Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 47 (2020) 203–210

Transportation Research Procedia

www.elsevier.com/locate/procedia

22nd EURO Working Group on Transportation Meeting, EWGT 2019, 18-20 September 2019, Barcelona, Spain

Learning a Precipitation Indicator from Traffic Speed Variation Patterns

Yu Feng*, Claus Brenner, Monika Sester

Institute of Cartography and Geoinformatics , Leibniz University Hannover, Appelstr. 9A, 30167 Hannover, Germany

Abstract

It is common sense that traffic participants tend to drive slower under rain or snow conditions, which has been confirmed by many studies in the field of transportation research. When analyzing the relation between precipitation events and traffic speed observations, it was shown that by using extra weather information, road speed prediction models can be improved. Conversely, traffic speed variation patterns of multiple roads may also provide an indirect indication of weather conditions. In this paper, we attempt to learn such a model, which can detect the appearance of precipitation events, using only road speed observations, for the case of New York City. With a seasonal trend decomposition model *Prophet*, residuals between the observations and the model were used as features to represent the level of anomaly as compared to the normal traffic situation. Based on the timestamps of weather records on sunny days versus rainy or snowy days, features were extracted from traffic data and assigned to the corresponding labels. A binary classifier was then trained on six-month training data and achieved an accuracy of 91.74% when tested on the remaining two-month test data. We show that there is a significant correlation between the precipitation events and speed variation patterns of multiple roads, which can be used to train a binary indicator. This indicator can detect those precipitation events, which have a significant influence on the city traffic. The method has also a great potential to improve the emergency response of cities where massive real-time traffic speed observations are available.

Keywords: Machine Learning; Traffic Speed Variation; Precipitation Events Detection; Gradient Boosting

1. Introduction

Many events can influence the speed of traffic. Local events, such as concerts, football matches, or traffic accidents, normally have a limited influence range around the location of the event. However, global events, especially inclement weather conditions, such as rain, snow, mist and haze can lead to a significant reduction of traffic speed for much larger areas. Such traffic variation patterns may repeat for similar events with similar severity, which makes the presence of such events predictable.

* Corresponding author. Tel.: +49-511-762-19437; fax: +49-511-762-2780.
E-mail address: yu.feng@ikg.uni-hannover.de

10.1016/j.trpro.2020.03.090

Many previous researches mainly focus on finding a general model, e.g. using a non-linear regression model (Lam et al., 2013), to represent the influence of precipitation events on traffic flow and density. Statistical analysis of the effect of weather conditions on vehicle speed has also been conducted (Jägerbrand and Sjöbergh, 2016). The impact of weather conditions on macroscopic urban travel times was investigated regarding different intensities of rain, snow and temperature levels (Tsapakis et al., 2013). The precipitation information was also used to improve the prediction of macro traffic state with LSTM (long short-term memory) networks (Jia et al., 2017). A recent approach applied correlation analysis, principle component analysis and LASSO (Yang and Qian, 2019) to predict travel time with additional weather information. In most of the cases above, the goal of the research is to extract the correlation between traffic speed and the amount of precipitation. Some others aim to predict the road speed or travel time with additional precipitation data. Conversely, there is only little research about extracting precipitation information from moving vehicle data. Our previous research has employed motorcars as moving rain gauges. Using windscreen wipers' activities as sensors, the precipitation amount can be estimated and the spatial resolution of precipitation data was enhanced (Haberlandt and Sester, 2010). In recent years, some data driven approaches have also been proposed to learn the rainfall or weather condition from observed data. Prasad et al. (2013) learned a tree-based rainfall indicator from weather records, such as humidity, pressure, temperature, etc. Sathiaraj et al. (2018) learned a random forest classifier regarding normal and abnormal weather condition from the hour of day, temperature, precipitation, visibility and wind speed. Here, the situation is considered to be abnormal if the normal traffic volume is exceeded by one standard deviation.

Nowadays, real time traffic speed observations are available in many cities, with a high spatial and temporal resolution. Thus, there are large amounts of road speed data available. In this work, a proof-of-concept research was conducted to learn a precipitation indicator based on the traffic speed variation patterns of multiple road segments. To the best of our knowledge, this is the first attempt to learn the precipitation information from the road speed observations directly. We present first results on the training of a binary precipitation indicator, which can detect precipitation events directly from traffic speed observations. However, the purpose of this work is not to replace the weather stations, but rather to prove that such an indicator could be learned from the road speed variation patterns. We are also not expecting this could achieve a performance similar to a set of distributed weather stations. Rather, we see this as a potential side product, which could be extracted from huge amounts of currently available traffic data.

Furthermore, the traffic speed data is not only collected by the transportation departments based on sensor measurements, but is also derived from crowd-sourced GPS trajectories collected by navigation service providers, such as Google and Tomtom, with much less expense. A recent research analyzed road speed data of nine stormy days in Shenzhen, China (Li et al., 2017), where the average speed of roads was estimated based on map-matched taxi trajectories. For some less developed countries with fewer weather stations, such models could be helpful to detect potential disasters and global weather events, in real time.

This paper is organized as follows. In Section 2, the data and method used for this research are introduced. Section 3 presents the results and evaluations of the proposed method. In the last section, we conclude and give an outlook on future work.

2. Methodology

In this section, the data used for this research is introduced and our proposed method for training a precipitation indicator is explained.

2.1. Data

The traffic speed observations used in this paper are available at New York OpenData[1], which is provided by the Traffic Management Center (TMC) of New York City Department of Transportation (NYCDOT). It covers mostly the major arterials and highways within the city limits, where NYCDOT has installed traffic speed detectors. We used a subset of this traffic speed data, from Aug. 8th, 2017 to Apr. 25th, 2018. In this dataset, 135 road segments

[1] `https://data.cityofnewyork.us/Transportation/Real-Time-Traffic-Speed-Data/qkm5-nuaq`.

were observed (shown in Fig. 1 left). Each road is associated with one road id. Two of them were not used in our analysis because they did not contain enough observations for the time series analysis. Therefore, the data from 133 road segments are used in this study. The observations are generally in 15-minute intervals and significant gaps can be observed for most of the road segments. For the same time range, precipitation intensity data and textual weather descriptions from Central Park Station in New York (shown in Fig. 1 right) was retrieved via the WeatherUnderground API[2]. In this study, we considered only the weather information from this single station. This dataset has a lower and unevenly distributed sampling rate, ranging from 10 minutes (minority) to one hour (majority).

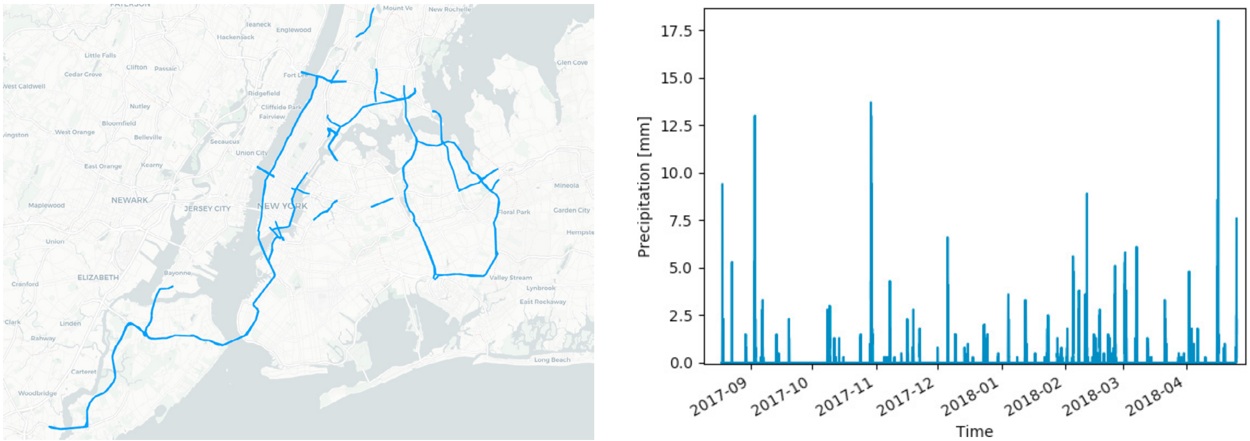

Fig. 1. Spatial distribution of road segments in New York City (left, basemap: OpenStreetMap) and precipitation data retrieved from WeatherUnderground (right).

2.2. Method

Since the traffic speed observations contain strong seasonal effects, we modeled the traffic speed data with *Prophet* – an open source software from Facebook (Taylor and Letham, 2018). *Prophet* is a time series analysis tool based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality. It is regarded as a tool that is robust against missing values and outliers, which should make it well suited for our dataset containing gaps and noise. Observations of each road segment were modeled separately, an example is shown in Fig. 2. The time series were decomposed into trend, seasonal and residual signals. The extracted residuals represent the difference between the actual observations and the periodic model, which can indicate the anomaly level, as compared to the normal traffic state. Since the given data is within one year, we considered both the weekly and daily seasonality. Similar to the example in Fig. 2, we observed in most cases that the model fits nicely to the data.

The precipitation records have non-uniform sampling intervals (mostly 1 hour), however, the traffic speed data has a higher sampling rate in general. Therefore, pre-processing for both datasets was necessary. We searched for all of the speed residual records 15 minutes before and after each weather record timestamp. The residual values of all road segments are averaged to represent the anomaly state compared to the normal state. If no record can be found within this range, NaN will be filled in. In this mean, for each timestamp from the weather records, there are 133 values corresponding to the 133 road segments, representing the anomaly state of the whole area. This is exactly the input feature vector for training the models. Using classical machine learning methods, binary classifiers were trained to generate predictions of precipitation events based on these features.

[2] https://www.wunderground.com/weather/api/

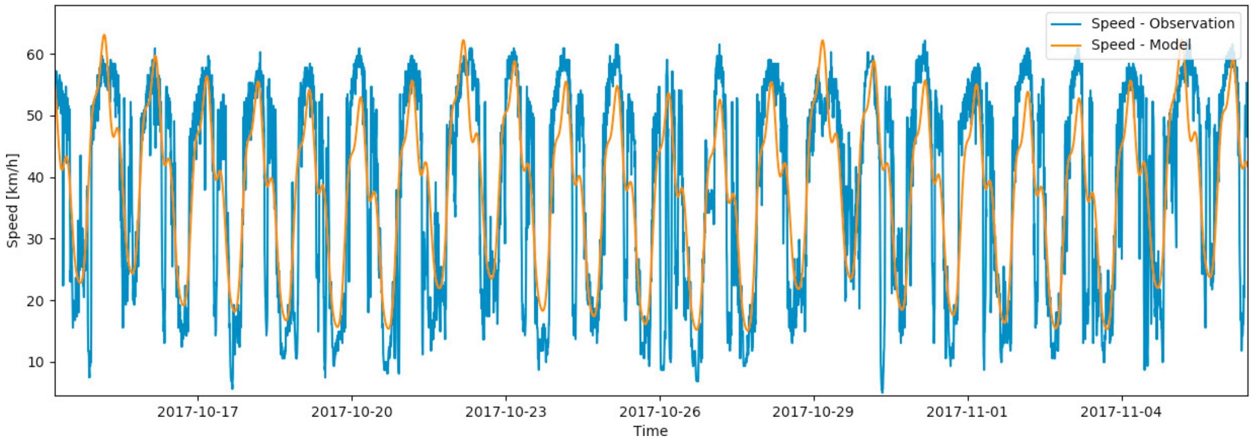

Fig. 2. Example results achieved by *Prophet*, speed observations versus predictions, for one road segment.

3. Experiment and Results

We separated the training and test set at the date Feb. 28th, 2018. The observations before that date were used for training, the remaining part for testing. The training set covers almost 6 months. Very light precipitation events normally do not have significant impact on the traffic speed. Therefore, we selected only the weather records above 0.5 mm precipitation as positive examples. From the training set, 408 weather records were extracted and the same number of negative examples were randomly selected from the records with sunny or cloudy weather conditions based on the textual weather descriptions. In this way, we built a balanced dataset for training our models. Features were extracted based on the timestamp of each weather record. We compared several classical machine-learning methods, such as Nearest Neighbors, SVM (Support Vector Machine) with RBF (Radial Basis Function) kernel, Random Forest, AdaBoost, Xgboost, and others for this binary classification task. The implementations with default settings from scikit-learn (Pedregosa et al., 2011) were used to train the binary classifiers. The achieved overall accuracy and F1-score on the positive class are shown in Fig. 3.

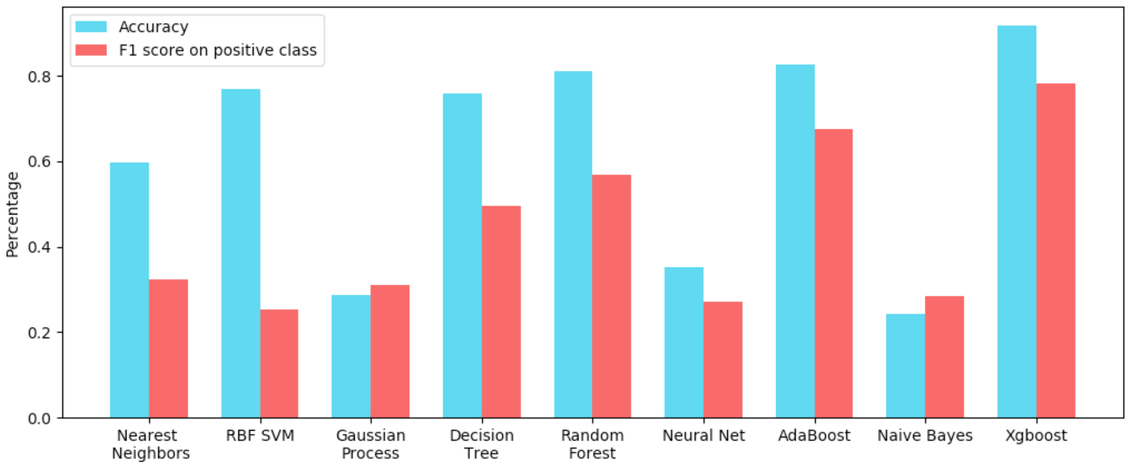

Fig. 3. Comparison of some of the classic machine learning methods.

From the results above, Xgboost (Chen and Guestrin, 2016) outperformed the others and achieved an overall accuracy of 91.74% and F1-score of 78.34% on the positive class, on the two-month test set. More details regarding the evaluation of this model are shown in Table 1 and 2. From the results below, the model has achieved a high precision and recall on the negative class, with slightly lower precision and recall on the positive class. The number of false positives and false negatives are small compared with the true positives and true negatives.

Table 1. Precision, recall and F1-score on test set for Xgboost model.

		Precision	Recall	F1-score
No Precipitation	0	0.93	0.97	0.95
Precipitation	1	0.85	0.73	0.78

Table 2. Confusion matrix on test set for Xgboost model.

		Prediction - 0	Prediction - 1
True Label	0	1312	45
	1	96	255

Besides precipitation, there may be other reasons which lead to a slowdown of traffic. Global events, such as mist and haze may also lead to similar traffic speed variation patterns. Therefore, we further analyzed the weather conditions of the false positive predictions. Out of the 45 false positive predictions, 29 are associated with a description of light rain, light snow, haze or mist, which all indicate adverse weather conditions, but still result in 0 mm precipitation records. For the true negative predictions, 81 out of 96 are less or equal to 1 mm precipitation and no record above 3.3 mm precipitation was missing.

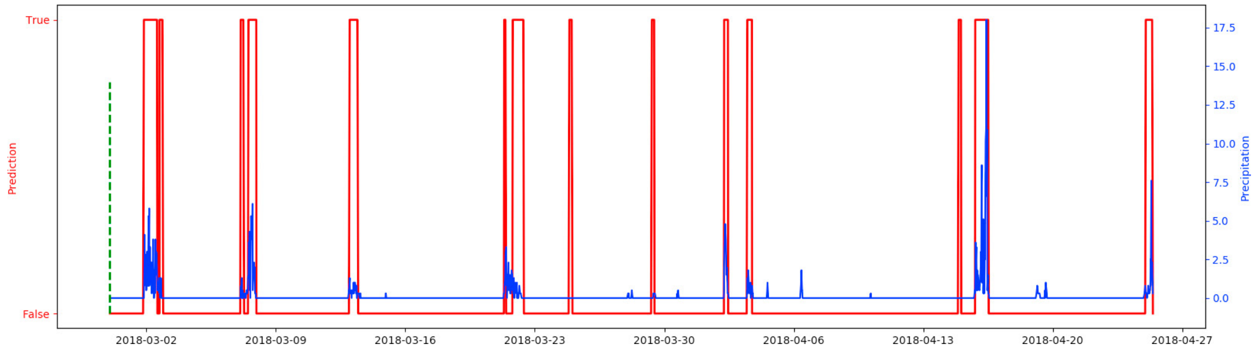

Fig. 4. Predictions on 2-month traffic speed test dataset with Xgboost. The blue line indicates the precipitation amount in millimeter and red line indicates the prediction from our Xgboost model.

Fig. 4 shows the predictions of the Xgboost model on the 2-month test data compared with the precipitation amount. This set is totally independent of the training data. As can be seen, almost all of the precipitation events are covered by our positive predictions. The model also ignored the events with very little precipitation. Only two false alarms were given, both of them being very short.

A second output of the Xgboost classifier is the feature importance. Since each feature corresponds exactly to one road in the dataset, the importance of each road for the overall classification result can be determined. For the Xgboost classifier, this is shown in Fig. 5. The highlighted roads have a higher importance than the others. Therefore, we considered these roads to be more related to the precipitation events than the others, because they play a more important role for the model to make a reasonable prediction. These highlighted roads can be further observed and the reasons for their importance should be analyzed in further work.

Figure 6 and 7 show examples of two precipitation events on Apr. 25th and 15th, 2018. The color indicates the speed observation, slower (blue) or faster (red) than the *Prophet* model, for each individual road (identified in the figure by their individual road id). The black and green lines represent the start and end time of the precipitation event based on the textual weather description. On the right side, the corresponding precipitation amounts and binary prediction from our Xgboost model are compared. In both examples, a significant slow down of the traffic can be observed when the precipitation event happens, which is in line with our common sense. Comparing start and end time extracted from textual descriptions, our model makes only positive predictions when the precipitation amount increases. The time range of significant precipitation events was successfully identified in both cases.

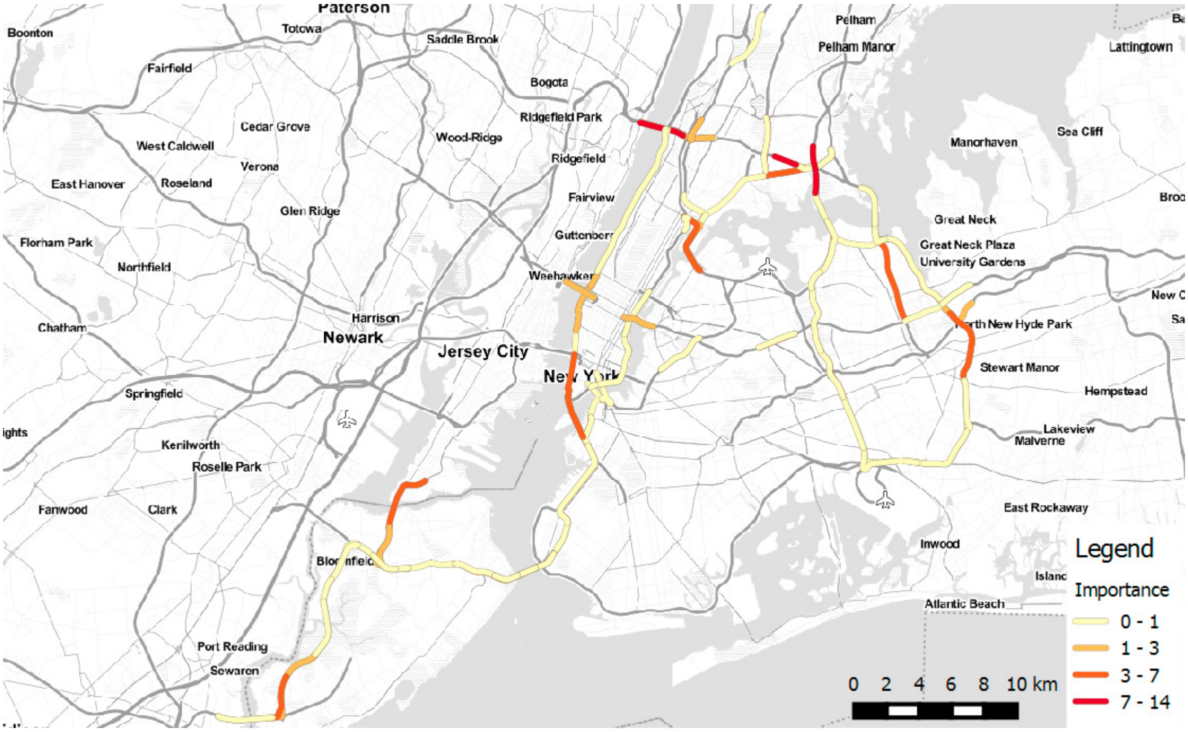

Fig. 5. Importance of each road, as determined by Xgboost (basemap: OpenStreetMap).

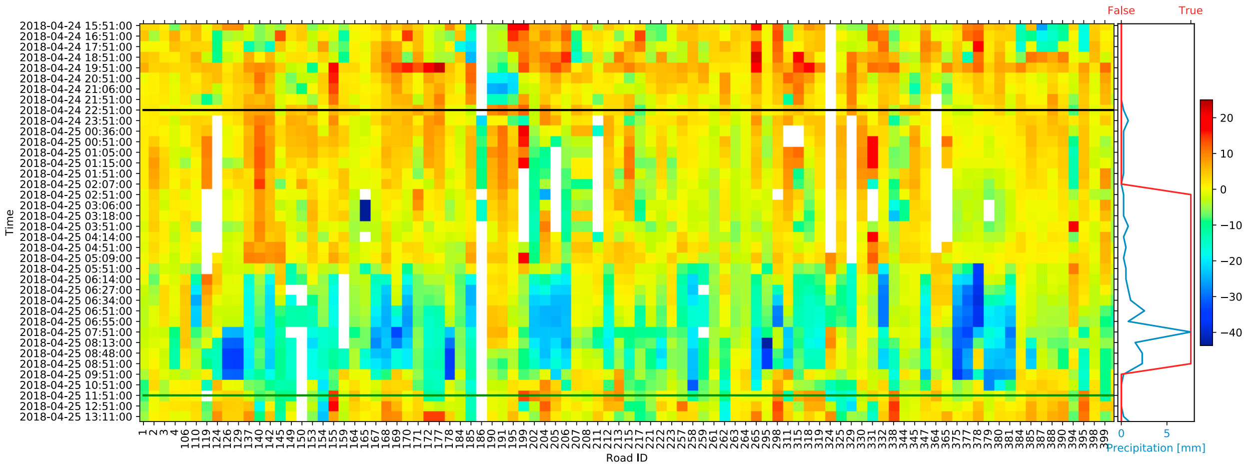

Fig. 6. Speed variation pattern for a precipitation event on Apr. 25, 2018. The color indicates the speed observation, slower (blue) or faster (red) than the *Prophet* model. The black and green lines represent the start and end time of the event based on the text description. On the right side, the blue line indicates the corresponding precipitation amount and the red line indicates the prediction from our xgboost model.

4. Conclusions and Outlook

In this paper, we proposed a proof-of-concept approach, which can indicate precipitation events based on traffic speed variation patterns. Seasonal trend decomposition was used to eliminate the daily and weekly periodic effects of the traffic observations. Residuals of this model were used as features which indicate the anomaly level of the traffic as compared to the normal traffic state. Several machine learning methods were compared, and finally gradient boosting

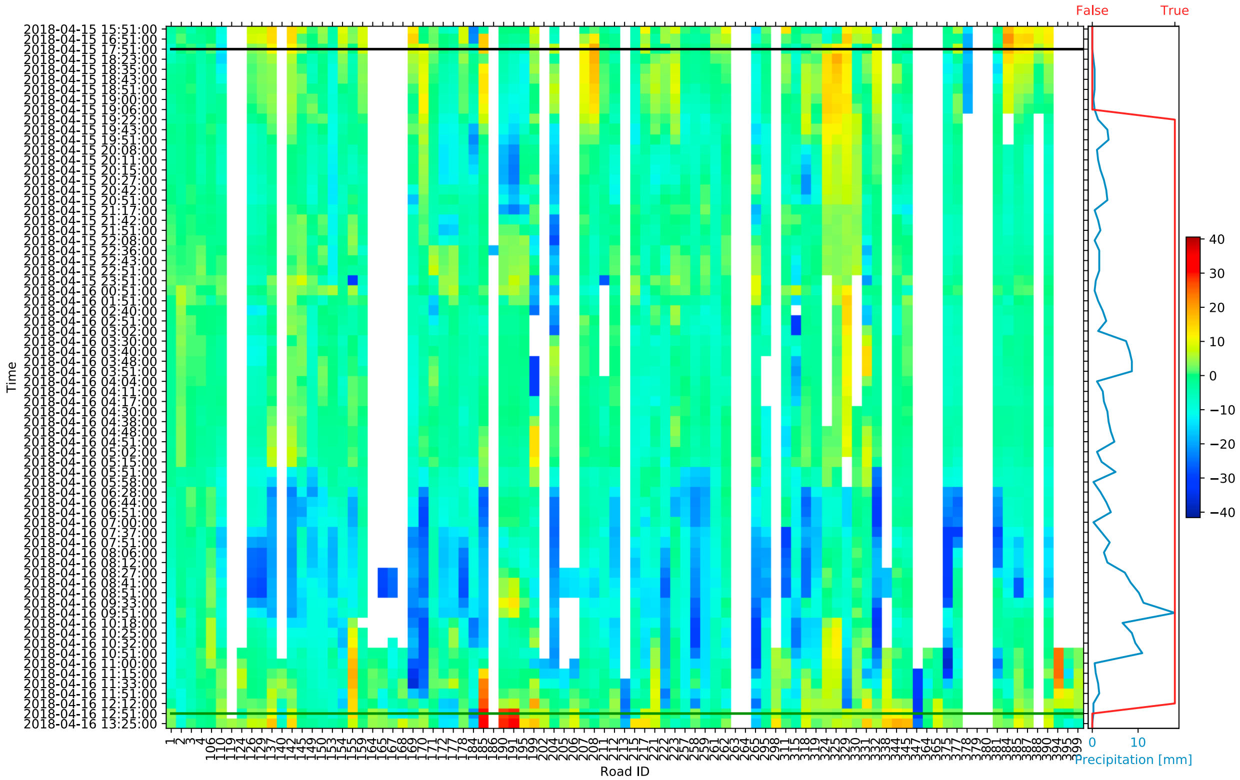

Fig. 7. Speed variation pattern for a precipitation event on Apr. 15, 2018. The color indicates the speed observation, slower (blue) or faster (red) than the *Prophet* model. The black and green lines represent the start and end time of the event based on the text description. On the right side, the blue line indicates the corresponding precipitation amount and the red line indicates the prediction from our Xgboost model.

was chosen to train a classifier, which makes predictions with respect to precipitation events based on these features. Since only a limited amount of positive examples were available within our observation range, we confined ourselves to a binary classifier. It has achieved a promising performance, an accuracy of 91.74% and F1-score of 78.34% for the positive class. Furthermore, we generated visualizations of precipitation events. Our indicator was able to successfully identify most of the precipitation events during the two months for testing.

If longer observation periods containing more positive examples are available, this approach can be further investigated, whether it is possible to learn an indicator which can estimate the severity level of precipitation events. With dense road speed observations from multiple regions, separately trained models with respect to each region can help to identify where such precipitation events happen and begin to affect traffic. Not all precipitation events have a significant influence on traffic, therefore, this system could be used to detect such events and to provide alerts to drivers only when it is necessary.

Acknowledgements

The authors would like to acknowledge the support from BMBF funded research project “EVUS – Real-Time Prediction of Pluvial Floods and Induced Water Contamination in Urban Areas” (BMBF, 03G0846A).

References

Chen, T., Guestrin, C., 2016. “Xgboost: A scalable tree boosting system,” 22nd ACM SIGKDD international conference on knowledge discovery and data mining. San Francisco, CA, USA, 785–794.

Haberlandt, U., Sester, M., 2010. Areal rainfall estimation using moving cars as rain gauges - a modelling study. Hydrology and Earth System Sciences 14(7), 1139–1151.
Jägerbrand, A. K., Sjöbergh, J., 2016. Effects of weather conditions, light conditions, and road lighting on vehicle speed. SpringerPlus 5(1), 505.
Jia, Y., Wu, J., Xu, M., 2017. Traffic flow prediction with rainfall impact using a deep learning method. Journal of advanced transportation, 2017.
Lam, W. H., Tam, M. L., Cao, X. Li, X., 2013. Modeling the effects of rainfall intensity on traffic speed, flow, and density relationships for urban roads. Journal of Transportation Engineering 139(7), 758–770.
Li, Q., Hao, X., Wang, W., Wu, A., Xie, Z., 2017. Effects of the rainstorm on urban road traffic speed-A case study of Shenzhen, China. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 71-75.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.; Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Prasad, N., Reddy, P. K., Naidu, M. M., 2013. "A Novel Decision Tree Approach for the Prediction of Precipitation Using Entropy in SLIQ," 2013 UKSim 15th International Conference on Computer Modelling and Simulation. Cambridge, UK, 209–217.
Sathiaraj, D., Punkasem, T. O., Wang, F., Seedah, D. P., 2018. Data-driven analysis on the effects of extreme weather elements on traffic volume in Atlanta, GA, USA. Computers, Environment and Urban Systems 72, 212–220.
Taylor, S. J., Letham, B., 2018. Forecasting at scale. The American Statistician 72(1), 37-45.
Tsapakis, I., Cheng, T., Bolbol, A., 2013. Impact of weather conditions on macroscopic urban travel times. Journal of Transport Geography 28, 204–211.
Yang, S., Qian, S., 2019. Understanding and predicting travel time with spatio-temporal features of network traffic flow, weather and incidents, arXiv.