lots of pubs, nightclubs, shopping malls, and crowded with vehicles and people. Figure 5 (b), (c), (f) and (g) illustrated the predicted and real risk map at morning and afternoon rush hours, and the most of predicted results are coincident with the real data and common sense as well. It can be seen that region in two red circles have more risk areas than other time of the day because a number of working people commute into or out of these regions every workday.

2) Evaluation Metrics: To evaluate the precision of the simulation results, we selected mean absolute error (MAE), mean relative error (MRE) and root-mean relative error (RMSE) as our metrics. They are defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|r_i - \widehat{r_i}| \tag{2}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|r_i - \widehat{r_i}|}{r_i} \tag{3}$$

$$RMSE = \frac{1}{n}\Big[\sum_{i=1}^{n}(r_i - \widehat{r_i})^2\Big]^{\frac{1}{2}} \tag{4}$$

where n is the sample size, r_i and $\widehat{r_i}$ are real and predicted risk level respectively.

3) Baseline Models: We selected several traditional machine learning models and a Stack denoise Autoencoder (SdAE) based on deep learning [7] as our baseline models to compare prediction performance with our TARPML method. The traditional machine learning models we used are Logistic Regression (LR) [29], Support Vector Machine (SVM) [30], Decision Tree (DT) [31].

4) Performance Evaluation: We compared the predictability of our model with that of baselines, and Table I and TableII demonstrates their MAE, MRE and RMSE values at different time horizons (30-min and 60-min). These tables shows that our model outperforms Stack denoise Autoencoder [7], Logistic Regression, Decision Tree and Support Vector Machine, and have smaller prediction errors.

C. Granger Causality Analysis

The black box characteristics of neural network have been criticized by many researchers as it is difficult to interpret how the networks work and what knowledge the network discovered from data. Recent researches have found that, the so called Granger causality, which is capable of characterizing the causality according to incremental predictability, can be used to understand the causal-relationship behind neural networks [32], [4]. In this subsection, we utilized Granger causality analysis to rank the predictive power of different factors and features.

Firstly, we ranked different factors based on their predictive power. These factors are traffic accident (A), traffic flow (F), geographical position (G), weather (W), air quality (Q), time period (T) and holiday (H). Because weather, air quality, time period and holiday are not regional specific factors, we just consider their predictive power in combination, namely

TABLE I
PERFORMANCE FOR 30-MIN PREDICTION

Method	MAE	MRE	RMSE
TARPML	0.55	0.35	0.63
SdAE	0.69	0.49	0.75
LR	0.90	0.52	0.93
DT	1.01	0.65	1.04
SVM	0.95	0.61	0.98

TABLE II
PERFORMANCE FOR 60-MIN PREDICTION

Method	MAE	MRE	RMSE
TARPML	0.58	0.38	0.69
SdAE	0.73	0.51	0.79
LR	0.95	0.55	1.00
DT	1.02	0.69	1.07
SVM	0.97	0.64	1.03

W+Q+T+H. That means these factors are identical in all regions of a city at the same time, and can not reflect regional characteristics to predict risk for different regions. In addition, because the performance of 30-min prediction and 60-min prediction just have little difference, we only chose 60-min prediction as the task of Granger analysis, and compared performances of TARPML with different factors. Table III gives the predictive power of different factors or their combination, it can be noticed that the combination of all factors achieved the best performance, but it just performed a little better than the combination of A+F+G, this indicated that the combination of W+Q+T+H have only a little predictive power. As for predictive power of single factor, traffic flow showed better performance than traffic accident and geographical position. Therefore, we can establish the order of predictive power as traffic flow $>$ traffic accident $>$ geographical position $\gg$ weather + air quality + holiday + time period.

TABLE III
PREDICTIVE POWER OF DIFFERENT FACTORS

Factor	MAE	MRE	RMSE
A	0.66	0.45	0.79
F	0.62	0.42	0.74
G	0.69	0.47	0.83
A+F+G	0.60	0.39	0.70
A+F+G+W+Q+T+H	0.58	0.38	0.69

Secondly, we ranked the predictive power of different features. According to the selected hyper-parameters, we set the time window length of different feature with equal value, namely $n = n_d = n_w = 4$. Then, we performed the prediction with standalone short-term feature (S), daily periodic feature(D) or weekly periodic feature (W) independently, and tested the predictive power of their combination (S+D+W). The results is given in Table IV, and it revealed the following patterns: 1. The predictive power of single feature can be ordered as daily periodic feature $>$ short-term feature $\approx$ weekly periodic feature. It indicated that traffic pattern of the same time last day are more similar and valuable than that of last week or last hours for traffic accident risk prediction. 2. The combination of features led better predictive power than standalone feature.

TABLE IV
PREDICTIVE POWER OF DIFFERENT FEATURES

Feature	MAE	MRE	RMSE
S	0.67	0.44	0.79
D	0.62	0.42	0.74
W	0.68	0.44	0.81
S+D+W	0.58	0.38	0.69

VI. CONCLUSION

In this paper, we collected big and heterogeneous data related with traffic accident, and built a deep learning model with temporal-spatial information for predict traffic accident risk. The visualization of predicted result shows the effectiveness and predictive power of our model. Based on Granger causality method, we ranked the predictive power of traffic related factors as traffic flow $>$ traffic accident $>$ geographical position $\gg$ weather condition + holiday + air quality + time period. We also discovered daily periodic feature have more predictive power than weekly feature and short-term feature. This study therefore indicates that benefits gained from temporal-spatial feature, heterogeneous data and deep recurrent neural network can bring about improvement of traffic accident risk prediction. Our method can be easily applied to the traffic accident warning system and help people avoiding traffic accident by choosing safer regions.

However, due to the complexity of traffic accident, our study has some limitations in following aspects. First, although the factors we utilized in this paper can reveal and predict some traffic accident risk patterns, they are far from complete, and other factors, such as driver behavior, road characteristic, light conditions and special events, are important as well. Second, our prediction results are coarse-grained, and can not provide road level accident risk prediction.Therefore, future work combined with structure of urban road network and comprehensive factors related with traffic accident will be promising to make better prediction result.

REFERENCES

[1] G. Zhang, K. K. W. Yau, and G. Chen, "Risk factors associated with traffic violations and accident severity in China." *Accident; analysis and prevention*, vol. 59, pp. 18–25, Oct. 2013.

[2] C. Oh, J.-S. Oh, S. Ritchie, and M. Chang, "Real-time estimation of freeway accident likelihood," in *80th Annual Meeting of the Transportation Research Board, Washington, DC*, 2001.

[3] M. Abdel-Aty, N. Uddin, A. Pande, F. Abdalla, and L. Hsia, "Predicting freeway crashes from loop detector data by matched case-control logistic regression," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1897, pp. 88–95, 2004.

[4] Y. Wu and H. Tan, "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework," *arXiv preprint arXiv:1612.01022*, 2016.
[5] C. Bai, Z. Peng, Q. Lu, and J. Sun, "Dynamic bus travel time prediction models on road with multiple bus routes," *Computational Intelligence and Neuroscience*, vol. 2015, pp. 432 389–432 389, 2015.
[6] F. Toqué, E. Côme, M. K. El Mahrsi, and L. Oukhellou, "Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks," in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. IEEE, 2016, pp. 1071–1076.
[7] Q. Chen, X. Song, H. Yamada, and R. Shibasaki, "Learning deep representation from big and heterogeneous data for traffic accident inference," 2016.
[8] Y. Lv, S. Tang, and H. Zhao, "Real-time highway traffic accident prediction based on the k-nearest neighbor method," in *Measuring Technology and Mechatronics Automation, 2009. ICMTMA'09. International Conference on*, vol. 3. IEEE, 2009, pp. 547–550.
[9] M. Hossain and Y. Muromachi, "A bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways," *Accident Analysis & Prevention*, vol. 45, pp. 373–381, 2012.
[10] S.-h. Park, S.-m. Kim, and Y.-g. Ha, "Highway traffic accident prediction using vds big data analysis," *The Journal of Supercomputing*, vol. 72, no. 7, pp. 2815–2831, 2016.
[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning." *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
[13] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
[14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.
[15] A. Ivakhnenko, "Cybernetic predicting devices," Tech. Rep.
[16] D. E. Rumelhart, G. E. Hinton, R. J. Williams *et al.*, "Learning representations by back-propagating errors," *Cognitive modeling*, vol. 5, no. 3, p. 1, 1988.
[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[19] A. Graves, G. Wayne, and I. Danihelka, "Neural turing machines," *arXiv preprint arXiv:1410.5401*, 2014.
[20] J. Weston, S. Chopra, and A. Bordes, "Memory networks," *arXiv preprint arXiv:1410.3916*, 2014.
[21] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2015.
[22] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *PloS one*, vol. 10, no. 3, p. e0119044, 2015.
[23] L. Li, Y. Lv, and F.-Y. Wang, "Traffic signal timing via deep reinforcement learning," *IEEE/CAA Journal of Automatica Sinica*, vol. 3, no. 3, pp. 247–254, 2016.
[24] *Supplemental Material, See Supplemental Material for additional results.*
[25] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.
[26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. Oct, pp. 2825–2830, 2011.
[27] F. Chollet, "Keras (2015)," *URL http://keras. io*, 2017.
[28] G. Hinton, N. Srivastava, and K. Swersky, "Rmsprop: Divide the gradient by a running average of its recent magnitude," *Neural networks for machine learning, Coursera lecture 6e*, 2012.
[29] S. H. Walker and D. B. Duncan, "Estimation of the probability of an event as a function of several independent variables," *Biometrika*, vol. 54, no. 1-2, pp. 167–179, 1967.
[30] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
[31] P. E. Utgoff, "Incremental induction of decision trees," *Machine learning*, vol. 4, no. 2, pp. 161–186, 1989.
[32] L. Li, X. Su, Y. Wang, Y. Lin, Z. Li, and Y. Li, "Robust causal dependence mining in big data network and its application to traffic flow predictions," *Transportation Research Part C-emerging Technologies*, vol. 58, pp. 292–307, 2015.