SynTIS: Synthetic Traveler Information in Smart City

Yohan Chang

Korea Research Institute for Human Settlements, South Korea (email: ycanns@krihs.re.kr)

Keywords*: Smart City, Safety, Machine Learning, Prediction, Data fusion*

Introduction

A traffic safety is one of the critical concerns for not only individual travelers but also agencies. Although rapidly advancing technologies including connected/autonomous vehicle (CAV) and Internet of things (IoT) allowed us to open a new door to the Smart City, we are still living in the world where has a crash in every 5-second in the United States and this statistic has been kept growing since 2011 (NHTSA). In transportation planning and operation domains, there have been lots of efforts to eliminate these dangerous trends with diverse ways including roadway detector data and historical crash data. What is still absent is to harmonize and synthetize those data sources into together to support a better decision making, especially in traffic crash predictions. A traffic is quite dynamic due to the complexity of the nature so that both user and provider sides need an extra attention to improve the traffic safety. In this research, we offer a framework, synthetic traveler information in smart city (SynTIS), to provide an accurate real-time traveler information for individual travelers and state governments uniting a slew of information into a single data source.

SynTIS utilized and synthesized several data sources and the data sources including road characteristics (i.e., number of lanes and speed limits), real-time travel time data powered by Nokia HERE, general transit feed specification (GTFS) data, real-time special events data such as work zone and sports games queried from Twitter, crash report data reported by state highway patrol, and weather data. Finally, SynTIS predicts a crash event using three emerging machine learning methods including random forest (RF), t-distributed stochastic neighbor embedding (t-SNE), and stacked autoencoder (SAE).

A suburban area of St. Louis, the U.S. was selected for this research. During between March 2017 and February 2018, a total of 891 crash events was identified along the test area; 334 crash events were with no specific injury, 54 crash events were with fatal or serious injuries, and 225 crash events were in sport game days during the period.

Methodology

The Six major data sources, i.e., road characteristics, travel time information, GTFS, special event information, crash report, and weather history, were collected and combined with based on location and time of day information for each data source. For example, real-time travel time data, crash record and road characteristics were obtained from Missouri State Highway Patrol (MSHP)[1] and Missouri Department of Transportation (MoDOT)[2] and these data sources are publicly accessible. Archived weather data can be obtained from National Oceanic and Atmospheric Administration (NOAA)[3], special event data can be extracted from MoDOT's Twitter postings[4], and GTFS data was obtained from the public[5]. All process for data collecting and analyzing were performed using open source tools: R and Python.

The machine learning techniques are applied to predict two different event outcomes. In scenario 1, a crash chance for different time of day and segment of the roadway is delivered. While in scenario 2, an injury severity is predicted for different time of day and segment of the roadway. Total dataset was divided into three subsets with rule of thumb – 50% for training, 25% for validation, and 25% for testing. Also, several hyper-parameters for each machine learning model are finetuned with simple optimization techniques; hyperparameters for RF model were calibrated using Tabu-search, hyperparameters for t-SNE was fine-tuned using Tabu-search in conjunction with K-nearest-neighbors (K-NN) algorithm, and genetic algorithm (GA) was used for optimizing SAE's hyperparameters (Chang and Edara 2017).

Results

Performance results of three machine learning models is compared with a Naïve classifier. A Naïve classifier classifies all events as a single outcome (i.e., safe event in scenario 1 and 50% of injury severity in scenario 2). The model performance on testing are shown in Table 1. The RF model achieved the highest accuracies for both types of event predictions, and SAE model also showed high accuracies in the next of the RF model.

[1] https://www.mshp.dps.missouri.gov/ (access date 2019. 12. 25)
[2] http://traveler.modot.org/ (access date 2019. 12. 25)
[3] https://www.noaa.gov/ (access date 2019. 12. 25)
[4] https://twitter.com/MoDOT (access date 2019. 12. 25)
[5] https://transitfeeds.com/ (access date 2019. 12. 25)

Table 1. Crash event prediction results for each scenario

	Naïve	**SAE**	**RF**	**t-SNE**
Scenario 1	79.04%	89.13%	93.41%	85.17%
Scenario 2	60.64%	85.67%	87.20%	81.80%

Conclusion

Traffic safety should be never underestimate since it deals with a human life. With respect to the increasing crash records of the United States, there have been wealth of efforts to minimize this concern to save more lives. Most of the existing studies focused merely a few key factors such as location and weather (Salas et al. 2017), travel time and speed (Yu et al. 2014; Chand et al. 2018; Hou et al. 2018), and detector information (Zhang et al. 2018), although a crash is the consequence of the complex environments. This study proposes a synthetic data fusion framework for predicting crash event with smart city environment. Six data sources, road characteristics, real-time travel time information, GTFS, special event data, crash event information, and weather information, were collected from the public and processed for this study using open source tools, R and Python. The results showed that RF model outperformed other models in terms of accuracies.

Three machine learning algorithms, RF, t-SNE, and SAE, have unique characteristics in terms of their training fashions. For instance, RF is supervised learning, t-SNE uses unsupervised learning, and SAE is semi-supervised learning fashion. Such a unique training fashion can be performed not only conventional crash prediction problems but also unconventional crash prediction problems such as new driving conditions of CAV.

References

1. Annual Fatality Analysis Reporting System (FARS) Final, NASS GES; Vehicle miles traveled (VMT), Federal Highway Administration, Traffic Volume Trends, monthly and yearly Highway Statistics (VM-1) (annual) (https://cdan.nhtsa.gov/SASStoredProcess/guest, access date 2019.12.25)
2. Salas, A., Georgakis, P., Nwagboso, C., Ammari, A. and Petalas, I., 2017, July. Traffic event detection framework using social media. In *2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC)* (pp. 303-307). IEEE.
3. Yu, W., Park, S., Kim, D.S. and Ko, S.S., 2015. An arterial incident detection procedure utilizing real-time vehicle reidentification travel time data. *Journal of Intelligent Transportation Systems*, *19*(4), pp.370-384.
4. Chand, S. and Dixit, V.V., 2018. Application of Fractal theory for crash rate prediction: Insights from random parameters and latent class tobit models. *Accident Analysis & Prevention*, *112*, pp.30-38.
5. Zhang, Z., He, Q., Gao, J. and Ni, M., 2018. A deep learning approach for detecting traffic accidents from social media data. *Transportation research part C: emerging technologies*, *86*, pp.580-596.
6. Hou, Y., Edara, P. and Chang, Y., 2017, October. Road network state estimation using random forest ensemble learning. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1-6). IEEE.
7. Chang, Y. and Edara, P., 2017, October. Predicting hazardous events in work zones using naturalistic driving data. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1-6). IEEE.