The Integrated Analysis of Accident Reports and Traffic Flow Data Sets With Early Traffic Disruption Detection and Segmentation

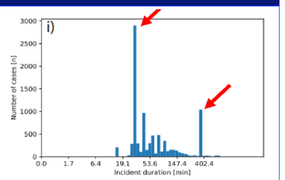
Artur Grigorev (Artur.Grigorev@student.uts.edu.au)
Dr Khaled Saleh, Dr. Adriana-Simona Mihaita, Dist. Prof. Fang Chen

INTRODUCTION

Many accident modelling studies still rely on the reported accident data logs as ground truth, which can be affected by reporting errors. To avoid these errors, the actual accident duration can be estimated from traffic state readings.

The ability to accurately predict how long accidents will last and which impact it will have on the traffic state can provide a significant benefit to both end-users (by providing better routes mitigating accident effects) and traffic operation managers (to determine how many resources need to be allocated in accident management).

The study focuses on the San Francisco area in Unites States (83 detectors stations, 9,275 accident reports, 1,792 accident reports with Vehicle-Detector-Station (VDS) association).



"Rounded" accident durations (30 and 360 minutes) observed in CTADS data set.

AIM

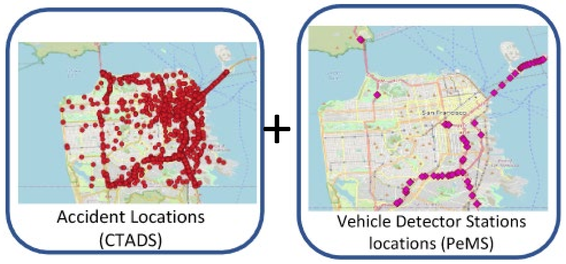
We aim at associating observed disruptions in traffic state (from Caltrans PeMS data set) with accident reports (from Countrywise Traffic Accident data set).

The proposed methodology includes:

1) Detection, segmentation and extraction of observed disruptions in the traffic speed.

2) Modelling of segmented and extracted accident impact together with traffic accident reports to predict the traffic accident duration and impact.

By measuring how each accident affects the traffic state, we can study the accident impact with precision.

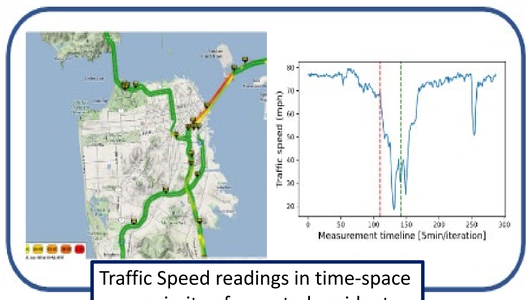


METHODOLOGY

Firstly, we associate the road vertexes with corresponding Vehicle Detector Stations (VDS) within 500m proximity from the Caltrans PeMS data set. Then we associate locations of reported accidents with road vertexes covered by nearby VDS (see Algorithms 1 and 2 proposed in the paper). This way, traffic accident reports get associated with the traffic flow, speed and occupancy readings from the VDS stations.

Secondly, we propose new algorithms relying on Wasserstein and Chebyshev metrics for disruption detection and segmentation. By segmenting the disruptions that occurred in time-space proximity of reported traffic accidents, we obtain the estimated traffic accident duration.

Our methodology provides more information to include in accident models than just the simple accident duration:
1) the accident duration estimated from the impact on the traffic speed, 2) the disruption shape in terms of altered speed profile.



Traffic Speed readings in time-space proximity of reported accident

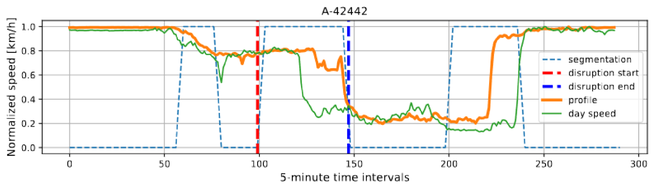
RESULTS & DISCUSSION

a) While the reported durations are distributed between 29 and 360 minutes (see Fig. 1), disruptions observed from traffic speed are much shorter in reality.

b) Estimated accident durations are distributed between 29 and 69 minutes (0.10 and 0.90 quantiles correspondingly) with a mean duration of 58 minutes, while reported accident duration are distributed between 29 and 360 minutes with a mean duration 108 minutes, which highlights the effect of reporting errors on accident statistics.

c) By applying the Wasserstein difference between monthly speed profile and daily traffic speeds with 30-minutes sliding window and extracting the corresponding disruption interval we can observe the shape of disruption impact, which may be used in further analysis of long-term and secondary disruptions produced by accident.

d) We further compare the regression model prediction performance on the CTADS data set by using both our estimated versus the reported accident durations



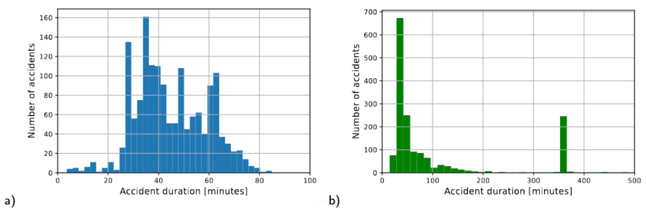


Figure 8: Distribution of accident durations for a) estimated and b) reported accident durations for the area of San Francisco

TABLE 1 Accident duration prediction results

Regression Model	$RMSE_{est}$	$RMSE_{rep}$	$MAPE_{est}$	$MAPE_{rep}$
KNN	**60.1**	85.7	46.0	**32.9**
RF	**57.8**	81.1	40.4	**28.6**
LR	**60.5**	129.6	**40.8**	72.7
SVM	**61.9**	113.8	**36.1**	50.3
GBDT	**59.3**	69.2	41.4	**31.1**

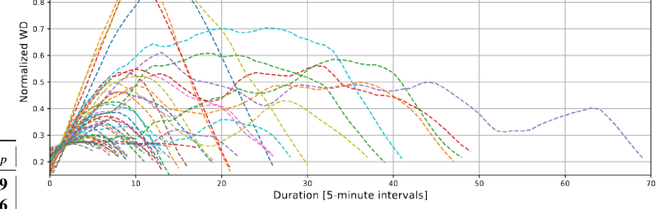


Figure 10: Normalized Wasserstein distance plot for disruption shapes extracted for segmented intervals

CONCLUSIONS

We propose: 1) a fusion methodology of two large data sets (CTADS and PeMS) for a detailed traffic accident Analysis, 2) a novel methodology for the disruption mining using a combination of different metrics (Wesserstein and Chebyshev), 3) we evaluate multiple machine learning models by comparing both the reported and the estimated accident duration predictions, 4) we show that reported accident durations can mismatch with the actual accident duration estimated from traffic speed.

Traffic authorities can integrate the accident duration estimation from traffic state into accident duration prediction models and decision-making systems.

References

[1] S. Fukuda, H. Uchida, H. Fujii, and T. Yamada, "Short-term prediction of traffic flow under incident conditions using graph convolutional recurrent neural network and traffic simulation," IET Intelligent Transport Systems, 2020.
[2] Moosavi, Sobhan & Samavatian, Mohammad & Parthasarathy, Srinivasan & Ramnath, Rajiv. (2019). A Countrywide Traffic Accident Dataset.

Acknowledgments

This work has been done as part of the ARC Linkage Project LP180100114. The authors are highly grateful for the support of Transport for NSW, Australia. This research is funded by iMOVE CRC and supported by the Cooperative Research Centres program, an Australian Government initiative."

http://www.fmlab.org/