Inferring Traffic Incident Start Time with Loop Sensor Data

Mingxuan Yue  
Computer Science  
Department  
University of Southern  
California, Los Angeles, CA  
mingxuay@usc.edu

Liyue Fan  
Integrated Media Systems  
Center  
University of Southern  
California, Los Angeles, CA  
liyuefan@usc.edu

Cyrus Shahabi  
Integrated Media Systems  
Center  
University of Southern  
California, Los Angeles, CA  
shahabi@usc.edu

ABSTRACT
Traffic incidents and their impacts have been largely studied to improve road safety and to reduce incurred life and economic losses. However, the inaccuracy of incident data collected from transportation agencies, especially the start time, poses a great challenge to traffic incident research. We present INFIT, a system that infers the incident start time utilizing traffic data collected by loop sensors. The core of INFIT is IIG, our newly developed inference algorithm. The key idea is that IIG considers the traffic speed at multiple upstream locations, to mitigate the randomness in traffic data and to distinguish among multiple impact factors. INFIT includes an interactive interface with real-world incident datasets. We demonstrate INFIT with three exploratory use cases and show the usefulness of our inference algorithms.

Keywords
Impact Propagation; Traffic Data; Traffic Incidents

1. INTRODUCTION
According to National Safety Council [4], nearly 19,000 people died and 2.2 million were seriously injured in traffic incidents across the U.S. during the first six months of 2015, resulting a total cost of $152 billion. To reduce the life and economic losses, the impact of traffic incidents has been studied in various disciplines including transportation, civil engineering, policy, and operation research (e.g., [7, 5]).

In our previous study [3], we built ADMS, a large-scale database to maintain real-time and historical traffic data including incident data reported by transportation agencies in Southern California. We further studied in [5] the impact propagation of traffic incidents with ADMS. However, we find the agency-reported incident start time is often inaccurate, containing up to 15-minute error based on our estimation. In general, the agency reported incident time is derived from human estimate or the 911 call record, which may be the cause of errors as people must ensure their safety first after an incident happens.

For a reported incident in Figure 1a, we identify 4 upstream loop sensors, i.e., S1-S4, and retrieve the traffic speed measured 30 minutes before and after the reported time. We observe that the upstream traffic speed at S1 is significantly affected 10 minutes before the reported incident time, i.e., speed change ratio (defined in Section 2) exceeds 40%. This effect is present at other upstream locations consistently. Therefore, we estimate that the “actual” incident start time is around 10 minutes before the reported time.

Since the accuracy of incident data is essential to all incident related research, the goal of our study is to infer the start time of agency-reported traffic incidents using traffic data detected by loop sensors. Two major challenges arise when analyzing the traffic data: the first challenge is that the traffic data contains sensor noise and frequent fluctuations, which creates lags and gaps when performing temporal analysis; the second challenge is the co-existence of multiple incidents or external factors, such as weather or construction, the impact of which may overlap in time and space and thus is difficult to distinguish.

In this paper, we present INFIT, a system for inferring the start time of traffic incidents using loop sensor traffic data. INFIT visualizes the agency-reported incidents on a map and illustrates the inaccuracy of the reported incident time with upstream traffic speed data. It also demonstrates the effectiveness our newly developed inference algorithm, which recognizes the propagation of speed change through multiple upstream locations. The inferred time can be used for data cleansing in the ADMS database and to enable various research studies on incident impact.

Figure 1: Example of Traffic Incident Impact Propagation

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored.

For all other uses, contact the owner/author(s).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored.

For all other uses, contact the owner/author(s).
2. INFIT OVERVIEW

The framework of INFIT is presented in Figure 2. We have access to the ADMS database [3] that collects incident reports from transportation agencies and high-resolution traffic data from loop sensors. Our newly developed inference algorithm will generate accurate estimates of incident start time, which will be used to cleanse the noisy incident reports and to enable other traffic studies.

2.1 Datasets

We will utilize two of our in-house datasets collected from LA County highways and arterial streets. Incident Data. The incident dataset [1] is collected in real-time from various reporting agencies including California Highway Patrol, LA Department of Transportation, CalTrans. For each incident, 43 attributes are reported including time, location, type of incident, direction, etc. The type of incident various from traffic collision to road construction, and we only consider collisions in this paper.

Traffic Data. We also have access to historical and real-time traffic data collected from around 4,230 loop sensors in OC and LA County covering nearly 5000 miles [5]. The sensors report speed, volume, and occupancy values every 30 seconds. We only consider traffic speed in this paper. One-minute aggregation is performed to fill missing values.

2.2 Inference Algorithms

Existing Approach - Bayesian: The majority of previous studies [6, 8] addressed the automatic detection of traffic incidents by considering traffic data only, as they may not have access to the agency-reported incident data. In our case, we can utilize both traffic data and incident data, yielding 0% false positive rate. Singliar and Hauskrecht [9] studied the supervised realignment of incident time given the inaccurate reported time with Bayesian probabilistic models. However, they only considered traffic data, i.e., features, at the nearest loop sensor, and the inferred time is often a local optimum due to sensor noise and random fluctuations. Nonetheless, we adapted the Bayesian [9] algorithm to our datasets and implemented them for comparison in Section 3.

IIG Algorithm: We design an unsupervised approach, i.e., Impact Interval Grouping (IIG) algorithm, to infer the start time of a traffic incident by recognizing a series of impact intervals at multiple upstream locations.

For each sensor location, an impact interval is a consecutive time period during which the speed change ratio, compared to historical data, exceeds a predefined threshold \( \lambda \). Specifically, for each loop sensor, we generate its historical average speed over a month as in [5] for each minute of the day. Given a reported incident, we retrieve the traffic speed \( v[t] \) at each upstream location during a 60-minute window centered at the reported time, as shown in Figure 3a. We then compare the real-time speed with the historical average \( \bar{v}[t] \) to obtain the impacted time \( t \), i.e., \( |v[t] - \bar{v}[t]| \geq \lambda \).

The raw impact intervals, i.e., solid blue lines, are shown in Figure 3b, and they are relatively short due to sensor noise and fluctuations in traffic speed. Smoothing can be adopted, such as Moving Average and concatenation, to eliminate the noise and randomness (Figure 3c).

Another challenge is that multiple impact intervals are present at each upstream location within the time window, e.g., two significant impact intervals at the nearest upstream sensor in Figure 3c. By the nearest sensor alone, we may not correctly identify the interval impacted by the specified incident. Therefore, we propose to group the impact intervals throughout multiple upstream sensors that conform to the impact propagation behavior discovered in [5]. For simplicity, we adopt two heuristics, i.e., maximum overlap and nearest center, to identify the impact interval potentially resulted from that of the preceding sensor. Furthermore, the impact of multiple incident or external factors may coexist in the same time window, i.e., several impact interval groups circled by orange lines in Figure 3d. We preserve the nearest group to the report time and its start time is used to estimate the actual incident time (arrow in Figure 3d).

Hybrid Solution: Since the IIG algorithm depends on the predefined impact threshold \( \lambda \), lower \( \lambda \) values allow for more sensitivity to minor incidents but incur more noisy intervals, and vice versa. Therefore we design a hybrid approach that combines IIG with Bayesian [9] to eliminate the inference bias by \( \lambda \). We first apply IIG with a fixed \( \lambda \) value and identify a shorter time window, e.g., 20 minutes centered at the estimated time, and then apply Bayesian in the new window for more accurate inference.

3. DEMONSTRATION PLANS

We built INFIT with Django framework and utilized Google Maps API as well as a back-end SQLite database. INFIT incorporates our newly developed IIG algorithm in Python and the project source code is available for download [2]. We will demonstrate three interactive use cases: (1) exploring incidents and real-time traffic data at upstream lo-
cations; (2) exploring the IIG inference algorithm; (3) comparing the accuracy of the aforementioned algorithms.

Exploring Incident and Traffic Data: The main purpose of this scenario is to allow the audience to interact with the incident data provided by the ADMS database. They will observe the inaccuracy of the agency-reported incident time by looking at traffic data at upstream locations given the selected incident. On the main frame of INFIT as in Figure 4, the audience will find an interactive map showing the locations of reported incidents. The audience can then click on any incident and observe on the right its attributes retrieved from our database. Below the map, we will show the time series of traffic speed detected by multiple upstream loop sensors, for 30 minutes before and after the reported incident time. The audience can observe significant speed decrease at various upstream locations before the agency reported incident time, indicating the report may be inaccurate. On the bottom right, we include a 3D plot showing the spatiotemporal impact propagation of the selected incident.

Exploring IIG Algorithm: Once an incident is selected, the audience can navigate to “IIG: Under-the-Hood” page as in Figure 5 to explore our algorithm. The audience can specify IIG parameters and heuristics, such as λ and grouping heuristics, and observe intermediate as well as final results. The IIG demo will present the major steps to the audience, similar to Figure 3, given the specified incident: (1) the speed series of upstream locations, where each horizontal curve represents the speed series at one upstream location in the 60-minute window; (2) the raw impact intervals calculated using the real-time speed series and historical speed according to the specified the impact threshold λ; (3) the smoothed impact intervals, depending on the specified smoothing techniques, i.e., moving average or concatenation; (4) the interval groups, depending on the specified grouping heuristics, i.e., nearest start, nearest center, or maximum overlap; (5) the start time of the chosen interval group will be highlighted, identifying the estimated incident start time.

We studied the sensitivity of our IIG algorithm for 183 incidents from the dataset used in [5] with manually labeled incident start time. As can be seen in Table 1, the inferred time is less accurate when the impact threshold λ is too high or too low. We also conducted a comparative evaluation with the three algorithms and the results (omitted for brevity) confirmed the superiority of the Hybrid solution.

### Table 1: Average Error of IIG in Minutes

<table>
<thead>
<tr>
<th>λ</th>
<th>Grouping Heuristic</th>
<th>AvgErr (mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>Nearest Center</td>
<td>4.71</td>
</tr>
<tr>
<td>20%</td>
<td>Maximum Overlap</td>
<td>4.71</td>
</tr>
<tr>
<td>40%</td>
<td>Nearest Center</td>
<td>3.92</td>
</tr>
<tr>
<td>40%</td>
<td>Maximum Overlap</td>
<td>4.03</td>
</tr>
<tr>
<td>60%</td>
<td>Nearest Center</td>
<td>6.55</td>
</tr>
<tr>
<td>60%</td>
<td>Maximum Overlap</td>
<td>7.02</td>
</tr>
</tbody>
</table>

### 4. CONCLUSIONS

In this demo, we presented the INFIT system, which infers the incident start time using traffic speed data from loop sensors. We described the IIG algorithm, i.e., the INFIT core, for accurate incident time inference by identifying the traffic impact throughout multiple upstream locations of a given incident. IIG provides robust performance in the presence of noise in the traffic data and accurate inference even when multiple impact factors are present at the same time. Our interactive, graphical interface provided an exploratory experience to the audience with real-world incident datasets. We visualized the IIG internal procedures and demonstrated the usefulness compared to the existing solution.

Future work may include: (1) the integration of INFIT to TransDec [3] for data cleansing; (2) the improvement of IIG by considering the “severity” of impact, in addition to time and space. The severity can be evaluated by generating impact intervals with a range of impact thresholds for each sensor. It may improve the grouping of impact intervals and eliminate the algorithm’s dependency on the threshold λ.

### 5. ACKNOWLEDGMENTS

This research has been funded in part by NSF grant IIS-1115153, METRANS Transportation Center under grant from Caltrans DTRT12-G-UTC57, the USC Integrated Media Systems Center (IMSC), and unrestricted cash gifts from Oracle. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the sponsors.

### 6. REFERENCES

APPENDIX

Figure 4: INFIT Main Frame

Figure 5: IIG Demonstration

Figure 6: Comparison of Inference Algorithms