Mobility21

A USDOT NATIONAL UNIVERSITY TRANSPORTATION CENTER

Carnegie Mellon University

Penn University of Pennsylvania

The Ohio State University

CCAC

Real-Time Traffic Monitoring and Prediction for Cranberry Township

Final Research Report

Weiran Yao[1] and Zhen (Sean) Qian[1,2]

1 Department of Civil and Environmental Engineering, Carnegie Mellon University
2 Heinz College, Carnegie Mellon University

Contract # 69A3551747111

Smart Mobility Challenge Final Report

Real-Time Traffic Monitoring and Prediction for Cranberry Township

Weiran Yao[1] and Zhen (Sean) Qian*[1,2]

[1]Department of Civil and Environmental Engineering, Carnegie Mellon University
[2]Heinz College, Carnegie Mellon University

February 12, 2019

1 Introduction

Cranberry Township is a progressive municipality that works to maintain traffic efficiency on its transportation networks. Cranberry Townships unique geographical location at the junction of State Route 19, Interstates 79 and Interstates 76 (PA Turnpike) can pose tremendous challenges in coordinated traffic operations. Specifically, the Township operates a Coordinated Signal System that relies on historically generated signal timing plans, coupled with real time technology to manage day-to-day operations on the local network. Unfortunately, any planned or unplanned incidents (such as hazardous weather conditions, accidents, local events, etc.) on the Township's network can cause catastrophic traffic gridlocks that can hardly predicted or prevented by its current operations.

To proactively forecast incident-induced congestion and ultimately alleviate it, this project incorporates real time data inputs from crowdsourced data feeds, traffic sensors and weather reports in the regional proximity of the Cranberry Township to predict traffic delays in real time for 30 minutes in advance. These predictions, together with the current traffic conditions in Cranberry Township, are used to recommend the Township's Traffic Management Center (TMC) effective contingency signal timing plans using a well-tuned rule-based approach. A web-based traffic information system is then built for the Township to visualize the forecast, and alert TMC staff through dynamic online dashboards, email notifications and text messages. The prediction of congestion is made to each road segment in the Township, which can be used to recommend respective contingency signal plans ahead of actual traffic breakdowns. Consequently, this web-based system can alert TMC operators with foreseen traffic issues in the Township, and to allow for timely traffic management on the real-time basis.

2 System architecture

This section presents the three main components of our proposed system: traffic prediction model, signal plan recommendation system and online dashboards.

*Email: seanqian@cmu.edu

(a) Google satellite map of Cranberry Twp.

(b) Intersection at US-19 and Freedom Rd.

(c) Heavy traffic on highways in Cranberry Twp.

Figure 1: Project background: traffic network in Cranberry township (source: [1]).

2.1 Traffic prediction model

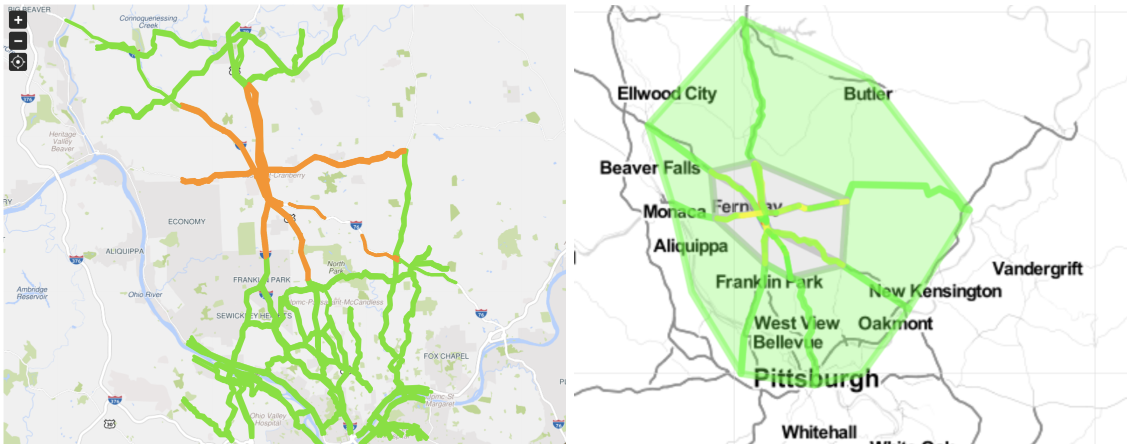
Predicting traffic beyond 5-10min ahead predictions is essentially hard for vanilla time-series methods (e.g., autoregressive models [2]). As these time-series methods rely primarily on correlations between future and past states, they could fail for longer-term prediction if additional factors have more impacts on prediction targets. Traffic on a road segment can change drastically due to traffic incidents, weather hazards or atypical traffic patterns in its proximity. In these cases, past traffic dynamics on target road segments may have little useful information implying their future traffic states if predicted with longer forecasting horizons (e.g. 30 minutes in advance).

A widely used solution found in literature [3, 4, 5] is to take into account spatiotemporal correlations between target road segments and nearby segments. Because it takes time for traffic to propagate from nearby segments to targets, abnormal traffic observed on nearby segments can serve as longer-term predictors for traffic on target roads. Additionally, traffic incidents and weather information are assumed to have influence on traffic conditions. Having considered these factors, we apply the state-of-the-art Long Short-Term Memory (LSTM) encoder-decoder neural networks [6] to predict the future traffic flow sequence up to 30 minutes with a resolution of 5 minutes. A feature section step is first conducted to select critical predictors influencing future traffic on target roads. This model architecture encodes past and real-time lagged road traffic conditions, traffic incidents and weather information using recurrent neural networks and decodes the encoded vector auto-regressively using its own predictions through an LSTM decoder. An overview of the proposed model architecture is shown in Figure 3.

2.1.1 Data processing

The model inputs include probed traffic speed data provided by INRIX traffic [7], traffic incident records provided by PennDOT Road Closure Report System (RCRS [8]) and weather information scraped from Weather Underground [9]. The INRIX traffic data were reported every 5 minutes for road segments georeferenced by Traffic Message Channel (TMC) code. Each data record includes the TMC code, time stamp, observed speed (mph), average speed (mph), reference speed (mph) and two parameters for the confidence of the speed, namely confidence score and confidence value. RCRS incidents are grouped by unique event

ID. Each update of an incident status was reported as a data sample, whose data fields include road name, direction, cause, lane status, last update time, reported time, and latitude and longitude of the incident location. Weather Underground data were reported every one hour. Each report includes time stamps, temperature, dew point, humidity, wind speed, precipitation, visibility, etc.



(a) INRIX segments within 30 min driving distance. (b) Incident blocks within 30 min driving distance.

Figure 2: Traffic data used in this project.

INRIX traffic speed We transform raw traffic speeds into congestion rate to reduce the noise of traffic sensors. Define i to index all road segments in this project. Because the largest forecasting horizon is set as 30 minutes, all major roads within 30-minute driving distance from the center of Cranberry Township under free-flow condition are selected as candidate predictors. As shown in Figure 2a, a total of 694 Traffic Message Channel (TMC) segments are included in this study. The reference (free-flow) speed v_i^{ref} of road segment i is calculated as the 85 percentile of observed speed on that segment for all time periods (Eq. 1), which is a commonly-used way to determine reference speed from probe-based speed data [10, 11, 12]. Let v_i^t be the observed traffic speed on road i at time t. Congestion is described by congestion rates (r_i^t). The congestion rate on a road segment is defined in Eq. 2 as the percentage decrease from the free-flow (reference) traffic speed of the road v_i^{ref} to the observed speed v_{it}. For each time index t, a congestion rate vector $C_t = [r_0^t, r_1^t, ..., r_n^t]$ is generated.

$$v_i^{ref} = P_{0.85}(v_i^t) \tag{1}$$

$$r_i^t = 1 - v_{it}/v_i^{ref} \tag{2}$$

PennDOT RCRS incident report We filter traffic incidents data from PennDOT RCRS data feed using the same method as in the process of speed data, i.e, traffic incidents reported within 30-minute driving distance from Cranberry Township are selected. Unlike traffic speed, a traffic incident may happen anywhere along the road so the closed road segments associated with the incident can be very flexible. This means the feature space can be huge if without special feature encodings. Also, incident data are temporally-sparse if compared with probed traffic speed. Hence, we characterize the traffic incident data with self-defined incident blocks, which are shown in Figure 2b. The incident area is first split by major roads passing Cranberry Township (I-79, I-76, US-19, I-279 and Freedom Rd), and further split by two closed circles indicating the occurrence distance to Cranberry Township. Because most traffic incidents happened on major roads, we split these major roads a by incident blocks and use the incident counts in each road segments to encode the spatial information of these incidents. For incidents happening on other places in these blocks, the closed road is first associated with intersected incident blocks. The angle between the road direction and the center of Cranberry Township is further used to decide whether its abnormal traffic dynamics will "flow into" or "flow out of" Cranberry Township. Hence, each incident block b has two variables: b_{in}^t and b_{out}^t to count the incident occurrences. For each time index t, an incident vector $I_t = [a_0^t, a_1^t, ..., b_{in}^t, b_{out}^t]$ is generated.

Weather information Weather variables included in the project are apparent temperature T_{AT}^t, precipitation intensity P^t and rain R^t, snow S^t, fog F^t probabilities observed at t. Apparent temperature is computed by combinations of Heat Index (T_{HI}^t), which measures "how hot it really feels when relative humidity is factored in with the actual air temperature [13]", and Wind Chill Temperature (T_{WC}^t), which measures "the lowering of body temperature due to the passing-flow of lower-temperature air [13]". T_{HI}^t and T_{WC}^t are calculated using Meteorological Calculator provided on [13]. As T_{HI}^t is used when air temperature T^t is higher than 80 F, and T_{WC}^t is defined when temperatures are below 50 F and wind speeds WS^t are above 3 mph, we apply these two measures as T_{AT}^t if their conditions can be met. Otherwise, we use air temperature directly as in Eq. 3. Precipitation intensity P^t measures the precipitation in the past 1 hour. For each time t, a weather vector $W_t = [T_{AT}^t, P^t, R^t, S^t, F^t]$ can be generated.

$$T_{AT}^t = \begin{cases} T_{HI}^t, \text{if } T \geq 80\text{F} \\ T_{WC}^t, \text{if } T \leq 50\text{F} \wedge WS^t \geq 3 \text{ mph} \\ T^t, \text{otherwise} \end{cases} \tag{3}$$

2.1.2 Model construction

We define a temporal vector $L_t = [C_t, I_t, W_t]$ for each time index t. The model inputs are time-lagged feature matrix $X_t = [L_{t-p}, L_{t-p+1}, ..., L_t]$ where $p = 30$min is maximum time-lags considered in this project. We normalize each variable to have "zero mean and unit variance." The model outputs are congestion rates on target road segments up to the next 30 minutes, i.e., $y_{t+h} = \{r_i^{t+h}\}$ where $i \in C^{tar}$.

We first perform L1-regularized feature selection [14] on the feature set to get rid of unrelated features. For each target road segments in Cranberry Township and each forecasting horizon h (5 min - 30 min), we built an LASSO regressor to predict its congestion rate r_i^{t+h}. These predictors learn weights ω_{ih} such that:

$$\min_{\omega_{ih}} \sum_t (r_i^{t+h} - \omega_{ih}^T X_t)^2 + \alpha_{it} ||\omega_{it}||_1 \tag{4}$$

We applied 4-fold cross-validation to tune α_{it}. Feature columns which are not selected in any regressors are removed from feature matrix. We then fed the selected time-lagged features into the LSTM encoder model in Figure 3. To explicitly add temporal information, we concatenated time-of-day, day-of-week, holiday and month variables with the encoded feature vector as auxiliary inputs to the LSTM decoder.

When training the encoder-decoder network, we use "teacher forcing" strategy. Teacher forcing works by using the actual or expected output y_{t+h} from the training dataset at the current time step as input in the next time step rather than the output $\tilde{y}_{t+h}$ generated by the network. During prediction phase, LSTM decoders use their own outputs from the last time step $\tilde{y}_{t+h}$ as model inputs at the next time step $t + h + 1$.

2.2 Contingency plan recommendation

TMC operators in the Cranberry Township use rules in Table 1 to manually change traffic signals in response to traffic incidents, which are currently verified with real-time traffic sensors such as cameras, loop detectors, etc. The verification of incidents usually takes longer time than it would need before a timely timing plan change to effectively alleviate congestion.

With the help of our 30-min ahead prediction model, congestion growth in Cranberry Township can be predicted well before traffic actually breaks down. This technique enables the Township's managers to efficiently locate possible road closures and program traffic signal timing changes proactively. To support traffic operations, this project further develops a recommendation system for recommending contingency signal plans that following the congestion predictions. When incidents are observed and atypical traffic patterns are predicted by our model, the system alerts TMC managers of possible traffic breakdowns and recommend contingency plans.

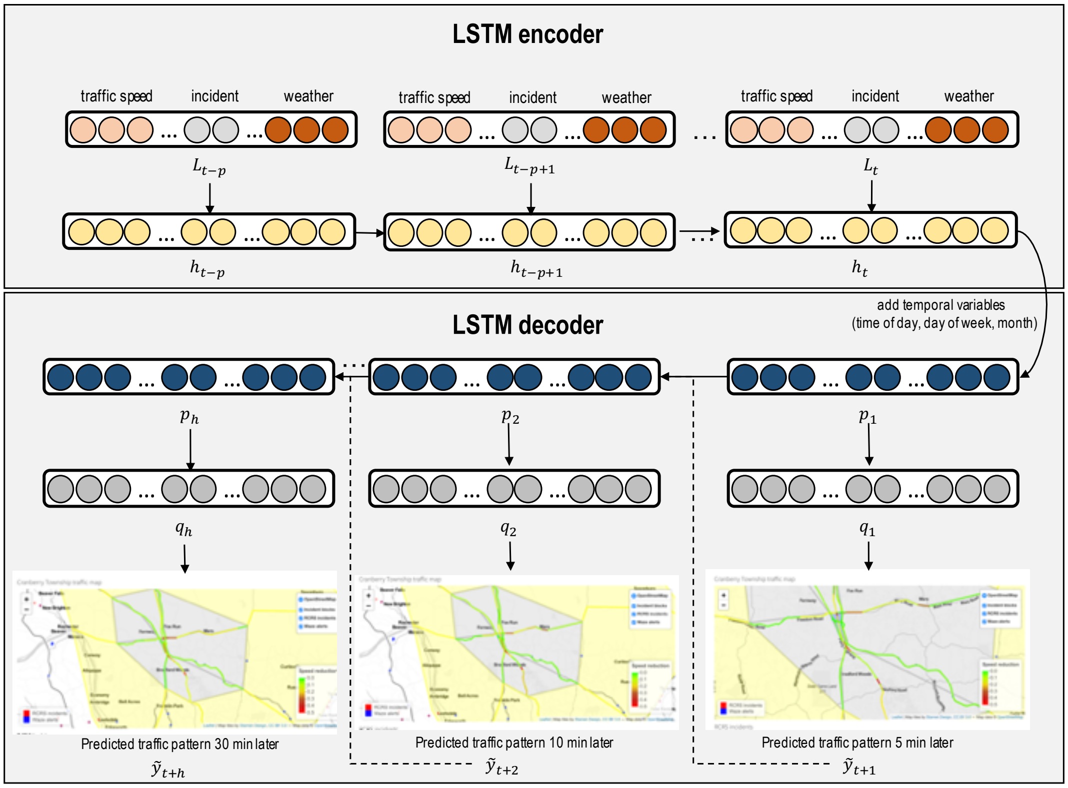


Figure 3: LSTM encoder-decoder neural network used in this project.

2.2.1 Rule-based approach for contingency plan recommendation

It is natural to use a rule-based approach to encode the decision-making process in Table 1. However, the difficulty in this approach is to set the priority of these self-defined rules and to detect "Full Closure" and "Partial Closure" conditions. We design a signal recommendation algorithm (shown in Algorithm 1) to tackle these two issues. Define a rule vector $M_r = (S_r, H_r, C_r, R_r)$ for each row of Table 1, where S_r is the signal plan number on row r, H_r is the effective hours on row r, C_r is the road closure condition that triggers the plan and R_r are the set of associated road segments for this rule. Detailed considerations are as follows.

Rule priority We have two categories of information to decide which signal plans to recommend: (1) current traffic speed, incidents and weather information, and (2) traffic speed predictions up to 30 minutes. We assume for managing traffic signal plans, missing targets, i.e., not changing signal plans in time, is much severer than false alarms. Hence, in our design, any traffic breakdowns implied by (1) or (2) will trigger the system to send alerts to traffic managers. Furthermore, recommendations made by (1) has priority over (2). For example, if this road closure is observed at the present but traffic speeds will be normal 30 minutes later, no doubt the system will recommend action plan 81. If current traffic is normal in the morning but predicted traffic speed shows road closures north of Route 228 in Cranberry Twp 30 minutes later, the system will recommend action plan 81. In addition, the action plan number has a priority order by default, i.e., plans with smaller action number value have priority. For example, if both I-79 Southbound and I-76 Eastbound are partially closed in the morning, the system will recommend action plan 85.

Tuning thresholds for road closure detection Our predicted congestion rates $\tilde{r}_i^t$ is the percentage reduction of observed speed to free-flow traffic speed and can be used to describe congestion levels for all road segments. However, it is still hard to define a threshold for road closures simply from r_i^t. To make it worse, because INRIX traffic speed data are collected by averaging travel speeds of vehicles on roads, full closure of the roads can prevent INRIX vehicles from entering the roads and the collected speed is unreliable.

Table 1: Cranberry township contingency plans [15].

Action Plan Number	Description	
	Full Closure	
81 (AM) 82 (PM)	(A) I-79 Southbound	Closure north of Route 228
83 (AM) 84 (PM)	(B) I-79 Northbound	Closure north of Route 228
	Partial Closure	
85 (AM)	(C) I-79 Southbound	Closure south of Route 228
86 (MID/PM)	(D) I-79 Northbound	Closure south of Route 228
87 (MID/PM)	(E) I-76 (PA Turnpike) Eastbound	Closure east of Cranberry Twp
88 (MID/PM)	(F) I-76 (PA Turnpike) Westbound	Closure west of Cranberry Twp
	Weather Related Plans	
89 (MID/PM)	Incident Weather Timing Plan (Snow)	
	Holiday Shopping Timing Plans	
92-98	Holiday Shopping Timing Plan	

Algorithm 1: Signal plan recommendation

Data: current congestion rates r_i^t, predicted congestion rates $\tilde{r}_i^{t+h}$,
effective incidents I_i^t, W^t, rules vector $M_r = (S_r, H_r, C_r, R_r)$
Result: signal plan S^t
Initialize $S^t = \emptyset$, $ActTime = \emptyset$;
for *row* $r = 1$ **to** *the end of Table 1* **do**
  for *horizon* $h = 0$ **to** *the largest forecasting horizon* **do**
    foreach *segment* i *of the associated segments* R_r **do**
      roadCondition, actTime=`DetectClosure(`r_i^{t+h}, I_i^t`);`
      if *actTime in* H_r **and** *roadCondition meets or severer than* C_r **then**
        return S^t, $ActTime$;
      end
    end
  end
end
return $\emptyset, \emptyset$

To deal with the two issues, we analyzed the archived signal system activity logs and tune the rules for detecting full/partial road closures. Four contingency plan activities were identified between Jan, 2017 and Feb, 2018, with three using plan 85 triggered by partial road closure and one using plan 81 triggered by full road closure. We found that for the three partial closure activities, the congestion rates $r_i^t \geq 0.35$ for all associated road segments before the changing of traffic signals. Thus, we use $r_i^t \geq 0.35$ as the threshold to detect partial road closures. For full closure, we found the INRIX traffic speed data are missing during this activity period for affected segments and an RCRS incident along I79 SB was reported. Hence, we use RCRS incident to detect full closure. To avoid missing targets, if $r_i^t \geq 0.8$ but no incidents are detected, we still identify it as full road closures.

2.3 A web-based traffic information system

A web-based traffic information system is built to visualize predicted traffic information and signal plan recommendations, and to send alerts to the Township's managers if traffic breakdowns are predicted to trigger any of the contingency plans.

2.3.1 Dashboard

The dashboard is the front-end of both predictions and timing plan recommendation. This design has the following highlighted functionalities.

Interactive visualizations As shown in Figure 4, the dashboard visualizes current and predicted traffic speed for each road segment in the Township, location and duration of current traffic incidents provided by RCRS and Waze in terms of both map highlights and a list. The dashboard also shows a recommended signal plan if applicable. The Township's TMC managers can visualize predicted travel time/speed for a road segment by clicking this segment on the map. The visualizations can be shown in terms of interactive plots in the left lower corner and traffic animation in the map.

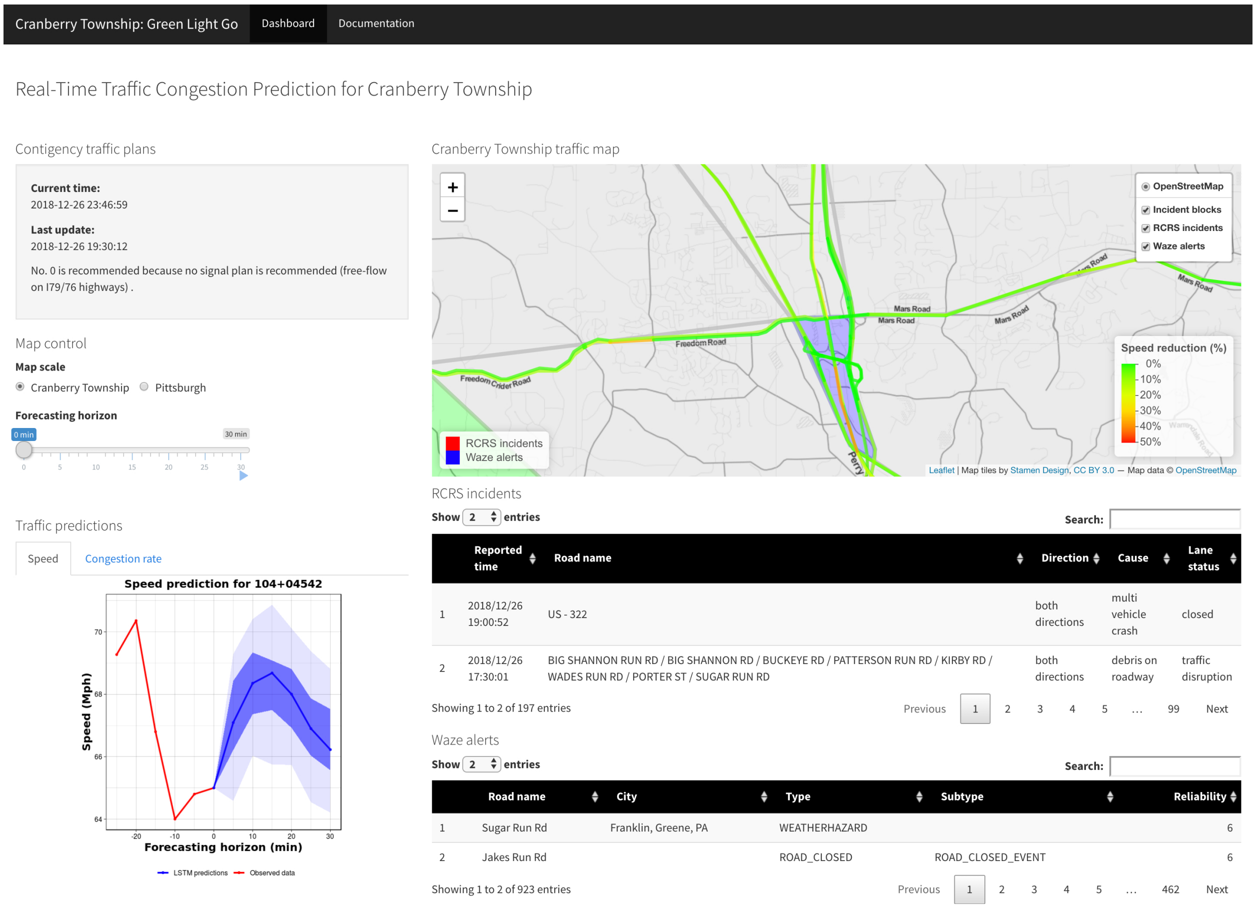


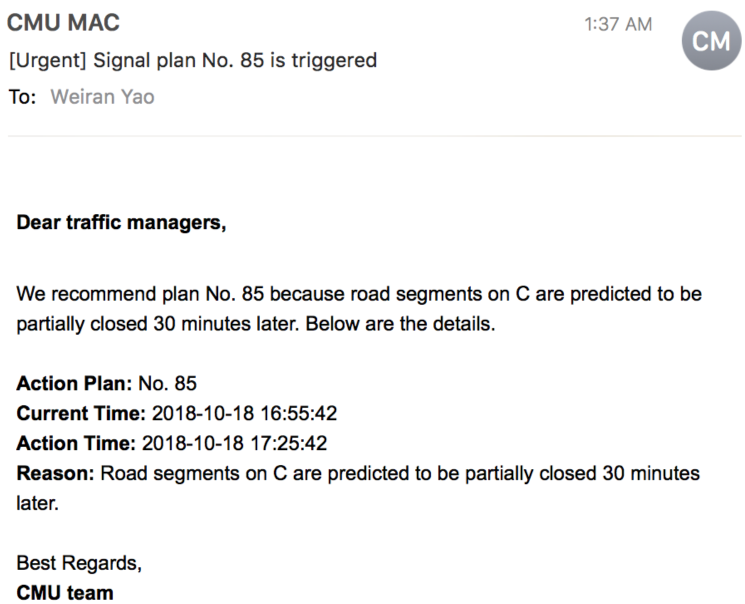
Figure 4: Overview of the dynamic dashboard design.

- **Traffic speed:** Both current and predicted traffic speeds are visualized on the map. Users can use "Forecasting horizon" slider under the "Map control" to select the forecasting time horizon (0 min - 30 min) and visualize the predicted traffic on the map. Note that the default forecasting time horizon is set as 0 to show the real-time traffic. An animation of the predicted traffic evolution in the next 30 minutes can be generated by clicking the "Play" button positioned under the slider. Users can further examine the traffic predictions of a certain road segment by clicking the segment on the map. Interactive prediction plots are shown in the "Traffic predictions" panel.
- **Traffic incidents:** RCRS incidents and Waze alerts effective at the present are visualized on the map and listed in the tables. Users can hover over the red dots (RCRS) and blue dots (Waze) to see the details (e.g. Reported time, Road name, Cause, Lane status, etc.) of these incidents. We also provide tabular views of traffic incidents to ease user searching. Users can do keyword search and sorting, change the number of rows, and go to selected pages in this dynamic table to look for incidents of interest.
- **Message board:** The dynamic message board shows the last update time of current traffic prediction, and the signal recommendation (if applicable) with reasons for making the recommendation.

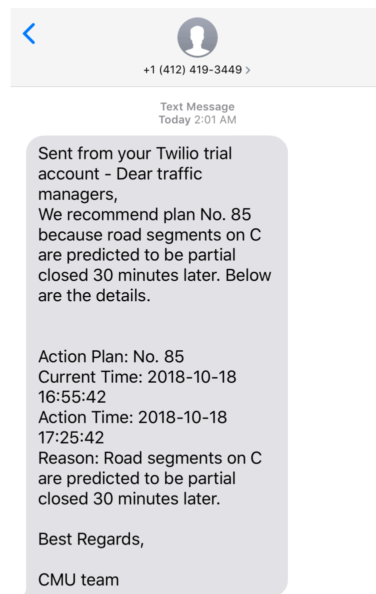
Real-time updates This dashboard is updated every 5 minutes 24/7 everyday to provide real-time decision support for the Township's managers.

2.3.2 Alerts

The traffic information system sends alerts to the Township's managers by emails and by text messages if any traffic information, incidents or weather trigger the recommendation of a contingency timing plan. An email notification from the system email account (cmumactrack@gmail.com) will be sent to managers in the Cranberry Township. A sample email alert is shown in Figure 5a. The system also sends alerts by text messages from +1 (412) 419-3449 to the Township's managers. A sample text alert is shown in Figure 5b. Note that alerts will only be sent if a recommended change in signal timing plan is triggered.



(a) Email notification sample.



(b) Text notification sample.

Figure 5: Real-time alerts sent by system.

The app employs responsive web designs and can be accessed by web browsers of desktop and laptop computers, smartphones and tablets. As shown in Figure 6, the easy access from multiple user-end platforms, such as smartphones, enables real-time examinations and monitoring of traffic conditions after notifications and signal recommendations are received.

3 Results and discussion

To assess our system performance, we first evaluate our model predictions on a separate test data set. A real-world demonstration is then made to show how the system could manage to recommend reasonable contingency plans, and how much time in advance the system is able to notify traffic managers, using an example of a traffic accident occurred at the I79 Southbound on Oct 18, 2018 during evening rush hours.

3.1 Model performance

When testing our model performance, we first split the whole data set into training samples (80%) and test samples (20%). We train our model using those training data and evaluate the model on the test data. Two benchmarks are used in this study: historical average predictor and the LASSO predictor used for feature selection in Section 2.1.2. For the historical average predictor, we use a moving average of the past one month

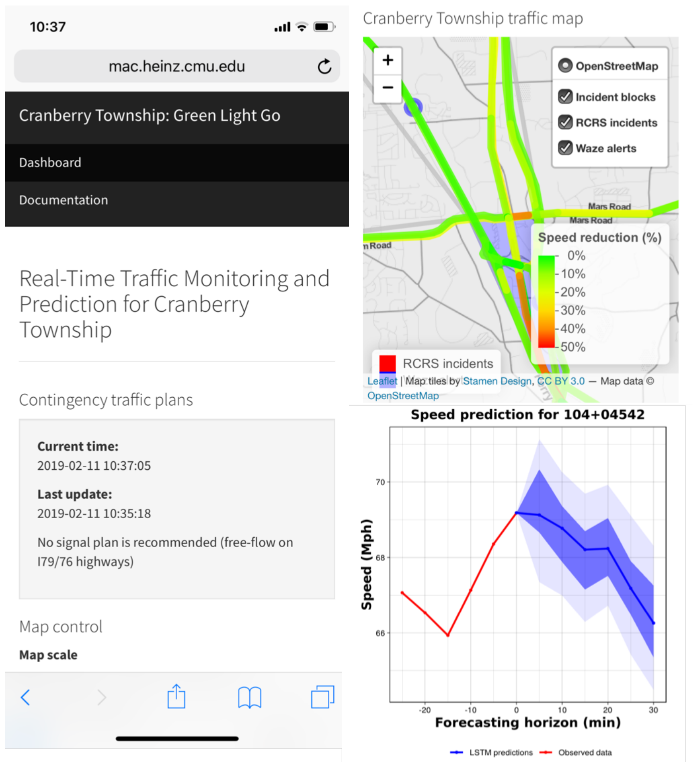


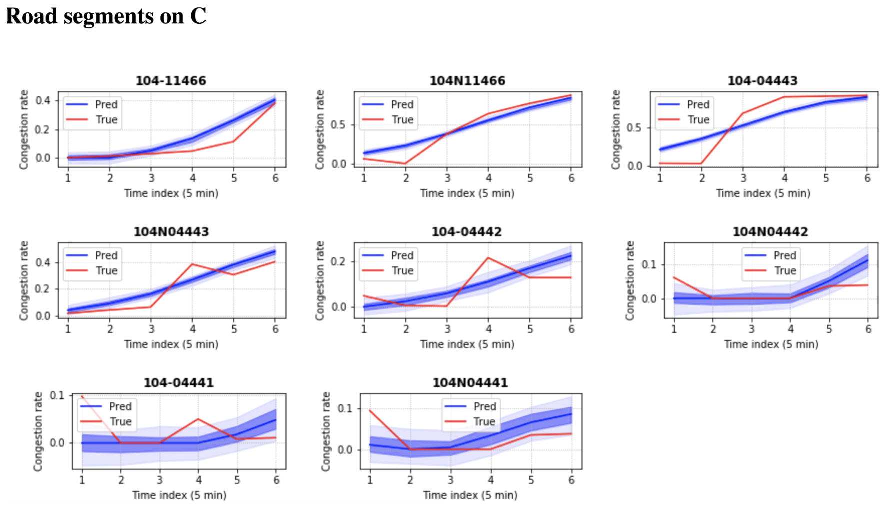
Figure 6: The web application accessed from smartphone platforms.

of traffic speed data to infer the time-of-day traffic speed. Note that the day-of-week variances are considered by filtering data of the same day-of-week in the moving window. This method does not include real-time traffic speed and can not predict non-recurrent/unscheduled traffic patterns. LASSO predictors in Section 2.1.2 are tuned by a 4-fold cross-validation on the training data set. This predictor consider the impacts of atypical traffic patterns in adjacent roads, incidents and weather conditions, but it cannot capture non-linear relationships of those factors in regards to the travel time. We use the Root Mean Square Error (RMSE) to measure model prediction errors on the test data set. RMSEs are computed by Eq. 5. The comparative results are shown in Table 2. The results show that our encoder-decoder neural network model substantially outperforms the other two benchmarks for predicting traffic patterns. Therefore, we train the prediction model with identical hyper-parameters settings (learning rate, number of neurons, training iterations, etc.) based on the whole dataset for the deployment of this project (namely deployed as part of the web-based online information system).

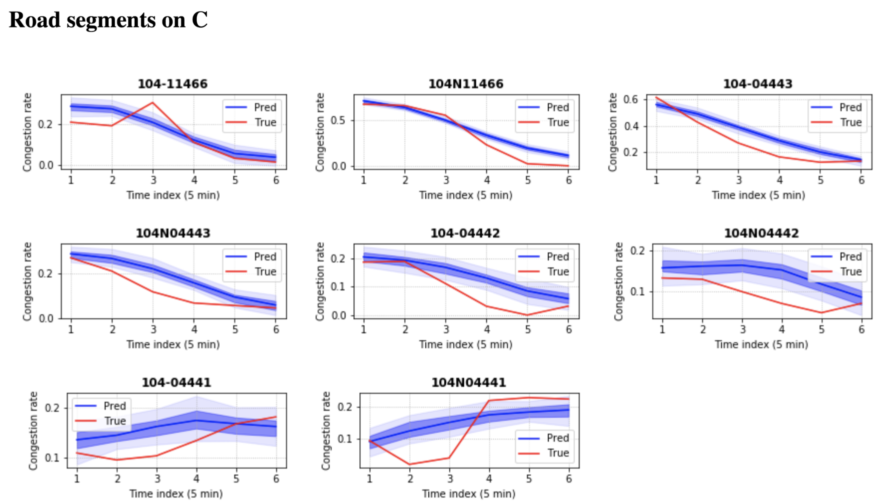
$$RMSE = \sqrt{\frac{\Sigma_{t=1}^{T}\Sigma_{i=1}^{N}\Sigma_{h=1}^{H}\left(\tilde{r}_i^{t+h} - r_i^{t+h}\right)^2}{TNH}} \tag{5}$$

Table 2: Model comparisons.

Method	RMSE (congestion rate)
LSTM encoder-decoder network	0.0032
LASSO predictor	0.0103
Historical average predictor	0.0153



(a) Model predictions at 4:55 PM for Road segment C as specificed in Table 1 (Alert: start contingency plan No. 85 half an hour later).



(b) Model predictions at 5:25 PM for Road segment C as specificed in Table 1 (Alert: stop contingency plan No. 85 half an hour later).

Figure 7: Traffic prediction for an incident at the I79 SB on Oct 18, 2018

3.2 System demonstration

3.2.1 Case I: Crash on I79 SB on Oct 18, 2018

We now use a real-world case study to demonstrate the model effectiveness. This case study involves an incident occurred at the I79 Southbound on Oct 18, 2018 during evening rush hours It occurred at interconnect between the I-76 Turnpike and SB 79, and largely impacted the traffic conditions in the Cranberry Township. The traffic incident was reported to the Township's managers between 5:45 PM and 6:00 PM. This was the time a contingency timing plan would be considered. However, using the model developed in this project, we would be feeding this model with real-time incidents and travel speed data. Then, we evaluate how much time ahead and how well our system is able to predict the traffic breakdowns in Cranberry Township as if it were running in real time.

The prediction, if this model were running, made at 4:55 PM and 5:25 PM are shown in Figure 7a and Figure 7b. At 4:55 PM, as shown in Figure 7a, our system would have predicted the traffic congestion growth 30 minutes later at the road segment "104N11466" and "104-04443", which are the segments of Road C as indicated in Table 1. This condition would have triggered our recommendation system and the system would have sent alerts to traffic managers for switching to a contingency plan at that time. Action plan No. 85 would be recommended. At 5:25 PM, our system would have predicted the congestion alleviation 30 minutes later and then recommended managers to stop the contingency plan No. 85 at 5:55 PM.

Compared to the actual traffic operation, namely receiving complaints at 5:45 PM and taking action afterwards, our system would have managed to detect the incident-induced congestion by 4:55 PM, 50min in earlier. Furthermore, this traffic breakdown could be predicted even earlier than 4:55 PM if the incidents report from Wave or RCRS were timely. The 30 minutes forecasting horizon would provide the Township managers sufficient time to verify the traffic conditions and examine the recommended signal timing plans.

3.2.2 Case II: Crash on I79 SB on Jan 30, 2018

The second case study was on an incident on I79 (1/30/19 at 7:50 am). Traffic backed up on I79 from Wexford to Route 228 in Cranberry Township. We also found an RCRS incident (a multi-vehicle crash) on the segments at 6:39 AM and a series of Waze alerts (crash, traffic jams, weather hazards, etc.) reported from 6:15 AM to 10 AM. We infer from Figure 8 that, if our program were up running, our method can indeed recommend the plan No. 85 to traffic managers no later than around 6:00 AM, when drastic declining trends in traffic speed are observed for all associated segments of Road C.

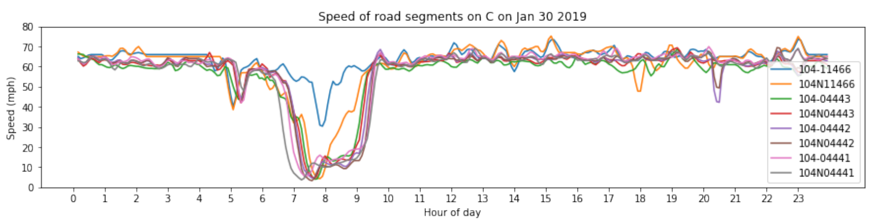


Figure 8: INRIX Speed data on Jan 30, 2019

4 Conclusion

In this research project, we use multi-source data to create a real-time travel time prediction model that can reliably predict travel time/speed by road segment. This system uses past and current spatiotemporal INRIX traffic speed, RCRS incidents and weather information to predict congestion on target road segments in the Cranberry Township up to 30 minutes in advance with a time-of-day resolution of 5 minutes. A rule-based signal plan recommendation system is designed to encode the rules provided by the Township's staff and to use traffic predictions to proactively recommend the traffic signal plan ahead of actual traffic breakdowns. To visualize our system outputs, an web-based online dashboard is developed to support interactive traffic monitoring, and the decision making of traffic management on the real-time basis. If a recommendation of a contingency signal plan is made or triggered, our system directs the signal plan recommendation to several ways, which include online dashboards, email notifications and text messages to the Township's managers. Results show that our prediction model is accurate and reliable, outperforming historical average baseline and LASSO predictors. In the real-world demonstration, our system proves to successfully alert the Township's traffic operators of the upcoming traffic gridlock 50 minutes in advance comparing to the actually reporting time, which, if implemented at the time, would allow prompt and effective traffic management.

References

[1] Cranberry Township Traffic Operation Center. Successfully managing cranberry township traffic using real-time traffic technologies, 2018.

[2] Jingrui He, Wei Shen, Phani Divakaruni, Laura Wynter, and Rick Lawrence. Improving traffic prediction with tweet semantics. In *IJCAI*, pages 1387–1393, 2013.

[3] Alireza Ermagun and David Levinson. Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews*, pages 1–29, 2018.

[4] Shuguan Yang and Zhen Sean Qian. Understanding and predicting roadway travel time with spatio-temporal features of network traffic flow, weather conditions and incidents. Technical report, 2018.

[5] Wanli Min and Laura Wynter. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19(4):606–616, 2011.

[6] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[7] INRIX. INRIX Traffic, 2018.

[8] Pennsylvania Department of Transportation. Road Closure Reporting System, 2018.

[9] Weather Underground. API Weather Underground, 2018.

[10] Kartikeya Jha. *Determining Reference Speed from Probe-based Travel Speed Data*. PhD thesis, 2017.

[11] Shawn Turner, D Schrank, and Lauren Geng. Twin cities metropolitan and mn statewide irc arterial travel time analysis: Use of private sector traffic speed data. *Minnesota Department of Transportation*, page 22, 2011.

[12] Andrew Meese and Wenjing Pu. Applying emerging private-sector probe-based speed data in the national capital region's planning processes. *Transportation Research Record: Journal of the Transportation Research Board*, (2243):17–26, 2011.

[13] National Weather Service. Meteorological calculator, 2018.

[14] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[15] Cranberry Township Traffic Operation Center. Cranberry township contigency traffic plans. Technical report, 2018.