Proactive Detection of Crash Hotspots using In-vehicle Driving Recorder

Jooyoung Lee
The Cho Chun Shik Graduate School of Green Transportation
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
mose585@kaist.ac.kr

Kitae Jang
The Cho Chun Shik Graduate School of Green Transportation
Korea Advanced Institute of Science and Technology
Daejeon, Republic of Korea
kitae.jang@kaist.ac.kr

Abstract— **Crash hotspot detection is important to reduce traffic crashes by allowing effective deployment of countermeasures in those locations. However, current hotspot detection methods rely mostly on crash occurrences and, therefore, countermeasures can be implemented only after a number of crashes have been occurred. To prevent crashes prior to their actual occurrences, crash precedents, also known as surrogate safety measures, are required. In this regard, driving behavior is recognized as a reliable precedent of crash occurrence because it reflects how human drivers respond to their driving environments. Therefore, the objective of this study is to develop a proactive crash hotspot detection method by evaluating relation between crash and aggressive driving behavior that was extracted from 5 taxi drivers' driving records. The spatial correlation analysis results showed that there was a distinct positive correlation between crash and aggressive behavior occurrence as evident from high correlation coefficient (which is about 0. 863.) This finding implies that aggressive driving incidents that were suggested in this research could be considered as a surrogate safety measure for proactively detecting crash hotspots.**

Keywords-crash hotspot detectiont; in-vehicle driving recorder; aggressive driving behavior; correlation analysis

I. Introduction

Traffic accidents are one of the main causes of fatalities in South Korea. Due to the advance in safety technology and continuous effort to improve roadway safety, the number of accidents has consistently decreased since 2000. In 2014, however, about 5,000 fatalities in 200,000 crashes were occurred. This was the highest rate with 11.3 deaths per 100,000 people among the OECD countries (Korean statistical information service, 2014).

A major approach to safety improvement has been detecting hotspots and implementing proper countermeasures in those areas. This approach is feasible only after the sufficient number of crashes were observed. It is limited to prevent crash occurrences because traffic crashes are so rare and uncertain event that it takes time to collect enough cases for statistically significant hotspot detection. Prior to actual crash occurrences, therefore, proactive approach that detects locations with high potential for crash hotspots is needed. Previous research is based mostly on an approach using real traffic crash data and cluster analysis – detecting locations where crashes are concentrated and inspecting factors that might have caused the crash occurrences [1-5]. However, this approach is limited for in-depth causal analysis as it has focus solely on a relation between crash occurrence and environmental factors while crashes occurred via far more comprehensive procedure, especially from the driver side. Furthermore, it is known that drivers are the major cause of crashes, accounting for more than 90 percent of vehicle accidents [6]. Hence, human factors should be taken into consideration in proactive detection of crash hotspots.

To this end, previous research has developed an approach employing surrogate safety measures that identifies crash precedents such as time to contact, speed fluctuations and shock wave frequency and computes crash likelihood before the crash actually occurred [7, 8]. Among various safety surrogate measures, [9] proposed that the driving behavior was better precedents to crash occurrence because it reflects footprints of how human drivers respond to driving factors. Some research has been done in this realm. However, two important points have not been evaluated – i) types of driving behavior that may have impact on crash occurrences; and ii) relation between such driving behavior and crash occurrences (i.e., verification of relation between driving behavior and crash occurrence). To answer these questions, we extracted aggressive driving behavior as a surrogate safety measure from in-vehicle driving records and examined spatial association between aggressive driving behavior and crash occurrences.

Recent technological advance in in-vehicle sensors facilitates the collection of information on driving behavior. In Korea, it is legislated that commercial vehicles were mandated to install in-vehicle driving recorders in 2011. Since the legislation, the number of commercial vehicles equipped with driving recorder has risen and large amount of data have been collected from these vehicles. Taxi drivers, among all the commercial vehicles, have relatively comparable driving pattern with general drivers; and the taxi drivers' aggressive driving behavior could be extracted by observing abruptly changed behaviors using driving recorder. Therefore, in this study, we suggest the proactive crash hotspot detection method using in-vehicle driving record by considering the aggressive driving behavior as a surrogate safety measure. We evaluate relation between aggressive driving behavior and crash occurrence to verify the causal effect of the aggressive driving behavior on crash occurrences.

The remainder of this paper is organized as follows: in section II, theoretical backgrounds about hotspot detection methods are explained. In section III, descriptions on taxi

978-1-5090-5753-5/16 $31.00 © 2016 IEEE
DOI 10.1109/APWC.on.CSE.2016.38

driving recorder data, crash data and study area are presented. In section IV, results of detected crash hotspots are displayed, and, in section V, the findings and implications are discussed.

II. Method

The procedure for proactive crash hotspot detection is twofold: A. aggressive driving behavior extraction using in-vehicle driving recorder, and B. Geographic Information System (GIS) analysis using crash data and extracted aggressive driving behavior data. The first procedure for aggressive driving behavior extraction was composed of three steps: driving event extraction from driving record data, time-series feature extraction, and cluster analysis. At each step, a different technique was utilized. Abrupt change detection (ACD) algorithm was used for driving event extraction; Auto-encoder (AE) was used for driving event feature extraction; two-level clustering was used for clustering analysis. The second procedure, GIS analysis, was composed of kernel density analysis and statistical correlation analysis. Feasibility of proactive crash hotspots was evaluated based on the correlation with crash and aggressive driving behavior.

A. Aggressive Driving Behavior Extraction

The aggressive driving behavior extraction method used in this study is abruptly changed driving pattern clustering method that was developed by [10]. The following summarizes the method. For more details, please see [10].

1) Abrupt change detection: Since driving record data have a long time-series data format, it is hard to derive driving characteristics directly. Thus, the record data need to be processed to characterize driver behavior by extracting driving events using the ACD algorithm when drivers control their vehicles in respond to the events. The ACD is the method for statistically detecting change points in time series. In this study, we used relative unconstrained least-squares importance fitting method which was possess a better performance than other change point detection algorithms [11]. Then, 15-sec time intervals that over 500 abrupt change scores were defined as driving events.

2) Auto-encoder: The extracted driving events are too diverse because there are many similar patterns and such diversity reduces clustering performances. Thus, before clustering, the feature of driving event need to be extracted. In this study, auto-encoder was used for feature extraction. The auto-encoder is a kind of neural network algorithm and provides more abstracted values then conventionlly-used statistical values [12]. Input and output data were same as values of driving record data: speed, acceleration and yaw rate. During training, the hidden representations were calculated. These hidden representations indicated that the abstracted values of the driving event patterns which were extracted using ACD, and became the feature of driving events. The extracted feature was used for the input of clustering analysis.

3) Two-level clustering: The two-level clustering algorithm was used to classify aggressive driving behaviors among extracted driving events. The two-level clustering algorithm is combination of two clustering algorithms: Self Organized Map (SOM) and *K*-means clustering. This algorithm applied the SOM first for initializing and dimension reduction, and then applying *K*-means clustering for organizing optimized clusters. The benefit of this method was it could reduce the computational cost, minimize the local minimum risk [13]. Then, Davies-Bouldin index was used to find the best cluster numbers of *K*-means clustering. Finally, the potential aggressive driving behavior was classified and categorized among organized clusters by analyzing each cluster's features: statistics of driving speed, acceleration and yaw rate.

B. GIS Analysis

GIS analysis was conducted to verify spatial correlation with extracted aggressive driving behavior and crash occurred spots. We used Esri's ArcGIS 10.2 software for this study which is widely-sued for spatial visualization and analysis applications. Another purpose of GIS analysis was observing and analyzing geographical patterns of aggressive driving behaviors distribution and crash hotspots; thus, the GIS analysis was done based on quadrat analysis. The quadrat analysis was performed on point data and divided study region into same size of cells called quadrats. This analysis has a merit of extraction representative values from point values in GIS. The optimal size of the quadrat is determined following (1) which was suggested on [14].

$$Q_{size} = 2 \times \text{study area} \,/\, \#\text{ of samples}. \tag{1}$$

1) Kernel density analysis: Kernel density analysis deduced the density map by calculating kernel function of aimed events within unit area. This analysis used for visual representation of aimed events' dense zone because high severity zone had higher density and colored in density map. The analysis was done using predefined cell on GIS as basic units. Each cell density was calculated by adding the value of overlaid other cell's kernel surfaces. ArcGIS 10.2 used quartic kernel function form presented in [15]. The kernel function was one of the probability density function. The basic form of quartic kernel function showed on (2) and its shape was described in Fig.1. The predicted value of not occurred area could be calculated by using density map because kernel analysis smoothed the spatial distribution of aimed events. In this study, kernel density analysis was done both crash data and extracted aggressive driving behaviors; and compared their properties.

$$K(u) = \frac{15}{16}(1-u^2)^2 * \delta(u),\ \ \delta(u) = \begin{cases} 1 & if\ |u| \le 1 \\ 0 & otherwise \end{cases}. \tag{2}$$

1) Statistical correlation analysis: The correlation analysis is the method for statistically estimating the linear relationship between two variables. The correlation was presented as correlation coefficient. This coefficient did not explain the causal relationship. In this study, simple correlation analysis was conducted for estimating the correlation between the crash and the extracted aggressive driving behaviors using Pearson

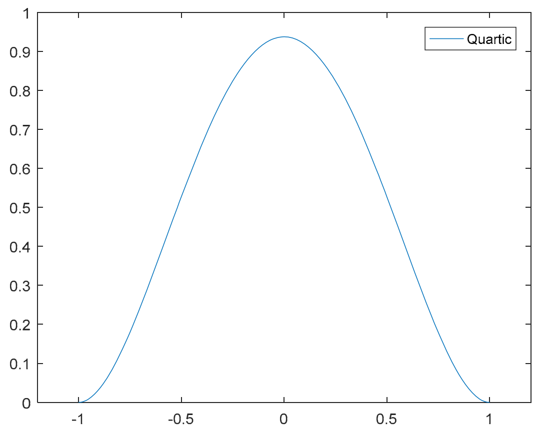


Figure 1. Quartic kernel function graph.

correlation coefficient which generally used to calculate two variable's correlation [16]. The Band Collection Statistics toolbox in ArcGIS calculated the coefficeint between density maps. This coefficient r was calculated using (3), (4).

$$cov(Crash, Agg) = \frac{\sum_{k=1}^{N}(Z_{crash,k}-\mu_{crash})(Z_{Agg,k}-\mu_{Agg})}{N-1}. \quad (3)$$

$$r = \frac{cov(Z_{crash}, Z_{Agg})}{\sqrt{var(Crash)}\sqrt{var(Agg)}}, \quad (4)$$

where, Z = *value of cell,*
μ = *the mean of a layer,*
N = *the number of cells.*

III. Data Description

In this study, in-vehicle driving recorder data of taxis that drove in Korea metropolitan cities was used for aggressive driving behavior extraction; and, 6-year taxi accident data were used for developing proactive crash hotspot detection method. The in-vehicle driving recorder data was composed of two types: 39 taxi drivers' driving recorder was used for constructing the aggressive driving behavior cluster map, and another 5 taxi drivers' driving recorder used for correlation analysis based on GIS; because clustering analysis needed many drivers driving data for generalization of potential aggressive driving behavior clusters, and long term driving data of selected region is needed for correlation analysis to draw statistically significant result.

The aggressive driving behaviors were classified among each of 39 driver's driving events which were extracted from driving recorder using ACD algorithm and 39 taxi drivers drove taxis in three regions (Seoul, Daegu and Busan) in Korea. The driving recorder data was 39 taxi drivers' one week driving records from Jan. 2nd. 2012 to Jan. 8th. 2012; 20 drivers in Seoul, 10 drivers in Daegu, 9 drivers in Busan. The 5 taxi drivers' driving recorder for correlation analysis were preprocessed as same with 39 drivers' data. It collected from two taxi companies located in southeast Seoul. Total record date were 156 days from Jun. 2013 to Sept. 2013. In 5 drivers' driving recorder, the GPS based coordinate value of each record point was collected so projection on GIS map was possible. The detail record lists of both driving recorder data are summarized in TABLE I. The detail record lists of both driving recorder data are summarized in TABLE 1. Study area was selected based on the location of 5 drivers' taxi companies because the general one-day trip of each taxi driver was started from company's garage and finished to company's garage; so the driving data was concentrated near taxi company. Since all taxi companies in this study were located on southeast Seoul, study area was selected in southeast Seoul: Yangcheon-gu, Youngdeungpo-gu, Dongjak-gu, Gwanak-gu, and Geumcheon-gu. Fig.2 shows the study area (gray area) and the location of two taxi companies (red mark). The 6-year taxi crash data was collected from Transport Workers Management System in Korea from 2010 to 2015, and the total number of crash in study regions was 8,533 which included all crash type. This crash data has GPS position of crash occurred spot; so we could observe the geographically distribution of crash hotspots.

IV. Results

A. Aggressive Driving Behavior Extraction

1) Driving events extraction from driving record: Using ACD, all 15 seconds of driving period that were regarded as aggressive driving behavior were extracted. The ACD calculate the change score using 15 seconds time window, which is assumed the driving event time, and sliding it by one seconds until covering whole driving record. The change scores are increased when sensor value changed. The threshold of the change point score for discriminating change points was 500 which was about the top five percent.

TABLE I. Description of DTG Data

Type	Category	Description
Clustering analysis	Period	7 days (2012.01.02. ~ 2012.01.08.)
	Region	Seoul, Daegu, Busan
	# of Driver	39
GIS analysis	Period	156 days (2013.06. ~ 2013.09.)
	Region	Seoul
	# of Driver	5

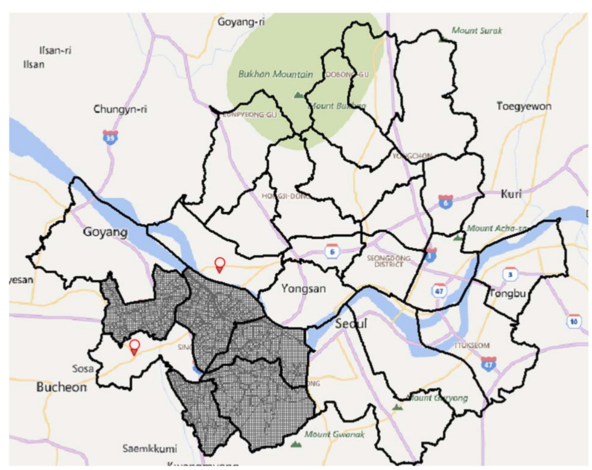


Figure 2. Study area of GIS analysis.

With the change point as the center of driving event, 15 seconds of driving record is extracted as driving event. The total number of extracted events was 124,793.

2) Time-series feature extraction: Before clustering the detected driving events, to improve performance of clustering, there needs to be feature extraction process for driving events. Combining 15 seconds of three sensor values to constitute one-dimension vector, the input of auto-encoder was made. In this study, to construct the auto-encoder, 30 hidden representations, logistic sigmoid encoder and linear function decoder were used; and transfer functions of each neuron were set as sigmoid for encoder and linear for decoder. In order to extract more generalized features, additional constraints such as 0.001 of L2 weight regularizer, 4 of sparsity regularizer, and 0.05 of sparsity proportion were used for training the auto-encoder. After the training, 30 hidden representations had abstracted value of driving events and engaged all the abstracted value, the feature matrix was derived.

3) Clustering result: The feature of driving events was clustered by projecting to the cluster map. Criteria for aggressive driving behavior needed to be set to extract the aggressive driving behavior. In this study, we used the clustering based aggressive driving behavior criteria which introduced in [10]. These criteria were derived from similar three steps, which we used to analysis the driving record data. TABLE II represents the detailed criteria for each aggressive driving behavior cluster. In high speed range, many aggressive behaviors were rapid deceleration, and in row speed range, many behaviors were rapid acceleration. Since the turning behaviors were not occurred independently, these type of behaviors were placed at middle range of speed. The rapid acceleration behaviors had less proportion than rapid deceleration however rapid acceleration with turning behaviors were higher than deceleration with turning. The extracted driving events were projected to the cluster map, which constructed using the criteria, and the aggressive driving behaviors of each driver was distinguished. The result of clustering is extracted aggressive driving behaviors. Total number of extracted aggressive driving behavior is 18,377.

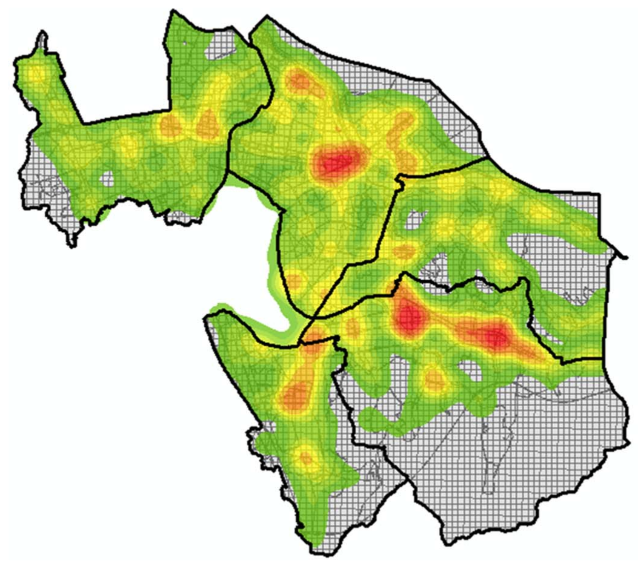
B. GIS Analysis

Using aggressive driving behavior projected result, we conducted the GIS analysis of study area: Yangcheon-gu, Youngdeungpo-gu, Dongjak-gu, Gwanak-gu, and Geumcheon-gu in southwest of Seoul. The accident data from this region was also projected onto GIS. The kernel density analysis was carried out on both aggressive driving behavior and crash; then, derived the event hotspots and comparing them. Fig. 3 presents the kernel density results of them. The region we focused on is red-colored area, which implies higher event occurrence rate than other area. Distinct three crash hotspots are observed on Fig. 3(a) and aggressive driving behavior hotspots have more hotspot regions, including crash hotspot region.

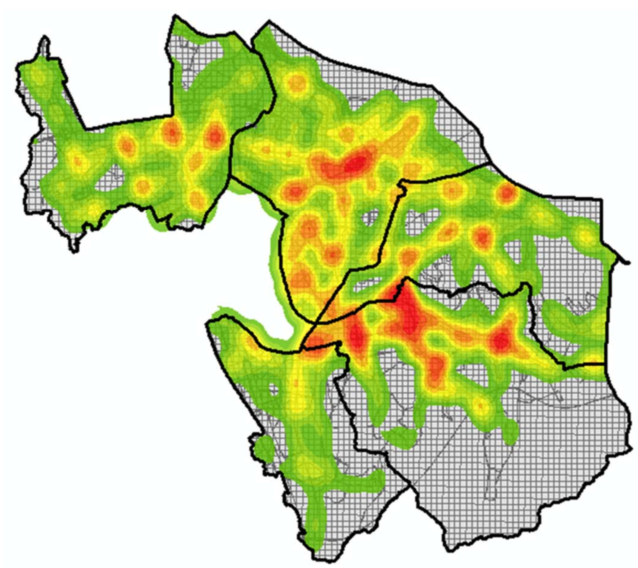
TABLE II. CLUSTER CRITERIA OF AGGRESSIVE DRIVING BEHAVIOR

Cluster	Speed (km/h)	Acceleration (km/h/sec)	Yaw Rate (Degree/sec)	Description
1	under 17	-9.04	2.97	Rapid Deceleration
2	17 ~ 55	-7.72	2.62	Rapid Deceleration
3	17 ~ 55	-5.56	6.57	Rapid Deceleration with Steering
4	17 ~ 55	7.74	-3.40	Rapid Acceleration with Steering
6	17 ~ 55	3.79	-13.69	Rapid Steering
7	17 ~ 55	-5.90	5.89	Rapid Deceleration with Steering
8	over 55	-4.47	0.87	Rapid Deceleration
11	17 ~ 55	6.20	5.51	Rapid Acceleration with Steering
13	under 17	6.28	-0.07	Rapid Acceleration
16	over 55	4.99	-0.95	Rapid Acceleration
17	17 ~ 55	5.91	5.74	Rapid Acceleration with Steering
18	17 ~ 55	-7.75	1.48	Rapid Deceleration
21	17 ~ 55	-7.91	0.43	Rapid Deceleration
23	under 17	7.06	-4.17	Rapid Acceleration with Steering
26	17 ~ 55	-7.57	3.61	Rapid Deceleration
29	17 ~ 55	6.57	-3.26	Rapid Acceleration with Steering
31	under 17	-5.94	7.82	Rapid Deceleration with Steering

This shows that the aggressive driving behavior has high potential for crash precedent. To verify this, correlation coefficients with both events are deducted. We calculate the coefficient with crash and total aggressive driving behavior, and then, with crash and each aggressive driving behavior. TABLE III shows the result of correlation analysis. The results show that there is distinct positive correlation between crash and total aggressive behavior occurrence because the correlation coefficient was 0.863. Also, almost every aggressive driving behavior, except cluster 8 and 16 that occurred in high speed range, have distinct positive correlations. Therefore, we could consider some spots as proactive crash hotspots if some region has a high incidence rate of aggressive driving behavior. Finally, proactive crash hotspots of the study area are derived using quadrat analysis. The size of quadrat is 140m calculated from (1); the study area: 99,269,976m^2, and the number of samples: 21,083, which is a weighted average of crash and aggressive driving behavior. Then, calculate the number of both events each cell and shows Fig. 4's graph. The left-upper section refers to high aggressive driving rate areas. According to this study, the cells, included in this section, could be regarded as proactive crash hotspots. The most significant areas were selected and projected on the crash hotspot map as red squares (See Fig. 5). Crash prevention plans can be established on selected areas by investigating geometric design and traffic conditions.



(a) Crash hotspots



(b) Aggressive behavior hotspots

Figure 3. Kernel density analysis results.

V. Conclusion

Prior work has documented the importance of proactive crash hotspot detection and the positive correlation between traffic accidents and aggressive driving behavior. However, these studies did not focus on detection of crash hotspot using aggressive driving behaviors. Data acquisition of aggressive driving behavior is challenging and indistinct because of sensor limits. Other research on analysis of aggressive driving behavior using in-vehicle recorder lacked consideration of influential factors such as road environments, vehicle type and driver characteristics.

TABLE III. Coefficient result of both events

Correlation analysis	Pearson coefficient
Crash	1
Total aggressive behavior	0.863
Cluster 1	0.825
Cluster 2	0.751
Cluster 3	0.763
Cluster 4	0.745
Cluster 6	0.697
Cluster 7	0.770
Cluster 8	0.269
Cluster 11	0.765
Cluster 13	0.805
Cluster 16	0.529
Cluster 17	0.790
Cluster 18	0.772
Cluster 21	0.621
Cluster 23	0.781
Cluster 26	0.764
Cluster 29	0.815
Cluster 31	0.773

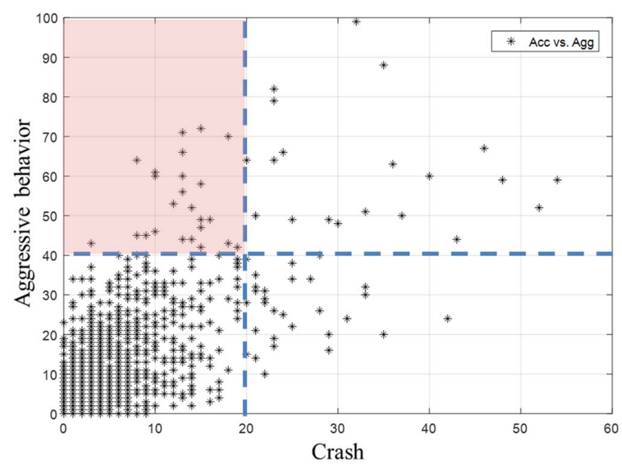


Figure 4. Relation of crash and aggressive behavior.

In this study, we suggested a proactive crash hotspot detection method using in-vehicle driving recorder. Aggressive driving behaviors, which were extracted from the driving recorder, were considered as a surrogate safety measure. Since taxi drivers tend to drive longer distances and a wider area than the average driver, large-scale driving data can be collected from various regions. We collected 5 taxi drivers driving data and verified the causal effect of the aggressive driving behavior by showing the correlation with crash and aggressive driving behavior; and, among southwest of Seoul, the proactive crash hotspots were detected based on the GIS analysis. We suggested the process of proactive crash hotspot detection which could be used for accident prevention.

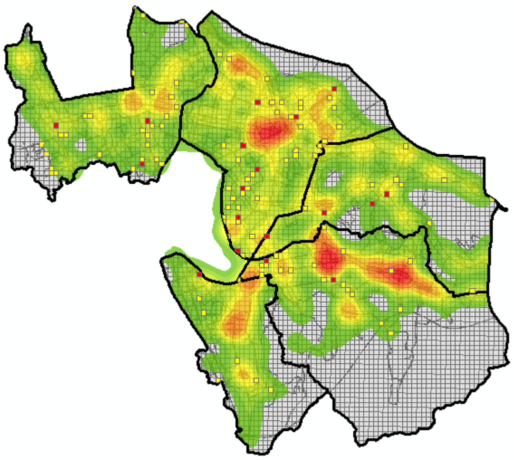


Figure 5. Proactive crash hotspots.

In particular, our study can contribute to suggesting an application technique of large-scale driving record data which has been collected from commercial vehicles in South Korea. However, there are some limitations in this study. The correlation analysis of aggressive driving behavior was conducted for two taxi companies. Although the findings in this study are a limited representative of the selected proactive crash hotspots, expanding the number of test companies and analysis of whole regions of Seoul will support our method. Future work will find the cause factors of difference between crash hotspot and aggressive driving hotspot by analyzing environmental factor of traffic and road geometry.

REFERENCES

[1] B. Depaire, G. Wets, and K. Vanhoof, "Traffic accident segmentation by means of latent class clustering," *Accident Analysis & Prevention,* vol. 40, pp. 1257-1266, 2008.

[2] T. K. Anderson, "Kernel density estimation and K-means clustering to profile road accident hotspots," *Accident Analysis & Prevention,* vol. 41, pp. 359-364, 2009.

[3] V. Prasannakumar, H. Vijith, R. Charutha, and N. Geetha, "Spatio-temporal clustering of road accidents: GIS based analysis and assessment," *Procedia-Social and Behavioral Sciences,* vol. 21, pp. 317-325, 2011.

[4] K. Chung, K. Jang, S. Madanat, and S. Washington, "Proactive detection of high collision concentration locations on highways," *Transportation research part A: policy and practice,* vol. 45, pp. 927-934, 2011.

[5] S. H. Park, K. Jang, D. Kim, S. Kho, and S. Kang, "Spatial analysis methods for identifying hazardous locations on expressways in Korea," *Scientia Iranica. Transaction B, Mechanical Engineering,* vol. 22, p. 1594, 2015.

[6] H. Lum and J. A. Reagan, "Interactive highway safety design model: accident predictive module," *Public Roads,* vol. 58, 1995.

[7] T. E. Boyce and E. S. Geller, "An instrumented vehicle assessment of problem behavior and driving style:: Do younger males really take more risks?," *Accident Analysis & Prevention,* vol. 34, pp. 51-64, 2002.

[8] K. Chung, K. Jang, S. Oum, Y. Kim, and K. Song, "Investigation of attributes of kinematic waves preceding traffic collisions," in *17th ITS World Congress*, 2010.

[9] B. A. Jonah, "Accident risk and risk-taking behaviour among young drivers," *Accident Analysis & Prevention,* vol. 18, pp. 255-271, 1986.

[10] J. Lee, "The study of large-scale digital tachograph analysis for machine learning-based aggressive driving behavior classification," M.S. thesis, GSGT, KAIST, Daejeon, 2016.

[11] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-point detection in time-series data by relative density-ratio estimation," *Neural Networks,* vol. 43, pp. 72-83, 2013.

[12] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters,* vol. 42, pp. 11-24, 2014.

[13] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *Neural Networks, IEEE Transactions on,* vol. 11, pp. 586-600, 2000.

[14] J. Lee and D. W. Wong, *Statistical analysis with ArcView GIS*: John Wiley & Sons, 2001.

[15] B. W. Silverman, *Density estimation for statistics and data analysis* vol. 26: CRC press, 1986.

[16] J. Lee Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician,* vol. 42, pp. 59-66, 1988.