IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 30, 2019, accepted June 13, 2019, date of publication June 19, 2019, date of current version July 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923663

A Special Event-Based K-Nearest Neighbor Model for Short-Term Traffic State Prediction

HAIYANG YU[1,2], NAN JI[1], YILONG REN[1,2], AND CAN YANG[3]

[1]Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China
[2]Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China
[3]Hefei Innovation Research Institute, Beihang University, Beijing 230013, China

Corresponding author: Yilong Ren (yilongren@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51878020.

ABSTRACT Recently, short-term traffic state prediction for urban transportation networks has become a popular topic. However, due to the uncontrollable and unpredictable elements of special events, it is difficult to get abundant data and desired predictions under such condition. As k-nearest neighbor (KNN) has a competitive advantage over other approaches, it could predict traffic state based on a small correlative part of data. Thus, a special event-based KNN (SEKNN) model is proposed for the short-term traffic state prediction with three key points presented in this paper. First, the evolution of the traffic states is redefined as a multipart object, state unit, which includes the benchmark state and the trend vector. Second, to select the nearest neighbors, the state distances of the state units are designed to be compatible with the benchmark states and the trend vectors by fusing the Euclidean distance and the cosine distance. Finally, the prediction results are forced to adjust the benchmark states based on the prediction function using the Gaussian weighted method. The proposed SEKNN is implemented in the district of the Beijing Workers' Stadium (257 links), where special events occur frequently. The results show that the proposed model performs significantly better under special events than the other traditional machine-learning approaches and state-of-the-art deep-learning approaches.

INDEX TERMS Intelligent transportation systems, k-nearest neighbor, short-term traffic state prediction, special events, urban road network.

I. INTRODUCTION

Traffic state prediction contributes to foreknowledge of the variation of traffic states on different future time scales, from minutes to hours or even days. Short-term traffic state prediction is a vital real-time decision-making tool of intelligent transportation systems for traffic managers and travelers who must make decisions in minutes. For instance, in intelligent transportation systems, advanced traffic management systems and advanced traveler information systems depend on timely and accurate predictions of traffic states. There have been numerous studies on this topic, but only a small percentage of them have paid attention to prediction under special events. In addition to traffic accidents, adverse weather and land closures, special events also refer to important unexpected events or incidents that cause nonrecurrent congestion [1]. Although these unexpected events or incidents are relatively rare, they can have a severe negative impact on transportation systems [2]. They often exert heavy pressure on traffic managers owing to the difficulty of managing and controlling traffic under these conditions compared with managing daily recurrent congestion.

Special events are public events, including concerts, sport events, and parades, that may attract large crowds into transportation systems in a certain place [3]. In other words, special events are a special case of travel demand fluctuations in which a significant difference from the typical pattern in the vicinity of the events can be observed [4]. During the short period when large crowds converge on a certain place, the excess travel demand usually induces congestion and then propagates between the bottlenecks in the surrounding transportation network. In this study, we consider the impact of concerts and sport events on the transportation network surrounding Beijing Workers' Stadium in Beijing, China, which is one of most congested districts in Beijing owing to these two kinds of events, according to media reports.

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Awais Javed.

Unlike day-to-day recurrent conditions, traffic states during special events are usually difficult to predict. On the one hand, the influence of special events on transportation systems may be immeasurable because of uncertain and uncontrollable elements. Special events of different types and scales generate different magnitudes of traffic demand. The variation of the duration, timing and location of such events also affects the traffic of the surrounding district [5]. Most advanced commercial solutions perform well in predicting recurrent traffic situations but are incapable of predicting nonrecurrent traffic in advance [6]. On the other hand, the infrequent occurrence of special events increases the difficulty of obtaining adequate traffic data. Previous studies have highlighted the data quality in short-term traffic state prediction [7]. However, there are limited high-quality historical data that can be leveraged to describe dramatic event-caused traffic changes in a timely manner, which indicates that the data have low predictive quality [8]. These limitations have an adverse impact on traditional data-driven prediction approaches.

K-nearest neighbor (KNN) is a data-driven prediction approach, and data quality is a critical factor on prediction performance. Nonetheless, KNN predicts traffic states via the most correlative part of data instead of entire known historical data, which is categorized as instance-based learning. While most of data-driven methods, e.g. deep-learning methods, are constructed based on the entire dataset that the special events data and the general data are mixed together. The prediction performance is influenced by the attributes of datasets, e.g uneven sample distribution. For the short-term traffic state prediction under special events, KNN has the potential to overcome this barrier because KNN has the ability to find out the most similar historical patterns and ignore other dissimilar patterns of the dataset. And it is expected to further enhance the accuracy by adjusting the key parameters and making other improvements. However, in common with most other traditional machine-learning approaches, KNN faces the curse of dimensionality in network-wide traffic prediction [9]. There are hundreds or thousands of links in a transportation network, and the states of these links change every second. Thus, an unexpectedly high-dimension solution space must be searched [10].

In this research, we aim to propose a SEKNN model to alleviate such problems and enhance prediction performance under special events. We attempt to solve these by exactly finding the limited relevant data in historical dataset with less inputs and setting a baseline to avoid the impact due to few historical data under special events. There are three key points in the proposed model for the purposes. First, the evolution of traffic states is redefined as the state unit, a multipart object, consisting of the benchmark state and trend vector. The state unit contributes to reducing the dimensions of input and highlighting the dramatic traffic tendencies. Second, to select the nearest neighbors, the state distances of the state units are designed to be compatible with the benchmark state and trend vector by fusing Euclidean distance and cosine distance. This distance metric of the state units helps to find the correlative part of historical data. Third, the prediction results are forced to adjust the benchmark states based on the prediction function using the Gaussian weighted method. It is beneficial to keep the results on track and relax the constraint of the parameter k, which reduces the impact of scare data on the event duration.

The rest of our paper is organized as follows. Section II discusses the literature on short-term traffic state prediction and the existing approaches to prediction under special events. Section III proposes the SEKNN model and presents the state units, the state distance and the proposed prediction function. Section IV elaborates the numerical experiments using the traffic states depicted from GPS trajectory data (vehicle speeds) in the Beijing Workers' Stadium district. In addition, other traditional machine-learning models (original KNN, support vector regression (SVR), random forest (RF), and gradient boosting decision tree (GBDT)) and the current popular deep learning models (stacked autoencoder (SAE) and spatiotemporal recurrent convolutional network (SRCN)) are compared to demonstrate the advantage of the proposed SEKNN. The last section concludes the study and discusses future work.

II. LITERATURE REVIEW

A. SHORT-TERM TRAFFIC STATE PREDICTION

A great deal of research has focused on short-term traffic prediction, and the existing approaches can be divided into two categories: parametric approaches and nonparametric approaches [11].

1) PARAMETRIC APPROACHES

Parametric approaches predetermine the structures of the model on the basis of theoretical or physical assumptions regarding the time evolution of traffic [11]. Thus, these approaches are considered to not match reality because of their strong dependence on theoretical considerations [12].

Analytical models and traffic simulation models with parameters and inputs tuned to real-time data are a critical aspect of parametric approaches [11], and they depend on theoretical mathematical models to simulate traffic evolution. Frequently cited simulators, such as DynaMIT [13], DynaSMART-X [14], and TRANSIMS [15], predict traffic states based on dynamic traffic assignment, traffic flow models, car-following models, cellular automata or complex network theory [16].

Some statistics-based approaches are also categorized as parametric approaches, and they usually rely on statistic assumptions. The autoregressive integrated moving average (ARIMA) and its variants are the most common parametric approaches. Hamed *et al.* [17] attempted to develop a simple ARIMA of order (0,1,1) to predict traffic volume in urban arterials. Williams *et al.* [18] and Williams and Hoel [19] applied seasonal ARIMA (SARIMA) to predict urban freeway traffic flow and obtained better performance. Ding *et al.* [20] proposed a space-time ARIMA (STARIMA) to predict the traffic volume in urban areas five minutes in advance. The Kalman filter (KF) [21] and KF-based approaches, such as extended KF [22] and adaptive KF [23], also have important applications in short-term traffic prediction. In addition, Markov chain [24], exponential smoothing and other approaches have been mentioned in previous studies, respectively.

2) NONPARAMETRIC APPROACHES

Nonparametric approaches rely only on a sufficient mass of data rather than a priori knowledge and strict assumption [25], [26]. The majority of popular data-driven short-term traffic state prediction approaches fall into this category [27], which is founded on statistical learning theory (including most traditional machine-learning approaches) and artificial intelligence (AI, mainly deep learning) algorithms. In recent decades, these approaches have been developed rapidly due to abundant data attached to extensive traffic sensors and advanced big-data processing technology.

In the domain of machine-learning approaches, KNN, SVR, RF and artificial neural network (ANN) are typical approaches for short-term traffic state prediction. Cai *et al.* [28] proposed a spatiotemporal correlative KNN to enhance forecasting accuracy. Yu et al. [29] developed a multi-time-step KNN prediction algorithm based on road link. Yu *et al.* [30] proposed an RF-based approach into which the idea of KNN was integrated. Vlahogianni *et al.* [10] developed an improved ANN model to predict traffic flow using univariate and multivariate traffic data in urban arterial.

Deep learning approaches for short-term traffic prediction have become extremely popular and successful in recent years because of their powerful learning and generalizing ability. Lv *et al.* [31] applied a deep architecture model (i.e., an SAE model) to predict traffic flow for the first time. Ma *et al.* [32], [33] employed a long short-term memory (LSTM) neural network and convolutional neural network (CNN) to large-scale transportation network speed prediction. For richer and more accurate spatiotemporal features, deep hybrid architectures consisting of CNN and LSTM have become acclaimed. For example, Yu *et al.* [12] proposed an SRCN model and achieved accurate both short-term and long-term traffic state prediction, and Yang *et al.* [34] developed a convolutional long-term memory neural network based on critical road sections to overcome the problem of structural missing data.

B. SHORT-TERM TRAFFIC STATE PREDICTION UNDER SPECIAL EVENTS

Although Kumar and Vanajakshi [35] considered the availability of limited data using SARIMA similar to the problem faced by prediction under special events, the predicted traffic flow was still normal regular patterns in urban arterials. Previous studies [7], [34], however, have indicated that nonparametric approaches are preferable to offer more convincing results than parametric approaches under unstable traffic conditions. The advantage of flexibility means that nonparametric approaches remain the mainstream strategy for prediction under special events. The following two aspects are the strategies to enhance the performance in existing studies: improve the quality and quantity of data; utilize more powerful prediction approaches.

From the data perspective, the solutions use sufficient data from long-term observations or other information sources. With the advantages of flexibility, adaptability, learning and generalizability, deep learning theory was considered to predict Chicago highway traffic flow during special events in [36]. To capture the influence between downstream and upstream in a highway, they employed up to 180-day traffic data in their experiments. On the one hand, a high requirement for long-term traffic data collection is indicated. On the other hand, this study suggests a problem that is mentioned in other studies [37]: deep learning approaches rely heavily on feeding a great quantity of high-quality training data; otherwise, the model will probably face the problem of overfitting. Ni *et al.* [38] extracted effective information from social media to understand the attention and opinions of the public in relation to prior-event traffic prediction. With the aid of other information sources, more features of the event-caused impact on traffic flow can be captured.

Utilizing more powerful predictors or combining multiple predictors are the available methods to enhance prediction performance. In this respect, some researchers have attempted to employ novel approaches, such as Online-SVR [39], in prediction under special events. And others have combined multiple predictors to reduce the instability of single predictors. Kwoczek *et al.* [6] proposed a prediction model by fusing two different machine-learning models, and Wu *et al.* [40] employed a gradient boosting technique to combine multiple KNN models.

Overall, pursuing more effective utilization of existing information is the main idea among previous works in addressing prediction under special events. Admittedly, it is straightforward to follow the strategy of improving the quality and quantity of data, but more powerful prediction approaches work on a fundamental level when faced with situations of unmodifiable distribution of traffic detectors or limited data sources.

III. METHODOLOGY

Due to the simple but unusual structures and the flexible parameter settings, KNN plays an important role to predict short-term traffic states under special events. In this section, we introduce KNN for short-term traffic state prediction and then propose SEKNN for prediction under special events.

A. KNN FOR SHORT-TERM TRAFFIC STATE PREDICTION

KNN is a nonparametric approach that is widely applied in classification and regression tasks. For short-term traffic state prediction, the goal of KNN is to identify the similarities between current and historical traffic states and integrate generations of the most similar k historical states as prediction results. The most similar k historical states are referred to as k nearest neighbors. A typical KNN model for

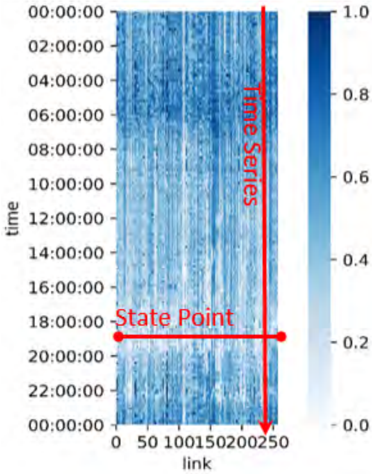

FIGURE 1. Visualization of the space-time matrix of normalized speeds in one day.

short-term traffic state prediction consists of three fundamental elements [41]: state series(or vector), distance metric and prediction function. As for network-wide traffic prediction, these three fundamental elements are described in the following sections.

1) DEFINE THE STATE SERIES

Traffic data can be acquired in various ways and expressed as time series. The evolution of multiple road segments or networks can be represented by a multidimensional series called an $m \times n$ space-time matrix [9], in which there are m state points (rows) for m time steps and n time series (columns) for n road links. The dimensions of the state point hinge on the number of traffic detectors. For example, Fig.1 shows a visual space-time matrix of normalized speeds from network of Beijing Workers' Stadium district in one day.

Assume that l is the size (i.e. the length of the time lag) of the influential time steps to be aggregated as a process of network-wide traffic evolution. The state series is usually defined as the input of traffic predictors using these l observed states, which is a portion of the space-time matrix employed for the once prediction process. In other words, the state series is a subseries of a certain vital traffic variable v during the time lag, such as road average speeds in this study. Thus, a state series can be described as consisting of the state point $S(t)$ at current time t, and $(l-1)$ its previous state points $S(t-1), S(t-2), \cdots, S(t-l+1)$. From another perspective, state series can be decomposed to n sub-time series $T(l_1), T(l_2), \cdots, T(l_n)$ from road link $l_1, l_2, \cdots, l_n$. The state series $x(t)$ defined at time t is expressed as follows:

$$x(t) = \begin{pmatrix} S(t-l+1) \\ \vdots \\ S(t-1) \\ S(t) \end{pmatrix} = \begin{pmatrix} T(l_1) & T(l_2) & \cdots & T(l_n) \end{pmatrix} = \begin{pmatrix} v_{l_1}(t-l+1) & \cdots & v_{l_n}(t-l+1) \\ \vdots & \ddots & \vdots \\ v_{l_1}(t) & \cdots & v_{l_n}(t) \end{pmatrix} \tag{1}$$

The archived state series are organized in the historical dataset $\mathcal{H}$ for the selection of the nearest neighbors. $\mathcal{H}$ is shown as follows:

$$\mathcal{H} = \{x_{hist}(t_i) | i = 1, 2, \cdots, N\} \tag{2}$$

where $x_{hist}(t_i)$ denotes the i-th state series in historical dataset $\mathcal{H}$ defined at t_i, and N is the size of the simples in $\mathcal{H}$.

2) MEASURE THE DISTANCE BETWEEN STATE SERIES

Euclidean distance (ED) is usually employed to measure the similarities between the predicted state series and the archived state series in traditional KNN models [29], [42] as follows:

$$ED_i = \|x(t) - x_{hist}(t_i)\|_2, x_{hist}(t_i) \in \mathcal{H} \tag{3}$$

where x denotes the predicted state series at t, and ED_i is the Euclidean distance between $x(t)$ and $x_{hist}(t_i)$, which is calculated by the second-order norm with the notation $\| \cdot \|_2$.

3) PREDICT THE GENERATIONS

The generation is the predicted state after f time steps by the (1), i.e., $S(t+f)$, written as the simplified notation y. After the recognition of k nearest historical state series $x_{hist}(t_{c_1}), x_{hist}(t_{c_2}), \cdots, x_{hist}(t_{c_k})$ according to the distance measure, the desired generation $\hat{y}$ can be calculated by utilizing these nearest neighbors' generations $y_{c_1}, y_{c_2}, \cdots, y_{c_k}$ in the direct average method in the original KNN as (4). However, this is not reasonable because the simple direct average method homogenizes all candidates with the same weights.

$$\hat{y} = \frac{1}{k} \sum_{c_j=1}^{c_k} y_{c_j} \tag{4}$$

B. SEKNN CONSTRUCTED WITH STATE UNITS AND STATE DISTANCE

Short-term traffic is considered a complex, dynamic, and nonlinear system with stochastic and chaotic characteristics impacted by uncertainties [43], [44]. It is a challenge to build an impeccable prediction model, especially for complicated network-wide traffic under special events. With the intention to exploit SEKNN building with state units and state distance preferable predictors for short-term traffic under special events, the limited historical information need to be captured more validly. Thus, we propose a SEKNN with the improvements of all three fundamental elements.

1) REDEFINE THE PROCESS OF TRAFFIC STATE EVOLUTION VIA STATE UNITS

The consideration is two-fold to redefine the expression of the input of KNN model: highlight the mutational traffic evolution tendency and use as less dimensions as possible at same time. We propose the state unit, a multipart object, to describe the traffic state evolution under special events by

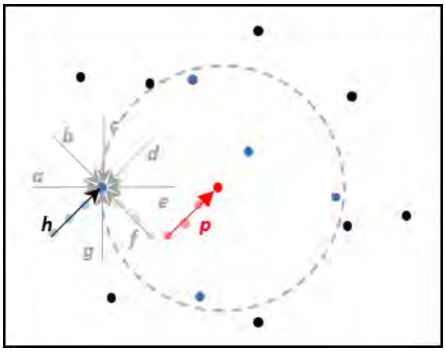

FIGURE 2. **Two-dimensional schematic diagram of the SEKNN model.**

decomposing the state series. The state unit is defined as two parts in (5): the benchmark state and its trend vector.

$$x'(t) = \{S(t), \vec{\tau}(t)\} \tag{5}$$

The benchmark state is the last state of a state series, which is the most relevant, influential and considerable one. We propose the trend vector to directly indicates how the benchmark forming, and its formula is shown in (6). The drastic changes of traffic evolution under special events are more prominent in such a form of expression, and the homogenized arrangements of state points in the state series are discarded. The parameter of l will be determined by a large number of experiments using real-world data, as discussed in Section IV-B.

$$\vec{\tau}(t) = S(t - l + 1) - S(t) \tag{6}$$

For example, state points are visualized as two-dimensional points in Fig.2. Specifically, the blue points denote candidates, the black points represent other historical states, and the red point indicates the predicted state. The benchmark state as an intermediate state may evolve from either network-wide congestion forming or dissipating, which results in two distinct generated states. Therefore, the processes of dramatic event-caused traffic changes are considerable. We redefine the formation process of the current benchmark state as its trend. The trends are represented by arrows pointing to the last known states. If the expression of state series was adopted, the most similar with $\vec{p}$ among the nearest neighbors may be $\vec{e}$ or $\vec{f}$ among $\vec{a}$ to $\vec{h}$ when trend evolutions are unknown, whereas the most exactly similar with $\vec{p}$ is $\vec{h}$ when considering trend similarity. However, different trends usually indicate different practical meanings in traffic evolution. It is necessary to add a section to show a definite trend of dramatic traffic evolution instead of enumerating traffic states as the state series.

The adoption of the state units means the representation of short-term traffic evolution changes from scalar to vector; and the object of the distance metric goes from calculating the distance between multiple state points, i.e. state series, to combining the last state point, i.e. benchmark state, with its evolution tendency.

2) DESIGN THE STATE DISTANCE TO MEASURE THE DISTANCE BETWEEN STATE UNITS

For the first part of the state unit, benchmark state S(t), ED is still employed as an effective metric. The formal Euclidean distance for the benchmark states is adjusted as (7).

$$ED_i(t) = \|S(t) - S_{hist}(t_i)\|_2, \quad i = 1, 2, \cdots, N \tag{7}$$

The method to measure the distance between another part of the state unit, i.e. the trend vector, use cosine distance (CD). Cosine similarity measures the similarities between vectors by the cosine of their angle, which has been one of most practical similarity measures applied to text document retrieval and clustering [45]. Non-positive cosines usually conclude the opposite trends. In this case, we define the CD to reflect the similarity of traffic evolution trends, as shown in (8). CD takes 1 minus the cosine similarity; thus, it is bounded by $[0, 2]$. The CD increases with decreasing cosine similarity, and cosine similarity decreases with the increasing angle between $\vec{\tau}(t)$ and $\vec{\tau}_{hist}(t_i)$. The CD is 0, which means that the angle between $\vec{\tau}(t)$ and $\vec{\tau}_{hist}(t_i)$ is zero and that these two trend vectors are regarded as identical; meanwhile, the CD is 2, which means that the angle between $\vec{\tau}(t)$ and $\vec{\tau}_{hist}(t_i)$ is 180°, and these two trend vectors are regarded as opposite.

$$CD_i = 1 - \frac{\vec{\tau}^T(t)\vec{\tau}_{hist}(t_i)}{\sqrt{\vec{\tau}^T(t)\vec{\tau}(t) + \vec{\tau}_{hist}^T(t_i)\vec{\tau}_{hist}(t_i)}} \tag{8}$$

The EDs between the current benchmark state and historical states are mapped to the scale of $[0, 2]$ as same as CD using the min-max scaling method. They can be combined as a novel distance measure used for state units, called the state distance (SD), as follows:

$$SD_i(t) = \alpha \frac{2(ED_i - \min\limits_{i=1,\cdots,N}\{ED_i\})}{\max\limits_{i=1,\cdots,N}\{ED_i\}} + (1-\alpha)CD_i \tag{9}$$

where α is the equilibrium factor to determine which metric should be more emphasized. The closer α is to 1, the more emphasis is placed on ED; in contrast, the closer α is to 0, the more the metric focuses on CD. The equilibrium factor makes that the SD is flexible to fit the current benchmark state. When facing dramatic traffic changes under special events, the trend vectors can be underline with a small α.

3) ADJUST THE METHOD TO PREDICT THE GENERATIONS BASED ON THE GAUSSIAN WEIGHTED METHOD

The weighted average method based on the Gaussian function mentioned is employed to predict generations in this study, which reduces the impact of the severe value of k in our experiments. That means SEKNN is able to adapt to the absence of limited historical data under special events. In a variation from (4), we adjust the result of the calculated generation by adding the current benchmark state and integrating the Gaussian weighted increments instead of directly integrating the generations of the Gaussian weighted nearest neighbors.

This adjustment is made to ensure that the prediction results do not deviate from the benchmark states, lest the predictive performance be severely impacted by the few similar historical state units in $\mathcal{H}$. The relevant equation is written as follows:

$$\hat{y} = S(t) + \sum_{c_j}^{c_k} \frac{w_{c_j}}{\sum_{c_j}^{c_k} w_{c_j}} \Delta y_{c_j} \tag{10}$$

where the increment acquired by $\Delta y_{c_j} = y_{c_j} - S(t_{c_j})$, w_{c_j} is the Gaussian weight determined by the formula $w_{c_j} = \exp\{-\frac{ED_{c_j}^2}{2\sigma_2}\}, j = 1, 2, \cdots, k$.

IV. CASE STUDY

A. DATA PREPARATION

Because of rich spatiotemporal information and easy accessibility, trajectory data has been used for various practical applications and services [46] with the rapid development of intelligent transportation and vehicle systems [47]–[49], e.g. human mobility pattern mining [50], dynamic shuttle bus route planning [51]. In this section, we introduce the GPS trajectory data into the task of short-term traffic state prediction as follow.

1) DATA SOURCE

To evaluate the predictive performance of the SEKNN, link average speeds, which were processed from probe vehicle speed data collected by taxis equipped with GPS devices in Beijing, China, are applied in this study. The sampling frequency of the GPS devices is 2 minutes, and the data of all the days are utilized due to nonnegligible lively nocturnal business in this district. As a result, 720 states can be observed per day. The data were collected from June 30th, 2015 to July 31st, 2015 (duration 32 days). The scope of the data is the road network around Beijing Workers' Stadium and its gymnasium, which encompasses 257 links and has an approximate total road length of 22.73 km and an area of 4.52 km^2. An aerial image of this district is shown in Fig.3. Particularly in summer, special events are frequently held in the stadium and the gymnasium. Two football games and concerts were held at Beijing Workers' Stadium and Gymnasium in July 2015, which increased the burden of this traffic network more frequently than usual. The dataset includes the observation of several special events in the district of interest.

According to the proportion 2 : 1 : 1, the data are divided into the historical dataset, validation dataset and test dataset. The historical dataset is used to query the nearest neighbors; the validation dataset is applied to parameter calibration; and the test dataset is utilized for performance evolution. Thus, the number of each dataset sample is 11520, 5760, and 5760, respectively. The number of special events archived in each dataset is intentionally 2 for the fair-to-train, validate and test models. Information about the special events is shown in Table 1.

FIGURE 3. The road network of Beijing Workers' Stadium district.

TABLE 1. Information on the datasets.

Dataset	Number of samples	Special events	
		Date	Special events
Historical dataset	11520	July 11st	Concert
		July 12nd	Football game
Validation dataset	5760	July 17th	Concert
		July 20th	Football game
Test dataset	5760	July 25th	Concert
		July 26th	Concert

2) DATA PROCESSING

Data processing should be completed before prediction because outliers, missing values and other unfavorable factors in raw data impact the performance of short-term traffic state prediction models [52]. Therefore, we followed four steps to process the raw data.

First, we removed the abnormal values and imputed the missing values. The abnormal values were identified and removed by the interquartile range method, and the missing values were imputed by the linear interpolation.

Second, the data were normalized according to the urban road speed limit. Because of the differences in the road hierarchy and other factors, the speed limit of each link as a crucial varying attribute determined the range of vehicle speeds. Hence, the data were normalized as the ratios of observed speed to the speed limit of each link. The normalized speeds greater than 1 were assigned a value of 1.

Third, a loess filter of the time lags was applied to reduce noisy and estimated trends. Smoothing/filtering techniques are a critical class of denoising methods that are commonly used to suppress the influence of random noises and smooth raw data [28], [36], [53]. The reason for choosing the loess filter is its merits of robustness to outliers, high flexibility and independence of any assumption [54], [55]. The details of this method are presented in the literature [53].

Finally, correlation analysis was used as the feature selection method in this study. Treating whole road links as features means that a large number of sensors must be set in the study district, which usually induces wide-area correlation and high superfluous dimensionality of data [56]. Thus, it is important to take advantage of feature selection technology to reduce the dimensions. The Pearson correlation between the series $T(l_i)$ and $T(l_j)$ is calculated using (11). Then, a correlation matrix can be constructed from the Pearson correlation

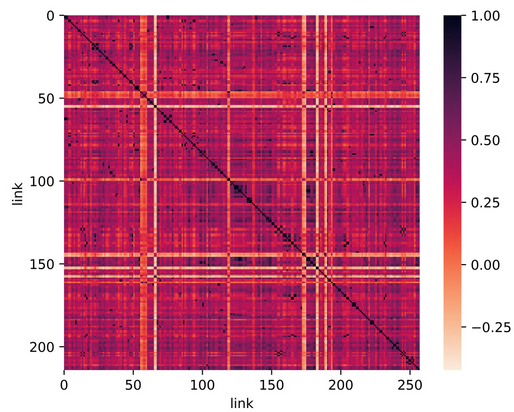

FIGURE 4. Correlation matrix of road links.

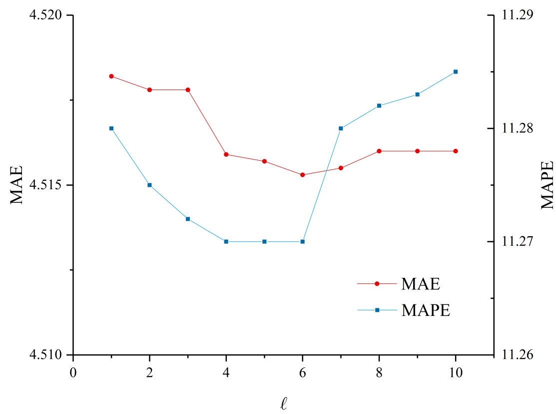

FIGURE 5. Influence of l on MAE and MAPE of SEKNN.

of the data of each of the two links, as shown in Fig.4. One of each pair of highly correlated links was eliminated in this study. Thus, the feature dimension decreased from 257 to 94 in the condition of high correlation ($\rho_{ij} \geq 0.8$, referring to [57] and it is acceptable on the basis of the discussion in [56], which suggests that the input features of predictors should be limited within 100. As a result, the average speeds from these 94 links will be input into the predictor, and entire traffic states (the average speeds from entire 257 links) will be output.

$$\rho_{ij} = \frac{cov(T(l_i), T(l_j))}{\sqrt{D(T(l_i))}\sqrt{D(T(l_j))}} \tag{11}$$

B. PARAMETER CALIBRATION AND SENSITIVITY ANALYSIS

The parameter combinations in the SEKNN were determined by plentiful experiments using the validation dataset according to the prediction performance. The performance can be evaluated by the indicators: mean absolute error (MAE, km/h) and mean absolute percentage error (MAPE, %), as shown in (12) and (13), respectively:

$$MAE = \frac{1}{N \times n} \sum_{i=1}^{N} \sum_{l=1}^{n} |\hat{y}_l^i - y_l^i| \tag{12}$$

$$MAPE = \frac{1}{N \times n} \sum_{i=1}^{N} \sum_{l=1}^{n} \frac{|\hat{y}_l^i - y_l^i|}{y_l^i} \times 100\% \tag{13}$$

where $\hat{y}_l^i$ and y_l^i are the predicted value and actual value at i-th time at the l-th link, respectively; n denotes the number of links in the road network; and N denotes the number of samples in the validation or test dataset. As described in Section IV-A.1, the value of n is 257, and the value of N is 5760 in the experiments.

The parameters to be calibrated include the length of time lag l, the number of nearest neighbors k and the value of equilibrium factor α on the condition of difference prediction steps, namely, predicting the next 2 to 10 minutes (f ranges from 1 to 5). The following content illustrates the relation between the parameters and the performance indicators.

The length of time lag l is regarded as an important parameter to determine the input of SEKNN. It is generally recognized that significant changes in traffic states can occur within a few minutes. To determine which value of l achieves the optimal performance, we observed the influence of different l values in a wider range of up to 20 minutes ($l = 10$) and discovered that $l = 6$ has the least influence on MAE and MAPE. For example, Fig.5 shows the results of the performance when $f = 2$ and other parameters are fixed.

The value of k is the key parameter affecting the accuracy of the KNN-based prediction model. In general, a much higher or lower k will cause worse prediction performance [58], such as the greater MAPE shown in Fig.6(a). Reference [28] shows, however, that the Gaussian weighted prediction function integrates the generations of nearest neighbors so effectively that MAPE does not noticeably increase or even gradually decreases when k increases by more than a certain value. Similar results of our experiments are shown in Fig.6(b). Although the impact of k on model performance is reduced due to the Gaussian weight in terms of the distance, a greater value of k generally leads to more program running time, as shown in Fig.6(c). Thus, the appropriate k is supposed to minimize MAPE and be as small as possible. In the experiments, we observed the influence of both k and α on MAE and MAPE. For example, for a single-step prediction, the results of the experiments are shown in Fig.7(a) and Fig.7(b). According to the figures, the optimal value of k is 18, and α is within an acceptable range from 0.4 to 0.7 to minimum MAE; the optimal value of k is 97, and α is 0.9 to minimum MAPE. On the one hand, MAE decreases slowly with the increase in k when k is greater than 18; on the other hand, MAE changes indistinctly with α (e.g., it is difficult to tell the difference between $\alpha = 0.6$ and $\alpha = 0.9$ from the right curves of Fig.7(a)), while MAPE changes more obviously with α. Thus, the calibration results of the parameters referred principally to their influence on MAPE shown as other subplots in Fig. 7.

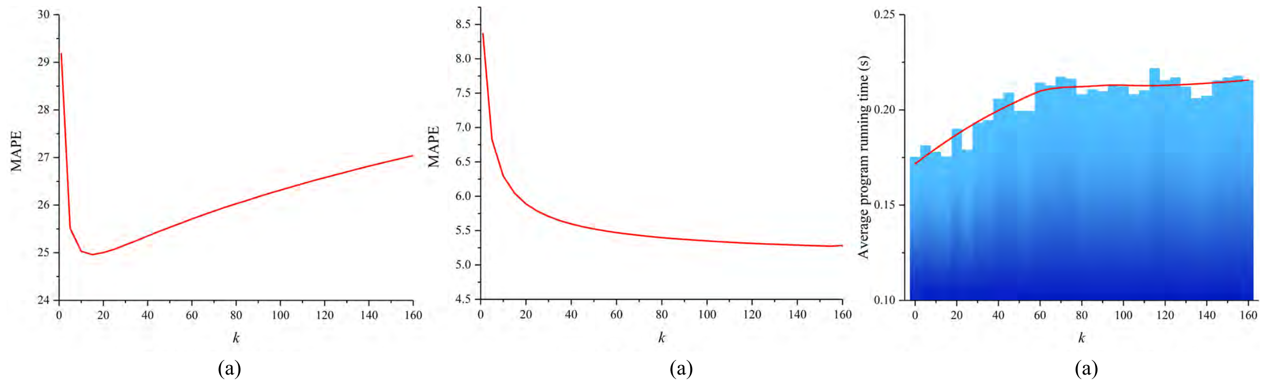

FIGURE 6. Influence of k on prediction performance in SEKNN. (a) MAPE in original KNN. (b) MAPE in SEKNN. (c) Average program running time for once prediction.

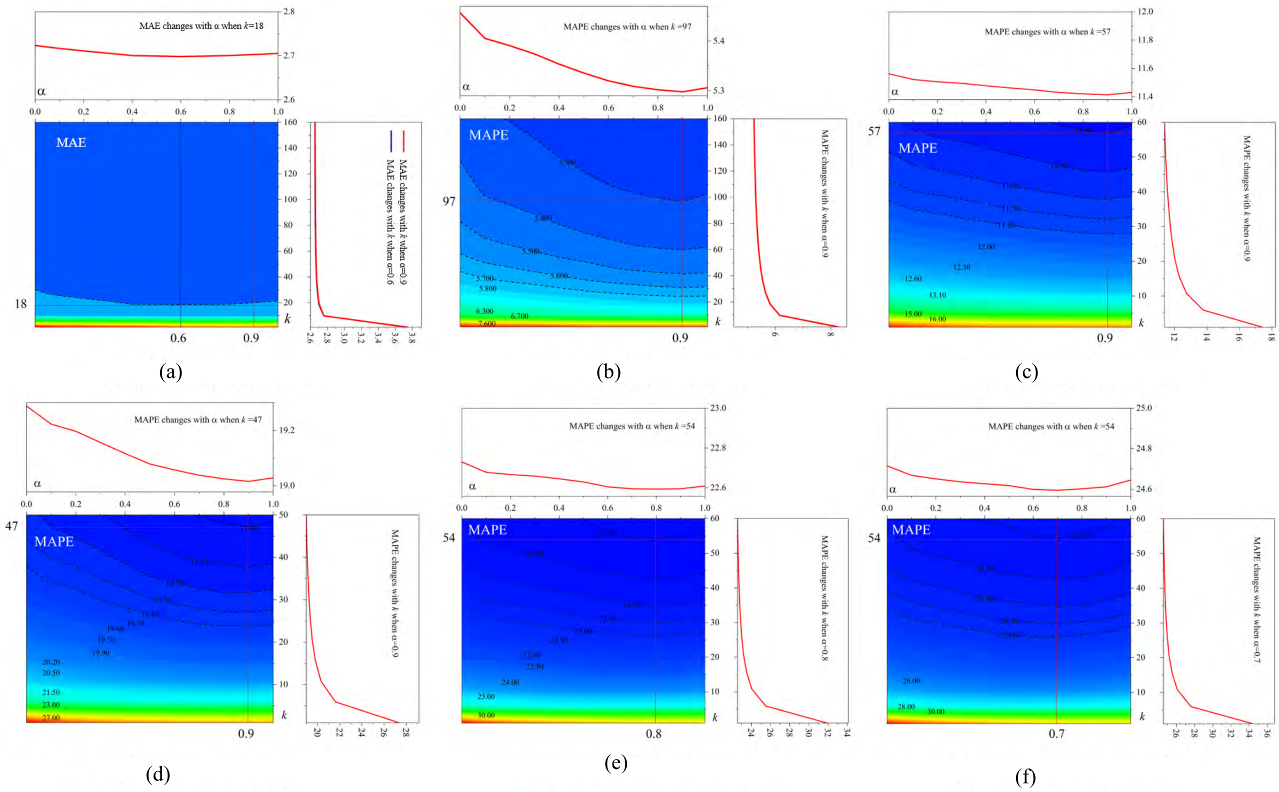

FIGURE 7. Influence of k and α on prediction performance in SEKNN. (a) MAE in SEKNN when $f = 1$. (b) MAPE in SEKNN when $f = 1$. (c) MAPE in SEKNN when $f = 2$. (d) RMSE in SEKNN when $f = 3$. (e) MAPE in SEKNN when $f = 4$. (f) MApE in SEKNN when $f = 5$.

Once the optimum parameters were identified in the validation dataset, the SEKNN was evaluated to predict on the test dataset. The calibration results of the different prediction steps are listed in Table 2. From the calibration results, k and α decrease with the increase in prediction steps. It can be explained that when the prediction steps increase, on the one hand, traffic states are more variable because of longer time-varying process, and thus it is more difficult to capture similar state units; on the other hand, the impact of the benchmark state weakens gradually, and thus the distance metric is supposed to decrease the proportion of the benchmark state and increase the proportion of the trend vector by means of adjusting the value of α.

TABLE 2. The calibration results of the parameters in SEKNN.

f	k	α
1	97	0.9
2	57	0.9
3	57	0.8
4	54	0.8
5	54	0.7

C. IMPLEMENTATION

In this section, the implementation details of SEKNN and other compared models were described. The parameters of each compared model were optimized based on validation dataset. And all of these experiments were performed on

TABLE 3. Comparison of different models using the test dataset.

Model \ f	1		2		3		4		5		Average error	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
SEKNN	**2.52**	**11.28**	**4.14**	**18.95**	**4.84**	**22.56**	5.17	**24.55**	5.35	**25.67**	**4.41**	**20.60**
Original KNN	4.74	25.84	4.90	26.74	5.06	27.80	**5.14**	28.41	**5.19**	28.82	5.00	27.52
MFCP	3.05	14.56	4.50	22.39	5.01	25.49	5.22	27.00	5.33	27.84	4.62	23.46
SVR	4.17	21.31	5.06	26.07	5.30	27.60	5.42	28.37	5.48	28.99	5.09	26.47
RF	5.31	29.13	5.37	29.66	5.41	30.10	5.42	30.26	5.43	30.52	5.39	29.94
GBDT	3.99	23.79	4.79	27.97	5.08	29.80	5.21	30.80	5.29	31.42	4.87	28.76
SAE	5.01	26.29	5.04	27.94	5.07	28.16	5.15	28.22	5.21	29.11	5.10	27.94
SRCN	4.92	26.51	5.00	27.56	5.12	29.04	5.22	29.32	5.20	29.85	5.09	28.46

The units of MAE and MAPE are km/h and %, respectively.

a computing platform with an Intel Core i7-5930K CPU (3.50 GHz), NVIDIA GeForece GTX TITAN X GPU, 32.0 GB memory, and running Windows 10 Education.

1) IMPLEMENTATION DETAILS OF SEKNN

The state units were archived in datasets. And the historical dataset was used to match the nearest neighbors according to SD (as (9)). Rolling prediction was adopted to imitate real change of the traffic states with time elapsing. The SEKNN with calibrated parameters was implemented to predict the traffic states using the test dataset. The tasks to evaluate SEKNN included one to five step-ahead prediction, i.e. $f \in \{1, 2, 3, 4, 5\}$. The SEKNN model was written in Python 3.5 programming language.

2) IMPLEMENTATION DETAILS OF COMPARED MODELS

The same predictive tasks were performed by the following compared models. And the implementation details of these models were introduced briefly as follows:

- Original KNN: The details of original KNN has been described in Section III-A. The parameter of k was determined by grid search method; and as shown in Figure 6 (a), $k = 14$ was adopted in experiments.
- Multiple factors combined prediction (MFCP): The speeds in the different spatiotemporal positions were regarded as the multiple factors to develop a MFCP model according to [59].
- SVR: The optimization of SVR was performed by grid search method on validation dataset, and the parameters were determined with the kernel of Gaussian radial basis function, $C = 10$ and $\gamma = 0.01$.
- RF: RF is a machine-learning model based on ensembles of decision tree and bagging algorithm. The optimized parameter (i.e. number of trees is 100) was adopted in experiments by grid search method.
- GBDT: GBDT also is an ensemble model of decision trees but based on boosting algorithm. Scikit-learn package helps to build model. And we optimized the parameter setting, including the number of trees was 100, the depth of trees was 5, and the learning rate was 0.01.
- SAE: SAE is a novel deep-learning model for traffic flow prediction proposed in [31]. We constructed a SAE model in Keras framework. Its architecture was listed in appendix A. And we added dropout layer and used early-stopping method to prevent overfitting problem. The SAE was trained in a greedy layerwise way, in which, firstly, each hidden layer was pre-trained respectively, and then all layers were fine-tuned by back-propagation algorithm.
- SRCN: SRCN is a typical state-of-the-art deep learning hybrid (CNN+LSTM) architecture for short-term traffic state prediction proposed in [12]. Appendix B shows more architectural details. The same measures were implemented in modeling to prevent overfitting. RMSprop optimizer were employed for training. Above implementation of SRCN was aid of Keras framework. In addition, for better performance, we also preprocessed the inputs using grid-based network representation method as same as [12].

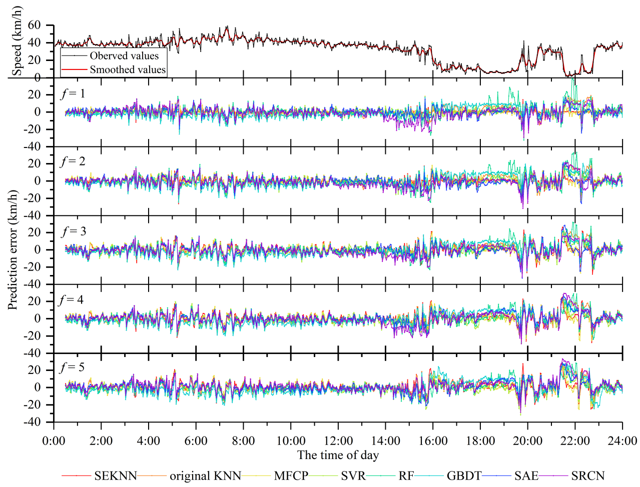

FIGURE 8. The observed values and the performance of prediction error on July 25th at link *No*.128894.

D. COMPARISON

In this section, all models, including our SEKNN and seven compared models, were conducted on multi-step-ahead prediction from $f = 1$ to $f = 5$, and we compared their performance as follows.

The overall prediction performance using the test dataset is displayed in Table 3. SEKNN obviously outperformed the other seven models, with the lowest MAPE in the condition of each different prediction steps. When f ranged from 1 to 3,

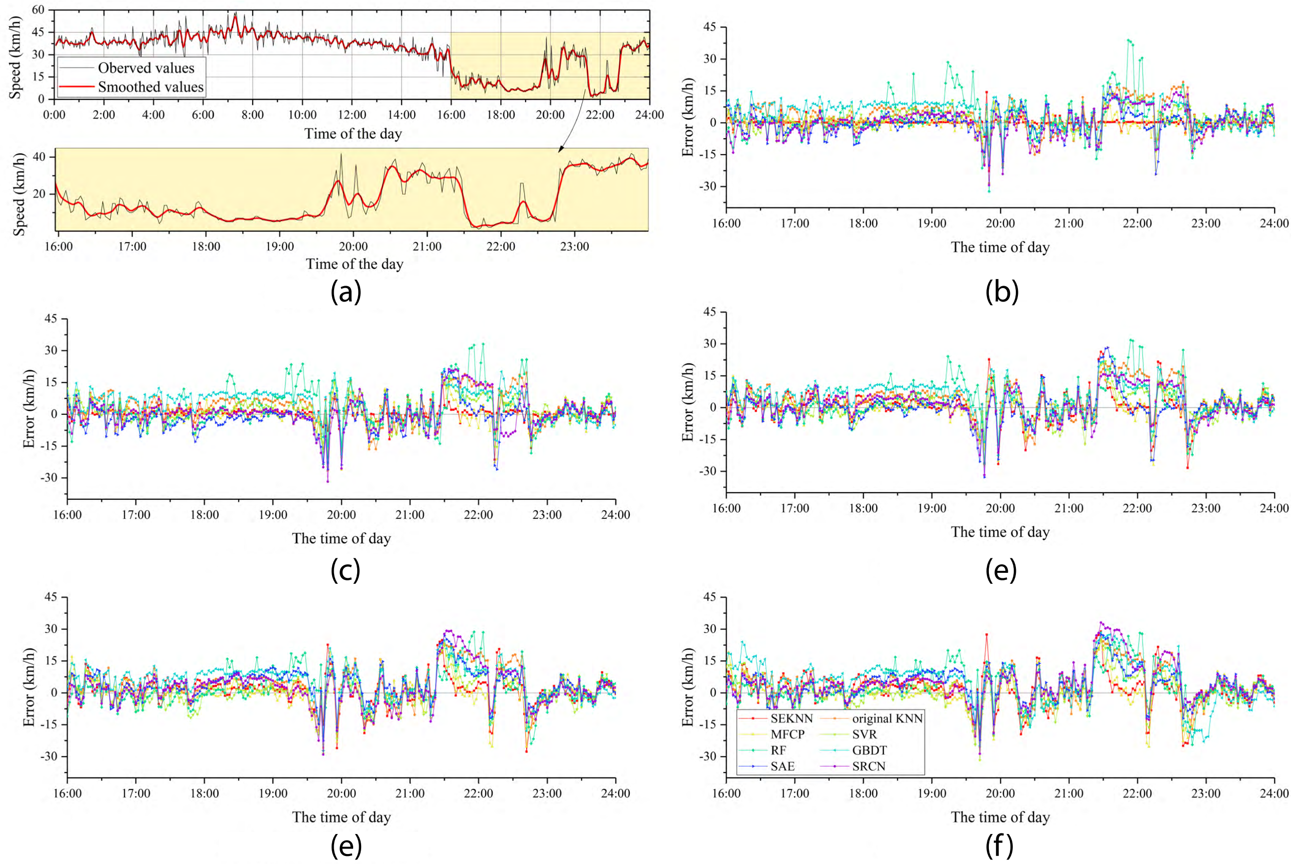

FIGURE 9. The prediction error under special event on July 25th at link *No*.128894. (a) Observed values. (b) Prediction error when $f = 1$. (c) Prediction error when $f = 2$. (d) Prediction error when $f = 3$. (e) Prediction error when $f = 4$. (f) Prediction error when $f = 5$.

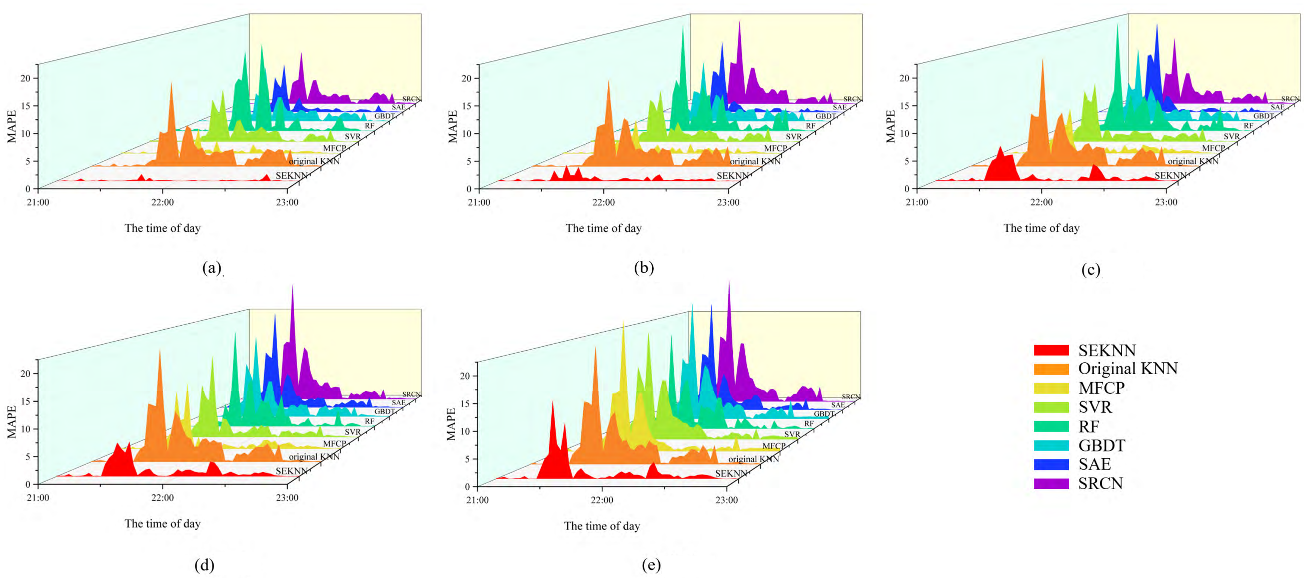

FIGURE 10. MAPE under the impact of the concert from 21:00 to 23:00 at link *No*.128894. (a) $f = 0$. (b) $f = 1$. (c) $f = 2$. (d) $f = 4$. (e) $f = 5$.

SEKNN was superior to the other models in terms of MAE, although the advantages gradually narrow with increasing prediction steps. When f was 4, SEKNN fell behind original KNN only approximate 0.03 km/h; When $f = 5$, SEKNN fell about 0.02 to 0.19 km/h behind original KNN, MFCP, GBDT and SRCN. With the increase in f, it was indeed difficult to predict because of the more complex and uncertain variation. The average MAE values of SEKNN for the

TABLE 4. The details of SAE architecture.

Layer	Type of layer	Number of units
0	Input	6×257
1	hidden layer 1	400
	relu	-
2	hidden layer 2	400
	relu	-
3	hidden layer 3	400
	relu	-
4	dropout	0.2
5	output	257

TABLE 5. The details of SRCN architecture.

Layer	Type of layer	Channels	Size
0	input	1	162×224
1	convolution 1	16	(3,3)
2	max-pooling 1	16	(2,2)
	relu	-	-
	batch normalization	-	-
3	convolution 2	32	(3,3)
4	max-pooling 2	32	(2,2)
	relu	-	-
	batch normalization	-	-
5	convolution 3	64	(3,3)
	relu	-	-
	batch normalization	-	-
6	convolution 4	64	(3,3)
	relu	-	-
	batch normalization	-	-
7	convolution 5	128	(3,3)
8	max-pooling 3	128	(2,2)
	relu	-	-
	batch normalization	-	-
9	flatten	-	-
10	fully connection	-	257
11	LSTM	1	800
	thah	-	-
12	dropout	-	0.2
13	LSTM	1	800
	thah	-	-
14	dropout	-	0.2
15	output	1	257

other models, in the order listed in Table 3, decreased by nearly 11.8%, 4.55%, 13.36%, 16.33%, 9.45%, 13.53% and 13.36%, respectively. For average MAPE values, SEKNN for the other models decreased by approximately 25.15%, 12.19%, 22.18%, 31.20%, 28.37%, 29.93%, and 27.62%, respectively. In general, SEKNN exhibited the best performance on the test dataset, with lower MAE and MAPE than the other models, especially for relatively short-term prediction. Original KNN, MFCP, and GBDT obtained not-bad MAEs but higher MAPEs than SEKNN. And deep-learning models, i.e. SAE and SRCN, seemed mediocre for this issue. A probable reason is that, the complex special-event evolution in training data was limited for SAE and SRCN.

For example, Fig.8 shows the observed values and the prediction errors of SEKNN, original KNN, MFCP, SVR, RF, GBDT, SAE and SRCN at link *No*.128894 (a part of the major avenue) on Saturday, July 25th, 2015, when a concert was held. In the upper part of the figure, the black line is the observed values; and in the lower part of the figure, the prediction errors, namely the residuals of predicted values deviating from the observed values, are described as the values of colored line deviating from the horizontal line of zero. Form the figure, the speed was relatively smooth from 0:00 to 16:00. In general, all models predicted stably except the slight fluctuation in morning peak hours. During that afternoon (after 16:00), traffic states changed distinctly, with the speeds decreasing and low speeds being maintained until the concert began (approximately 19:30). Two hours later (approximately 21:30), the traffic congestion occurred again at the end of the concert. This series of event-caused traffic evolution was also predicted by all models with different prediction steps.

In fact, what concerned more is the performance under the influence of special events (approximately from 16:00 to 23:00). Fig.9 shows the zoomed Fig. 8 during this period. In the Fig.9 (a), the black line is the observed values; the red line is the smoothed values; and the influenced period of the concert is zoomed on purpose. In the subplots from Fig.9 (b) to (f), the higher the value taken by f was, the stronger the fluctuations yielded by all of the models. The error of SEKNN was mostly closer to zero when f was small. With the increase in prediction steps, SEKNN still performed well relative to the other models, although it produced fluctuations. We focus on the performance during the event-caused congestion at the end of the concert in which traffic states changed rapidly, and Fig.10 shows the MAPE under the influence of the concert from 21:00 to 23:00 at link *No*.128894. Because the lower speed value under special events easily causes the higher MAPE, we can clearly realize the abilities of different models to predict traffic state under special events. From Fig.10, SEKNN still outperformed the other models for the different multi-step-ahead prediction tasks. It is noteworthy that the models, e.g. MFCP, GBDT, and SRCN, who had considerable MAE's performance on the overall test dataset when $f = 5$, while they had the high MAPE under special events, as shown in Fig.10(e).

Generally, SEKNN showed the best short-term prediction performance compared with the other seven models, i.e. original KNN, MFCP, SVR, RF, GBDT, SAE, and SRCN, in both the overall dataset and special-event conditions, especially for relatively short-term prediction tasks. Therefore, it can be said that SEKNN provides a more accurate prediction of short-term traffic states under special events.

V. CONCLUSION

Short-term traffic state prediction is an important management tool for traffic guidance and control and an effective decision-making tool to help travelers plan en-route trip and avoid congested road sections. However, prediction under special events faces enormous challenges because of its non-recurrent nature and limited available data to describe dramatic traffic changes. In this paper, SEKNN is improved based on the original KNN in all three basic elements around network-wide traffic state evolution, including the multiapart state unit, a novel distance metric and prediction function.

We used real-world traffic speed data collected from GPS devices in taxis in the Beijing Workers' Stadium road network with 257 links to test the proposed model. The dataset contains valuable information on network-wide traffic states and evolution processes during football games and concerts. The experimental results indicate that SEKNN is the most effective model to predict the dynamics of traffic evolution under special events compared with original KNN, RF, SAE and SRCN. However, with the increase in prediction steps, the great lead of SEKNN gradually disappears. This suggests that SEKNN has a scope of application in which it predicts more accurately than other models.

In future studies, SEKNN can be usefully developed as a submodel affiliated with a fusion approach, so that SEKNN and other submodels can give full play to their respective strengths and make up for each other. Every short-term traffic prediction approach has unique characteristics. Fusion-based approaches use the framework and strategies to generate a final prediction by combining the output of two or more individual predictors; thus, complementary predictive characteristics of different approaches can be leveraged [60]. Consequently, multiple prediction approaches should be used in a proper fusion framework to fully exhibit their advantages and enhance overall prediction performance in later research and practical applications.

APPENDIX A
ARCHITECTURE OF SAE

We developed a SAE model according to the literature [31]. The architecture of SAE consisted of one input layer, three hidden layers and one output layer. The number of hidden units was 400 in each layer. We added the dropout layer to prevent overfitting before the final output layer. The details of the architecture were shown in Table 4.

APPENDIX B
ARCHITECTURE OF SRCN

The SRCN was constructed referring to the literature [12]. As shown in Table 5, there were one input layer, five convolutional layers with (3,3) kernels, three max-pooling layers with (2,2) kernels, two LSTM layers with 800 units, and one output layer. Between the sections of convolution-based spatial features capturing and LSTM-based temporal features capturing, the flatten layer and fully connected layer were used to reshape the output of previous layers and prepare for the later LSTM layers. Relu and the hyperbolic tangent (tanh) are the activation functions after convolutions and LSTMs, respectively. Two dropout layers were employed to prevent overfitting. And batch normalization was applied to accelerate training.

REFERENCES

[1] J. Kwon, M. Mauch, and P. Varaiya, "Components of congestion: Delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1959, no. 1, pp. 84–91, 2006.

[2] T. Thomas and E. C. V. Berkum, "Detection of incidents and events in urban networks," *IET Intell. Transp. Syst.*, vol. 3, no. 2, pp. 198–205, Jun. 2009.

[3] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 306–328, Apr. 2017.

[4] A. Chen and Z. Zhou, "The α-reliable mean-excess traffic equilibrium model with stochastic travel times," *Transp. Res. B, Methodol.*, vol. 44, no. 4, pp. 493–513, May 2010.

[5] B. Anbaroğlu, T. Cheng, and B. Heydecker, "Non-recurrent traffic congestion detection on heterogeneous urban road networks," *Transportmetrica A, Transp. Sci.*, vol. 11, no. 9, pp. 754–771, 2015.

[6] S. Kwoczek, S. Di Martino, and W. Nejdl, "Predicting and visualizing traffic congestion in the presence of planned special events," *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 973–980, Dec. 2014.

[7] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[8] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C Emerg. Technol.*, vol. 19, no. 4, pp. 606–616, 2011.

[9] 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, 2015, Art. no. e0119044.

[10] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transp. Res. C, Emerg. Technol.*, vol. 13, no. 3, pp. 211–234, 2005.

[11] H. van Lint and C. van Hinsbergen, "Short-term traffic and travel time prediction models," *Artif. Intell. Appl. Critical Transp.*, vol. 22, pp. 22–41, Nov. 2012.

[12] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 7, p. 1501, 2017.

[13] M. Ben-Akiva, M. Bierlaire, H. Koutsopoulos, and R. Mishalani, "DynaMIT: A simulation-based system for traffic prediction," in *Proc. DACCORD Short Term Forecasting Workshop*, Delft, The Netherlands, 1998, pp. 1–12.

[14] H. S. Mahmassani, "Dynamic network traffic assignment and simulation methodology for advanced system management applications," *Netw. Spatial Econ.*, vol. 1, nos. 3–4, pp. 267–292, Sep. 2001.

[15] L. Smith, R. Beckman, and K. Baggerly, "TRANSIMS: Transportation analysis and simulation system," Office Sci. Tech. Inf., Los Alamos Nat. Lab., Los Alamos, NM, USA, Tech. Rep. LA-UR–95-1641, 1995. doi: 10.2172/88648.

[16] J. Barros, M. Araujo, and R. J. F. Rossetti, "Short-term real-time traffic prediction methods: A survey," in *Proc. Int. Conf. Models Technol. Intell. Transp. Syst.*, 2015, pp. 132–139.

[17] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in urban arterials," *J. Transp. Eng.*, vol. 121, no. 3, pp. 249–254, 1995.

[18] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction: Application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1644, no. 1, pp. 132–141, Jan. 1998.

[19] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.

[20] Q. Y. Ding, X. F. Wang, X. Y. Zhang, and Z. Q. Sun, "Forecasting traffic volume with space-time ARIMA model," *Adv. Mater. Res.*, vols. 156–157, pp. 979–983, Oct. 2010.

[21] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, 1984.

[22] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: A general approach," *Transp. Res. B, Methodol.*, vol. 39, no. 2, pp. 141–167, 2005.

[23] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 50–64, Jun. 2014.

[24] Y. Qi and S. Ishak, "A hidden Markov model for short term prediction of traffic conditions on freeways," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 95–111, Jun. 2014.

[25] P. Dell'Acqua, F. Bellotti, R. Berta, and A. De Gloria, "Time-aware multivariate nearest neighbor regression methods for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3393–3402, Dec. 2015.

[26] D. Xu, Y. Wang, P. Peng, S. Beilun, Z. Deng, and H. Guo, "Real-time road traffic state prediction based on kernel-KNN," *Transportmetrica A, Transp. Sci.*, vol. 38, no. 6, pp. 1–15, 2018.

[27] A. Ermagun and D. Levinson, "Spatiotemporal traffic forecasting: Review and proposed directions," *Transp. Rev.*, vol. 38, no. 6, pp. 786–814, 2016.

[28] P. Cai, Y. Wang, G. Lu, C. Peng, C. Ding, and J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 21–34, Jan. 2016.
[29] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "κ-nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *J. Transp. Eng.*, vol. 142, no. 6, 2016, Art. no. 04016018.
[30] B. Yu, H. Wang, W. Shan, and B. Yao, "Prediction of bus travel time using random forests based on near neighbors," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 4, pp. 333–350, 2018.
[31] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.
[32] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.
[33] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.
[34] G. Yang, Y. Wang, H. Yu, Y. Ren, and J. Xie, "Short-term traffic state prediction based on the spatiotemporal features of critical road sections," *Sensors*, vol. 18, no. 7, p. 2287, 2018.
[35] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal arima model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, p. 21, 2015.
[36] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.
[37] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.
[38] M. Ni, Q. He, and J. Gao, "Using social media to predict traffic flow under special event conditions," in *Proc. 93rd Annu. Meeting Transp. Res. Board*, 2014, pp. 1–23.
[39] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6164–6173, 2009.
[40] T. Wu, K. Xie, D. Xinpin, and G. Song, "A online boosting approach for traffic flow forecasting under abnormal conditions," in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, May 2012, pp. 2555–2559.
[41] Z. Zheng and D. Su, "Short-term traffic volume forecasting: A κ-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 143–157, Jun. 2014.
[42] S. Wu, Z. Yang, X. Zhu, and B. Yu, "Improved κ-nn for short-term traffic forecasting using temporal and spatial information," *J. Transp. Eng.*, vol. 140, no. 7, 2014, Art. no. 04014026.
[43] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, "A distributed spatial-temporal weighted model on MapReduce for short-term traffic flow forecasting," *Neurocomputing*, vol. 179, pp. 246–263, Feb. 2016.
[44] L.-Y. Zhang, Z.-R. Peng, D. J. Sun, and X. Liu, "Rule-based forecasting of traffic flow for large-scale road networks," *Transp. Res. Rec.*, vol. 2279, no. 1, pp. 3–11, 2012.
[45] R. M. Aliguliyev, "Performance evaluation of density-based clustering methods," *Inf. Sci.*, vol. 179, no. 20, pp. 3583–3602, 2009.
[46] X. Kong, M. Li, K. Ma, K. Tian, M. Wang, Z. Ning, and F. Xia, "Big trajectory data: A survey of applications and services," *IEEE Access*, vol. 6, pp. 58295–58306, 2018.
[47] R.-H. Zhang, Z.-C. He, H.-W. Wang, F. You, and K.-N. Li, "Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system," *IEEE Access*, vol. 5, pp. 6649–6660, 2017.
[48] X. Sun, H. Zhang, W. Meng, R. Zhang, K. Li, and T. Peng, "Primary resonance analysis and vibration suppression for the harmonically excited nonlinear suspension system using a pair of symmetric viscoelastic buffers," *Nonlinear Dyn.*, vol. 94, no. 2, pp. 1243–1265, Oct. 2018.
[49] H. Xiong, X. Zhu, and R. Zhang, "Energy recovery strategy numerical simulation for dual axle drive pure electric vehicle based on motor loss model and big data calculation," *Complexity*, vol. 2018, Aug. 2018, Art. no. 4071743.
[50] X. Kong, M. Li, G. Zhao, H. Zhang, and F. Xia, "COOC: Visual exploration of co-occurrence mobility patterns in urban scenarios," *IEEE Trans. Comput. Social Syst.*, vol. 6, no. 3, pp. 403–413, Jun. 2019.
[51] X. Kong, M. Li, T. Tang, K. Tian, L. Moreira-Matias, and F. Xia, "Shared subway shuttle bus route planning based on transport data analytics," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1507–1520, Oct. 2018.
[52] L. Li, Y. Li, and Z. Li, "Efficient missing data imputing for traffic flow by considering temporal and spatial dependence," *Transp. Res. C, Emerg. Technol.*, vol. 34, pp. 108–120, Sep. 2013.
[53] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 61–78, May 2015.
[54] W. S. Cleveland and S. J. Devlin, "Locally weighted regression: An approach to regression analysis by local fitting," *Publications Amer. Statist. Assoc.*, vol. 83, no. 403, pp. 596–610, 1988.
[55] E. S. Park, S. Turner, and C. H. Spiegelman, "Empirical approaches to outlier detection in intelligent transportation systems data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1840, no. 1, pp. 21–30, 2003.
[56] Y. Su, "On feature selection for traffic congestion prediction," *Transp. Res. C, Emerg. Technol.*, vol. 26, no. 1, pp. 160–169, 2013.
[57] A. Guitton, A. Skordylis, and N. Trigoni, "Utilizing correlations to compress time-series in traffic monitoring sensor networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2007, pp. 2479–2483.
[58] H. Chang, D. Park, S. Lee, H. Lee, and S. Baek, "Dynamic multi-interval bus travel time prediction using bus transit data," *Transportmetrica*, vol. 6, no. 1, pp. 19–38, 2010.
[59] J. Wang, X. Kong, W. Zhao, A. Tolba, Z. Al-Makhadmeh, and F. Xia, "STLoyal: A spatio-temporal loyalty-based model for subway passenger flow prediction," *IEEE Access*, vol. 6, pp. 47461–47471, 2018.
[60] F. Guo, J. W. Polak, and R. Krishnan, "Predictor fusion for short-term traffic forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 90–100, Jul. 2018.

HAIYANG YU received the Ph.D. degree in traffic environment and safety technology from Jilin University, China, in 2012. He is currently an Associate Professor with the School of Transportation Science and Engineering, Beihang University, China. His research interests include traffic big data, traffic control and simulation, and intelligent vehicle infrastructure cooperative systems.

NAN JI received the B.S. degree from Jilin University, in 2016. He is currently pursuing the M.S. degree with the School of Transportation Science and Engineering, Beihang University. His research interests include the algorithms and applications of short-term traffic state prediction.

YILONG REN received the B.S. and Ph.D. degrees from Beihang University, in 2010 and 2017, respectively, where he is currently an Assistant Professor with the School of Transportation Science and Engineering. His research interests include urban traffic operations and traffic control and simulation.

CAN YANG received the M.S. degree in transportation engineering from Beihang University, in 2015, where she is currently an Engineer with the Hefei Innovation Research Institute. Her research interests include urban traffic operations and control.