Forecasting Spatiotemporal Impact of Traffic Incidents on Road Networks

Bei Pan, Ugur Demiryurek, Cyrus Shahabi
Integrated Media System Center
University of Southern California
Los Angeles, United States
{*beipan,demiryur,shahabi*}*@usc.edu*

Chetan Gupta
Hewlett Packard Labs
Palo Alto, United States
chetan.gupta@gmail.com

***Abstract*—The advances in sensor technologies enable real-time collection of high-fidelity spatiotemporal data on transportation networks of major cities. In this paper, using two real-world transportation datasets: 1) incident data and 2) traffic data, we address the problem of predicting and quantifying the impact of traffic incidents. Traffic incidents include any non-recurring events on road networks, including accidents, weather hazard, road construction or work zone closures. By analyzing archived *incident* data, we classify incidents based on their features (e.g., time, location, type of incident). Subsequently, we model the impact of each incident class on its surrounding traffic by analyzing the archived *traffic* data at the time and location of the incidents. Consequently, in real-time, if we observe a similar incident (from real-time incident data), we can predict and quantify its impact on the surrounding traffic using our developed models. This information, in turn, can help drivers to effectively avoid impacted areas in real-time. To be useful for such real-time navigation application, and unlike current approaches, we study the dynamic behavior of incidents and model the impact as a quantitative time varying spatial span. In addition to utilizing incident features, we improve our classification approach further by analyzing traffic density around the incident area and the initial behavior of the incident. We evaluated our approach with very large *traffic* and *incident* datasets collected from the road networks of Los Angeles County and the results show that we can improve our baseline approach, which solely relies on incident features, by up to 45%.**

***Keywords*-intelligent transportation, traffic forecast, traffic incidents, impact analysis, spatiotemporal data**

I. Introduction

The Texas annual Transportation report [2] estimates that 5.5 billion hours and 2.9 billion gallons of fuel are wasted due to the problem of traffic congestion in the United States in 2012. According to [16], approximately 50% of the freeway congestions are caused by non-recurring incidents, such as traffic accidents, weather hazard, special events and construction zone closures. Hence, our goal is to predict and quantify the impact of traffic incidents on the surrounding traffic. This quantification can eliminate the significant financial and time lost by traffic incidents, for example it can be used by city transportation agencies for providing evacuation plan to eliminate potential congested grid locks, for effective dispatching of emergency vehicles, or even for long-term policy making.

The McKinsey report [1] predicts a worldwide consumer saving of more than $600 billion annually by 2020 for location-based-services, where the biggest single consumer benefit will be from time and fuel savings from navigation services tapping into real-time traffic data. Therefore, for the remainder of this paper, we focus on a next generation consumer navigation system (in-car or on smart phone), called *ClearPath*, as a motivating application, which can help drivers to effectively plan their routes in real-time by avoiding the incidents' impact areas. That is, suppose an accident is reported in real-time (by crowdsourcing [22] or through agency reports or SIGALERTS [19]) in front of a driver but the accident is 20 minutes away. If we can effectively quantify the impact of the accident, ClearPath would know that this accident would be cleared in the next 10 minutes. Thereby, ClearPath would guide the driver directly towards the accident because it knows that by the time the driver arrives the area, there would be no accident.

To be more specific, consider another example illustrated in Figure 1. In this figure, the caution mark, the directed solid red lines, and the dashed blue lines represent the incident location, the congested region caused by the incident, and the route a driver plans to follow, respectively. Without prediction, but with the knowledge of the incident, a typical navigation application, such as Waze [22], may suggest the route shown in Figure 1(a) to the drivers. If the driver follows this route, he would be stuck in the traffic congestion caused by the incident, as illustrated in Figure 1(b), due to the fact that the congested region has grown. On the other hand, if we can predict how the impacted spatial span (i.e., congested region) evolves over time, ClearPath could calculate the route that can effectively avoid the congestion from the beginning, as shown in Figure 1(c).

The problem of predicting traffic incident's impact has been widely studied by researchers in multiple disciplines, including in transportation science, civil engineering, policy planning, and operational research (e.g. [23]). In the past, without real-world traffic data, most researchers resorted to mathematical models, simulation studies and field surveys (e.g., [10]). However, these theoretical methodologies cannot accurately infer the impact of incidents in real-world scenarios and the spatial transferability of their models is

(a) (b) (c)

Figure 1. (a) route calculated based on current incident's impact (b) time-varying expansion of impacted region as driver approaches the incident location (c) route calculated based on accurate prediction of impact

limited. In recent years, due to the sensor instrumentation of road networks in major cities as well as the vast availability of auxiliary commodity sensors (e.g., CCTV cameras, GPS devices), for the first time a large volume of real-time *traffic* and *incident* data at very high spatial and temporal resolutions has become available. In this paper, we use such datasets that we have been collecting and archiving in the last three years in the LA County and Orange County, towards predicting the impact of traffic incidents to the surrounding traffic.

For our motivating navigation application, ClearPath, to be effective, we need to predict specific values of speed changes and backlog lengths over the lifetime (i.e., temporal) and impact-area (i.e., spatial) of an incident. This is in contrast to previous application scenarios where forecasting abstract or aggregate values was sufficient. In particular, consider the following three aspects that we need to forecast.

First, we need to predict the exact values of speed changes and backlog lengths. There are two major approaches to measure the impact of incidents: 1) *qualitative* approaches (i.e., classify incident's impact into conceptual categories such as "severe" or "non-severe", and "significant delay" or "slight delay"); 2) *quantitative* approaches (i.e., providing numeric measurement such as 45% speed decrease, and 3.2 miles of congested backlog). In the past, most studies focused on *qualitative* approaches for measuring impact, which makes the impact easier to predict (e.g., [13]). The qualitative measurement may be sufficient for general decision-making or response analysis, however, not precise enough for ClearPath. In this paper, we describe the impact from a *quantitative* perspective, and provide numeric measurements of the impact to the surrounding areas.

Second, since the impact region of an incident evolves over time and space (as shown in Figure 1), we need to predict the spatiotemporal behavior of the impact. In previous studies, it was sufficient to predict an incident's impact as a single or a set of aggregate values. For example, in [15], the impact is predicted as average speed decrease or average of the backlog length. In this paper, the outcome of our prediction approach is the exact length of time varying backlogs (i.e., evolution of congested spatial span) with different scales of speed changes.

Third, we need to predict the sudden speed changes caused by incidents in a faraway future (e.g., the next 30 minutes). The occurrence of incidents always involves two phenomenon: 1) *abrupt* speed changes; for example, it is very common for the traffic speed to drop 60% when an incident occurs on freeways in LA; and 2) *long*-lasting propagation of the speed changes; for example, a closer sensor to the incident may report speed decrease in 3^{rd} minute after its occurrence, however, a farther sensor may report similar decrease in 30^{th} minute. Since traditional prediction approaches rely on the immediate past data to predict the future, they cannot effectively predict the *abrupt* speed changes and how they propagate over a *long* term, which is important for ClearPath to successfully navigate drivers around the incident impact area. Towards this end, we analyze the correlations between archived incident data and traffic data. Specifically, we first classify incidents based on their features (e.g., time, location, type of incident), which are correlated with their impact to the surrounding traffic. Next, we improve the classification by incorporating traffic density and the initial behavior of incident. By utilizing such models, we can effectively predict the *abrupt* speed change and the propagation over a *long* term by identifying similar classes of incidents mined from archived dataset.

In sum, the contributions of our paper are as follows:

- We present a novel method to quantify the impact of incidents as a time varying spatial span, showing significant advantages over static impact measurements, e.g., revealing the affected spatial region more precisely, and enabling next-generation navigation applications.
- For impact prediction, we leverage incident features, traffic density, and the initial incident behavior to improve the accuracy in forecasting of time-varying spatial spans. Consequently, our approaches significantly reduce the prediction error.
- We validate our approaches using a large-scale, real-world traffic and incident datasets, which constitute the data collected from 4,230 sensors and 6,811 incidents on road network. Our results show that our baseline approach that relies only on incident features is already superior to the state-of-the-art as it can quantify impact over space and time. Moreover, once we incorporate the traffic density and the initial behavior of incident into our prediction model we can improve the baseline approach even further by up to 45%.

II. Related Work

In the last decade, the impact of traffic incidents has been widely studied in multiple disciplines. Most of these studies are based on theoretical modeling and simulations, which can be classified into three groups: 1) deterministic queuing theory or shockwave theory (e.g.,[10], [23]); 2) heuristic methods and simulations (e.g.,[14]); 3) microscopic

modeling of driver's behavior (e.g., [6], [21]). However, the outcome of these studies relies on theoretical simulations of road network traffic instead of the real-world collected traffic data. Also, none of these studies use a source of incident data with description variables and reporting techniques, and their spatial transferability is limited. In this work, we use a very detailed high resolution traffic dataset and incident dataset.

Recently, with the availability of real-world data, a variety of data mining approaches have been applied for the prediction of incident's impact, such as decision trees [13], classification trees [8], [20], as well as Bayesian classifier [3] and nearest neighbor classifier [9]. In most of these studies, the focus is to predict the general behavior of incident's impact (e.g., severe or not severe [7]). Thereby, they always categorize the incident's impact into different classes, and utilize classification models for the prediction. However, our problem is to provide *numerical* results in both spatial (i.e., affected region) and temporal (i.e., traffic speed decrease) aspects as the predicted impact. for example, the region of 60% travel time delay is 3.2 miles in 20^{th} minute. Therefore, their classification models are not suitable for our problem.

The set of most relevant studies to our study are the models proposed in [15], [12], [5]. In these studies, they considered both spatial and temporal aspects to quantify the impact. However, their quantification strategy are designed to capture the one-time impact of the incident, instead of the time varying nature of impact at different locations. As illustrated in the example of Figure. 1, the impact of traffic is not always an one-time phenomenon, in fact, it follows a growing/shrinking pattern after the occurrence of incident. Towards this end, in this paper, we quantify the impact of an incident as a time varying spatial span. Hence, instead of predicting static parameters for the one-time impact, our approach predict the behavior of the time varying spatial span of an incident.

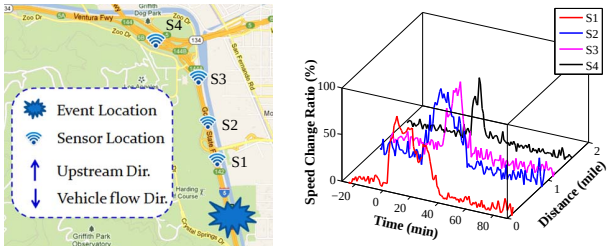
III. Preliminaries

To explain the preliminaries, consider a sample incident that occurred on the freeway I-5 South as illustrated in Figure 2(a). Sensor *S1*-*S4* represents the four affected sensors located on I-5 South upstream of the incident location[1]. In the rest of this paper, we use this scenario as a running example to explain our approach.

Definition 1: (Speed Change Ratio) The speed change ratio (Δv) at a specific location (l) and time (t) is defined as decreased ratio of current traffic speed (v_c) compared with normal traffic speed at l and t, as shown in Equation (1).

$$\Delta v(l,t) = \frac{v_r(l,t) - v_c(l,t)}{v_r(l,t)} \times 100\% \quad (1)$$

Here, the normal speed (v_r) is calculated as the historical average value at location l of same time t in the past. Figure 2(b) shows the corresponding time varying speed change ratios for four sensors depicted in Figure 2(a). Here, the axis labeled as *Time* refers to the elapsed time after the occurrence of the incident, where the negative values refers to the time stamp before the incident occurs. The axis labeled as *Distance* refers to the road network distance between sensor location and incident location.



(a) Scenario map (b) Speed change ratio

Figure 2. Sample traffic incident on I-5 South

In our problem, the key to predicting the time varying spatial span is to predict the speed changes of all sensors over time. One intuitive solution is to apply traditional time series prediction approach on the speed time series. Towards this end, we need to predict the speed changes for each sensor. However, this solution has a few limitations and drawbacks. In the following, we provide a brief explanation of its limitations through two critical observations made from Figure 2(b).

Observation 1: For all sensors, the speed decreases *abruptly* after the occurrence of a traffic incident, suggested by the sudden increase of speed change ratio.

For example, for sensor $S1$, the speed dropped from 67 MPH to 18 MPH within 2 minutes after the occurrence of incident. The time series prediction approaches[15] (e.g., auto-regressive models) cannot effectively predict abrupt variation in time series because most of them relies on the data in the immediate past. Thereby, according to observation 1, traditional time series prediction techniques cannot effectively predict the traffic time series at the beginning of a traffic incident.

Observation 2: The abrupt speed change for each sensor starts at different time stamps after the incident's occurrence.

In our running example, sensor *S3* reports the abrupt speed decrease at 12^{th} minutes, while sensor *S4* reports at 19^{th} minutes after the incident's occurrence. Hence, in this scenario, given the incident just occurred, we need to predict the speed changes in 12 or 19 minutes ahead. This task requires a multi-step prediction strategy for time series prediction approaches. However, according to the study in [4], multi-step time series prediction suffers from error accumulation problem when the prediction period is long. Thus, the time series approach cannot accurately predict the speed changes in a long term, for example, 30 minutes in advance for general cases.

To conclude, we argue that traditional time series prediction technique cannot effectively predict the speed decrease

[1]In this study, we focus on the impact on the upstream direction of incident location for incidents occurred on freeways.

for all sensors impacted by an incident. To address this issue, in the following, we propose a modeling strategy towards incidents' impact and corresponding prediction techniques.

IV. Impact Modeling

First, we define whether a location is impacted according to the magnitude of speed changes as follows:

Definition 2: (Impacted Threshold λ) λ is defined as an impact parameter related to the magnitude of the speed changes. Given a time stamp t and a location l, if the speed change $\Delta v(l,t)$ satisfy the following inequality, we denote the location l as impacted at time t.

$$\Delta v(l,t) \geq \lambda \tag{2}$$

In the experiments, we will study the effects of λ values in the prediction accuracy of propagation behavior.

Consider λ as 60%, and cut the 3D Figure 2(b) horizontally with Δv=60%. We will obtain a series of scatter points in a 2D space of distance and time, as depicted in Figure 3. Each point (x, y) in Figure 3 represents a specific sensor located at y miles from the incident location with 60% speed decrease at x-minute after the incident occurrence time. For the four points on the left side (with $x < 20$), their x-axis value indicates the time stamp when a sensor *starts* to get impacted, which is referred as *propagation phase*. For the other four points, their x-axis value indicates when a sensor *ends* from getting impacted, which is referred as *clearance phase*. As a byproduct, the impact duration of a sensor can be derived as the time difference between the points in the propagation phase and clearance phase. In this study, we focus on predicting the impact in propagation phase.

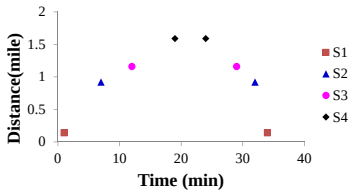


Figure 3. Intersecting Figure 2(b) with λ = 60% in speed change

As shown in Figure 3, we observe that the closer a sensor is to the incident location, the earlier it starts to get impacted. Intuitively, if a sensor s get impacted at a time t, all the sensors closer than s should be impacted before t. Therefore, the impact backlog (i.e., spatial span) of traffic incident is defined as follows:

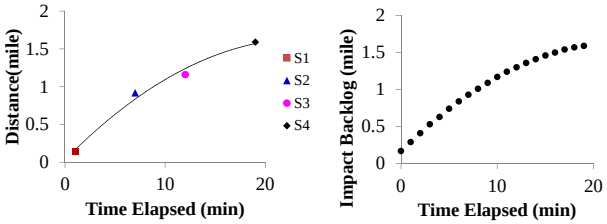
Definition 3: (Impact Backlog) Given an incident location on freeway l and occurrence time t_0, and impact threshold λ, the impact backlog b at time t (b_t), is the road network distance between the occurrence location and the furthest impact location (with $\Delta v(l,t) \geq \lambda$), along the upstream direction (i.e., the opposite direction of the vehicle flow).

In the following, we will use the example in Figure 3 to explain how to calculate b_t, with λ=60%. In this example, sensor *S2* (0.9 miles from the incident) starts to get impacted at 8^{th} minute after the incidents. Therefore, the impact backlog at the 8^{th} minute is 0.9 miles. If we consider the granularity of time stamp(t) in the definition as 1 minute, we could derive b_8=0.9. Similarly, we could derive b_1, b_{12} and b_{19} from the sensor *S1*, *S3* and *S4*, according to the time they get impacted and their distances to the incident location. With the notation of impact backlog, the time varying spatial span of incident impact in terms of propagation behavior is defined as follows:

Definition 4: (Propagation Behavior) Given an incident (e) at location l occurred at time t_0, and λ, e's propagation behavior is defined as a time series of impact backlog after t_0 and before it reaches the maximum impact backlog. Assuming e reaches the maximum impact backlog after t minutes, its propagation behavior is represented as $\vec{b}$ or $\{b_0, b_1, ..., b_t\}$, where the subscript i for b_i represents the time units after t_0. Here, b_i is the distance from the incident location that is "impacted" at time t_i.

To calculate the propagation behavior for an incident, one naive way is to record the speed changes on all the possible upstream locations. However, this method requires a fairly dense placement of sensors. In most sensor networks, the sensors reporting traffic speed are always placed with a certain distance interval (e.g., 0.5 mile). Therefore, due to the limited availability of sensor data, we can only derive impact backlog from the locations equipped with sensors. To create a continuous propagation behavior, we utilize a fitting strategy. The overall modeling strategy is summarized as follows:

1) We utilize the distance of a sensor from the incident location to represent the impact backlog at time t, which is the stamp they start to get "impacted".
2) Consequently, we plot the derived impact backlogs into 2D space (e.g., the scatter points in Figure 4(a)), and train a polynomial function to fit the plotted discrete points.
3) Finally, we utilize the learned fitting function and interpolate the backlogs at missing time stamp and generate a complete propagation behavior. Figure 4(b) shows the propagation behavior for our running example, where the impact backlog $\{b_0, b_1, ..., b_{19}\}$ is plotted at each minute.



(a) Fitting result (b) Interpolation result

Figure 4. Sample propagation behavior

There are alternative modeling approaches, such as the use of coefficients in polynomial fitting function. The superiority

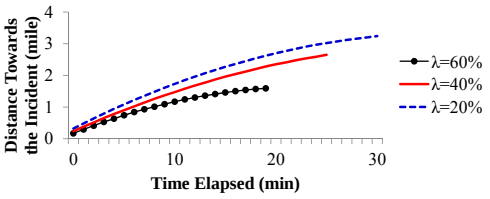


Figure 5. Propagation behaviors under different λ

of our modeling strategy over this approach is as follows: when we construct the propagation behavior, we only use the fitting function to interpolate the missing impact backlogs, for existing impact backlogs we still use the original data. However, if we rely on the coefficient vectors of the fitting function, we may introduce fitting error into the original data, which may result in inaccurate representation of the propagation behavior.

Applications of Propagation Behavior: The prediction of propagation behavior can enable intelligent route planning, effective transportation policy making, and faster traffic emergency responses. In the following, we will briefly detail how to use the propagation behavior within route planning applications. For each traffic incident, we can predict multiple propagation behaviors based on different λ values. For example, Figure 5 illustrates the propagation behavior for the running example under different λs. The value of λ can be tuned according to the preference of the end users of a route planning application.

For a specific time, by utilizing the combination of propagation behaviors, we could derive the set of affected road segments under different magnitude of speed changes. For example, in Figure 5, at 15^{th} minute after the occurrence of incident (i.e., x=15), the location at 1.3 mile towards the incident is at least 60% speed decrease, the location at 1.7 mile is at least 40% speed decrease and the location at 2.0 mile is at least 20% speed decrease. Similarly, for a specific location, by fixing the y values, we could derive the time when starts to get 20%, 40% or 60% speed decrease. Such predictive information are crucial to generate travel time weight for road segments near incident locations, further to be utilized in the fastest path calculation in route planning.

V. Impact Prediction

In this section, we will explain our proposed techniques for predicting the impact of incidents on road networks in terms of propagation behavior. First, we will discuss a baseline approach for grouping similar incidents based on their attributes to estimate the impact. However, in some particular cases, although two incidents have similar attributes, their impacts are still highly different from each other. Therefore, we introduce a new prediction model that addresses the shortcomings of the baseline approach by incorporating traffic density measures such as volume and occupancy. Finally, we explain a multi-step prediction approach that takes into account initial behavior (i.e., sub-pattern of propagation behavior) of an incident to further improve the prediction accuracy.

Our dataset includes three years of historical sensor readings (i.e., speed, volume and occupancy) referred to as *traffic data (D)*. Specifically, the volume reading represents the number of cars passed by a sensor within a sampling interval, and the occupancy reading represents the percentage of time a sensor is occupied. In addition, we also include the dataset of incident reports that includes set of 43 attributes, such as fatality, number of lanes affected etc., referred as *incident data (R)*. Our impact prediction problem is defined as follows:

Problem Definition: Given an incident e ($e \in R$) occurring at time t_0, and the dataset D collected before t_0 (i.e., $[t_0 - T, t_0]$, where T is the duration of the datasets), to predict propagation behavior of e in the next t time stamps, i.e., $\{b_1, b_2, ..., b_t\}$.

A. Baseline Approach

In this section we introduce a baseline approach that classify incidents solely based on their attributes for prediction. In particular, we assort historical incident to different classes based on their attributes. The main intuition here is that the incidents within the same class should be strongly correlated, and hence given an incidents e with certain attributes may follow the similar impact. The detailed steps of the baseline approach is as follows: 1) given historical incidents and all their attributes, apply a feature subset selection algorithm to identify the set of relevant features that are maximally correlated with their propagation behavior; 2) classify all historical incidents into different groups according to their values of selected features. For example, if the incident location (e.g., I-5 South) is one of the selected features, all incidents occurred on freeway I-5 South should be put into one group. within each group, we use the average propagation behavior as the representative for prediction. In this way, when a new incident occurs, we extract its correlated feature values, use them to identify the group it belongs to, and use the representative in that group as predicted propagation behavior.

With our dataset, we observe that the feature subset selection algorithm determines the following attributes: street name (e.g., I-5 South), start time (i.e., occurrence time), affected number of lanes (i.e., number of lanes blocked by the incidents), and incident type (such as traffic collision, etc). Therefore, we will use these attributes to classify the incidents in the 2) step. Note that the length of propagation behavior might be different from each other, thereby, during the calculation of average propagation behavior, its length equals the shortest propagation behavior in one cluster.

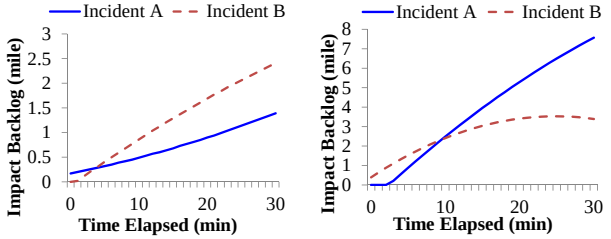
B. Prediction with Traffic Density (PAD)

In the baseline approach, we assumed that incidents with similar attributes may follow similar impact, and hence classified the incidents based on the values of selected

attributes. However, our observations from the real-world datasets show that in some cases, even two incidents have similar attributes, their impact propagation behavior can be significantly different from each other. This is particularly notable when two incidents occurred on the same street but different road segments. For example, consider two incidents (with same attributes) that occur at a rush-hour on two different segments passing through downtown area and rural area (significantly less crowded). Obviously, the impact of these accidents will be different. Therefore, we argue that traffic "density" around the incident is correlated with its propagation behavior, and hence can improve the prediction accuracy. In the rest of this section, we will present two selected case studies to verify our hypothesis and propose an approach that utilizes traffic density.

We quantify the traffic density using two traffic measures: volume (the number of cars passing from a sensor location) and occupancy (the percentage of time the sensor is being occupied) from the sensors that on the same streets close to the incident location. As we discussed these measures are available in our sensor dataset. Below we explain the effect of each measure in turn.

Effect of Volume: To illustrate the correlation between volume and propagation behavior, we present two real-world incidents (e_A and e_B) that occurred on I-405 S with similar incident attributes, but different volume values (i.e., low volume for e_A, high volume for e_B). Their propagation behavior are depicted in Figure 6(a). As shown, for e_B, as the vehicles accumulated quickly (due to large traffic volume), the impact propagates very fast after a few minutes. On the other hand, the propagation speed of e_A (with lower volume) is not as fast as e_B. Hence, it is likely that different volume values can result in different propagation behavior.



(a) Effects of Volume (b) Effect of Occupancy

Figure 6. Case studies on traffic environment

Effect of Occupancy: Similar to volume case study, we will show the impact of occupancy using an example. In this case, we choose two incidents that occurred on I-5 S with different occupancy values. Figure 6(b) shows the propagation behavior for e_A (with higher occupancy value) and e_B (with lower occupancy value). Obviously, the average propagation speed (average curve gradient) for e_A is higher than that of e_B. This means that the incident impact propagates faster on more occupied locations, and hence occupancy is also correlated with propagation behavior.

As illustrated in the above two case studies, the traffic density (measured by volume and occupancy) are very important parameters to predict the propagation behavior of an incident. Therefore, we incorporate traffic density into our prediction model. In particular, for each incident, we create a two-dimensional feature vector composed of volume and occupancy values and cluster incidents based on this vector. Our Prediction approach that combines incident Attributes and traffic Density (PAD) is summarized as follows:

- **Training Phase:** 1) Classify the historical incidents into groups according their correlated attributes trained in the baseline approach; 2) Within each group, cluster all incidents on the feature space composed by the volume and occupancy value, i.e., $< v, o >$;
- **Prediction Phase:** For a newly occurred incident e, 1) we identify its group based on its correlated attributes, and use its volume and occupancy value to find the cluster (C) it belongs to; 2) we select all the archived incidents inside the C and use the average of their propagation behaviors for the impact prediction of e.

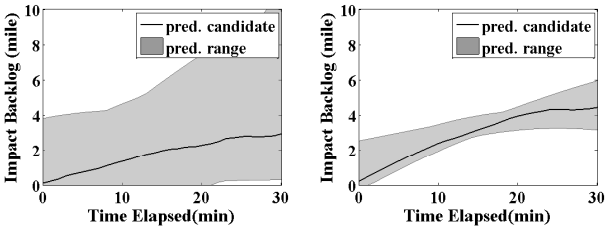
To ensure the cluster quality we maximize the number of clusters (k) while guaranteeing the quality of each cluster, which is measured by average silhouette coefficient (s) defined in [18][2].

C. Prediction with Initial Behavior (PADI)

In the previous section we discussed PAD model that improves the accuracy of the baseline approach by using traffic density information. However, there are still other impact correlated features that PAD does not take into consideration, such as weather conditions or other information that are not available in our dataset. Therefore, in some cases, the accuracy of PAD still can be improved. Figure 7(a) shows one such case for a sample cluster learned by PAD. In this figure, prediction candidate refers to the average propagation behavior with similar attributes and traffic density, and prediction range (i.e., the gray area) is calculated based on the maximum deviation of each instance towards the candidate. If we use this candidate for predicting propagation behavior for incidents with same attribute and density, the prediction error would be non-trivial. To shrink the range for the prediction candidate, we cluster all the propagation behavior within a group of incidents (under same attributes and traffic density), and generate multiple prediction candidates. This eliminates the need to rely on the candidate in terms of average propagation behavior for the prediction. Figure 7(b) shows a sample candidate and its range after the clustering on propagation behavior.

We elaborate the training procedure for this method as a hierarchy structure illustrated in Figure 8. Level I, II and III indicates the successive grouping of incidents based on

[2]Specifically, we choose the maximum number of clusters while constrain s to stay in the range (0.5, 0.7], which indicating the reasonable evidence for clustering result.



(a) PAD approach (b) PADI approach

Figure 7. Sample prediction comparison on I-405 S.

attributes, density and propagation behavior. One may think of merging all three levels into one level containing all three types of information (i.e.,attributes, environment, propagation), and conduct clustering algorithm only once. However, it is difficult to balance the weight for the features of the three types of information during clustering. Therefore, the hierarchical structure helps us to avoid potential problems in weight tuning step.

During the prediction step, for a given incident, we use its attributes and traffic density to search in the first two levels. To identify a suitable cluster in Level III, we relax the prediction problem, and use initial behavior of the accident to match the cluster centroid, which is defined as follows:

Definition 5: (Initial Behavior) Given an incident (e) and its propagation behavior $\vec{b}$, i.e., $\{b_1, ..., b_t\}$, its initial behavior is defined as the first h time stamps in $\vec{b}$ (i.e., $\{b_1, ..., b_h\}$), where h is defined as *forward lag*, and $h < t$.

In particular, with the help of initial behavior, when a new incident e occurs, we match its initial behavior with the first h times stamp (i.e., $\vec{b}_{1...h}$) among the corresponding propagation behavior centroids in the Level III, and identify the closest centroid as the candidate for predicting $\vec{b}_{h+1...t}$. Note that initial behavior can be learned from traffic data. Therefore, by considering the initial behavior as input, we *relax* our prediction problem by knowing the traffic in the first a few minutes after the occurrence of incidents.

To illustrate the use of initial behavior, consider the example in Figure 9. The prediction candidates (i.e., cluster centroids on propagation behavior) for incidents that occurred on freeway I-405 South with similar attributes and traffic density is illustrated as solid lines in Figure 9(a). The black dash line in Figure 9(b) represents the initial behavior in the first 5 minutes for a newly occurred incident. By matching $\{b_0, ..., b_5\}$ between its initial behavior and the five prediction candidates, we select the closest cluster centroid

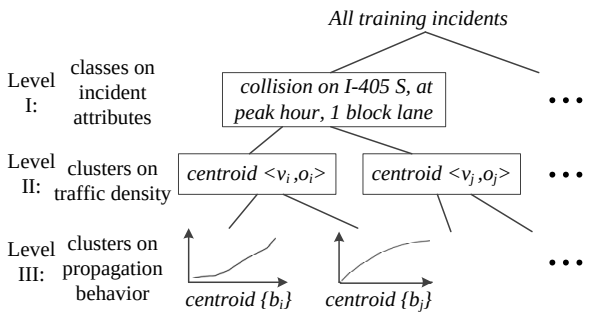
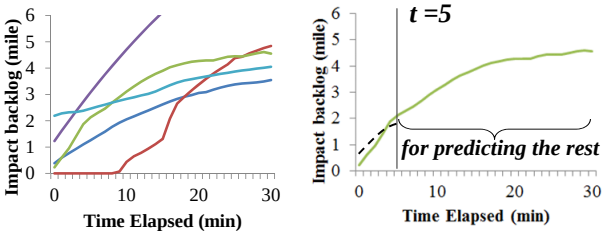


Figure 8. Hierarchy structure for training



(a) Prediction candidate (b) Selected candidate

Figure 9. Sample prediction on I-405 S.

to predict the propagation behavior after b_5, as depicted in Figure 9(b).

It is important to note that, there exists various metrics to evaluate the "closeness" between initial behavior and first h stamps in cluster centroids. In our approach, we use both Euclidean distance and Mahalanobis distance [11] to measure the closeness. The Mahalanobis distance differs from Euclidean distance in that it takes into account the correlations in the dataset and is scale-invariant. To measure the differences between propagation behavior $\vec{b_1}$ and $\vec{b_2}$, the Mahalanobis distance is calculated as follows:

$$d_M(\vec{b_1}, \vec{b_2}) = \sqrt{(\vec{b_1} - \vec{b_2})^T S^{-1} (\vec{b_1} - \vec{b_2})} \quad (3)$$

where S is the covariance matrix between $\vec{b_1}$ and $\vec{b_2}$. We evaluate the prediction accuracy for both Euclidean distance and Mahalanobis distance in Section VI.

D. Discussion

To conclude, in this section we discuss the strategy of using traffic density and initial behavior to predict the propagation. Last but not the least, we want to complete the discussion by providing a solution to transportation system where the measurement of traffic density is either not available or inaccurate. In particular, for transportation systems based on GPS data and crowd sourcing, although they can still have access to incident reports and speed changes, but it is generally challenging for them to have accurate traffic density measurement such as volume and occupancy around the incidents. Therefore, for these systems, we provide a similar Prediction strategy by only using incident Attributes and Initial behavior (PAI). Specifically, instead of the three levels as shown in Figure 8, we only have the first and third level in this approach.

VI. Experiments

A. Experimental Setup

We conducted experiments with real-world datasets under various parameters (see Table II) to evaluate our proposed impact prediction techniques. We measure prediction effectiveness using impact threshold (λ), forward lag (h) (i.e., the length of initial behavior), and distance metric.

1) Data Set: At our research center, we maintain a very large-scale and high resolution (both spatial and temporal) dataset collected from entire LA County highways and arterial streets [17]. We have been continuously collecting

Table I
DATASET DESCRIPTION

Traffic data	data duration	Jun. 1st - Jul. 7th
	# of sensors	4,230
	sensor sampling rate	1 reading/30 secs
	temporal aggr. interval	1 min
	spatial range	OC & LA County
Incident data	# of incident	6,811
	# of attributes	43
	updating rate	1 min
	spatial range	OC & LA County

and archiving the data for the past three years. We use this real-world dataset to create and evaluate our techniques. This dataset includes:

1) *Traffic data*: collected from traffic sensors covering approximately 5000 miles. The sensors report occupancy, volume and speed values.
2) *Incident data*: collected from various agencies including California Highway Patrol (CHP), LA Department of Transportation (LADOT), and California Transportation Agencies (CalTrans).

The statistics about this dataset is given in Table I.

2) Evaluation Method: With our experiment, we first use two case studies to reveal the effectiveness of traffic density and initial behavior in the prediction of impact. Then, we evaluate the overall prediction accuracy under various system parameters, which is listed in Table II. For each set of experiments, we only vary one parameter and fix the remaining to the default values. Without loss of generality, in the experiments, we set t to 30 as the default value to evaluate the results and the granularity of time stamps as one minute. This means we evaluate our approach by forecasting the time series $\{b_1\ ,\ b_2\ ,\ ...,\ b_{30}\}$, where b_i refers to the backlog at i^{th} minute after t_0.

The prediction accuracy is measured by root mean square error between the predicted propagation behavior $\widehat{\vec{b}}$ (i.e., $\{\widehat{b_i}\}$) and actual propagation behavior $\vec{b}$ (i.e., $\{b_i\}$).

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(b_i - \widehat{b_i})^2} \tag{4}$$

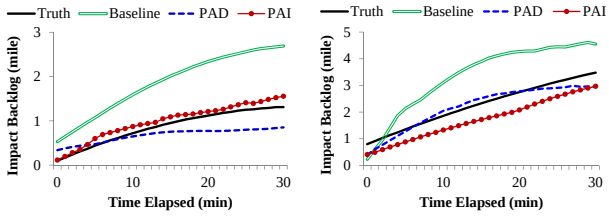
In the experiments, we will compare within the following techniques: prediction with attributes only (Baseline), Prediction with Attributes and traffic Density (PAD) and the Prediction with Attributes, Density and Initial behavior (PADI), and the Prediction based on Attributes and Initial behavior only (PAI) for transportation system without density information.

B. Results

1) Case Studies: In this section, we select two traffic incidents (i.e., collision accident) and compare the prediction accuracy of baseline approach with PAD, and PAI, to examine the effectiveness of traffic density and initial behavior independently. The results are shown in Figure 10 where the solid black line indicates the actual propagation behavior interpolated from the actual sensor readings. Figure 10(a) and 10(b) depict the traffic collision incidents that occurred on I-405 North freeway and on I-5 South freeway, respectively. The results show that in case one, PAI approach yields the best prediction accuracy (i.e., with predicted pattern closest to the actual one). In the second case, PAD yields the best accuracy. The observation indicates that 1) the use of traffic density and initial behavior can improve the prediction accuracy compared with the baseline approach; 2) both of them are *necessary* for the improvement of prediction accuracy, since the results reflect that they are functioning in different ways towards the improvement of prediction accuracy in different cases.

Table II
EVALUATION PARAMETERS

Parameters	Default	Range
Impact threshold (λ)	20	20, 40, 60 (%)
Forward lag (h)	5	0, 2, 5, 10 (min)
Distance metric	Euclidean	Euclidean, Mahalanobis

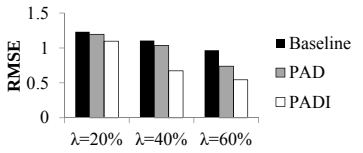


(a) Sample incident on I-405 N. (b) Sample incident I-5 S.

Figure 10. Case studies on two sample incident

2) Effects of Impact Threshold (λ): In this set of experiments, we compare the prediction accuracy under different λs. Figure 11(a) depicts the average of prediction error on 905 incidents in the test data for the three approaches with available traffic sensor dataset. As shown, both PAD and PADI outperforms baseline approach and the percentage of their improvement over baseline is listed in the Table 11(b). In addition, as illustrated in Figure 11(a), as λ increases, the prediction error decreases regardless of which approach is used. To investigate the reason of this phenomenon, we conduct an case study based on an incident occurred on I-405 South during off-peak hours (see Figure 12).

In fact, when we increase λ, the number of impacted sensor decreases as well. Figure 12(a) shows the interpolation result when we create propagation behavior with respect to different λ values. Each scatter point (x,y) represents a sensor located at y starts to get impacted at time x. The dashed lines represent the fitted curves for the corresponding set of scatter points. Table 12(b) shows the average fitting error for each fitted curve. As illustrated, when λ is large, the sensor noise can hardly affect the precision in creating propagation behavior, reflected by the minimum noise in the



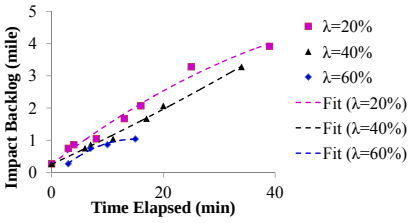
(a) Average prediction error

	PAD	PADI
λ=20%	4.2%	11.3%
λ=40%	7.4%	39.1%
λ=60%	23.9%	45.8%

(b) Improv. over baseline

Figure 11. Effect of impact threshold (λ)

fitting result. As the propagation behavior is quantified in a more accurate way, the prediction accuracy is also higher. When λ is small, the impact is less significant and hence the result can be more easily affected by the noise in the sensor speed readings, which yields lower prediction accuracy. Furthermore, we also observe that the larger the λ values cause shorter propagation behavior. This is because, given an incident, the significant speed decrease normally propagates a shorter distance than that of trivial speed changes. Thereby, it is easier to predict the propagation behavior with less time duration under large λ value. In sum, with larger λ, the propagation behavior is modeled more accurately (i.e., less fitting error), and hence easier to predict.



(a) Fitting graphs

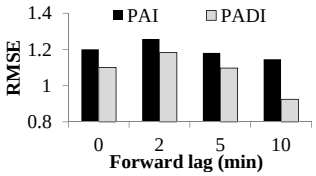
	avg(ε_f)
λ=20%	0.21
λ=40%	0.03
λ=60%	0.01

(b) Fitting error

Figure 12. Case study on impact threshold

3) Effects of Forward Lag (h): In this set of experiments, we study the effect of forward lag (h) length over the prediction accuracy (see Figure 13). We only evaluate the prediction accuracy based on PAI and PADI as there is no initial behavior pattern matching step in the Baseline and PAD approaches. It is important to note that when h=0, PAI, PADI are reduced to Baseline and PAD, respectively. Figure 13(a) depicts the average prediction accuracy of PAI and PADI by varying the forward lag from 0 to 10. Here, the unit of h is minute. Table 13(b) shows the improvement of PADI over PAI regarding different values of h. In general, as h increases, the prediction accuracy of both PAI and PADI increases. This is because the longer time using initial behavior as indicator yields better estimation. However, for some cases, there is an slight increase in prediction error (e.g., when h increases from 0 to 2 minutes). One explanation for this case is that the propagation behavior for the first 2 minutes is noisy, which may due to the difference in immediate reactions of the drivers to the incidents. For example, at the very beginning of the incidents, whether to stay of the road or move to the shoulder to take an exit may greatly affects the incident propagation behavior.

4) Effect of Distance Metric: In this set of experiments, we compare the prediction accuracy by choosing the distance metric by matching the initial behavior in PADI. Figure 14(a) illustrates the prediction accuracy for top six freeways



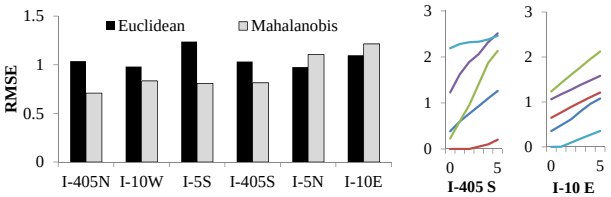
(a) Average prediction error

	Improv.
h=0	4.2%
h=2	6.0%
h=5	7.2%
h=10	19.2%

(b) PADI over PAI

Figure 13. Effect of forward lag (h)

with most incident occurrences using Euclidean distance metric and Mahalanobis distance metric. As shown, the performance of Euclidean and Mahalanobis distance metrics are variant, i.e., changes based on highways. For example, while Mahalanobis distance yields better results on I-405 South and I-405 North, Euclidean distance is better for I-10 East and I-5 North.



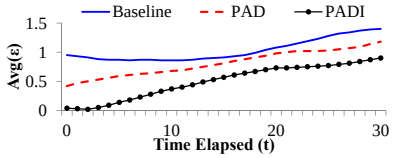
(a) Prediction accuracy on freeways (b) Sub-predictor

Figure 14. Effects of distance metric on PADI approach

To investigate the reason, we plot the first 5 minutes of training results under two selected freeways (see Figure 14(b)). Specifically, we choose the two clusters for I-405 S and I-10 E to represent the cases with better prediction in Mahalanobis and Euclidean distance metric, respectively. As shown in Figure 14(a), the five minutes of cluster centroids in I-405 S present distinct patterns from each other. Thereby the Mahalanobis distance metric is more helpful in selecting the centroids for prediction, due to it measures the correlative distance between two variables. However, first five minutes of cluster centroids in I-10 E follow the similar pattern (i.e., curves with similar gradient), which means they are already highly correlated with each other. In this case, the correlation is no longer a good metric, we need to utilize scale information to distinguish them from each other. Therefore, the Euclidean distance metric introduces lower prediction error in this case. To effectively select the distance metric in our techniques, we evaluate the degree of pattern correlation in the first h minutes of the cluster centroids trained by PADI approach, and set specific thresholds to decide the better metric accordingly.

We conclude the section of experiments using Figure 15, which illustrates the overall performance of the three prediction approaches (i.e., Baseline, PAD and PADI) over time. In this experiments, λ and h is set to 60% and 5 minutes respectively. To compute the prediction result, we directly calculate the differences of actual propagation behavior and predicted ones at each time stamp. That is, for each incident at time t, the ε_t is defined as $|b_t - \widehat{b_t}|$, where b_t refers to the impact backlog for its actual propagation

behavior at time t, and $\widehat{b_t}$ refers to the impact backlog for the predicted behavior at t.



(a) Effects of time

	PADI
T=15	91.7%
T=20	84.2%
T=30	72.4%

(b) Accur. summary

Figure 15. Overall results

As shown in Figure 15(a), the prediction error increases as t increases. For example, for the prediction of impact backlog in 10^{th} minute, the accuracy is higher than the same prediction in 30^{th} minute. In addition, at any time stamp, PADI outperforms both PAD and Baseline. To calculate the percentage of incidents that are accurately predicted before time T, for each incident occurred at t_0, we consider its impact as accurately predicted if the following inequality is satisfied:

$$avg(\varepsilon_{[t_0,\,T]}) \leq \gamma \tag{5}$$

where γ is set to 0.5 mile according to the sensor placement configuration on Los Angeles freeways (i.e., the average sensor placement interval is 0.5 mile). Since our approach is based on the interpolation of traffic between sensors, the average estimation error brought by the availability of sensor data is also 0.5 mile. Under this circumstances, if the average error for an incident i before T is no more than the internal estimation error, we define the impact of the incident i is accurately predicted. Table 15(b) summarizes the percentage of incidents that is accurately predicted under different time interval T, from our best approach PADI. As shown, for predicting the spatial span with 60% travel time delay in 15^{th}, 20^{th} and 30^{th} minutes after the occurrence of incidents, our best solution reaches the prediction accuracy of 91.7%, 84.2% and 72.4%, respectively.

VII. Conclusions

To enable next-generation navigation systems, in this paper, we quantified an incident's spatiotemporal impact as a time varying spatial span and predicted it with certain speed changes for recently occurred incidents. Based on evaluation with real-world traffic and incident datasets, we showed that our proposed prediction algorithm utilizing the traffic density and the initial behavior significantly improves the prediction accuracy of baseline approaches by up to 45%. In future, we plan to extend our impact quantification strategy by considering the traffic behavior in the clearance phase of incidents and also on the surrounding arterial roads.

VIII. Acknowledgements

This research has been funded in part by NSF grant IIS-1115153, a contract with Los Angeles Metropolitan Transportation Authority (LA Metro), the USC Integrated Media Systems Center (IMSC), HP Labs and unrestricted cash gifts from Google, Northrop Grumman, Microsoft and Oracle. Any opinions, ndings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reect the views of any of the sponsors such as the National Science Foundation or LA Metro.

References

[1] Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*, 2011.

[2] Texas transportation institute (tti), annual urban mobility report and appendices. 2012.

[3] S. Boyles, D. Fajardo, and S. T. Waller. Naive bayesian classifier for incident duration prediction. In *TRB'07*.

[4] H. Cheng, P.-N. Tan, J. Gao, and J. Scripps. Multistep-ahead time series prediction. In *PAKDD'06*.

[5] Y. Chung and W. W. Recker. A methodological approach for estimating temporal and spatial extent of delays caused by freeway accidents. In *IEEE Transactions on Intelligent Transportation Systems*.

[6] C. F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28:269–287, 1994.

[7] G. Giuliano. Incident characteristics, frequency, and duration on a high volume urban freeway. *Transportation Research Part A: General*, 23(5):387–396, Sept. 1989.

[8] W. Kim, S. Natarajan, and G.-L. Chang. Empirical analysis and modeling of freeway incident duration. In *11th International IEEE Conference on Intelligent Transportation Systems, 2008. ITSC 2008*, pages 453–457, 2008.

[9] J. Kwon, M. Mauch, and P. P. Varaiya. Components of congestion : delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. In *TRR'06*, pages 84–91.

[10] T. W. Lawson, D. J. Lovell, and C. F. Daganzo. Using the input-output diagram to determine the spatial and temporal extents of a queue upstream of a bottleneck. *Trans. Res. Rec*, 1572:140–147, 1997.

[11] P. C. Mahalanobis. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India 2*, 1:49–55, 1936.

[12] M. Miller and C. Gupta. Mining traffic incidents to forecast impact. In *UrbComp ' 12*.

[13] K. Ozbay and P. Kachroo. *Incident management in intelligent transportation systems*. Artech House, 1999.

[14] R. Pal and K. C. Sinha. Simulation model for evaluating and improving effectiveness of freeway service patrol programs. *Journal of Transportation Engineering*, 128:355–365, 2002.

[15] B. Pan, U. Demiryurek, and C. Shahabi. Utilizing real-world transportation data for accurate traffic prediction. In *ICDM'12*.

[16] F. M. Report. http://www.metro.net/board/Items/2012/03March/20120322RBMItem57.pdf. Last visited Feb 14, 2013.

[17] RIITS. http://www.riits.net. Last visited December, 2011.

[18] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[19] SIGALERT. http://www.sigalert.com. Last visited May, 2013.

[20] K. W. Smith. and B. L. Smith. Forecasting the clearance time of freeway accidents. In *Center Transp. Studies, Univ. Virginia, Charlottesville, VA, Rep. STL-2001-01, 2001*.

[21] Z. Wang and P. M. Murray-Tuite. A cellular automata approach to estimate incident-related travel time on interstate 66 in near real time. *Virginia Transportation Research Council*, 2010.

[22] WAZE. http://www.waze.com. Last visited May, 2013.

[23] S. C. Wirasinghe. Determination of traffic delays from shockwave analysis. *Transportation Research*, pages 343–348, 1978.