Temporal Impulse of Traffic Accidents in South Korea

HYUNKYUNG SHIN1 AND JAEHO LEE2

1Department of Financial Mathematics, Gachon University, Seongnam 13120, South Korea
2Computer Education, Gyeongin National University of Education, Anyang 21054, South Korea

Corresponding author: Jaeho Lee (jhlee@ginue.ac.kr)

ABSTRACT With the emergence of urban computing technology, the development of smart cities has gained much attention as a means to improve citizens’ quality of life. As traffic accidents constitute a major problem that affects the quality of life, an effective solution to address this problem can significantly increase the level of intelligence of smart cities. This paper presents the development of a mathematical model for accurate analysis of big data to promote the effectiveness of policy decisions, thereby largely advancing the intelligent transportation systems (ITS) of smart cities. Temporal impulse was designed as a novel and measurable quantity to analyze traffic accidents by identifying the hidden patterns, such as varying causes and diverging impacts of traffic accidents. Based on the big data produced by the South Korean National Police Agency, we analyzed traffic accidents over three years by applying the temporal impulse. The research results suggested that the temporal impulse not only helped in identifying the varying influence of weather and driver conditions but also facilitated the establishment of sophisticated policies in the implementation of smart cities with the use of urban computing technology. As presented in the section VII, our simulation outputs indicated that our temporal model was predictive within the parameter space comprising driver’s dynamic behaviors, day of the week, and environmental factors including weather, road surface condition, and road type.

INDEX TERMS Big data, traffic accident data, intelligent transportation systems, smart city technology, stochastic process, time series, and temporal impulse.

I. INTRODUCTION

With the rapid progress of urbanization, problems such as traffic congestion, energy consumption, and environmental pollution are also increasing. These problems can be addressed by using traffic flow, human movement, and geographical data in urban computing, which is a representative interdisciplinary field connecting computer science, mathematics, urban engineering, traffic engineering, environmental science, ecology, social science, and urban policy. Thus, urban computing collects and analyzes the data acquired by various sensors installed and operated throughout the city [1].

Urban computing implements Internet of Things (IoT) concepts by connecting sensors and actuators embedded in various devices via smartphones. It processes and deploys huge amounts of data generated in various environments. Numerous efforts have been made to explore other possible applications of urban computing; one of these is an attempt to apply large and small urban computing technology in the wild, an environment completely different from the city [2].

With the emergence of urban computing technology, there has been an increased interest in constructing smart cities with the fundamental objective of improving the quality of life. It is therefore necessary to address traffic problems, which negatively affect the quality of life. Failure to properly control traffic can lead to serious traffic accidents and traffic congestion that can consequently deteriorate air quality, which in turn affects the quality of life. Consequently, to successfully construct and operate a smart city, a system that can appropriately control the traffic environment must be built.

Research on urban computing has focused on improving the quality of urban life. With the rapid diffusion of mobile devices such as smartphones, current research aims at effectively utilizing GPS, WiFi networks, RFID, sensors, and big data. Recently, studies have been conducted focused on urban computing methods (optimized for cloud environments) aimed at improving the technical aspects of urban life. From a technical standpoint, it is assumed that the cloud...
computing environment is optimal in that it supports the elemental technologies used in urban computing [3]. One method to achieve this objective was the construction of intelligent transportation systems (ITS) and currently, the performance of ITS is fairly satisfactory. The development of ITS began in the 1970s, and was intended to connect and create harmony between humans, vehicles, and roads and establish a wider range, fully efficient, real time, and accurate information management system [4]. As a result of the increase in productivity owing to the power of computing, and the merging of computing technology and telecommunications, we can now communicate more freely, which ensures comfort in workflow and other aspects of life. In the future, the convergence of fields such as telecommunications, computing, wireless, and transportation technology is expected to enhance both communication and transportation [5].

Transportation’s greatest dangers include issues such as high accident rates, traffic congestion, and air pollution from traffic and carbon emissions, among others. To boost the transportation sector, which is used for legally moving resources from one place to the other, studies have been conducted to integrate transportation with virtual technologies. Further development of ITS is expected play a vital role in overcoming transportation dangers [6]. Road accidents are one of the leading causes of preventable deaths, especially in developing countries, and is therefore a growing concern of governments. However, ITS can now aid on-road safety with the emergence of smartphones, a new platform that facilitates the application of sensor networks, driver-assistance systems, among others [7].

To improve the application of the aforementioned technologies, we must focus on vehicle-based monitoring systems such as driving behavior and style recognition, accident detection and road condition monitoring systems. The core technology of ITS relies on the analysis of past traffic history data and extremely short-term prediction using the current traffic situation. Generally, the extremely short-term unit used in ITS is ‘minutes’ and ‘seconds’ [8], [9]. Consequently, extremely short-term forecasting using current traffic conditions does not make a significant contribution to the traffic environment control. However, prediction of traffic situation by ‘hour’ unit or even ‘day’ unit can be extremely advantageous for application in smart cities.

How the traffic situation in units of ‘hour’ or ‘day’ be predicted? What type of data can be used to control this situation? An example of a suitable solution to the aforementioned questions is the use of weather forecast information. This proposition is based on the assumption that the traffic conditions change depending on the weather conditions. It can therefore be predicted if the weather conditions will affect the traffic situation, however, finding cases that provide accurate grounds is a challenge. Even if the big data about traffic accidents is analyzed based on frequency, there is a limitation. Consequently, developing a mathematical model that can accurately analyze big data will help to develop ITS that can effectively operate in smart cities.

Based on the abovementioned logic, we developed a novel mathematical model that effectively analyzed the big data related to traffic accidents managed by the South Korean National Police Agency. We subsequently analyzed the big data related to traffic accidents for three years using the developed mathematical model. The main features of this study are summarized as follows. First, the data analyzed in this paper are related to traffic accidents in Korea from 2015 to 2017, and the number of traffic accident data items analyzed was 491,545. Second, the temporal impulse was defined via mathematical modeling. Third, to validate the temporal impulse developed in this study, a statistical analysis based on frequency was performed. Fourth, we used R to statistically analyze the traffic accident data. Fifth, a temporal impulse analysis was performed to identify the unknown content of the abovementioned statistical analysis based on a frequency analysis.

The remainder of this paper is organized as follows. In section II, the previous studies related to our work are summarized. In section III, the mathematical description of our model is presented. In section IV, traffic accident data of South Korea are presented in detail. In section V, our methods based on mathematical modeling are presented. In section VI, factor analysis with time derivatives are presented. In section VII, we discuss the effect of traffic accident related issues on smart city technologies.

II. PREVIOUS RESEARCHES

Owing to the advancement of sensing and communication technologies, ITS were initiated by the EU in 2010. Among the underlying technologies of ITS, real time traffic accident detection has been one of the major research topics. Iranitalab and Khattak [10] survey various statistical and machine learning methods applied to crash severity prediction models. The detection component module of an ITS has been studied by various researchers as follows: Dogru et al. adopted anomaly detection methods from data mining techniques [11] and the random forest classifier to build vehicular ad-hoc networks (VANETs) [12]; Zeng et al. applied D-S evidence theory to detect traffic abnormalities [13]; Xiao devised a method of ensemble learning [14]; and Zhang et al. [15] surveyed data-driven frameworks.

Furthermore, studies have been conducted on the acquisition of field data from actual highways to build real time traffic accident detection models. Yu and Abdel-Aty [16] utilized car speed data on I-70 to train a support vector machine (SVM) classifier; Sheu [17] collected data from I-880 to apply a pattern recognition approach; Lu et al. [18] built the training dataset for a neural network classifier from highway A12 (in the Netherlands); Yasmin et al. [19] collected data from SR-408, 487, and 528 in Florida to build an econometric framework using multinomial logit model; and D’Andrea et al. [20] used GPS tracking data from the city of Pisa for traffic congestion detection.

The following notable techniques have been applied to traffic accident detection: a dynamic Bayesian network by
FIGURE 1. Samples of frequency graphs of daily car accident reports by hour of day. The graphs are obtained from the data reported during the period between February 1 (Sunday) and February 8 in 2015. The X axis represents the hour of day, the Y axis represents the frequency of accidents.

Sun and Sun [21]: a data mining technique for injury severity prediction by Chong et al. [22]; Bayesian logistic regression for crash risk analysis by Wang et al. [23]; and fuzzy deep learning for urban traffic incidence detection by Hatri and Boumhidi [24].

Lytras et al. [25]–[29] considered policy-making with sustainability, provided a roadmap toward the evolution of big data, and raised the issue of “normative bias” problem on Smart city research. Benevolo et al. [30] organized various aspects of the smart city including traffic, pollution, energy, and waste. Persaud et al. [31] and Pandey at el. [32] discussed issues with autonomous vehicles in smart cities. Recently, deep learning methods applied to smart city techniques have attracted attention from numerous researchers, which include Dotoli at el. [33] who proposed a mathematical model for a signal timing plan, Du et al. [34] and Puri et al. [35] who presented knowledge discovery with large data analysis techniques for the traffic optimization problem, Lv et al. [37] who proposed big data analysis techniques for traffic flow prediction, de Melo and Varde [38] who proposed a scalable learning scheme.

II. MATHEMATICAL MODELING

Several preliminary requirements exist for big data analysis [39]: perfectly processed ground truth datasets must be ready, the dataset must have a smaller number of features, and researchers must be sure of what they are looking for in the data. In this study, we work on real-world data from police reports of public traffic accidents in South Korea from 2015 to 2017.

Moreover, we focus on the time dependency of traffic accidents using time series analysis [40]–[42]. From the big data acquired as mentioned above, for a given date, X[t] denotes the frequency of car accidents occurring between the hours of t and t + 1.

Figure 1 illustrates the graphs of X[t] for the sampled dates. The daily data used in the figure were selected from 8 consecutive days between February 1 and February 8 in 2015. February 1, 2015 was a Sunday.

A police report on a traffic accident contains various information fields. As described in the section IV, in Korea, police reports contain 14 columns, including date, time, road type, surface condition, weather, and driver behavior. See the section IV for more details. We define a tentative descriptive model, M, for car accidents below.

\[ M := M(t; \Psi), \text{ where } \Psi = \Psi(\zeta_1, \zeta_2, \zeta_3, \ldots, \zeta_{14}). \] (1)

The parametric arguments \( \zeta_1, \zeta_2, \zeta_3, \ldots, \zeta_{14} \) denote 14 independent variables representing the 14 columns in the police report. The output of the model M at time t is the frequency depicted in the graphs in Figure 1, where the independent variables are considered as one global variable, \( \Psi \).

Typical factor analysis can show the significance of each variable through correlation estimation between the output and the variable of interest [43]. In this study, our interest is in utilizing big data to discover the hidden pattern, which requires time derivatives because the majority of the car accidents have a similar pattern, which may override other crucial elements. For example, most of the car accidents occur on straight roads with paved surfaces when the weather is sunny. However, it makes no sense to say that straight and paved roads in sunny weather are the main cause of car accidents due to the high frequency of accidents in these conditions. Thus, we attempt to investigate the dynamic changes in the influence of factors through derivatives. In this manner, we can reveal the hidden patterns in the causality of traffic accidents.

The multivariate Taylor series expansion of the model M is as presented Eq.2 below.

\[ M(t; \Psi) = M(t_0) + \frac{dM}{dt}(t_0) + \frac{d^2M}{dt^2}(t_0) + R(t_0). \] (2)

where \( |t - t_0| < \epsilon \), d, d^2 M, and R denote the first and second order derivatives, and remainder terms, respectively. By applying the chain rule [46], the time derivate terms are rewritten as Eq.3 and Eq.4 for the first and the second order derivatives, respectively:

\[ \frac{dM}{dt} = \frac{dM}{d\Psi} \frac{d\Psi}{dt} \]
\[ = \frac{dM}{d\Psi} \prod_{i=1}^{14} \frac{\partial\zeta_i}{\partial t} \] (3)

\[ \frac{d^2M}{dt^2} = \frac{d^2M}{d^2\Psi} \frac{d^2\Psi}{dt^2} \]
\[ = \frac{d^2M}{d^2\Psi} \prod_{j=1}^{14} \prod_{i=1}^{14} \frac{\partial^2\zeta_i}{\partial t^2} \] (4)

Here, the variable \( \zeta_i \) is defined as above. The frequency is the count represented by an integer value, which is not appropriate for application in derivative involving digitization error. To make matters worse, bias can occur. Consider the following two cases for a frequency function f(t) defined in Eq.5.

\[ \text{Case 1: } f(0) = 1, \quad f(1) = 2 \]
\[ \text{Case 2: } f(0) = 101, \quad f(1) = 102 \] (5)
In both cases, the derivatives are same as
\[
\frac{df}{dt} = 1. 
\]
In our analysis of traffic accidents, case 2 is more important than case 1. The term \(dM/d\Psi\) in Eq.3 is necessary in order to suppress this type of bias. The following paragraph explains the definition of \(\Psi\) in our model \(M\).

We constructed the model using the theory of conservation law on estimating velocity which states that, for a conserved physical quantity \(\rho\), the dynamic system (called conservation law) leads to the following Eq.6.

\[
\frac{d\rho}{dt} = \nabla \cdot F(\rho),
\]
where \(\nabla\cdot\) is the divergence operator, and flux function \(F\) is convex.

For simplicity of the model, if we set \(F(\rho) = \rho v\) then the above equation turns into Eq.7.

\[
\frac{d\rho}{dt} = \nabla \cdot (\rho v) = v \cdot (\nabla \cdot \rho),
\]
where the variable \(v\) should satisfy \(\nabla \cdot v = 0\) by the law of conservation of mass.

We assume that the frequency of traffic accidents \(f\) is a physical quantity representing the mass as in the mass of gas in air dynamics [44]. From the above equations, the time derivative of this frequency \(\frac{df}{dt}\) can be approximated by velocity. In dynamics, the momentum, \(m\), is a particularly important quantity formulated by

\[
m := \rho \cdot v, \quad (8)
\]
where \(v\) denotes the velocity of the entity having mass \(\rho\). The time derivative of the momentum is the impulse denoted by

\[
J := \frac{dm}{dt}. \quad (9)
\]
In this study, we used the temporal impulse to analyze the time series of the traffic accidents rather than using the raw derivatives \(\frac{df}{dt}, \frac{d^2f}{dt^2}\), which was useful in suppressing the bias from rare case data and in achieving the noise-stabilizing numerical derivative property.

**IV. DATA DESCRIPTIONS**

The data used in this study was obtained from the Road Traffic Authority in South Korea, after being pre-processed to avoid violation of the Personal Information Protection Act and to protect the deployment strategy of the police department, personal information, and police station information. The data consisted of police report data that had been integrated in Excel format.

Table 1 presents the processed data consists of 14 columns: date, time, type of accident, detailed type of accident, shape of road, detailed shape of road, type of road, detailed type of road, surface condition of road, weather, type of behavior 1, detailed type of behavior 1, type of behavior 2, and detailed type of behavior 2.

| Table 1. List of data columns in the police report for car accidents in Korea. The count of columns is 14. |
|---|---|
| 1. Date | 2. Time |
| 3. Type of accident | 4. Detailed type of accident |
| 5. Shape of road | 6. Detailed shape of road |
| 7. Type of road | 8. Detailed type of road |
| 9. Surface condition of road | 10. Weather |
| 11. Type of behavior 1 | 12. Detailed type of behavior 1 |
| 13. Type of behavior 2 | 14. Detailed type of behavior 2 |

**TABLE 2. Specification of the car crash data in South Korea. 491,545 counts from 2015 to 2017.**

- **Begin date**: 2015.01.01
- **End date**: 2017.12.31
- **Area**: South Korea
- **Total count**: 491,545
- **Data source**: Korean National Police

**FIGURE 2.** Total frequency graph of police traffic accident reports by hour of day during 2015 – 2017 in Korea. X-axis is hour of day and Y-axis is frequency of accidents. This plot shows two peaks occurring at 09:00 AM and at 07:00 PM.
TABLE 3. Groups of time series used in this study. All time series comprised 24 time steps from hour 0 to hour 23. Each time step indicated the cumulative count of traffic accidents that occurred within the next 1 hour.

| Time series NAME       | Frequency of traffic accidents at hour of day. 24 time steps. |
|------------------------|------------------------------------------------------------|
| TS_24HOUR_YEAR         | Yearly total. 3 time series for years 2015, 2016, and 2017. |
| TS_24HOUR_MONTH        | Monthly total. 12 time series for months (1, 2, 3, ... 12) |
| TS_24HOUR_WEEK         | Weekly total. 7 time series for days (M, T, W, T, F, S, and S). |
| TS_24HOUR_WEATHER      | Weather related. 5 time series: sunny, cloudy, rain, fog, and frost. |
| TS_24HOUR_ROADTYPE     | Road type. 7 time series: crosswalk, intersection, on overpass, on bridge, straight road, parking garage, in tunnel. |
| TS_24HOUR_SURFACE      | Road surface condition. 4 time series: dry, frost, snow, wet. |
| TS_24HOUR_A_TYPE       | Accident type. 6 time series: head-on collision, parking, driving, rear-end collision, side collision, reversing. |
| TS_24HOUR_BEHAV1       | Behavior of assails. 8 time series: turn, U-turn, passing, parking, go straight, lane change, starting, reversing. |
| TS_24HOUR_BEHAV2       | Behavior of victim. 8 time series: same as TS_24HOUR_BEHAV1 |

In the section VI, Analysis, we perform factor analysis of $\Psi$ in terms of $\zeta$ to determine the significant correlative factors of the traffic accidents.

V. METHODS

For time series data defined as discrete time steps $f(t_n)$, where $\{t_1, t_2, \ldots, t_N\}$ were uniformly distributed real numbers representing time values in an interval, we considered it a discretization of the continuous and smooth function $f(t)$, where $t_1 \leq t \leq t_N$. A numerical analysis proved the time derivative $\frac{df}{dt}$ could be approximated by finite difference methods. For a smooth function $f$, the Taylor series expansions at $x+h$ and at $x-h$ around $x$, are shown below.

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f^{(3)}(x)$$  (11)

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f^{(3)}(x)$$  (12)

Subtraction of Eq.12 from Eq.11 leads to Eq.13 below.

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{h^3}{6}f^{(3)}(\xi)$$  (13)

where $\xi$ is located in $[x-h, x+h]$. This implies that if the function $f$ is smooth, that is $|f^{(3)}(\xi)| < Ch^3$, then

$$\frac{|f(x+h) - f(x-h)|}{2h} < Ch^2$$  (14)

Similarly, we used the second order finite difference by adding the two Taylor series expansion formulas above.

$$f(x+h) + f(x-h) = 2f(x) + \frac{h^2}{2}f''(x) + \frac{h^4}{24}f^{(4)}(\xi)$$  (15)

This can be written to show the finite difference scheme for the second order derivative:

$$\frac{|f(x+h) + f(x-h) - 2f(x)|}{h^2} < Ch^2$$  (16)

Denoting frequency of traffic accidents as $\phi$, we introduced a new variable for representing (pseudo) mass $\rho = e^{1-\phi}$. Our reasoning assumed a compressible gas in a closed tube, where collisions between gas particles depend upon the density of gas. Therefore, the mass of traffic accidents is proportionally related to the inverse of frequency $\phi$. With the definition of $\rho$, we formulated the momentum of traffic accidents as

$$m = \rho \frac{d\phi}{dt}$$  (17)

The term $\frac{d\phi}{dt}$ in Eq.17 is evaluated by applying Eq.14 with the periodic boundary condition:

$$\frac{d\phi}{dt}(x) = \frac{\phi(x+1) - \phi(x-1)}{2} \text{for } \phi(24) = \phi(0).$$

Considering Newtonian physics and following the definitions of temporal impulse and the time derivative of momentum, we constructed the temporal impulse of a traffic accident as $\frac{dm}{dt}$. The two variables, momentum and temporal impulse, required time derivation operations. We adopted the central differencing scheme finite difference method described above to evaluate the derived terms.

The police report data on traffic accidents in Korea is itemized by 14 factors $\Psi = \Psi(\zeta_1, \ldots, \zeta_{14})$. Some are duplicative. In this paper, we generated 8 different groups of time series $\Psi = \Psi(\zeta_1, \ldots, \zeta_8)$ and described them in Table 3. To convert the records to time series, the R programming language was used [45].

VI. ANALYSIS

As mentioned in the section I, the objective of this study is to introduce and describe a new, quantified variable, namely the “temporal impulse” of traffic accidents. For example, consider the time series TS_24HOUR_YEAR (as seen in Table 3). Figure 3 shows the four output graphs for TS_24HOUR_YEAR. They indicate frequency $\phi$, pseudo-mass $\rho$, temporal impulse $i$, and momentum $m$ starting from the top left corner in a clockwise direction.
The frequency $\varphi$ was obtained via filtered counting. For a given period of time (a year in this case), for a selected hour of day (from 0 to 23), $s$, we traverse and read the big data to count the accidents that occurred between the hour $s$ and the hour $s + 1$.

The pseudo mass $\rho$ is estimated by our formula $e^{1-\varphi}$. The relation is not established from actual data, since there is no available data in police reports how many cars are on the road when a car accident occurs, but it is established from simulation. The momentum $m$ is graphed at the bottom left corner as $m = \rho \frac{d\varphi}{dt}$. As seen in Figure 3, the peaks coincide with those of the graph of frequencies $\varphi$. However, there is a one-hour time lag between the peaks of frequency and momentum. There are also deep valleys after the peaks.

The temporal impulse $i$ is shown at the bottom right corner. The traditional definition is used as seen in Eq.18.

$$i := \frac{dm}{dt}$$

In the graph, we can observe two strong and several weak temporal impulses between the two peaks.

In Figure 3, the three curves are obtained from the three years (2015, 2016, and 2017). There is essentially no discrepancy between the frequencies of those years indicating that traffic accidents are constantly repeating in terms of time of day. We examined the consistency of the frequency of traffic accidents on a yearly basis. We further scaled down our analysis to the monthly level. Figure 4 shows 12 curves, with each one representing a month. The darkest color represents January and the lightest color represents December.

There were two observable points: frequency dependency and temporal impulse dependency on the month. First, at the top of Figure 4, the graphs indicate that the frequency of traffic accidents depended on the month. Later months of a year (lighter colored curves) had greater frequencies than the earlier months (darker colored curves). Second, as can be seen at the bottom of the figure, morning traffic temporal impulses (occurring at the hour of 8) were common throughout the months, whereas the late afternoon traffic temporal impulses varied both in magnitude and hour.

We inserted the raw table values of the temporal impulse curves in Figure 5. The values exceeding 0.1 are red and those exceeding 0.05 are colored blue. At hour 8, all the temporal impulse values are greater than 0.05, indicating a high temporal impulse. On the contrary, for the afternoon traffic hour (at approximately hour 18), there is no consistency in the peaks of temporal impulses. During winter (October to February), the peak of temporal impulse is at hour 18, and during summer (May to July), the peak is at hour 17. In March and September, the peak of temporal impulses appears at hour 19. This indicates that after-office hour traffic patterns vary seasonally. We are, however, unable to explain this seasonal dependency, and therefore, for now, we only suggest the existence of seasonal dependency of afternoon traffic accidents.

Having examined the frequency of traffic accidents on yearly and monthly bases, we scaled down further to the weekly level. The layout of Figure 6 is identical to that of Figure 3 and Figure 4, except that the graphs indicate weekly data.

The most interesting observation in this graph is the strong shape discrepancy between weekdays and weekends. In this figure, we note two interesting points. One is the significantly smaller count of accidents on a Sunday, which is 75% of that of a Friday (see Figure 7). The other is the difference in the shapes and positions of peak points between weekdays and weekends. The temporal impulse graph (bottom) indicates an obvious first peak around hour 8 for weekdays (Monday through Friday) but no significant peaks for Saturday and Sunday. In contrast, the second peak around hour 18 occurs commonly for all days.

Yearly-, monthly-, and weekly-based analyses have been described. Now, we will present a structural analysis with the following five factors: weather, road type, road surface, accident type, and driver behavior.

First, we examined environmental factors such as weather, road type, and road surface condition. A common property of environmental factors is a huge bias of frequency in elements. For example, in terms of weather, most accidents occurred on sunny days; in terms of road surface condition, the majority of accidents occurred in dry conditions; in terms of road type, intersections and single lane roads commonly experienced accidents. As seen at the top of Figure 8, the weather related temporal impulse graph shows the huge impact of ‘fog’ in the early morning. This implies that the presence of fog imposes an extensive temporal impulse on traffic accidents. ‘Snow’ has a relatively stronger temporal impulse in the morning peak time. Except for these two behaviors, the behaviors are regular. As seen in the middle of Figure 8, the road surface
related temporal impulse graph shows two significant impacts of ‘frost/ice’ and ‘snow’ in the early morning. Except for these two morning bursts of temporal impulses, the behaviors are regular.

As seen at the bottom of Figure 9, there are no huge temporal impulses from road type; however, there are non-regular patterns from ‘parking lot’ and ‘tunnel’. The morning time temporal impulse from parking lot is two hours later than the regular one. The temporal impulse from the tunnel differs significantly from those for the other factors.

Hitherto, we examined temporal (yearly, seasonal, day of week) and environmental (weather, road surface and type) factors. Finally, we investigated human behavioral factors: accident type and driver behavior.

First, we described the result of the analysis on accident types. There are six different types of traffic accidents. The majority of accidents are broadside collisions, as seen in the top of Figure 10. The temporal impulse analysis shows the irregularity of collisions while reversing. However, we believe that the scarcity of data for ‘reversing’ leads to numerical errors. The rest of the factors have a regular temporal impulse pattern.

As seen in the two graphs in Figures 10-11, the top three behaviors at the moment of a car accident are ‘straight’, 'reversing' and 'parking lot'.
FIGURE 8. Frequency and temporal impulse of traffic accidents due to environmental factors. Three environmental factors are considered including weather, road surface condition, and road type. X-axis for both of the left and right panels denotes hour of day. Y-axis indicates frequency (f) on the left and impulse (i) on the right.

‘turn’, and ‘change lane’ (assailant) and ‘straight’, ‘parking’, and ‘turn’ (victim). The temporal impulse time graphs of driver 1 (assailant) are regular except for the pink line (‘reversing’). We believe that this irregularity is due to the scarcity of data (similarly for collisions while reversing). Barring these, the rest are regular.

The temporal impulse time graphs of driver 2 (victim) are irregular. To show the irregularity of the graphs, Figure 12 is presented. In this figure, the shapes of the closed curves differ between the left and the right. The closed curves at the left have wider and smoother interiors while exhibiting two significant peaks at hours 9 and 19. In contrast, the closed curves on the right have narrow and discontinuous shapes. We believe that this indicates the randomness of becoming a victim of a traffic accident.

A. VALIDATION

In this section, we present various analytical outputs via graphs, as depicted in Figure 3 (year), Figure 4 (month), Figure 7 (Week), Figure 8 (environmental factors including weather, road condition, and road type), Figure 9 (accident type), Figure 10 (driver’s behavior), and Figure 11 (victim’s behavior). On each graph, multiple curves are plotted to determine discrepancy or similarity. Furthermore, we performed the Kolmogorov-Smirnov test to guarantee the statistical discrepancies between data using the two measures, D-value and p-value. D-value = 0 means that the discrepancy was zero, whereas p-value = 1 implies the D-value was 100% guaranteed.
For clarity of presentation, we gathered all the test results in this subsection as detailed in Tables 4–8. In Table 4, KS (Kolmogorov-Smirnov) tests on yearly data are presented. There are no discrepancies observed between them, i.e., temporal impulses for car accidents repeated on a yearly basis.

In Table 5, KS tests on weekly data are presented. The table shows strong pairwise discrepancies ($D > 0.3$) for (Mon, Sun), (Tue, Sun), (Wed, Sun), (Thu, Sun), (Fri, Sun), and (Sat, Sun) while weak pairwise discrepancies ($D > 0.2$) for (Mon, Sat), (Tue, Fri), (Tue, Sat), (Wed, Fri), (Wed, Sat), and (Thu, Fri). The graphs of [Mon, Tue, Wed, Thu] are all similar each other: $D < 0.125$ and $p > 0.9$. Friday has no discrepancies with Monday and Saturday, weak discrepancies with Tuesday, Wednesday, and Thursday, and strong discrepancy with Sunday. The graph of Sunday shows strong discrepancies with all other day’s. The greatest discrepancy is between Friday and Sunday.

In Table 6, the impulse graph of fog is seen to be distinct from those of all other factors and that of snow is also observed to be different. Contrarily, graphs of sunny, rainy,
TABLE 8. Kolmogorov-Smirnov test for road types. The order of the headers is overpass (O/P), bridge, single lane (S/L), parking, tunnel, intersection (INTX), and cross-section. The maximum of the D-values and the minimum of the p-values are colored in red. The D-values greater than 0.2 and the p-values smaller than 0.7 are colored in green.

|                | D-value | p-value |
|----------------|---------|---------|
| O/P vs Bridge  | 0.081   | 0.666   |
| O/P vs S/L     | 0.291   | 0.262   |
| O/P vs Park    | 0.125   | 0.441   |
| O/P vs TUNN    | 0.208   | 0.686   |
| O/P vs INTX    | 0.125   | 0.994   |
| O/P vs CROSS   | 0.116   | 0.902   |
| Bridge vs S/L  | 0.208   | 0.686   |
| Bridge vs Park | 0.208   | 0.674   |
| Bridge vs TUNN | 0.208   | 0.666   |
| Bridge vs INTX | 0.208   | 0.686   |
| Bridge vs CROSS| 0.125   | 0.994   |
| S/L vs Park    | 0.291   | 0.239   |
| S/L vs TUNN    | 0.250   | 0.449   |
| S/L vs INTX    | 0.208   | 0.686   |
| S/L vs CROSS   | 0.333   | 0.139   |
| Park vs TUNN   | 0.125   | 0.992   |
| Park vs INTX   | 0.208   | 0.674   |
| Park vs CROSS  | 0.125   | 0.992   |
| TUNN vs INTX   | 0.125   | 0.994   |
| TUNN vs CROSS  | 0.125   | 0.994   |
| INTX vs CROSS  | 0.125   | 0.994   |

and cloudy factors are similar. The temporal pattern of car accidents can be expected to be different between foggy and snowy days. The maximum discrepancy was occurred between cloudy and foggy days.

In Table 7, discrepancy tests of the graphs produced by road surface conditions are presented. Ice/frost conditions on road surfaces behaved differently with dry. However, except ice/frost condition, road surface conditions do not affect the temporal impulse of car accidents. We can claim road surface conditions are the weakest factor since the maximum D-value is 0.25 which corresponds to the smallest compared with other factor groups except ‘year’.

In Table 8, the different road types are considered. S/L (single lane) is distinguishable from all the other types including O/P (overpass), Bridge, Park (parking lot), TUNN (tunnel), INTX (intersection), and CROSS (cross). The KS tests show that O/P, INTX, and CROSS are similar.

VII. DISCUSSION AND CONCLUSION

Intractable traffic accidents cause major damage to quality of life in cities and create significant economic loss due to traffic congestions. With the development of ITS and the assistance of ubiquitous sensors installed in the city, several cities have transformed into smart cities with improved efficiency and safety. Reducing traffic accidents is one of the most effective ways of increasing safety in and further advancing the smart cities. To this end, understanding the causes of traffic accidents is critical for the effective design of transportation systems.

With the aim of understanding the influence of different weather and drivers’ conditions and their varying influences on traffic accidents, this paper analyzed the temporal correlation of car accidents with various categorical factors. By defining a new physical quantity, temporal impulse, we measured the probability of car accidents under given environmental factors and drivers’ statuses. As a newly introduced and measurable quantity, the temporal impulse processed the traffic accident frequency based on a mathematical model and identified the hidden patterns of traffic accidents.

The temporal impulse mathematically verified the reliable patterns that could not be identified by traditional statistical analysis. It is well-known that bad weather, such as fog and snow, is strongly associated with traffic accidents. The existing statistical analyses can only identify the positive relationships and cannot confidently confirm the key factors of traffic accidents. The temporal impulse resolved the aforementioned problem by mathematically verifying the influence of varying factors, thus confirming that fog has the greatest effect on traffic accidents. The proposed quantity, temporal impulse, can therefore help policy makers establish sophisticated measures to realize smart cities by applying urban computing technology, which is not based on past experiences or personal intuition, but rather on mathematical verification. In particular, the obtained result can be used by policy makers for preventing and avoiding traffic accidents during foggy weather. The temporal impulse can also be used for identifying hidden and unexpected patterns in big data in various fields beyond traditional statistical analysis.

Finally, there are two major limitations in this study that can be addressed in future research. First, status of this study is early stage. This study was focused on theoretical perspective of mathematical formula proposed. However, it is lack of applicational development. For the next research stage, construction of mathematical prediction model will be considered by using machine learning paradigm. The physical quantity designed in this study, temporal impulse, will be key component of feature vector for training dataset. Second, limitation of data acquisition is obvious. This study used police reports for the three years in Korea. Privacy information act was major obstacle to collect data. With the goal of development of traffic accident prediction model, expansion of data size is necessary.

REFERENCES

[1] Z. Yu, C. Licia, W. Ouri, and Y. Hai, “Urban computing: Concepts, methodologies, and applications,” ACM Trans. Intel1. Syst. Technol., vol. 5, no. 3, pp. 1–55, Sep. 2014.
[2] F. Salim and U. Haque, “Urban computing in the wild: A survey on large scale participation and citizen engagement with ubiquitous computing, cyber physical systems, and Internet of Things,” Int. J. Hum.-Comput. Stud., vol. 81, pp. 31–48, Sep. 2015.
[3] A. Altmare, “Trajectory pattern mining for urban computing in the cloud,” IEEE Trans. Parallel Distrib. Stud., vol. 28, no. 2, pp. 1–10, May 2016.
[4] L. Qi, “Research on intelligent transportation system technologies and applications,” in Proc. Workshop Power Electron. Intell. Transp. Syst., Aug. 2008, pp. 529–531.
[5] F. J. Martinez, C.-K. Toh, J.-C. Cano, C. T. Calafate, and P. Manzoni, “Emergency services in future intelligent transportation systems based on vehicular communication networks,” IEEE Intell. Transp. Syst. Mag., vol. 2, no. 2, pp. 6–20, Oct. 2010.
[6] K. N. Qureshi and A. H. Abdullah, “A survey on intelligent transportation systems,” Middle-East J. Sci. Res., vol. 15, no. 5, pp. 629–642, Jan. 2013.
N. Doğru and A. Subaşı, “Comparison of clustering techniques for traffic forecasting: Where we are and where we’re going,” Transp. Res. C, Emerg. Technol., vol. 43, pp. 3–19, Jun. 2014.

A. Raikwar, R. R. Sadawarte, R. G. More, R. S. Gunjal, P. N. Mahalle, and P. N. Railkar, “Long-term and short-term traffic forecasting using holt-winters method: A comparability approach with comparable data in multiple seasons,” Int. J. Synth. Emotions, vol. 8, no. 2, pp. 38–50, 2017.

A. Irandtalab and A. Khatkat, “Comparison of four statistical and machine learning methods for crash severity prediction,” Accident Anal. Prevention, vol. 108, pp. 27–36, Nov. 2017.

N. Doğru and A. Subaşı, “Comparison of clustering techniques for traffic accident detection,” Turkish J. Electr. Eng. Comput. Sci., vol. 23, pp. 2124–2137, Jan. 2015.

N. Doğru and A. Subaşı, “Traffic accident detection using random forest classifier,” in Proc. 15th Learn. Technol. Conf. (L&T), Feb. 2018, pp. 40–45.

D. Zeng, J. Xu, and G. Xu, “Data fusion for traffic incident detector using D-S evidence theory with probabilistic SVMs,” J. Comput., vol. 3, no. 10, pp. 36–43, Oct. 2008.

J. Xiao, “SVM and KNN ensemble learning for traffic incident detection,” Phys. A, Stat. Mech. Appl., vol. 517, no. 1, pp. 29–35, Mar. 2019.

J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, “Data-driven intelligent transportation systems: A survey,” IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

R. Yu and M. Abdel-Aty, “Utilizing support vector machine in real-time crash risk evaluation,” Accident Anal. Prevention, vol. 51, pp. 252–259, Mar. 2013.

J.-B. Sheu, “A sequential detection approach to real-time freeway incident detection and characterization,” Eur. J. Oper. Res., vol. 157, no. 2, pp. 471–485, Sep. 2004.

J. Lu, S. Chen, W. Wang, and H. van Zuylen, “A hybrid model of partial least squares and neural network for traffic incident detection,” Expert Syst. Appl., vol. 39, no. 5, pp. 4775–4784, Apr. 2012.

S. Yasmin, N. Eluru, L. Wang, and M. A. Abdel-Aty, “A joint framework for static and real-time crash risk analysis,” Anal. Methods Accident Res., vol. 18, pp. 45–56, Jan. 2018.

E. D’Andrea and F. Marcelloni, “Detection of traffic congestion and incidents from GPS trace analysis,” Expert Syst. Appl., vol. 73, pp. 43–56, May 2017.

J. Sun and J. Sun, “A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data,” Transp. Res. C, Emerg. Technol., vol. 54, pp. 176–186, May 2015.

M. Chong, A. Abraham, and M. Paprzycki, “Traffic accident data mining using machine learning paradigms,” in Proc. 4th Int. Conf. Intell. Syst. Design Appl. (ISIDA), Aug. 2014, pp. 415–420.

L. Wang, M. Abdel-Aty, J. Lee, and Q. Shi, “Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors,” Accident Anal. Prevention, vol. 122, pp. 378–384, Jan. 2019.

C. El Hatri and J. Bournhidi, “Fuzzy deep learning based urban traffic incident detection,” Cognit. Syst. Res., vol. 50, pp. 206–213, Aug. 2018.

A. Visvizi, M. D. Lytras, E. Damiani, and H. Mathkour, “Policy making for smart cities: Innovation and social inclusive economic growth for sustainability,” J. Sci. Technol. Policy Manage., vol. 9, no. 2, pp. 126–133, Jul. 2018.

A. Visvizi and M. D. Lytras, “Rescaling and refocusing smart cities research: From mega cities to smart villages,” J. Sci. Technol. Policy Manage., vol. 9, no. 2, pp. 134–145, Jul. 2018.

M. Lytras and A. Visvizi, “Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research,” Sustainability, vol. 10, no. 6, p. 1998, Jun. 2018.

M. Lytras, A. Visvizi, and A. Sarirete, “Clustering smart city services: Perceptions, expectations, responses,” Sustainability, vol. 11, no. 6, p. 1669, Mar. 2019.

A. Visvizi and M. D. Lytras, Smart Cities: Issues and Challenges. Mapping Political, Social and Economic Risks and Threats, Amsterdam, The Netherlands: Elsevier, 2019.

C. Benevolo, R. Damari, and B. D’Auria, “Smart mobility in smart city, Action taxonomy, ICT intensity and public benefit,” in Proc. Empowering Organizations Conf., Jan. 2016, pp. 13–28.