Transportation Research Part C 122 (2021) 102862

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A shock wave diagram based deep learning model for early alerting an upcoming public event

Hanyi Yang [a], Lili Du [b,*], Jamshid Mohammadi [a]

[a] Illinois Institute of Technology, United States
[b] University of Florida, United States

ARTICLE INFO

Keywords:
Public event prediction
Traffic shock wave
Deep learning
Long-short term memory

ABSTRACT

Even though extensive efforts worked on traffic anomaly detection, this study noticed that little attention is given to predict public events before they start, like celebrations, sports games, and so on. Given a public event produces complicated spatiotemporal traffic impacts, existing data-driven approaches without involving traffic flow analysis show limited effectiveness to predict a public event. Motivated by this view, this study seeks to develop a domain-knowledge-based learning approach, which integrates shock wave analysis into deep learning models to predict the occurring of an upcoming public event (i.e., the SW-DLM approach). This integration raises new research challenging and calls for new approaches. Specifically, we develop an efficient algorithm to generate the shock wave diagrams to present the evolvement of the traffic anomaly, which expands whenever new traffic data is collected. Next, the shock wave diagram itself is not a well-coded input for feature extraction and learning. This study thus developed an innovative encoding approach, which transforms a shock wave diagram into an optimal pixel grid. Considering the features extracted from the encoded shock wave diagrams are fed into the long-short term memory (LSTM) model for the event prediction. The numerical experiments based on the field data indicate that the SW-DLM is able to predict a public event with 87% accuracy around 84 min before it starts in a day. It outperforms all data-driven machine learning models using point traffic data as inputs. Thus, we claim that using the shock wave diagrams can significantly improve the accuracy and efficiency of the learning approaches for predicting a public event. The SW-DLM will help develop preventive traffic control or route plans to avoid traffic congestion induced by public events.

1. Introduction

Many events, such as accidents, constructions, sports games, and local extreme weather, may cause traffic anomalies, such as road closures, traffic capacity drops, and congestion, thus have a critical impact on traffic efficiency (Wang et al., 2013). Among these events, public events, including various social activities like sports games, festival celebrations, and other ceremonies, will significantly affect (either increase or decrease) traffic demand in a local area temporarily, and they present different traffic effects from other events. While some of the public events can be scheduled and broadcasted ahead of time, many others are often spontaneous or narrowly announced to selective customers or communities. In addition, the influence of public events on traffic presents high

* Corresponding author.
E-mail address: lilidu@ufl.edu (L. Du).

https://doi.org/10.1016/j.trc.2020.102862
Received 4 May 2020; Received in revised form 15 October 2020; Accepted 26 October 2020
Available online 4 December 2020

uncertainty in practice. For example, it is hard to predict when and where significant congestion will start). Travelers en route or traffic managers often lack sufficient and timely information for upcoming or on-going those events and their real-time traffic impact in a big city. Correspondingly, they cannot respond timely and properly to avoid traffic congestion. Therefore, timely detecting induced traffic anomalies and alerting such traffic influential events can help develop proactive traffic management, control, and mitigate traffic congestion (Parkany and Xie, 2005).

Traffic anomaly detection/prediction raises many research interests in literature. Using various traffic data, such as loop-detector data, GPS vehicle trajectory data, and camera data, many research efforts have been devoted to developing algorithms for accident detection, but few attentions have been given to public events. Even though both traffic accidents and public events will cause traffic anomaly, this study notices that unlike traffic accidents, public events often do not directly interrupt road infrastructure but gradually attract a large amount of traffic demand to a local venue in a short period before the occurring of the events. The demand swarm will cause traffic congestion, which spreads over a big area and evolve in a relatively long time (such as 1–2 h). These complicated traffic effects are often captured by traffic flow analysis, such as shock wave analysis (see an example in Fig. 2). However, existing approaches, such as outlier detection (Ihler, 2006; Ivan and Sethi, 1998) and various machine learning approaches including Support Vector Machine (SVM) (Yuan and Cheu, 2003), Neural Network (Asakura et al., 2017), Kalman Filter (Witayangkurn et al., 2013), and Discrete Wavelet Transform (Jeong et al., 2011), are purely data-driven approaches without involving traffic flow analysis. Namely, they mainly rely on the features extracted from the raw traffic data collected at some points. Those point traffic data cannot fully recognize a temporal-spatial traffic flow evolvement in a local area resulting from a public event. Accordingly, those data-driven approaches cannot predict a public event accurately. Our numerical experiments observed the same issue. More importantly, these data-driven approaches focus on detecting occurred traffic accidents and they are not suitable for public event prediction. This is because the traffic anomaly induced by a public event often show very much before the start of the event. It is more meaningful to detect its early traffic impacts and alter this event rather than to report it while severe traffic congestions/interruptions already occur by the time close to the start of the event.

Motivated by the aforementioned research needs and gaps, the presented efforts seek to integrate the domain knowledge of traffic flow analysis into a deep-learning method so that we can detect the early traffic impact and then efficiently predict the occurring of an upcoming public event, even though we do not know its schedule. The novelty of this research raises the following research challenges, which demonstrate the unique contributions of our study later. First of all, the impact of a public event on traffic presents a dynamic evolvement process: gradually fade-in and then fade-out along with the variation of the attracted traffic demand. The full impacts of a public event will only be fully noticeable at the time of the event beginning. However, it is only beneficial to travelers or traffic managers if we can alter an upcoming public event by noticing the early traffic impact. Therefore, the learning approaches are required to recognize the occurring of the upcoming event by using incomplete features. Next, an upcoming public event will cause traffic congestion evolving in a local area linking to multiple branches in a relatively long period (1–2 h) before the start of the event. However, the traffic data collected at fixed points cannot fully present the effect. It is potentially beneficial if we use the domain knowledge to further interpret these point traffic data and provide more comprehensive features to feed machine-learning approaches. Last, even though we may implement dense sensors, such as cameras, to monitor the affected area online, it might be difficult to process such a big scale of image data in real time. Thus, the learning approaches should be computationally efficient.

To address the above research challenges and application requirements, this study developed a shock wave diagram based deep learning method (SW-DLM) to predict the occurring of an upcoming public event through detecting its early traffic impacts. The development of the SW-DLM contributes to the following ideas and methodologies. First, through shock wave analysis for a sample road segment, we confirmed that the shock waves (SW) rather than the point traffic volume data enclose more comprehensive features of the traffic effects resulting from a public event. It potentially enables us to do early traffic impact detection. Accordingly, we develop

Fig. 1. The Study Freeway Segment.

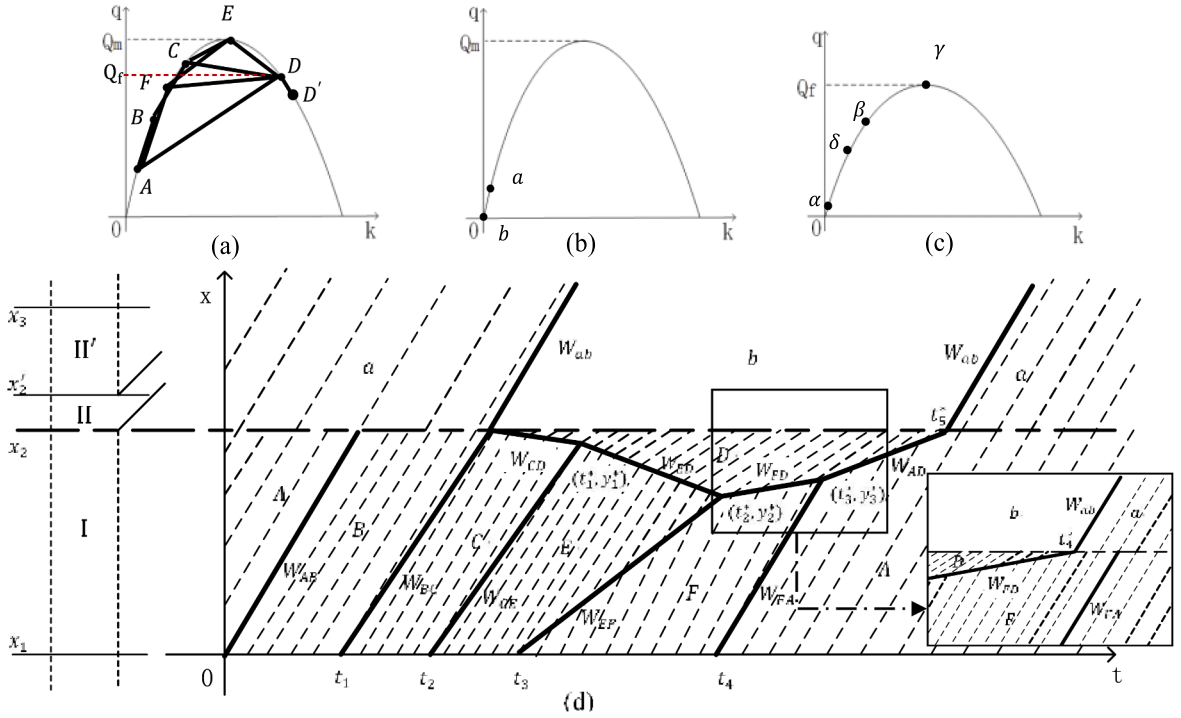

Fig. 2. Shockwaves Analysis under Free Flow State.

an online shock wave diagram generation algorithm (SWG) to generate and update the time-series shock wave diagrams at each timestamp according to the data collection rate. Next, we recognized that the shock wave diagrams are not well-coded inputs for a machine learning approach. This study, therefore, developed a novel encoding approach that meshes the shock wave diagrams by a well-designed pixel grid. Different to the pixel grid of a colored image, each pixel here holds the temporal-spatial traffic state therein (i.e., the location of the pixel and the average traffic flow and density). To balance the trade-off between the resolution and computation load, a constraint mathematical programming is developed to determine the optimal design of the pixel grid. Furthermore, this study constructs a customized "autoencoder" for the feature extraction and then takes its output as input for a deep learning method. Last, this study employs LSTM to learn the occurring of a public event since the shock wave diagrams are time-series data and recurrent neural network (RNN), especially LSTM, presents great success to analyze time-series traffic data (Bogaerts et al., 2020; Fu et al., 2016) in literature.

The performance of the SW-DLM is validated by the numerical experiments using the field data collected around exit 164B on I-5 interstate freeway northbound (orange curve), Seattle, WA. The experiment results show that the SW-DLM is able to predict the occurring of a public event around 84 min ahead with 87% accuracy in a day near the venue of a public event. More importantly, our experiments indicate that the machine learning model fed by the shock wave diagram inputs performs better than the models using raw traffic data.

In short, the novelty of this study lies in integrating the shock wave analysis into the machine learning approaches for detecting the early traffic impact and further predicting the occurrence of an upcoming public event without knowing its schedule. To ensure the success of this integration, we contributed the methodologies, including (i) the SWG algorithm to efficiently generate and update time-series shock wave diagrams; (ii) a new optimal encoding approach, which transforms a shock wave diagram into an optimal pixel grid so that more comprehensive features can be extracted and further fed into the LSTM approaches; (iii) the SW-DLM for efficiently predicting the occurring of an upcoming public event by detecting its early traffic impact. Therefore, the unique contribution of this study lies in providing an innovative approach, which enables us to use the domain knowledge in traffic flow theory to improve the performance of the deep learning approaches for alarming an unknown upcoming public event rather than developing a new approach in each individual domain.

The presented research efforts are organized as follows. Section 2 reviews the previous research efforts on traffic anomaly detection. Section 3 formally defines the proposed study and related preliminaries. Built upon that, Section 4 develops the SW-DLM, which integrates the traffic flow analysis, shock wave diagram generation and encoding, feature extraction, and deep learning algorithm. The SW-DLM is validated by the experiments built upon field data in Section 5. We summarize the research efforts, conclusions, and future study in Section 6.

2. Literature review

Public events are those activities that attract significant traffic demand during a time interval. They often cause traffic anomaly on the roads nearby the venue of the event. This study falls into the area of traffic anomaly detection. Our review shows that most of the existing efforts focus on traffic accident detection, but little attention was given to the public event prediction. In addition, the approaches for traffic accident detection do not work efficiently to predict the occurring of an upcoming public event, while the schedule of the event is unknown. The brief review below will demonstrate this research gap in detail and highlight the research needs and our contributions.

First of all, the literature shows that three main types of approaches have been used to detect traffic accidents by sensing traffic anomaly, including outlier detection, pattern classification, and traffic flow property recognition. The outlier detection approaches consider traffic accidents will lead to an irregular traffic state. The detection algorithms (Yuan and Cheu, 2003; Asakura et al., 2017) established probabilistic/statistical models for normal traffic state, such as volume or speed distribution on a road segment. And then these approaches report accident detection when outliers are noticed. Besides, many other studies considered traffic accident detection as a pattern classification problem. Accordingly, various machine-learning approaches were employed to detect traffic accidents. For example, the study of (Fuse and Kamiya, 2017) applied a Bayesian integration model to detect immediate downstream abnormal traffic conditions for a moving vehicle onboard video image data. The study in (Wang et al., 2013) used SVM to detect freeway and arterial incidents, taking input traffic data including speed, volume, and occupancy of consecutive time points from loop detectors on the upstream and downstream on every lane. Neural network models are also used to detect traffic accidents in many studies such as (Dia and Thomas, 2011; Zhang et al., 2018), using loop detector data and probe vehicle data as input. This review noticed that the outlier detection and pattern classification approaches mainly examine traffic variation at a particular point without considering traffic flow evolvement over a temporal-spatial area. This weakness degrades accuracy when they are used for public event detection. In addition, these approaches are more efficient to detect traffic anomalies resulted from the occurrence of a traffic event rather than predicting the occurring of the event, which gradually presents traffic impacts. Thus, they are not suitable for public event prediction. Having noticed this same issue, some studies analyzed the unique traffic flow features under traffic accidents and designed corresponding detection algorithms. For example, the study of (Asakura et al., 2017) detected the occurrence of traffic congestion by sensing the speed reduction and then used the ratio of the time headway before a vehicle runs into and after it leaves the congestion zone to detect a traffic accident. Shock wave theory was employed to calculate where and when the accident occurred. However, our shock wave analysis shows that the traffic anomaly resulting from a public event on a nearby segment is much more complicated (see Fig. 2 and Fig. 3). It is hard to use a single feature to capture.

Existing literature shows that only a few studies investigated public event detection through detecting the induced traffic anomaly. The study of (Ihler, 2006) developed Markov-Poisson models to learn the traffic pattern transformation from a normal state to a public event state, using time-series count data collected at an on-ramp around the venue of the event. Taking GPS traffic speed data as input,

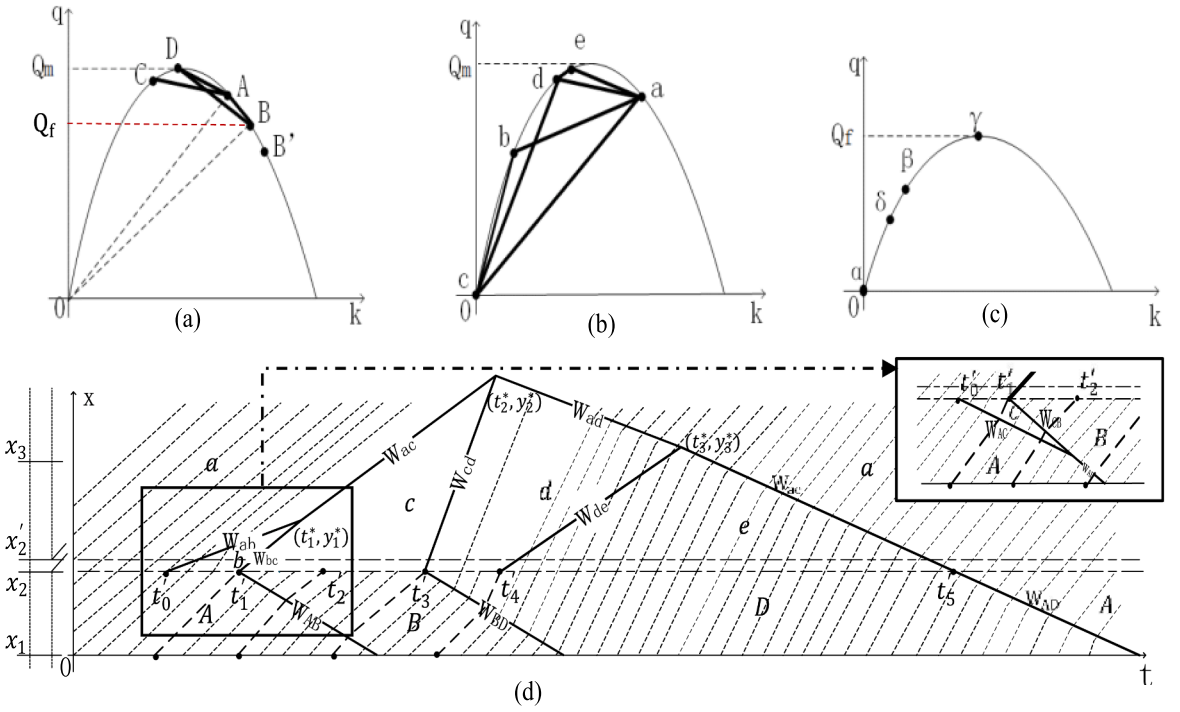

Fig. 3. Shockwaves Analysis under Congested Flow State.

Liu et al. (2019) used Restricted Boltzmann Machine (RBM) to recognize spatiotemporal traffic pattern and then detects general traffic anomaly, including traffic accidents or special events over a network. There are also some studies that try to detect public events through sensing abnormal human density. For example, the study of (Witayangkurn et al., 2013) applied the Hidden Markov Model (HMM) to detect public events in an area by analyzing the density of GPS points (not traffic data) in the area. Fuse and Kamiya (2017) applied a hierarchical Dirichlet process hidden Markov model to detect special events through human movement and density. A few studies such as (Zhang et al., 2018) used social media data to detect traffic anomaly and then inferred occurred traffic accidents or special events. Even though many studies developed deep-learning models for traffic data analysis (Bogaerts et al., 2020; Fu et al., 2016; Li et al., 2020; Zhang et al., 2019; Rahman and Hasan, 2018), few of them worked on detecting upcoming public events.

Therefore, the state-of-the-art shows that most of the existing studies focus on detecting an occurred event rather than predicting an upcoming public event. Thus, their outputs provide limited help in preventive traffic congestion control. In addition, the majority of the approaches are built upon the raw traffic state data without involving traffic flow analysis. Our experiments show that those approaches cannot provide satisfied prediction accuracy. The research will make up these research gaps, seeking to use the domain knowledge in traffic flow theory to improve the performance of the machine learning approaches for predicting the occurring of an upcoming public event.

3. Preliminaries

3.1. The sample road segment

This research considers a public event scheduled at a location where the traffic flow from several local streets connects to a highway segment through off-ramps or on-ramps. Using Fig. 1 as an example, we present two important observations below. First, even though a public event may widely affect traffic in its neighborhood area, the off-ramps, which are the connections between a highway and the local venue of the event will present the most significant effects. Second, the traffic congestion on a highway resulting from a public event may propagate further. The affected traffic flow presents special traffic flow characteristics, which are also affected by the features of the road segments. Therefore, without loss of generality, this research develops our approach based on the sample spatial area, which involves a highway segment and a connected off-ramp leading highway traffic to the venue of public events. Our traffic flow analysis will demonstrate the complexity of the traffic impact as well as the needed traffic data for event detection and prediction.

Using the example in Fig. 1, we formally define the sample area to facilitate our traffic flow analysis. Specifically, this spatial area includes a highway section from the location x_1 to x_3 with an off-ramp (such as 164B) at x_2 leading to the local traffic for the public event. The intersection between the highway section and the off-ramp is labeled from x_2 to $x_2^{'}$. Therefore, the entire study area is separated into sub-area I on the highway from x_1 to x_2, sub-area II on the highway from x_2 to x_3, and the sub-area III for the off-ramp. Within sub-area II, the road section from $x_2^{'}$ to x_3 is defined as sub-area II$^{'}$ to assist the traffic flow analysis. Sub-area I and sub-area II$^{'}$ are two closed road sections at the upstream and downstream of off-ramp separately. Note that all sub-areas are closed road segments without extra inside branches. The distance from x_1 to x_2 is L_1 and that from $x_2^{'}$ and x_3 is L_2. The traffic flow on the highway segment at time t is defined as $q_m(t,x)$, $x \in [x_1, x_3]$. The entering flow $q_m(t,x_1)$ at x_1 consists of the going-through flow denoted as $q_\mu(t,x|x \le x_3)$ entering in sub-area II, and the diverging flow denoted as $q_f(t,x|x \le x_2)$ exiting for the off-ramp. The lane capacity of the highway and off-ramp are denoted by Q_m and Q_f respectively.

3.2. Shock wave analysis for the traffic impact

Built upon the problem setup, this section captures the dynamic traffic evolution in the study area as extra traffic demand is attracted by an upcoming public event. The analysis starts with the scenario that the initial highway traffic is under free flow and will extend to the case with the initial congested flow. According to our observation in the field data, we pick a sample arriving flow profile in Eq. (1) for $q_m(t,x_1)$ at x_1 following the piece-wise constant function with the initial traffic flow $q_m(0,x_1) = q_A$.

$$\begin{cases} q_m(t,x_1) < Q_m & q_f(t,x_1) < Q_f & T_1 = [0,t_1) \\ q_m(t,x_1) < Q_m & q_f(t,x_1) > Q_f & T_2 = [t_1,t_2) \\ q_m(t,x_1) \ge Q_m & q_f(t,x_1) > Q_f & T_3 = [t_2,t_3) \\ q_m(t,x_1) < Q_m & q_f(t,x_1) < Q_f & T_4 = [t_3,t_4) \\ q_m(t,x_1) = q_A < Q_m & q_f(t,x_1) = 0 < Q_f & T_5 = [t_4,\infty) \end{cases} \tag{1}$$

More exactly, the piece-wise constant function is defined on five-time intervals (T_1, T_5). The arriving traffic increases slightly in the time interval T_1 and then reaches to the capacities of the off-ramp and the highway in the time interval T_2 and T_3 respectively. After that, it fades out to initial traffic in the time interval T_4 and T_5. This traffic flow profile at x_1 mimics a typical traffic demand peak before the start of a public event. It can be observed from our field data used in our experiments. Besides, the piece-wise constant function is consistent with the format of traffic data collected by a point sensor with a constant collection rate. Even though this analysis only considers the arriving flow before the occurrence of a public event, the same approach can be used to analyze the leaving flow after a public event. Last, without loss of generality, we assume that the off-ramp traffic is often under the free flow traffic condition if no public events occur.

The shock wave analysis with free flow initial condition for the study area is presented in Fig. 2, in which dashed lines represent vehicle trajectories, and solid lines indicate shock waves. The analysis to generate this shock wave diagram is presented in Appendix A

in detail. The shock wave analysis in the area while the initial highway traffic condition is under congestion is demonstrated in Fig. 3. The corresponding elaborate analysis procedure is provided in Appendix B. Note that A, $\boldsymbol{a}$, and $\boldsymbol{\alpha}$ are used to denote the initial traffic states for the three sub-areas. Since the study area is close to the highway-ramp intersection, we assume that all lane-change maneuvers are conducted before those vehicles enter the study area. Hence, we ignore the effect of lane changes. Accordingly, we only consider traffic flow on the lane adjacent to the off-ramp. The macroscopic fundamental diagram (MFD) such as the Greenshields model (Daganzo, 1997) is used to perform traffic analysis, even though other MFD models, like the triangle model, will lead to similar results.

Through studying the shock waves in both Fig. 2 and Fig. 3 we obtain important insights as follows. First of all, the traffic anomaly resulting from a public event may present different features under different initial traffic conditions, even though the point traffic data present the same traffic variation. Specifically, Fig. 2 shows complicated shock waves that occur in the sub-area I if initial highway traffic is under free flow. Fig. 3 indicates that the sub-area II presents more complicated shock wave features if highway traffic is initially under the congestion flow. More importantly, the shock wave diagrams indicate that traffic impacts resulting from a public event are quite complicated. It involves multiple shock waves corresponding to the fluctuation of arriving traffic demand. It is hard to use one or two simple features to characterize the impacts. Instead, the shock wave diagrams capsulate the complete information of the impacts. Then the shock wave enclose will be more comprehensive information than the point traffic data for pubic even prediction. This insight is also validated by our numerical experiments. Last, according to land use, it is noticed that most of the venues for a large-scale public event are located at a local area with good highway/main arterial connections. Therefore, combining with Remarks 1–2 in Appendix A and B, the selected sample area including the sub-area I and II possesses great significance to capture traffic anomaly resulting from the occurring of a public event.

Built upon the above observations, we will develop an online learning approach to predict a public event around the area of a highway-ramp intersection. The learning approach assumes that the highway traffic data such as volume and density can be collected in the study area adjacent to a venue (the dash-dot line box in Fig. 1) by traffic detectors. If traffic data are not available in this area but in other remote areas, it is out of the scope of this study. Our future study will address this general case.

4. Methodology

This section presents our efforts to develop the SW-DLM approach, which seeks to predict the occurring of a public event before it starts by detecting its early traffic impact on a selected road segment around the venue of the event. The SW-DLM consists of four critical components, including online shock wave diagram generation, encoding shock wave diagram, feature extraction, and event learning. More exactly, the online shock wave diagram generation algorithm uses the point traffic data (flow and density) collected by the loop detectors at each timestamp as the boundary condition input to generate time-series traffic shock wave diagrams, which captures traffic congestion evolution along the road segment due to traffic demand variation. Next, we develop our approaches to encode each shock wave diagram and further extract key features of an encoded shock wave by a designed autoencoder. The extracted features from the time-series shock wave diagrams are further used as the inputs to feed the LSTM model for public event prediction. The sections below show the technical details for each of them. Note that the main contribution of the methodology is integrating the domain knowledge of traffic flow analysis into the machine learning approaches to improve its interpretability rather than developing a new traffic flow theory or machine learning approach itself.

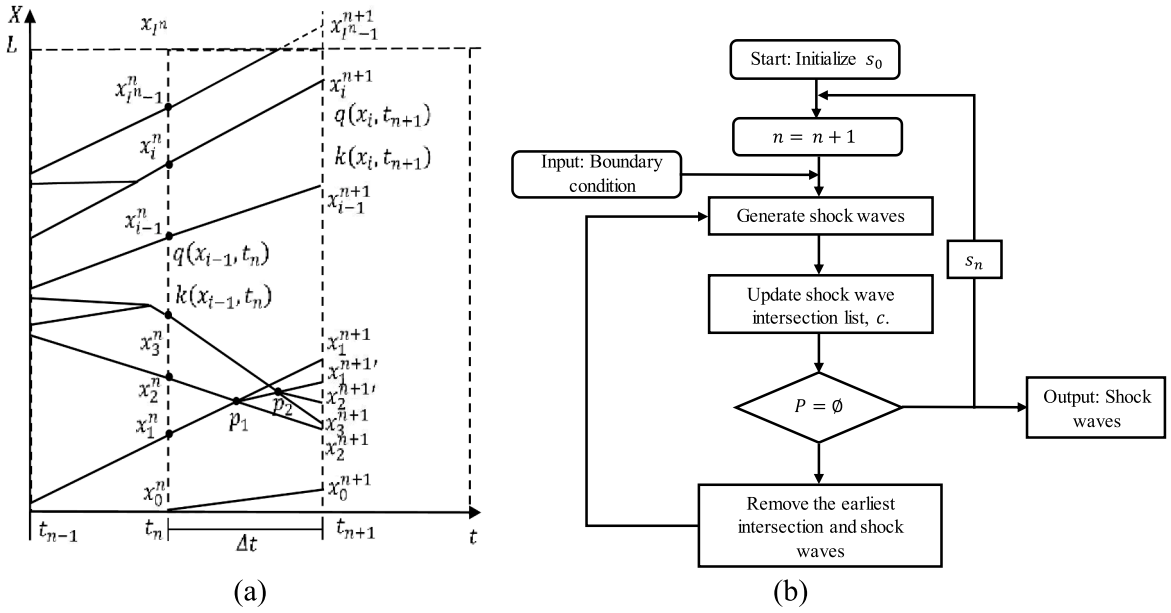

Fig. 4. SWG Algorithm.

4.1. Shock wave diagram generation

To capture the traffic effect capsulated in the shock wave diagrams and use it as the input to a machine learning approach for a public event prediction, we need an efficient algorithm, which can generate the time-series shock wave diagrams quickly, whenever new entering traffic data are collected for the study area. According to the existing traffic flow theory (Daganzo, 1997), this section develops an online shock wave diagram generation algorithm (labeled as SWG hereafter) to conduct this task. Mainly, it is recognized that the traffic data at boundaries is often collected at discrete time points with a certain rate by traffic detectors, such as the loop detectors. Taking it as the piece-wise constant input for the shock wave analysis, we will get time-series shock wave diagrams expanding at discrete time steps. It is known that with the piece-wise linear initial condition and boundary conditions, the shock-wave paths are themselves piece-wise linear (Daganzo, 1997). Moreover, the solution in the temporal-spatial plane is a set of polygons and their sides having the slope equal to $W_{ij}=\frac{q_j-q_i}{k_j-k_i}$ (Daganzo, 1997), which is considered as the shock wave between traffic state i and traffic state j.

According to the above theory, this study develops the SWG with the main idea illustrated in Fig. 4 Specifically, starting with an initial traffic state on the road segment, the SWG expands the shock waves (in black and solid lines) within a temporal-spatial cell with the size $\Delta t \times L$ at each discrete timestamp t_n, when new boundary traffic data (flow and density) are collected as input. For example, Fig. 4 (a) shows that we have a number of initial traffic states on the road at timestamp t_n, such as $\left(q\left(x_{i-1}^n, t_n\right), k\left(x_{i-1}^n, t_n\right)\right)$ at location x_{i-1}^n. The set of traffic states on the road segment at time stamp t_n are denoted by s_n in Fig. 4 (b). As new boundary traffic condition data are collected, the traffic state at the locations of the boundaries $x_0^n$and x_L^n will be first updated. This will further drive the SWG to expand the shock waves from the cell at t_n to the cell at t_{n+1}. More exactly, the SWG first generates all potential shock waves such as $x_i^n x_i^{n+1}$, $i \in I^n$ in Fig. 4 (a). However, it is noted that the shock waves, such as $x_1^n x_1^{n+1}$and $x_2^n x_2^{n+1}$, may intersect, disappear, and then generate a new shock wave at the intersection within the $\Delta t \times L$ space. To address this issue, the SWG will iteratively screen all intersections and search the earliest intersection, from which we remove the intersected shock waves and generate a new shock wave until no intersections left in the $\Delta t \times L$ space.

We illustrate this procedure using Fig. 4 as an example. The SWG can find that the earliest intersection at p_1, from which the SWG stops the shock waves $x_1^n x_1^{n+1}$ and $x_2^n x_2^{n+1}$ (removing the part from p_1 to $x_1^{n+1}$and that from p_1 to x_2^{n+1}), and then generates a new shock wave from p_1 to x_1^{n+1}, with the slope $\frac{q(x_3,t_n)-q(x_1,t_n)}{k(x_3,t_n)-k(x_1,t_n)}$. Note that the shock wave p_1 to x_1^{n+1}, may cause new intersections, which will be addressed in a new iteration. The SWG repeats this process until no intersections appear in the $\Delta t \times L$ space. These shock waves lead to the new traffic states on the road section at t_{n+1}, which will become the initial conditions for the next time step t_{n+2}. The shock wave diagram is expanded further as new traffic data is collected. Note that the SWG works on a closed road segment and no other branch for entering or exiting traffic flow. In addition, we assume that traffic information is accurate. As the error is involved, the SWG algorithm should be improved by involving data science approaches. We propose this extension as our future study.

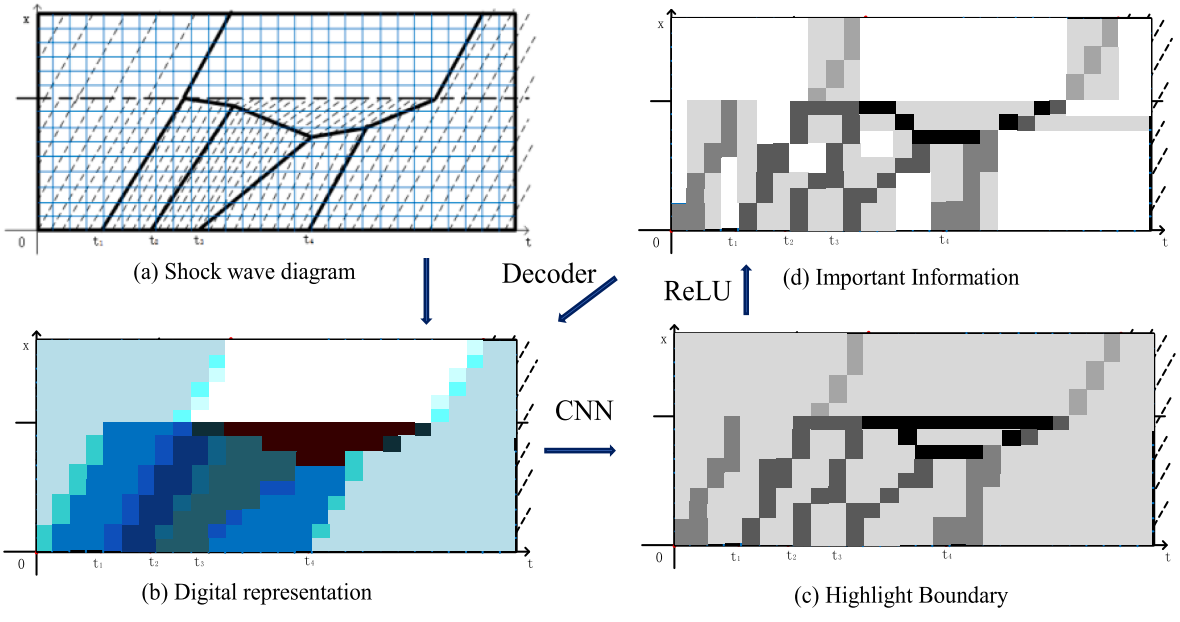

Fig. 5. Feature Extraction.

4.2. Encoding shock wave diagram

The SW-DLM takes the shock wave diagrams as input to feed the learning approach for public event prediction. However, the shock wave diagrams obtained from the SWG algorithm are mathematically presented by traffic states, slopes, and intersection points. Thus, the shock wave "diagram" is not a "visual image" in a grid of pixels and it is not a well-coded input for feature extraction and learning. This study, therefore, develops the approach to encode the information involved in a shock wave diagram by meshing it to a pixel grid.

Specifically, this study thinks that a shock wave diagram can be represented in the form similar to a BMP image. A BMP image is represented by a three-dimensional matrix, in which the first and second dimensions indicate the positions on a plane and the third dimension represents the color. The third dimension includes three elements corresponding to three primary colors, red, yellow, and blue, and the shade is indicated by the value of the corresponding element. By analogizing the representation of a BMP image, we mesh a shock wave diagram to a grid (see Fig. 5 (a)) and then use a matrix (M), say traffic state matrix, with $m \times n \times 2$ dimensions to store a shock wave diagram, in which the first two dimensions represent the temporal-spatial position of a pixel, and the third dimension represents the traffic state in the pixel, including density (k) and flow rate (q) (see Fig. 5 (b)).

However, the traffic state in a pixel of the grid presents a heterogeneous feature. When there is a shock wave going through a pixel, the traffic states are different at the two sides of the shock wave. To address this issue, this study uses the average density and flow calculated by Eq. (2) and Eq. (3) to represent the traffic state in each pixel.

$$k(A) = \frac{t(A)}{|A|} = \frac{\sum_{i=1}^{I} k_i \int_0^{\Delta t} l_i(t)dt}{\Delta l^* \Delta t} \tag{2}$$

$$q(A) = \frac{d(A)}{|A|} = \frac{\sum_{i=1}^{I} q_i \int_0^{\Delta l} t_i(x)dx}{\Delta l^* \Delta t} \tag{3}$$

where Δl and Δt are the length and interval of a pixel separately; I is the number of traffic states in a pixel; $l_i(t)$ is the length of the ith traffic state at time t; $t_i(x)$ is the time interval of the ith state at location x; $t(A)$ and $d(A)$ are the total travel time and travel distance respectively; $|A|$ is the area of a temporal-spatial pixel (Daganzo, 1997).

The traffic state matrix should be able to reflect the information of a shock wave diagram as much as possible, which needs higher resolution with more pixels. This, however, increases the dimension of the input data, which requires more computation resources and may also cause the curse of dimensionality. These issues will limit the applicability of the approach for this online event detection and prediction. Thus, we purse the trade-off between the accuracy and computation load. An optimization model is developed to address this difficulty.

$$min \frac{1}{m^* n} \sum_{u=1}^{m} \sum_{v=1}^{n} \sigma_{uv}^2 + m^* n \tag{4}$$

$$1 \le m \le \frac{Lk_j}{c}, \tag{5}$$

$$1 \le n \le \frac{H}{\tau}, \tag{6}$$

$$\sigma_{uv}^2 = \frac{1}{\Delta l^* \Delta t} \int_0^{\Delta l} \int_0^{\Delta t} (k_{uv}(t,x) - k_{uv}(A))^2 dtdx, \tag{7}$$

where $k_{uv}(t,x)$ is the density at the point (t,x) in a pixel (u,v); $k_{uv}(A)$ represents the average density in a pixel; σ_{uv}^2 represents the traffic density variance in a pixel.

As for the accuracy, we would like to minimize the loss of variance in density and flow. The existing study (Daganzo, 1997) shows that traffic flow and density are highly correlated. Therefore, minimizing the loss of density variance can also ensure the accuracy of the traffic state in terms of traffic flow. In the meantime, we also hope the values of m and n (the size of the grid), are as small as possible. Consequently, the objective of the optimization model is written in Eq. (4), in which m^*n is a penalty function. As for the constraints, we use the traffic flow features including jam density k_j and free flow traffic speed v_f and the length of the study area L to construct the formulation. The length of a pixel, Δl, should be large enough for the macroscopic traffic analysis (Daganzo, 1997). Thus, we assume that Δl should be larger than c times the spacing under jam state, $\Delta l \ge c/k_j$. c is an experience parameter (we discuss its potential values in our experiments). Correspondingly, the possible values of m should be an integer between 1 and Lk_j/c (see Eq. (5)). As for the value of n, we consider the minimum time is $\tau = c/(v_f k_j)$, which is the minimal travel time within the minimum length of a road section, c/k_j. Thus, the value of n should be an integer between 1 and H/τ (see Eq. (6)). H is the data collection interval. Taking the optimal results from the optimization model, we can decide the size of each pixel by $\Delta l = L/m$, $\Delta t = \frac{H}{n}$.

This model comes to a non-linear integer programming. Moreover, the integer decision variables (m and n) are the upper limits of the summation. Therefore, it is difficult to explore the global optimal solution rigorously. The heuristic algorithm, the Best First Search Algorithm (Szer and Charpillet, 2005), has demonstrated a good performance to search a locally optimal solution for discrete optimization efficiently. It is thus used to explore the optimal values of m and n. Briefly, according to constraints (5) and (6), the Best First

Algorithm first initializes a potential integer solution set with N number of solutions (it is named as OPEN). Then, the algorithm evaluates the quality of each solution by the objective function (see Eq. (4)) and determine the search direction to improve the existing solution set in the next step. More exactly, the immediate neighbors of the best one, e.g. 4 and 6 are the neighbors of 5, will be taken as the successors, and added to OPEN. The worst element will be removed from OPEN. This process is repeated until there are no new successors generated or the improvement of the objective becomes minor.

4.3. Feature extraction and learning

After meshing the shock wave diagram to a grid system (see Fig. 5. (b)), we noticed that many pixels store the same traffic states (e.g, the pixels with the same color in Fig. 5. (b)). This duplicate information is not very necessary for learning an upcoming event, but puts computation loads for the learning model. In contrast, the shock wave edges and the pixels around those edges are significant features. These observations motivated this study to conduct a dimension reduction (i.e., feature extraction) before feeding the encoded shock wave diagrams into the learning model. It is expected that the feature extraction keeps the most significant and unique features of the shock wave diagrams so that we can reduce the data dimension and the computation load for the learning process. There are several existing approaches for doing feature extraction such as Principal Component Analysis (PCA) (Polat and Güneş, 2008), Partial Least squares (Maitra and Yan, 2008), "autoencoder" (Wang et al., 2016), and so on. This study employs the "autoencoder" model with the structure shown in Fig. 6 since it has a well-proved performance to learn the important features in an image (LeCun et al., 2015). Its special structure can minimize the reconstruction error during the process of feature extraction. Specifically, the operation from an input $\boldsymbol{x}$ to the output $\boldsymbol{y}$ at the central layer is named as the encoder and that from $\boldsymbol{y}$ to the output $\boldsymbol{x}$' is named as decoder. The approach converts high-dimensional data such as $\boldsymbol{x}$ to a low-dimensional code $\boldsymbol{y}$ by training a multilayer neural network so that the output $\boldsymbol{x}$' to ensure that the critical features are not lost (Wang et al., 2016). In other words, training "autoencoder" seeks to well configure it to minimize the difference between $\boldsymbol{x}^{'}$ and the flattened $\boldsymbol{x}$.

There are different approaches and structures that can be used for encoder and decoder such as Back Propagation (BP) neural network (Wong, 1991), Hopfield neural network (Lee et al., 1998), CNN (Krizhevsky, 2012), and so on. We discuss our design of the encoder and decoder as follows, considering the particular features of the shockwave diagrams.

First, we use the CNN combined with ReLU (Rectified Linear Unit) to construct the encoder. This is because CNN achieved great success in recognizing objects (Krizhevsky, 2012) by highlighting the boundaries in an image through enlarging the values of elements around these boundaries and decreasing those at the other parts. The shock waves are the boundaries of different traffic states, which are the critical features of a shock wave diagram (see Fig. 5 (c)). The core part of the CNN is the convolution layer, which includes a kernel, with a small dimension of space, spreading through the entirety of the input, usually an image. During the spreading process, the dot production of the kernel and the corresponding region covered by it in the input data will be the corresponding element in the output. The parameters of the CNN model are the values of the elements in the kernel. Moreover, the output of CNN is taken as the input of ReLU, which can further convert the less important values to zero and keep the important ones by Eq. (8). Then, the encoder using CNN combined with ReLU will perform well to keep and highlight the most critical features of a shock wave diagram (see Fig. 5 (d)). More exactly, by inputting $\boldsymbol{x}$ vector, which represnts coded 3D shock wave diagrams, the encoder (CNN combined with ReLU) extracts features and flattens them into a 1D vector $\boldsymbol{y}$ so that it can be used to feed the deep learning algorithm.

$$max\{0, x - b_k\} \tag{8}$$

Next, we choose a traditional feedforward neural network (FNN) with full connection as the decoder. The decoder recovers the

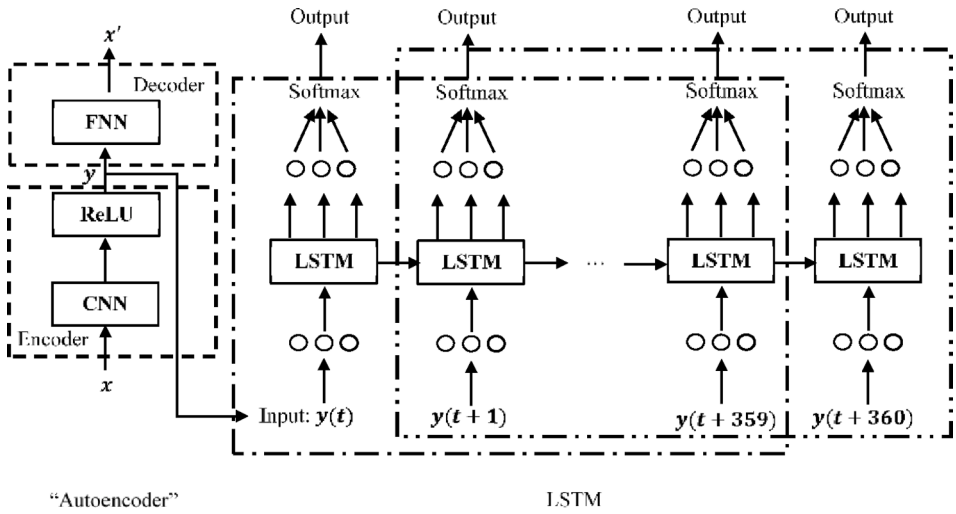

Fig. 6. "Autoencoder" and LSTM.

input feature vector y to a 3D matrix $x^{'}$ with high similarity to the original data x (see Fig. 5 (d)). FNN is a well-used decoder in literature and its performance has been proved (Wang et al., 2016; Hinton and Salakhutdinov, 2006). This "autoencoder" will be trained through historical data in our study and it performs well in our experiments (see an example in Fig. 5).

4.4. LSTM for public event prediction

Recall that the complete features inferring a public event only present right before its occurrence. To predict an upcoming public event, we can only rely on a subset of the features observed from the time-series shock wave diagrams generated according to traffic flow entering the study area. Existing studies already show that Long Short-Term Memory (LSTM), a variant of RNN achieved great success in learning time-series traffic flow data and it outperforms other time-series data analysis methods such as ARIMA (Tian et al., 2018; Bogaerts et al., 2020). This study thus applies the LSTM model (Fig. 6) to develop the learning function of the SW-DLM. Each LSTM cell has six correlated gates including input gate (i_t), forget gate (f_t), hidden gate (h_t), output gate (o_t), current system state (c_t) and g vector in Fig. 7. The good design of the gates ensures that the LSTM can address the time-series problem well. Appendix C explains the furcation of the gates in an LSTM cell in detail.

As for the public events prediction problem, the input of the LSTM model is the extracted features from "autoencoder" (see Fig. 6) and its output is the indicator referring if there is an upcoming public event or not. We will use historical data to train the LSTM for ensuring its prediction accuracy.

5. Case study

5.1. Experiment setting

This section validates the SW-DLM by a case study using field data. More exactly, the selected road segment is the right most lane of 0.89 miles (x_1 to x_3 in Fig. 1) around exit 164B on I-5 interstate freeway northbound, Seattle, WA. The distances from x_1 to x_2, x_2 and $x_2^{'}$, and $x_2^{'}$ to x_3 are 0.49 miles, 12feet, and 0.4 miles respectively. The exit 164B leads to two stadiums, CenturyLink Field and Safeco Field. Six month-long traffic data (May - October 2011) were collected per 20 s at three loop detectors, which are respectively implemented at positions $x_1^{'}$, x_2, and x_3. The collected raw data (including detector ID, timestamp, traffic volume, Scan count, weather, and traffic accidents) are published by the U.S. Department of Transportation Intelligent Transportation Systems Joint Program Office (JPO). The Scan_count data can be converted to traffic density using the formulation provided by the JPO data description (catalog.data.gov, 2018; May, 1990), assuming average vehicle length equal to 18 ft and detection zone length equal to 6 ft (Tian et al., 2018). In the meantime, the data published by (fbschedules.com, 2018; mlb.com, 2018) indicate that there are a total of 71 games occurred during the traffic data collection period, including baseball games (68) and football games (3).

To implement the SW- DLM, the experiment first applies the SWG algorithm to generate time-series shock wave diagrams by taking the following parameters and data as inputs. The shock wave diagrams of interests are generated for the road segment between x_1 and x_3 in Fig. 1 by using the traffic data (traffic flow and density) collected at positions x_1, x_2, $x_2^{'}$, and x_3 as boundary conditions while assume free flow initial condition on the road segment at 12 AM each day. Note that the traffic data can only be collected at $x_1^{'}$, x_2 and x_3. Thus, the needed data at x_1 and $x_2^{'}$ are estimated by the data collected at $x_1^{'}$, x_2 and x_3 using a linear regression model developed based on historical data. In addition, the travel time between x_2 and $x_2^{'}$ is ignored since it is too short given the short distance (12 feet) and high traffic speed (around 60 mph). Using Macbook Pro (2.2 GHz 6-core Intel Core i7 and 16 GB of 2400 MHz DDR4 onboard memory), the SWG takes 0.1 s to update shock wave diagrams with the given data collection interval (20 s). So, the computation performance of the SWG satisfies our application.

Next, the experiments encode a shock wave diagram by the matrix, $M_{m\times n\times 2}$ according to Eq. (2) and Eq. (3). Specifically, the optimal size of $M_{m\times n\times 2}$ (i.e., the optimal values of m and n) is determined by the Best First Search Algorithm, in which we set $m \in$ (Wang et al., 2013; Tian et al., 2018) and $n \in$ (Wang et al., 2013; Wong, 1991). $m = 25$ corresponds to $\Delta l \approx 100$ feet, which is around the length of 5 vehicles at jam density. $n = 20$ corresponds to $\Delta t \approx 1$, which is the minimum travel time within $\Delta l \approx 100$ feet. The algorithm explores optimal m's and n's for all shock wave diagrams generated from May 2011 to August 2011. Then, we take their

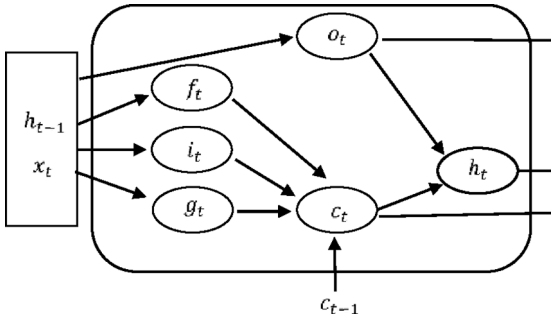

Fig. 7. LSTM cell structure.

average and set $m = 8$ and $n = 6$ for the study area. Thus, we end with 48 pixels (30 pixels in the sub-area I and 18 pixels in the sub-area II) in this study.

Furthermore, the experiment trains the "autoencoder" model, in which the "encoder" involves two CNN layers and one ReLU layer, and the "decoder" is designed including two layers of fully connected FNN. The kernel of both CNN layers is a matrix of $K_{2\times2}$ and the stride is 1. As for the "decoder" part, the first layer has 48 cells, and the second layer comprises of 96 cells, which corresponds to the dimensions of the input shock wave diagram, $m \times n \times 2 = 8 \times 6 \times 2$. The training data are the shock wave diagrams (i.e., $M_{m\times n\times 2}$) from May 2011 to August 2011. The model was trained by 500 iterations. The offline well trained "autoencoder" model will be used for the feature extraction online.

The final step is to implement the SW-DLM and to train the RNN model, which has two major parts, an LSTM layer, and a "SoftMax" layer. In addition, there is a layer of the FNN between the input and the LSTM layer and one between the LSTM layer and the "SoftMax" layer (see Fig. 6). The input data is a vector including an extracted feature vector and its corresponding timestamp in a day labeled by {1, …, 4320}, and the day of a week indicated by {1, …, 7}. The data in one day (24 h) includes 4320 vectors given the data collection rate is per 20 s. The training process sets "look back" length as 360 data points (2 h). It means that the input of the training process considers past 2-hour time-series shock wave diagrams (see Fig. 6**)**. In total, the batch size of the training process is 3960, which contains continuous traffic data in a day to reflect day-long traffic variation. The number of LSTM cells in the learning model is set as 200. The total training data is from May 2011 to August 2011. The training iteration is set to be 1000. In addition, ADAM (Adaptive Momentum Estimation) (Kingma and Ba, 2014) is used as the optimizer to train the model and the dropout regularization method is applied to prevent overfitting (Catania et al., 2012) with a 50% dropout rate. The well-trained LSTM model is used online public event prediction by monitoring the traffic flow variation in the studied sample area.

Please note that the "autoencoder" model and the LSTM are separately trained since they have different training objectives and are fed by different data. More exactly, as for "autoencoder", the training objective is to minimize the difference between the input vector of the encoder and the output vector from the decoder using coded shock wave diagrams. However, the training objective of LSTM is to minimize the difference between the predicted result and the ground truth (event or no-event), using the extracted features from the "autoencoder".

Our experiments showed that it took about 51 h to train the SW-DLM combined with the "autoencoder" but 71 h to train the SW-LDM without using the "autoencoder". The training time of the "autoencoder" (about 0.5 h) is much shorter than that of the LSTM. This is because the LSTM is fed by time-series data in 2 h, while the "autoencoder" is fed by a single shock wave diagram. Therefore, the "autoencoder" reduced about 20 h for training the LSTM. We conclude that it helps reduce the data dimension and then bring in the significant benefits to reduce the computation load.

5.2. *Experiment results and discussions*

This study conducts the experiments to demonstrate the merits of the SW-DLM, which uses the features extracted from the shock wave diagram as inputs. Given the advantages of the LSTM to other time-series learning models, such as ARIMA and ARMA, has been demonstrated by literature (Tian et al., 2018; Bogaerts et al., 2020), this study mainly focuses on showing the merits of the SW-DLM to existing traffic anomaly detection learning models. SVM is selected as the benchmark since it has achieved success in detecting traffic anomaly in extensive literature (Wang et al., 2013; Yuan and Cheu, 2003; Catania et al., 2012). Specifically, using the training data set collected from May 2011 to August 2011, and the testing data set collected from September 2011 to October 2011, the experiments trained and tested four machine learning models, including (i) the SW-DLM; (ii) the P-DLM (iii); the SW-SVM; (iv) the P-SVM. More exactly, the SW-methods (SW-DLM and SW-SVM) are fed by the features extracted from the shock wave diagrams, while the P-methods (P-DLM and P-SVM) are directly fed by the point traffic data collected at $x_1^{'}$, x_2 and x_3. In addition, the DLM methods (P-DLM and SW-DLM) use time-series data, while the SVM methods (P-SVM and SW-SVM) stack traffic data (point traffic data or shock wave features) rolling in 360 consecutive time intervals as one input matrix and then flatten it into a vector to feed SVM.

The experiments exam the performance of the machine learning models over days and in one-day. The performance is evaluated by the well-accepted metrics proposed in (Wang et al., 2013), including the prediction accuracy (PA: the ratio of the number of correct detections to the number of total detections), the detection rate (DR: the ratio of the number of the correctly detected event to the total number of the event), and the false alarm rate (FAR: the ratio of the number of the false alarms to the total number of the detections without events in truth).

(1) The average performance over the data set

We first compare the convergence performance of the SW-DLM and the P-DLM for the training data set. Fig. 8 shows that the prediction accuracy of the SW-DLM can reach around 93.5% and keep stable after 600 iterations. However, the prediction accuracy of the P-DLM can only reach around 90.5% and almost no improvement after 820 iterations. This result indicates that the shock wave diagram improves the convergence performance of the deep learning model significantly, even though both learning models can converge in the training process.

Next, we compare the prediction performance of the trained SW-DLM, P-DLM, SW-SVM, and P-SVM over the testing data set. The results shown in Table 1 indicate that the SW-DLM outperforms the P-DLM, SW-SVM, and P-SVM with a high accuracy demonstrated by all aspects of the PA, DR, and FAR in the evaluation metrics. Thus, we claim that the deep learning model integrating shock wave analysis is a preferred approach to predict public events. In addition, we noticed that the SW-SVM performs better than the P-SVM in

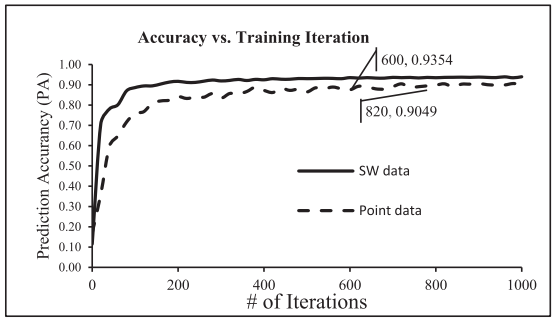

Fig. 8. Experiment results for training data (Average performance).

Table 1
Prediction Performance over testing data.

	Mean				STD			
	SW-DLM	P-DLM	SW-SVM	P-SVM	SW-DLM	P-DLM	SW-SVM	P-SVM
PA	0.92	0.80	0.72	0.65	0.014	0.021	0.023	0.022
DR	0.78	0.56	0.37	0.29	0.052	0.057	0.068	0.065
FAR	0.055	0.05	0.047	0.049	0.0032	0.0041	0.0057	0.0054

almost all aspects of the evaluation metrics. We also tested the performance of SW-LDM without autoencoder. It came with a similar prediction performance to the SW-LDM with autoencoder, but a longer training time. Both outperform P-DLM, SW-SVM, and P-SVM. These results reinforce our thought that the shock wave diagrams involve more comprehensive features than the point traffic data. Using shock wave diagrams will improve the accuracy for recognizing the occurring of a public event, while autoencoder extracts the most important features from the coded shock waves and it helps in dimension reduction. The SW-DLM is a very promising approach for traffic event prediction. The experiments below investigate its performance from more aspects.

(2) The evolvement of the average detection rate

This section discusses the evolvement of the average detection rate (DR) of the SW-DLM as the timeline moving close to the start time of public events. Fig. 9 (a) demonstrates that the average DR (i.e., PE|PE) of the SW-DLM keeps on improving as the leading time becomes shorter. The average DR can approach 76% around an hour before the games. We can expect that the average DR to reach around 80% right before the events. This is satisfactory performance. Our experiments also exam the FAR under the free flow (FF) and recurrent congestion (RC) initial traffic conditions given a public event occurs. Fig. 9 (a) shows that the FAR under free flow (i.e., FF|PE) is relatively high (30–60%) at the timestamp around 1.5 h or even earlier before the start of the games. But it quickly decreases to a rate of around 20% after that. The FAR under the congestion condition (i.e., RC|PE) is constantly low (less than 10%). These results indicate that the SW-DLM can predict the occurring of a public event early enough under both free flow and congestion flow.

We next look at the evolvement of the performance for the P-DLM. The results shown in Fig. 9 (b) indicate that even just before the games, its average DR can only reach around 55% (see the data label (1, 0.56)); its FAR under free flow condition (i.e., FF|PE) is more than 30% and the FAR under the congested flow (i.e., RC|PE) is around 12%. They are all worse than the corresponding performance of the SW-DLM. These results again indicate that the integration of the shock wave diagrams as inputs can improve prediction accuracy.

(3) One day performance evolvement

We last detect the prediction performance of the SW-DLM for a game on an individual day. First, we use the data on Oct 2, 2011, during which a football game occurs at 1:05 PM. Fig. 10 (a) indicates that the probability of the game keeps on increasing as the timestamp is moving closer to its start. Overall, the SW-DLM can predict the upcoming of the football game with a probability of 87% at the timestamp around 80 min before it starts. Moreover, it spends around 30 min (from 2 h to 1.5 h before the start of the game) to improve the probability from less than 10% to around 90%. This result indicates that the SW-DLM can effectively predict an upcoming public event.

Moreover, Fig. 10 (b) displays the traffic speed variation at the location of x_1 on a game day and non-game day. It indicates three significant speed drops occurring around 120, 90, and then 15 min before the start of the game. The SW-DLM is able to issue an alarm for the effect of the football game with the probability of 60% around the first speed drop (105 min before the game), and then with the probability of 80% around the second speed drop (94 min before the game). The P-DLM reports the occurring of the football game around 62 min ahead with the probability of 57%; the SW-SVM and the P-SVM alarm it around 46 and 38 min ahead of the game. At that time, the traffic has been under severe congestion for about 1.5 h. Thus, we claim that the SW-DLM captures the early effect of the public event much timely and accurately than the P-DLM, SW-SVM, and the P-SVM. It can more efficiently help reduce traffic

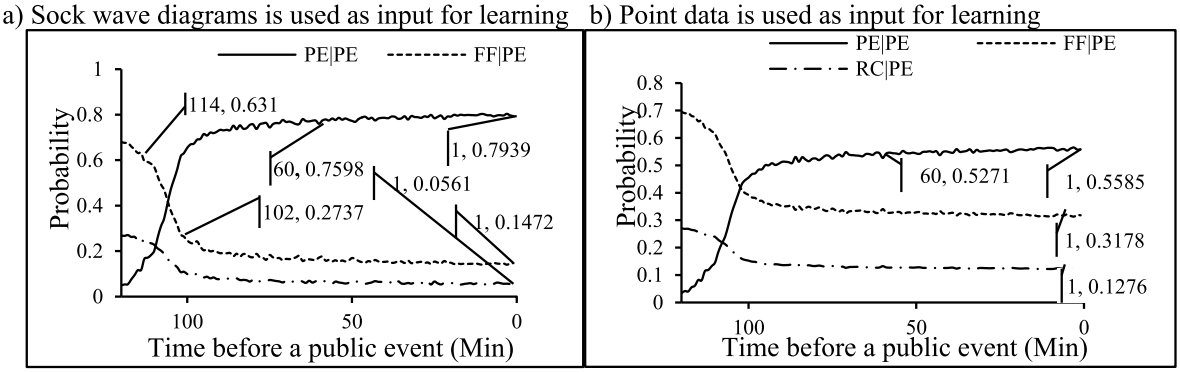

Fig. 9. Experiment results for testing data (Average performance).

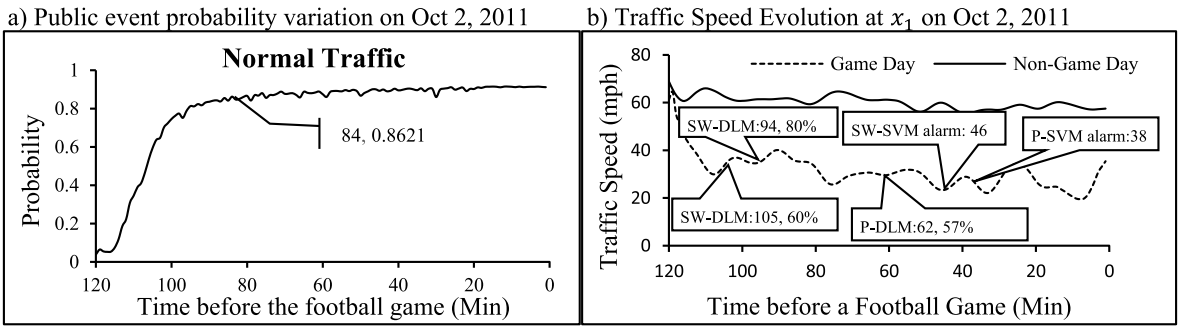

Fig. 10. Experiment results (One day performance).

congestion resulting from these types of public events.

Next, this study demonstrates the performance of the SW-DLM in a day with the interference of a traffic accident. We pick the data on Sep 25, 2011, with a football game at 1:15 PM and an accident occurring around 2.4 miles downstream of Exit 164B at 11:45 AM and cleared at 11:50. Fig. 11 illustrates that the predication probability reaches around 75% (and 80%) at a timestamp one and a half hours (and one hour) before the game. Therefore, the SW-DLM performs well even under the interference of a traffic accident occurring nearby its venue.

Finally, we exam the performance of the SW-DLM to differentiate traffic anomalies resulting from different events such as rain and sports events as well as the normal traffic scenarios such as free flow and congestion flow, according to the traffic data collected at the sample area. We implement the SW-DLM to monitor the traffic from 7:00 to 22:00 on Sep 25. The alarm report is provided in Table 2. The training set still consists of the data from May 2011 to August 2011. The weather station reported the rain at 11 AM and the football game began at 1:15 PM. Accordingly, this SW-DLM alarm table tells us that the traffic is in the free flow condition from 7 AM to 10 AM occurring of the sports game from 11 AM to 12 PM. After that, recurrent congestion started, and it transferred to free flow state from 2 PM to 4 PM. and the bad weather occurred at 11 AM. It also starts to alter the recurrent congestion appeared again from 5 PM to 7 PM. The traffic recovers to a free flow state after that. We can see the consistency between the ground truth and the SW-DLM report. It indicates that the SW-DLM can accurately monitor different traffic scenarios around the sports venue by sensing the traffic data in the selected sample area.

6. Conclusion and future work

Public events in an urban transportation system, such as sports, major promotions, and festivals often cause traffic anomalies and affect traffic efficiency significantly. It has been recognized that timely detecting those events and warning about their early traffic impact will enable traffic management centers to develop adaptive traffic control. On the other hand, travelers en route or CAVs in the future can take advantage of this information to make smart trip plans or reroute their trips to avoid traffic congestion. Therefore, predicting and informing such events and their impacts on travelers en route or traffic manager centers will potentially promote preventive/proactive traffic management and help mitigate traffic congestions. However, the impact of a public event on urban traffic presents complicated spatiotemporal dynamics. These unique features make existing traffic anomaly detection approaches, mainly developed for traffic accident detection, will not work efficiently for either detecting their early traffic impact or predicting the occurring of an upcoming public event.

Motivated by the above view, this study developed the SW-DLM approach, which integrates sophisticated traffic flow analysis and

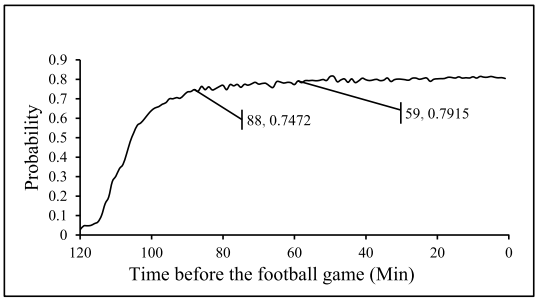

Fig. 11. Prediction on Sep 25, 2011 with a traffic accident.

Table 2
Traffic report on Sep 25, 2011.

Time	Free flow	Congestion	Event	Rain
7:00	Yes	No	X	X
8: 00	Yes	No	X	X
9:00	Yes	No	X	X
10:00	Yes	No	X	X
11:00	Yes	Yes	√	√
12:00	No	Yes	√	X
13:00	No	Yes	X	X
14:00	No	Yes	X	X
15:00	Yes	No	X	X
16:00	Yes	No	X	X
17:00	No	Yes	X	X
18:00	No	Yes	X	X
19:00	No	Yes	X	X
20:00	Yes	No	X	X
21:00	Yes	No	X	X
22:00	Yes	No	X	X

√: alarmed by SW-DLM.

advanced deep-learning approaches to predict the occurring of an upcoming public event through accurately recognizing its early traffic impacts. Mainly, our study contributes to the following methodologies. First of all, the SW-DLM is distinguished since it is a domain knowledge-based rather than purely data-driven approach. The learning approach is fed by the features extracted from the shock wave diagrams rather than directly from traffic data. Next, an automatic shockwave diagram generation algorithm, the SWG, is proposed to efficiently generate and update the time-series shock wave diagram. In addition, this study develops innovative approaches to encode a shock diagram by a three-dimension matrix with the optimal size, which addresses the trade-off between the resolution and computation load. It is recognized that the impact of a public event will generate time-series shock wave diagrams at discrete times. These diagrams expand over time and only involves partial features to capture a public event at the timestamps before the start of the event. By recognizing the characteristics, the "autoencoder" model and LSTM are constructed for feature extraction, dimension reduction, and learning due to their high performance in working on time-series data. Last, the experiment based on filed data shows that the SW-DLM taking shock wave diagrams as input presents timely and accurate predictions on the average as well as in an individual day with or without a traffic accident interference happening nearby the venue of a public event. It outperforms all other machine-learning models (P-DLM, SW-SVM, and P-SVM) using point traffic data as input. Moreover, the SW-DLM is able to successfully differentiate different traffic scenarios under free flow, recurrent congestion, public event, and bad weather.

The SW-DLM can be integrated into real-time traffic provision tools such as GPS to alarm travelers en route on-going events in a transportation network and help them to make a smart trip plan and avoid traffic congestion. Moreover, it is a great added-on capability to a current real-time traffic information system and promotes smart city development by monitoring traffic events in a transportation network.

This presented research represents our initial study for traffic event detecting/predicting in a transportation network. Several promising future research can be further developed. Our study noticed that the SWG algorithm is sensitive to data error. For example, missing data may stop the shock wave generation. A small data error at a timestamp may accumulate in preceding steps and drive a shock wave diagram expanding in a completely wrong direction. Our future study will integrate the SWG with data science approaches to make it adaptive to data imputation. Moreover, this study recognized that it might be hard to identify a significant sample area with available traffic data. Instead, traffic data on several branches around the venue of public events are of significance. Then, we need to develop approaches to identify the optimal data collection spots, considering their traffic flow correlation. We will investigate general approaches to address the above challenges in our near future work.

Author contribution

Hanyi Yang: Methodology developement, Data curation, Validatiion, Writing original draft. **Lili Du:** Conceptualization, Supervision, Formal analysis, Methodology, Writing -review & editing. **Jamshid Mohammadi:** Resources.

Acknowledgment

This research is partially supported by the National Science Foundation awards CMMI-1817346 and CMMI- 1818526. The authors would like to extend their gratitude to the reviewers and editor for their insightful comments, which have increased the quality of this paper.

Appendix A. Shock wave analysis under free flow initial state

Starting from $t \in T_1 = [0,t_1)$, the flow entering at x_1, $q_m(t,x_1)$, increases due to the attraction of the public event. But the increment is not significant such that the flow on the highway and the off-ramp stays $q_m(t,x_1) < Q_{fm}$ (state B in Fig. 2 (a)) and $q_f(t,x_1) < Q_f$ (state $\boldsymbol{\beta}$ in Fig. 2 (c)) respectively. Thus, it does not cause traffic congestion on either the highway or the off-ramp. Accordingly, the exiting flow at the off-ramp will be discharged without a delay. The highway going-through flow in the sub-area II will still keep at state $\boldsymbol{a}$ (see Fig. 2 (b)). Correspondingly, the interface of the states A and B in the sub area I generates the forward shock wave with the speed $W_{AB} = \frac{q^B - q^A}{k^B - k^A}$. It starts at $(0,x_1)$, and ends at $(\frac{L_1}{W_{AB}},x_2)$ where the traffic state A vanishes in the sub-area I (Fig. 2 (d)). Note that this study abuses the notation for a shock wave speed and the shock wave itself throughout the paper to avoid extra notations.

Starting from $t \in T_2 = [t_1,t_2)$, the flow arriving at x_1, $q_m(t,x_1)$, increases and results in state C on the highway in the sub-area I. Even though the highway flow is still under the highway capacity (i.e., $q_m(t,x_1) < Q_m$), the exiting flow is larger than the capacity of the off-ramp, i.e., $q_f(t,x_2) \geq Q_f$ (Fig. 2 (a)). The congestion will occur on the off-ramp and further spill out to the highway. Specifically, the interface of the states B and C in the sub-area I generates the forward shock wave $W_{BC} = \frac{q^B - q^C}{k^B - k^C}$, which starts at (t_1,x_1), and ends at $(t_1 + \frac{L_1}{W_{BC}},x_2)$, where this traffic state C arrives at the off-ramp and the traffic state B vanishes in the sub-area I. Given $q_f(t,x_1) > Q_f$, once the traffic flow in sate C reaches the intersection at x_2, it will make the off-ramp traffic reach to its capacity (denoted as state $\boldsymbol{\gamma}$ in Fig. 2 (c)), i.e., $q_f\left(t + \frac{L_1}{W_{BC}}, x_2 | t \in T_2\right) = Q_f$, and then the extra exiting flow will built up traffic congestion propagating upstream to the intersection x_2. We denote this queue traffic state in the sub-area I as state D. States C and D will meet in the sub-area I and generate a backward shockwave with speed $W_{CD} = \frac{q^C - q^D}{k^C - k^D}$ (Fig. 2 (d)), which starts from the point $(t_1 + \frac{L_1}{W_{BC}},x_2)$ and ends at a point where it intersects another shock wave later. Last, we discuss the traffic in the sub-area II highway. Here, we consider that most of the flow on the exiting lane in the sub-area I will exit for the off-ramp, and only little left for going-through to the sub-area II. Thus, we have the going-through flow approximately equals zero (i.e., $q_m\left(t + \frac{L_1}{W_{BC}} + \frac{L_2}{v_a}, x_3\right) \approx 0$), which is denoted as the state $\boldsymbol{b}$ (Fig. 2 (b)). Accordingly, it is noticed that a forward shockwave W_{ab} occurs at $(t_1 + \frac{L_1}{W_{BC}}, x_2)$ and propagates further downstream in the sub-area II (Fig. 2 (d)). We can see that the increase of the traffic may make the downstream traffic sparse because the exiting flow is blocked at the locationx_2. In reality, the traffic from the adjacent lanes may move in. Our analysis ignores this detail here.

Starting from $t \in T_3 = [t_2,t_3)$, the traffic arriving at x_1 continues increasing so that the total flow is above the highway capacity, $q_m(t,x_1) \geq Q_m$, and the exiting flow is above the off-ramp capacity,$q_f(t,x_1) \geq Q_f$, according to the profile in Eq. (1). Accordingly, the traffic congestion will be built up at the upstream of x_1 (i.e., x_1^-) and the traffic flow entering the sub-area I from x_1 satisfies $q_m\left(t,x_1^+\right) = Q_m$. We denote this traffic state in the sub-area I as E (Fig. 2 (a)). Then, three traffic states E, C, and D co-exist in the sub-area I in the increasing order of their distance away from x_1. Thus, three shock waves are generated. More exactly, the shock wave $W_{CE} = \frac{q^C - q^E}{k^C - k^E}$ is formed at (t_2, x_1) and ends at (t_1^*,y_1^*), where the state C and the shock wave W_{CD} diminish, and the state E meets the state D to start a new backward shockwave $W_{ED} = \frac{q^D - q^E}{k^D - k^E}$. The shock wave W_{ED} will end at a point where it meets the other shockwave later. We next analyze the traffic flow at the sub-area III and II. Given $q_f(t,x_1) \geq Q_f$, we still have $q_f\left(t,x_2^+\right) = Q_f$ and the state $\boldsymbol{\gamma}$ still occupies sub-area III. Regarding sub-area II, since the queue developed during $t \in T_2$ is not cleared in the sub-area I, we have $q_m(t,x_3|t \in T_3) = q_m(t,x_3|t \in T_2)$. Thus, the traffic stays at state $\boldsymbol{b}$.

Starting from $t \in T_4 = [t_3,t_4)$, the time just before the beginning of the public event, the flow arriving at x_1, $q_m(t,x_1)$, decreases to state F, satisfying $q_f(t,x_1) < Q_f$, and$q_m(t,x_1) < Q_m$ (see Eq. (1)). At this time, three traffic states F, E, and D co-exist in sub-area I in the increasing order of their distance away from x_1. (C may also not disappear yet. Then, the states are F, E, C, and D.) Accordingly, three shock waves are generated. Specifically, the interface of states F and E generates a forward shock wave $W_{EF} = \frac{q^E - q^F}{k^E - k^F}$, which starts at (t_3, x_1) and ends at (t_2^*, y_2^*), where state E and the shock wave W_{ED} vanish, and a new forward shock wave W_{FD} starts.[1] The forward shock

[1] Here we assume that traffic flow rate of F is less than that of the state D, otherwise, it will be a backward shock wave and the queue will continue growing. However, as the coming flow attracted by the public even continues decreasing, eventually it will reach a traffic state which has a flow rate less than the state D. The analysis ignores this detail.

wave W_{FD} will stop at some point. Depending on the length of the time interval of T_4, the shock wave $W_{FD} = \frac{q^D - q^F}{k^D - k^F}$ may or may not be able to reach x_2. We next discuss the traffic in sub-area II and III under these two scenarios.

Scenario 1, if the shock wave W_{FD} does not reach x_2 during T_4, but vanishes due to a new shock wave (such as W_{FA} later) generated inT_5 (see Fig. 2 (d)). Then the flow leaving sub-area I at x_2 during T_4 is all exiting flow in State D. Thus, we have $q_m(t, x_2) = q_f(t, x_2) = Q_f$ at the intersection. The traffic in sub-area III then stays at the state $\boldsymbol{\gamma}$ and the traffic in Subarea II stays at state $\boldsymbol{b}$.

Scenario 2 (see the cited box in Fig. 2 (d)), if the shock wave W_{FD} can reach the off-ramp at the time t_4^* before a new shock wave (such as W_{FA} later), it clears the queue of the exiting flow (state D) around the highway-ramp intersection. Namely, after t_4^*, the flow arriving at x_2 and entering the off-ramp is part of the flow under state F, which satisies $q_f(t, x_1) < Q_f$. Accordingly, the flow in the sub-area III is changed to a free flow denoted as state $\boldsymbol{\delta}$ in Fig. 2 (c); the going-through flow entering the sub-area II is a portion of $q_m(t, x_1)$ within the range of $(0, q_0]$. We approximate this traffic state by the initial traffic state $\boldsymbol{a}$. Then, a forward shock wave W_{ba} is generated.

Starting from $t \in T_5 = [t_4, \infty)$, the flow arriving at x_1, $q_m(t, x_1)$, decreases to the initial state A. Then, we have $q_m(t, x_1) < Q_m$, $q_f(t, x_1) = 0$ (see Eq. (1)). If the first scenario occurs in the previous stage, three traffic states A, F, and D co-exist in the sub-area I in the increasing order of their distance away from x_1. The interface of the states A and F generates the forward shock wave $W_{FA} = \frac{q^F - q^A}{k^F - k^A}$. The shock wave W_{FA} starts at (t_4, x_1) and ends at (t_3^*, y_3^*), where the shock wave W_{FA} intersects the shock wave W_{FD} so that the traffic state F and the shock waves W_{FD} vanish. At the same time, a new forward shock wave with the speed $W_{AD} = \frac{q^D - q^A}{k^D - k^A}$ is generated, which clears the queue of the exiting flow at the time t_5^*. After that, the traffic recovers to the initial traffic state. If the second scenario occurs in T_4, two traffic states co-exist in the sub-area I, including states A and F in increasing order of their shock wave W_{FA}, which starts at (t_4, x_1) in the sub-area I, and ends at $(t_4 + \frac{L_1}{W_{FA}}, x_2)$ due to the traffic diverge at the off-ramp. At the same point, the traffic state F vanishes.

Appendix B. Shock wave analysis under congested initial state

This appendix shows the effect of a public event when the highway in the study area is initially under congestion due to a downstream bottleneck. Recall we denote the left and right-hand side of the site x_2 by $x_2^-(i.e., x \in (x_2 - \varepsilon, x_2))$ and $x_2^+(i.e, x \in (x_2, x_2 + \varepsilon))$,[2] where ε a small positive value. Traffic flow arrives at x_2^- and leaves from x_2^+ at the highway-ramp intersection. Accordingly, we have relationships with Eq. (9) to Eq. (12) for the arriving and leaving flows at x_2.

$$q_m(0, x_1) = q_m\left(0, x_2^-\right) = q_m\left(0, x_2^+\right) = q_\mu\left(0, x_2^+\right) = q_m(0, x_3) = q_0 \tag{9}$$

$$q_m\left(t, x_2^-\right) \geq q_\mu\left(t, x_2^+\right) + q_f\left(t, x_2^+\right), \forall t \tag{10}$$

$$q_f\left(t, x_2^+\right) \leq min\left\{\theta(t) q_m\left(t, x_2^-\right), Q_f\right\}, \forall t \tag{11}$$

$$q_\mu\left(t, x_2^+\right) \leq min\left\{(1 - \theta(t)) q_m\left(t, x_2^-\right), Q_m\right\}, \forall t \tag{12}$$

where $\theta(t) = \frac{q_f(t, x_1)}{q_m(t, x_1)}$ is the proportion of the exiting flow to the off-ramp at the time stamp t, and denote $\theta(0) = \theta_0$; q_0 is the initial stable highway traffic flow. It represents the initial traffic state A in the sub-area I and state $\boldsymbol{a}$ in the sub-area II.

Specifically, we assume the initial exiting flow to the off-ramp is not significant, i.e., $\theta_0 \approx 0$, we have the initial traffic flow on mainline road satisfy Eq. (9). Given congestion may cause capacity drop on the highway, we consider$(q_0 \leq Q_m)$. Thus, Eq. (9) indicates that the highway traffic is initially running at the dropped capacity. Moreover, Eq. (9) and Eq. (11) together show that all arriving flow at x_2 can be discharged given the off-ramp is under free flow. But it starts to build up queue at x_2^- if the exiting flow rate is greater than the off-ramp capacity.

Next, Eq. (12) shows that the going-though flow leavingx_2 is limited by the arriving flow and highway capacity. With the above agreement, we conduct shock wave analysis for the arriving flow at x_1 fluctuates according to Eq. (1).

Starting from $t \in T_1 = [0, t_1)$, as the exiting flow $q_f(t, x_1)$ entering at the location x_1 slightly increases, so does the value of $(t)(>\theta_0)$ due to the attraction of the public event, but still satisfies $q_f(t, x_1) < Q_f$. Due to the capacity drop resulting from downstream traffic congestion, the sub-area I can only accommodate flow rate q_0 (i.e., stay at the state A in Fig. 3 (a)). The earliest traffic flow during the course of T_1 arrives at the intersection at x_2 by the time $t_0^{\cdot} = L_1 / v_A$, where v_A is the speed under traffic state A. Thus, the arriving flow at x_2 satisfied Eq. (13) is below.

$$q_m\left(t + t_0^{\cdot}, x_2^-\right) = q_m\left(t, x_1^+\right) = q_0, t \in T_1 \tag{13}$$

Next, this study analyzes the diverging of the arriving flow at x_2. Eq. (18), Eq. (10), and Eq. (13) indicate that the arriving flow at x_2 by the time $t_0^{\cdot}$ equals the flow leaving x_2, thus we have a stable flow at x_2, i.e., $q_m\left(t + t_0^{\cdot}, x_2^+\right) = q_0$. Given $q_f(t, x_1) < Q_f$, all the flow exiting to the off-ramp is accommodated, so no traffic queue is built up at the entrance of the off-ramp. We denote the new traffic state on the off-ramp by $\boldsymbol{\beta}$ in Fig. 3 (c). Furthermore, the increase of $q_f(t, x_2)$ will cause the decrease of the going-through flow in the sub-area II (see Eq. (14)), which leads to a new traffic state in the sub-area II denoted by the state $\boldsymbol{b}$ in Fig. 3 (b).

[2] i.e., $q_m(t, x_2^-)$ represents the arriving flow on the left side of x_2, and $q_m(t, x_2^+)$ the leaving flow on the right side of x_2.

$$q_\mu\left(t,x_2^+\right)=q_m\left(t,x_2^+\right)(1-\theta(t))=q_0(1-\theta(t))<q_0, t\in T_1 \tag{14}$$

Note that the state ***b*** results from the decrease of the arriving going through traffic demand. Thus, ***b*** is with a lower flow rate than the initial traffic state ***a***[3] and under free-flow traffic state. The co-existing traffic states ***a*** (=*A*), ***b*** on the highway will generate a forward shock wave $W_{ab}=\frac{q^a-q^b}{k^a-k^b}$.

Starting from $t\in T_2=[t_1,t_2)$, the arriving traffic flow at x_1 for the public event increases such that $q_f(t,x_1)\geq Q_f$, which also leads to an increase of $\theta(t)$. Due to the capacity drop, the traffic flow entering the sub-area I at x_1 is still limited at q_0 (i.e., stay at the state of *A*) but with an increased proportion of the flow exiting for the off-ramp. When the earliest flow arrives at the exit of the off-ramp during T_2 by the time $t_1^{'}$, it causes traffic congestion at the exit to the off-ramp since $q_f(t,x_1)\geq Q_f$. As a result, a traffic queue will be built up at the exit (i.e., the location x_2) such that $q_f\left(t,x_2^+\right)=Q_f$. This new traffic state on the off-ramp is denoted as ***γ***. The traffic queue on the off-ramp will propagate back to the highway and block the arriving flow so that we have Eq. (15) and Eq. (16), which respectively represent the traffic state *B* and state ***c*** shown in Fig. 3 (a) and (b) for the sub-area I and II. Note that the queue flow is mainly composed of the exiting flow for the public event, thus $\theta\left(t|t>t_1^{'}\right)=1$, and we only analyze the flow of the lane adjacent to the off-ramp.

$$q_m\left(t_1^{'},x_2^-\right)\geq q_m\left(t_1^{'},x_2^+\right)=Q_f \tag{15}$$

$$q_\mu\left(t_1^{'},x_2^+\right)\approx 0 \tag{16}$$

The traffic state *A* and *B* in the sub-area I will form a backward shock wave W_{AB} which starts at x_2 at the time $t_1^{'}$. In the sub-area II, the state ***c***, ***b,*** and ***a*** co-exist (please see the demonstration in Fig. 3 (d)), which forms three shock waves. More exactly, the shock wave W_{bc} is formed at point ($t_1^{'}$, x_2), and ends at point (t_1^*,y_1^*), where the state***b*** and the shock wave W_{ab} diminishes. Then a new shock wave, W_{ac}, starts and it will end at a point when it meets the other shockwave later.

Starting from $t\in T_3=[t_2,t_3)$, the arriving traffic flow at $x_1$for the public event continue increasing and make $q_f(t,x_1)\geq Q_f$. The traffic flow entering the sub-area I can be *A* or *B* depending on whether traffic state *B* (the queue) has propagated to x_1 before t_2 (see Eq. (1)) or not. Here, we assume that if the shock wave W_{AB} arrives at x_1 is later than the earliest time that the flow arrives at x_1 (i.e.,t_2 in Eq. (1)), i.e., $t_2\leq t_1^{'}+\frac{L_1}{W_{AB}}$. Then, we have $q_m(t,x_1)=q_0$ at x_1. Accordingly, there are two traffic state *A* and *B* in sub-area I and a backward shock wave W_{AB} continue propagating toward x_1. All the new arriving traffic will be stacked at the end of the queue. When it arrives at the exit at x_2, the congestion continues because of $q_f(t,x_1)\geq Q_f$. So (15) and (16) still holds. Accordingly, the traffic state ***c*** (shown in Fig. 3 (b)) continues in the sub-area II. The traffic state in the sub-area III is still at its capacity state, ***γ***.

Starting from $t\in T_4=[t_3,t_4)$, the arriving traffic flow at $x_1$for the public event decreases such that $q_f(t,x_1)<Q_f$. Due to the congestion at x_1, the entering traffic flow stays at A[4] ($q_m(t,x_1)=q_0$) like the discussion in the previous period but with a decreased proportion of exiting flow, i.e., a smaller value of $\theta(t)$. However, when the earliest time of this flow arrives at the point x_2 by the time $t_3^{'}$, the congestion at the exit (state *B*) will not continue since $q_f(t,x_1)<Q_f$. Given the vacancy (state *c*) in the sub-area II of the highway, all arriving going-through flow is able to be accommodated and speed up at x_2^+ and make $q_\mu\left(t_3^{'},x_2^+\right)$ increase to q^d, which may be slightly less than Q_m. This traffic state in the sub-area II is denoted as state ***d*** in Fig. 3 (b). Given $q_m\left(t_3^{'},x_2^+\right)=q_\mu\left(t_3^{'},x_2^+\right)+q_f$, the speeding up of the going-through flow will make the traffic flow rate leaving x_2 increase, i.e., $q_m\left(t_3^{'},x_2^+\right)\rangle q_m\left(t_3^{'},x_2^-\right)=q^B$. Denoting $q_m\left(t_3^{'},x_2^+\right)$ at the state *D* for the sub-area I. Note that the state *D* will evolve from congested state to free flow state along with the increasing of the proportion of going-through flow. Accordingly, this traffic change introduces two new shock waves formed at point ($t_3^{'}$, x_2) in the temporal and spatial diagram: a backward shock wave W_{BD} toward the sub-area I to clean the queue and a forward shock wave W_{cd} toward the sub-area II to diminish the vacancy. The backward shock wave W_{BD} propagates toward upstream and arrives at x_1 by the time $t_3^{'}+\frac{L_1}{W_{BD}}$, thus make traffic state *D* in the sub-area I. The forward shock wave W_{cd} propagates toward downstream and meets the previous shock wave W_{ac} at point (t_2^*,y_2^*). This intersection makes the state ***c*** vanish. Consequently, a new backward shock wave W_{ad} starts, which will end at a point where it meets the other shockwave later. Regarding the off-ramp, we have $q_f\left(t,x_2^+\right)=\theta(t)q_m\left(t,x_2^+\right)<Q_f$ and denoted it as state ***δ*** in Fig. 3 (c).

Starting from $t\in T_5=[t_4,\infty)$, the public event has started. Accordingly, the traffic flow arriving at x_1 for the public event decreases and leads to $(t)=\theta_0\approx 0$. The traffic flow at x_1 depends on the propagation of shock ware W_{AB} and W_{BD}. There are three potential scenarios. (i) Given the flow enters x_1 at $t\in T_5$,if$t<t_1^{'}+\frac{L_1}{W_{AB}}$, then the traffic flow enters x_1 before the shock wave W_{AB} arrives at x_1. Therefore, x_1 can only accommodate flow rate q_0. That is state *A* and $q_m(t,x_1)=q_0$. (ii) If $t\in[t_1^{'}+\frac{L_1}{W_{AB}},t_3^{'}+\frac{L_1}{W_{BD}}]$, the traffic flow enters x_1 in-between the timestamps that shock wave W_{AB} and W_{BD} arrives at x_1. Accordingly, we have $q_m(t,x_1)$ equal the flow under the state *B*. (iii) If $t>t_3^{'}+\frac{L_1}{W_{BD}}$, the flow enters x_1 later than the shock wave W_{BD}, then only the flow rate equal to state *D* can be accommodated. Namely, the traffic flow $q_m(t,x_1)$ is at the state *D*. Fig. 3 (d) presents scenario (ii). Thus, state *B*is up at x_1. Accordingly, two traffic states, *D* and *B*, co-exist in the sub-area I and the corresponding backward shock wave W_{BD} continues propagating toward x_1.

Under all the three scenarios at x_1, the earliest flow arrives at x_2 at $t_4^{'}$, which is always later than $t_3^{'}$ in T_4, by which the traffic state *D*

[3] The leaving of exiting flow provides going-through traffic space to speed up at x_2. It generates a state C with higher flow rate in sub-area I ($q_\mu>q_0$). Accordingly, a backward shock waves (W_{AC}, and W_{BC}) in Fig. 3.

[4] If $t_2>t_1^{'}+\frac{L_1}{W_{AB}}$, then it is state *B*

already occurs. Thus, we have traffic flow at the upstream of x_2 is under state D, i.e., $q_m(t, x_2^-) = q^D \le Q_m$. We next discuss traffic state at the downstream of x_2. It is noticed that little traffic exiting for off-ramp (i.e. $\theta(t) \approx 0$) and the sub-area II is under free flow state after $t_1^{'}$. Then, we have the flow rate entering the sub-area II will increase. i.e., $q_m(t_4^{'}, x_2^-) = q_m(t_4^{'}, x_2^+) = q^D \ge q^d$. This leads to a traffic state $\boldsymbol{e}$ in the sub-area II. Accordingly, there are three traffic states $\boldsymbol{e}$, $\boldsymbol{d}$, and $\boldsymbol{a}$ co-existing in the sub-area II, which form two shock waves. More exactly, the shock wave W_{de} is formed at point $(t_4^{'}, x_2)$ between state $\boldsymbol{d}$and $\boldsymbol{e}$. It ends at point (t_3^{*}, y_3^{*}), where W_{de} intersects W_{ad} and the state $\boldsymbol{d}$ and the shock wave W_{ad} diminishes. Then a new shock wave, W_{ae}, starts and it ends at x_2 and continues propagating toward x_1 labeled as W_{DA} in the sub-area I. The shock wave W_{DA} will continue propagating toward x_1 by the time $t_5^{'} + \frac{L_1}{W_{AC}}$. Then the studied road section returns to the initial state. In the sub-area III, the traffic state also comes to initial state $\boldsymbol{\alpha}$.

Remark 1. Even though the shock wave analysis for the off-ramp is not presented, the above shock wave analyses can reflect the traffic flow variation on the off-ramp. We demonstrate this point from the following aspects. First of all, according to flow conservation law, we have $q_m(t, x_2) = q_\mu(t, x_2) + q_f$. Thus, the traffic flow on the off-ramp (in the sub-area III) can be fully represented by the traffic flow in the sub-area I and II. Second, according to the shock wave analyses in Fig. 2 (d) and Fig. 3 (d), we can see that traffic variation on the off-ramp can be detected by the shock waves in either the sub-area I or sub-area II. Thus, the shock wave analyses for the sub-area I and II captures the traffic effect on the study sample area sufficiently. Also, the traffic data on off-ramp is not very necessary.

Remark 2. With minor modifications, the analysis can cover the effect of the traffic congestion occurring at the local infrastructure (such as local intersections and parking lots) around the venue, which is adjacent or close to the off-ramp. More exactly, local traffic congestion may propagate back to the off-ramp and further affect the highway traffic. We discuss this effect under two scenarios relevant to traffic flow at x_2. (i) Downstream local traffic congestion occurs first, propagates backward and causes off-ramp capacity to drop to $Q_f^{'}$, i.e. we have $Q_f^{'} < q_f(t, x_2) \le Q_f$ during T_1 in Eq. (1). Then, state D ($q^D = Q_f^{'}$) in Fig. 2 (d) (and B in Fig. 3 (d)) on the highway will occur earlier. (ii) The off-ramp traffic congestion occurs first at t_1 due to the rise of the upstream traffic demand, and the traffic on off-ramp reaches its capacity. Later, the local traffic congestion occurs at $t_1^+ > t_1$ may further block off-ramp traffic and cause the capacity drop on the off-ramp. Then the traffic state D in Fig. 2 (d) (and B in Fig. 3 (d)) will evolve to a new state $D^{'}$ ($B^{'}$ in Fig. 3 (d)) with $q^{D^{'}} = Q_f^{'}$ (or $q^{B^{'}} = Q_f^{'}$). The sequential shock waves further propagate according to states $D^{'}$ ($B^{'}$). Therefore, the shock wave diagram on the highway segment can reflect the effect of the relevant local traffic congestion to a certain extent. Thus, local traffic data is not very necessary either in this case.

Appendix C

There are three parts in an LSTM cell. They are respectively used to take previous information, memorize the current state, and generate output. In each time step, an LSTM cell will take the new collected data (x_t) and the information preserved from the previous states (h_{t-1}) as new input. This input will feed to the output gate (o_t) to generate current output. Meanwhile, this input will also feed to the input gate (i_t) Eq. (17), forget gate (f_t) Eq. (18), and g_t Eq. (20). Then, the current system state (c_t) can be calculated by Eq. (21), in which i_t is used to analyze the new input, forget gate (f_t) is used to determine how much previous system state (c_{t-1}) is kept, and g_t will indicate the weight of new input data**.** Finally, the information preserved to feed into the next step, (h_t), is calculated based on the current state, (c_t), and current output (o_t) through Eq. (19). Then, repeating this process, the LSTM can consecutively provide current output based on historical information.

$$i_t = \sigma(W_i[x_t, h_{t-1}] + b_i) \tag{17}$$

$$f_t = \sigma(W_f[x_t, h_{t-1}] + b_f), \tag{18}$$

$$o_t = \sigma(W_o[x_t, h_{t-1}] + b_o), \tag{19}$$

$$g_t = tanh(W_o[x_t, h_{t-1}] + b_g), \tag{20}$$

$$c_t = f_t{*}c_{t-1} + i_t{*}g_t, \tag{21}$$

$$h_t = o_t{*}tanh(c_t), \tag{22}$$

where $\sigma(x) = \frac{1}{1+\exp(-x)}$, W and b are parameters.

References

Asakura, Y., Kusakabe, T., Nguyen, L.X., Ushiki, T., 2017. Incident detection methods using probe vehicles with on-board GPS equipment. Transp. Res. Part C: Emerg. Technol. 81, 330–341.

Bogaerts, T., Masegosa, A.D., Angarita-Zapata, J.S., Onieva, E., Hellinckx, P., 2020. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. Transp. Res. Part C: Emerg. Technol. 112, 62–77.

Catania, C.A., Bromberg, F., Garino, C.G., 2012. An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection. Expert Syst. Appl. 39 (2), 1822–1829.
Daganzo, C.F., 1997. Fundamentals of Transportation and Traffic Operations Vol. 30, 67–160.
Dia, H., Thomas, K., 2011. Development and evaluation of arterial incident detection models using fusion of simulated probe vehicle and loop detector data. Inform. Fusion 12 (1), 20–27.
https://fbschedules.com/nfl-2011/team/seattle-seahawks. The latest access: July, 2018.
Fu, R., Zhang, Z., Li, L., 2016. Using LSTM and GRU neural network methods for traffic flow prediction. In: 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 324–328. IEEE.
Fuse, T., Kamiya, K., 2017. Statistical anomaly detection in human dynamics monitoring using a hierarchical dirichlet process hidden markov model. IEEE Trans. Intell. Transport. Syst. 18 (11), 3083–3092.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313 (5786), 504–507.
Ihler, A., Hutchins, J., Smyth, P., 2006. Adaptive event detection with time-varying Poisson processes. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 207–216. ACM.
Ivan, J.N., Sethi, V., 1998. Data fusion of fixed detector and probe vehicle data for incident detection. Comp-aided Civil Eng. 13 (5), 329–337.
Jeong, Y.-S., Castro-Neto, M., Jeong, M.K., Han, L.D., 2011. A wavelet-based freeway incident detection algorithm with adapting threshold parameters. Transp. Res. Part C: Emerg. Technol. 19 (1), 1–19.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inform. Process. Syst. 1097–1105.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Lee, K.Y., Sode-Yome, A., Park, J.H., 1998. Adaptive Hopfield neural networks for economic load dispatch. IEEE Trans. Power Syst. 13 (2), 519–526.
Li, P., Abdel-Aty, M., Yuan, J., 2020. Real-time crash risk prediction on arterials based on LSTM-CNN. Accid. Anal. Prev. 135, 105371. https://doi.org/10.1016/j.aap.2019.105371.
Liu, C., Zhao, M.o., Sharma, A., Sarkar, S., 2019. Traffic dynamics exploration and incident detection using spatiotemporal graphical modeling. J. Big Data Anal. Transp. 1 (1), 37–55.
Maitra, S., Yan, J., 2008. Principle component analysis and partial least squares: two dimension reduction techniques for regression. Appl. Multivar. Stat. Models 79, 79–90.
May, A.D., 1990. Traffic Flow Fundamentals. Prentice Hall, Englewood Cliffs, NJ.
Parkany, E., Xie, C., 2005. A complete review of incident detection algorithms & their deployment: what works and what doesn't. No. NETCR 37, NETC 00-7.
Polat, K., Güneş, S., 2008. Principles component analysis, fuzzy weighting pre-processing and artificial immune recognition system based diagnostic system for diagnosis of lung cancer. Expert Syst. Appl. 34 (1), 214–221.
Rahman, R., Hasan, S., 2018. Short-term traffic speed prediction for freeways during hurricane evacuation: a deep learning approach. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 1291–1296. IEEE.
Szer, D., Charpillet, F., 2005. An optimal best-first search algorithm for solving infinite horizon DEC-POMDPs. In: European Conference on Machine Learning, pp. 389–399.
Tian, Y., Zhang, K., Li, J., Lin, X., Yang, B., 2018. LSTM-based traffic flow prediction with missing data. Neurocomputing 318, 297–305.
https://catalog.data.gov/dataset/seattle-20-second-freeway. The latest access: June, 2018.
Wang, J., Li, X., Liao, S.S., Hua, Z., 2013. A hybrid approach for automatic incident detection. IEEE Trans. Intell. Transport. Syst. 14 (3), 1176–1185.
Wang, Y., Yao, H., Zhao, S., 2016. Auto-encoder based dimensionality reduction. Neurocomputing 184, 232–242.
Witayangkurn, A., Horanont, T., Sekimoto, Y., Shibasaki, R., 2013. Anomalous event detection on large-scale GPS data from mobile phones using hidden Markov model and cloud platform. In: Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, pp. 1219–1228. ACM.
Wong, F.S., 1991. Time series forecasting using backpropagation neural networks. Neurocomputing 2 (4), 147–159.
Yuan, F., Cheu, R.L., 2003. Incident detection using support vector machines. Transp. Res. Part C: Emerg. Technol. 11 (3-4), 309–328.
Zhang, W., Yu, Y., Qi, Y., Shu, F., Wang, Y., 2019. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. Transportmet. A: Transport Sci. 15 (2), 1688–1711.
Zhang, J., Zheng, Y.u., Qi, D., Li, R., Yi, X., Li, T., 2018. Predicting citywide crowd flows using deep spatio-temporal residual networks. Artif. Intell. 259, 147–166.
http://www.mlb.com/sea/downloads/y2011/printable_promo_schedule.pdf. The latest access: May, 2018.