Imputation Model of the Link Travel Speed Data for Incident Detection System

Yong-Kul Ki, Yong-Ho Kim, Yong-Chan Kim
Road Traffic Authority
Wonju City, Republic of Korea
kiyongkul@koroad.or.kr

Abstract—**Travel speed is an important parameter in the measurement of road traffic. Korea Traffic Information System (KTIS) has been used for measuring link travel speeds. However, KTIS incurs missing data, such as those caused by construction, detector failures, communication failures, and other factors. This paper describes an imputation model that uses the multiple regression to accurately estimate average roadway link travel speeds for the incident detection algorithm and Intelligent Transportation System (ITS). This model predicts link travel speeds using a robust data imputation method based on available information for neighbor links and the adjacent time periods. A field test showed that the variance of the percent errors of link travel speeds was reduced when they were measured using the new model. Therefore, it can be concluded that the proposed model significantly improves the accuracy of travel speed measurement**

I. INTRODUCTION

Traffic incidents occur frequently on roadways, resulting in congestion and delay. One traffic incident management strategy is to disseminate accurate incident information to travelers, who can then make more informed travel decisions. Another approach is to actively redirect traffic in a road network to avoid congestion. In both approaches, accurate incident detection is required. Incident management system generally uses travel speed data for incident detection. Therefore, accurate travel speed data is required for incident detection algorithm.

Korea Traffic Information System (KTIS) can provide valuable information about link travel speeds on urban roads, but KTIS, the primary source of real-time and archived traffic data, commonly fails to report data back to Traffic Information Center (TIC). These failures have many causes, including detector failure, communications network failure, and data archival system failure, among other factors. Because of the nature of TIC, some level of missing data is unavoidable [1]. The methods addressed here will impute traffic data collected by the KTIS.

This paper addresses the problem of obtaining reliable travel speed estimates from KTIS data by presenting a data imputing algorithm. The paper begins with a description of the infrastructure system, and the imputation methods used by existing systems. Problems with the current systems are discussed, and a new method is presented that was developed to measure travel speed observations under real road conditions. A more extended evaluation of the proposed system and algorithms is then performed for arterial street segments in the cities of Incheon and Bucheon. Finally, the main conclusions of the evaluations are presented

II. RELATED WORK

Owing to the increasing traffic congestion in the major metropolitan areas, many organizations are choosing to provide travel information to the public. This information allows drivers to intelligently choose the times and routes for their trips, thereby improving the efficiency of their trips and reducing congestion and delay. Travel speeds are probably the most challenging data to provide, because of the complexities involved in generating accurate estimates [2-4]. Travel speed must be accurate to be useful; inaccurate travel information can be worse than no information at all.

Although KTIS provides valuable information about link travel speeds on urban streets, it sometimes fails to report data back to the TIC. Replacing or imputing missing values in travel speed data is important for many applications. This is especially true for real-time applications, such as the Automatic Incident Detection System and Advanced Traveler Information System.

Missing traffic data are typical in most traffic detection systems. A study in San Antonio, Texas found that missing data ranged from 5-25%, even though only 5-15% of this missing data comes from loop detector failure [5]. A study in Georgia found an average rate of missing data of 4-14% for GA 400 [6]. In a study done in Virginia, 25-30% of the single-loop detectors at intersections were found to be off-line at any given time [7].

Several methods for missing data treatment have already been used in transportation applications. These methods include conditional mean substitution, time-series models, and regression models [7]. These transportation studies have focused on replacing the missing values (occupancy, flow, or speed) with imputed values so as to construct a complete set of traffic data. Most of these studies applied single imputation methods which mainly impute the missing data from the means and distributions of the observable data set, while only a few used multiple imputation methods [8].

Ni, et al. developed a stochastic method based on Bayesian networks that is able to produce unbiased estimates and preserve the natural characteristics of the raw data [6]. Zhong, et al. developed two genetic algorithms, locally weighted regression and time delay neural networks, to estimate imputation accuracy for one-hour [9]. Chen, et al. compared the imputation accuracy of an advanced neural network model, a hybrid algorithm that integrates the historical and average and time series analysis, and various simple methods. The authors suggest that the time-delay neural network is less applicable than other methods [10].

We developed an incident detection algorithm and imputation model for missing values that is based on real-time information provided by KTIS. The imputation method for link travel speed measurement uses multiple regressions.

III. INCIDENT DECTION ALGORTHM

A. Korea Traffic Information System

As shown in Fig. 1, KTIS is an in-vehicle advanced traveler information system that operates in South Korea. It is designed to provide origin-destination shortest-time route guidance to a vehicle based on an on-board static (fixed) data-base of average network link travel times by time of day, combined with real-time information on traffic conditions provided by radio frequency communications with a Korea Traffic Information Center (KTIC).

Fig. 1. Photograph of the KTIC

B. An Incident Detection Algorithm

In 2019, a new algorithm is suggested to detect traffic incident using the link travel speed data of KTIC [11]. The proposed algorithm uses neural networks. With capabilities of learning, self-adaptation, and fault tolerance, the Artificial Neural Networks (ANNs) approach has demonstrated good performance in many pattern classification applications. In addition to the extracted features, a five-layered ANN model for incident detection was developed (Fig. 2). The proposed model showed improved performance in incident detection.

According to the investigation, an incident is likely to create congestion in the upstream road section and reduce flow in the downstream road section; this leads to a high velocity difference between two traffic states. The new ANN model used this feature for incident detection.

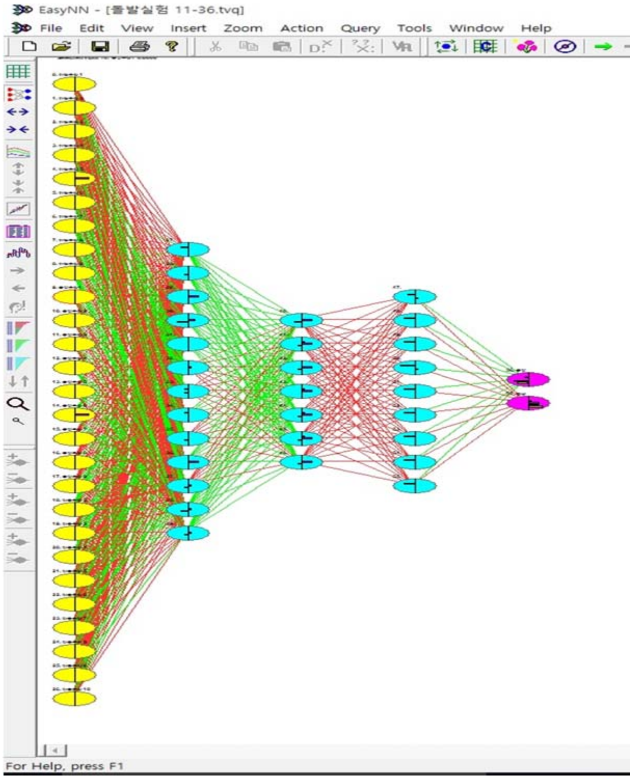

Fig. 2. Incident detection model using an ANN

IV. IMPUTATION ALGORITHM

Widely used traffic detectors for roadway monitoring or for any intelligent transportation system application often produce various patterns of missing data that consequently degrade the quality of the control operations. This inevitable encounter with the issue of missing data further complicates the challenging task of travel-time prediction, especially when only sparsely distributed detectors are available for the collection of real-time traffic conditions. In view of the quality of detectors in the existing market and their associated communication issues, it seems essential for any intelligent transportation system being considered for deployment to have a function that can effectively contend with missing data.

A. Applications of Data Imputation Methods in the Transportation Study

Imputation is the process of filling the holes in data that are caused by missing data values. In this regard, it has been found that the travel speeds of adjacent links and time periods are highly correlated. Therefore, if traffic data are missing for a lane, the parameters of the adjacent time periods and the neighboring lanes would provide data that can be used to estimate the missing parameters. In travel speed prediction, the following methods have been generally used to predict one parameter from another parameter. These methods are both easy to implement and fast to run.

B. Average of Adjacent Time (AAT) Periods

This method averages the values from the 15-min intervals before the missing value (equation 1). The application of this approach is limited because it can be used only when both the preceding periods do not have missing data.

$$\hat{S}_{i,j,t} = \frac{S_{i,j,t-1}+S_{i,j,t-2}+S_{i,j,t-3}}{3} \quad (1)$$

Where

$\hat{S}_{i,j,t}$ = estimated speed of detector i of station j at time t,

$S_{i,j,t}$ = actual speed of detector i of station j at time t,

$S_{i,j,t-1}$ = actual speed of detector i of station j at 5-min before time t,

$S_{i,j,t-2}$ = actual speed of detector i of station j at 10-min before time t,

$S_{i,j,t-3}$ = actual speed of detector i of station j at 15-min before time t,

C. *Average of Neighboring Links (ANL)*

The next method averages the values from the neighbor links if they are functional. Figure 3 shows data that were extracted from Stations A, B, and C. Imputation was performed on Station B, and Stations A and C were used for additional input data.

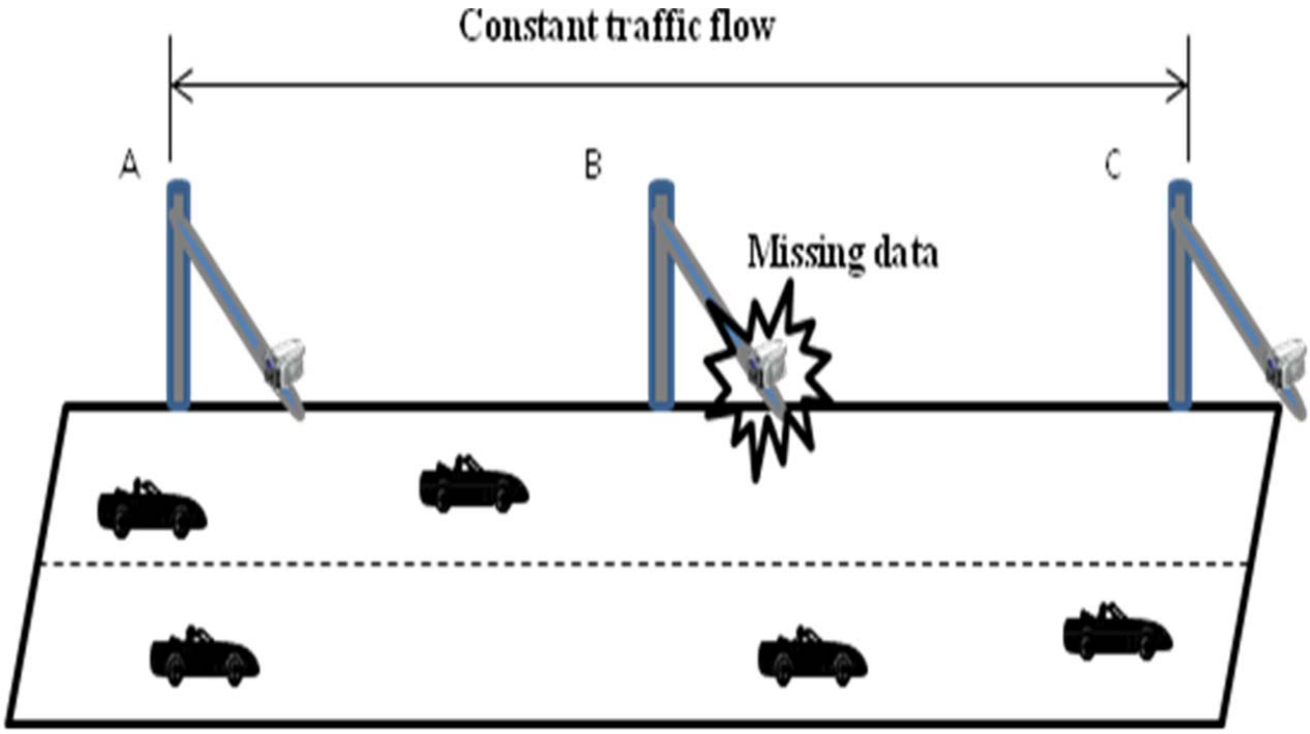

Fig. 3. Imputation method using average of neighboring links

D. *New Imputation Approaches for Travel Speed prediction*

After investigation, we found that the travel speeds of the neighbor links and the adjacent time periods are highly correlated. This high correlation means that multiple regression is a good way to predict one parameter from the other. The multiple regression model can be written as $V_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + N_t$ where V_t is the *t*th observation of the dependent variable, and $X_{1,t}, \cdots, X_{k,t}$ are the corresponding observations of the explanatory variables. The parameters $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ are fixed but unknown, and N_t is the unknown random error term.

In this research, we propose a multiple imputation model, called MI, to supplement a real-time travel speed prediction system developed for use in a sparsely distributed detection environment. The key feature of MI is its integration of the missing data imputation with the travel speed prediction, and its direct estimation of the missing travel speed based on available information of the neighbor links and the adjacent time periods. Independent data sets that were chosen from the KTIC on June 18 and August 6, during the morning peak period and the afternoon off-peak period, were used for regression analysis.

The independent variables used in the multiple regressions were the travel speeds of the upstream and downstream links of the missing value, and the travel speeds for the 15-min intervals before the missing value. In the regression analysis, the coefficient of determination (R^2) was 0.818, which is a reasonable value, and equation 2 was a regression function.

The travel speed of the downstream link of the missing value was not found to be statistically significant. This indicates that the travel speed of the downstream link of the missing value has no predictive ability with respect to travel speed in the neighbor link. Therefore, the downstream link variable was not used in developing the regression model.

$$V_t = 2.53 + 0.34 \times X_1 + 0.43 \times X_2 + 0.16 \times X_3 \quad (2)$$

where V_t is the t^{th} link travel speed

X_1 ; link travel speed before 5-min

X_2 ; average of the link travel speeds at 10-min and 15-min before the missing value

X_3 ; travel speed of the upstream neighbor link of the missing value

V. TEST AND EVALUATION

A. Test and Analysis Methods

We conducted validation tests to confirm the accuracy of the proposed model. Data were analyzed using a Microsoft Excel spreadsheet program developed by the research team. The spreadsheet program imported the traffic data and processed it before producing the output statistics. We also videotaped the test process at the same time for later verification of travel speeds, using the same equipment used by the Road Traffic Authority to measure link travel speeds in order to provide ground truth data.

To measure the effectiveness of the different imputation techniques, records from complete traffic data sets were artificially removed and saved. Then, using the imputation techniques, the removed data were imputed and compared to the ground truth. As shown in Fig. 4, traffic data were extracted from links A, B, and C, imputation was performed on link B, and links A and C were used for additional input data.

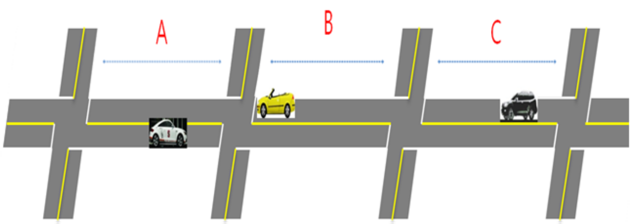

Fig. 4. Imputation method of link travel speed

We conducted validation tests to confirm the accuracy of the proposed model. When evaluating model performance, we used stochastic methods such as comparison of range, and Mean Absolute Percentage Error (MAPE). The variance of the percentage difference is a useful statistic for evaluating the deviation between the device and baseline data. Thus, we mainly used the variance of the percentage errors and MAPE to evaluate accuracy.

The MAPE is shown in Equation 3. It represents the average difference between the percentage of the observed value (in this case, the travel speed in a link) and the predicted value (in this case, the predicted travel speed in a link).

$$\text{MAPE} = \frac{1}{n}\sum_i^n \frac{|y_i - y_0|}{y_0} \times 100\% \quad (3)$$

Where y_i is the predicted value (i.e., travel speed in a link), and y_0 is the observed value (i.e., travel speed in a given link)

B. Experimental Test

We developed a travel speed prediction model using a new imputation algorithm (MI) to evaluate the algorithm's performance and we conducted field tests at two sites in the cities of Incheon and Bucheon. Travel speed data were collected from the KTIC on August 4 and August 11, during the morning peak period and the afternoon off-peak period.

As shown in Fig. 5, the following two study segments were selected for the experimental test: Southbound Central Park Street from the YMCA building to the Central library intersection, with a length of 187 m; Westbound National Highway 39 from the Hawoo hill to the Bucheon station, with a length of 172 m.

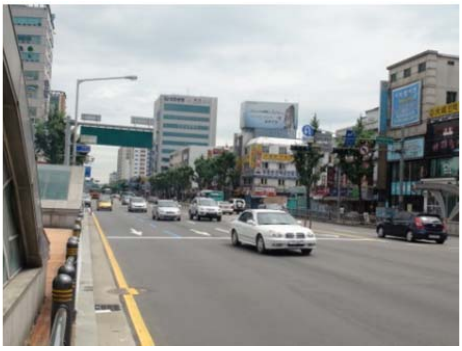

Fig. 5. Photograph of test sites

On August 4, we performed 2 h (08:00 ~ 09:00 and 13:00 ~ 14:00) of experimental testing on Central Park Street in Incheon and collected data. To test the performance of the imputation algorithms, artificial gaps were created in the testing data from the center link by removing data points from the original testing dataset. The test results showed a MAPE of

6.9%, which indicates that the proposed imputation algorithm works effectively.

On August 11, we conducted experimental tests on National Highway 39 from the Hawoo hill to the Bucheon station in Bucheon. In the field test, the new algorithm showed a MAPE of 5.9%. Overall, the new model using the new filtering algorithm significantly reduced the variance of percentage errors when measuring travel speeds over the test period. Figure 6 shows that the new filtering algorithm is very efficient compared with two conventional models (AAT and ANL models).

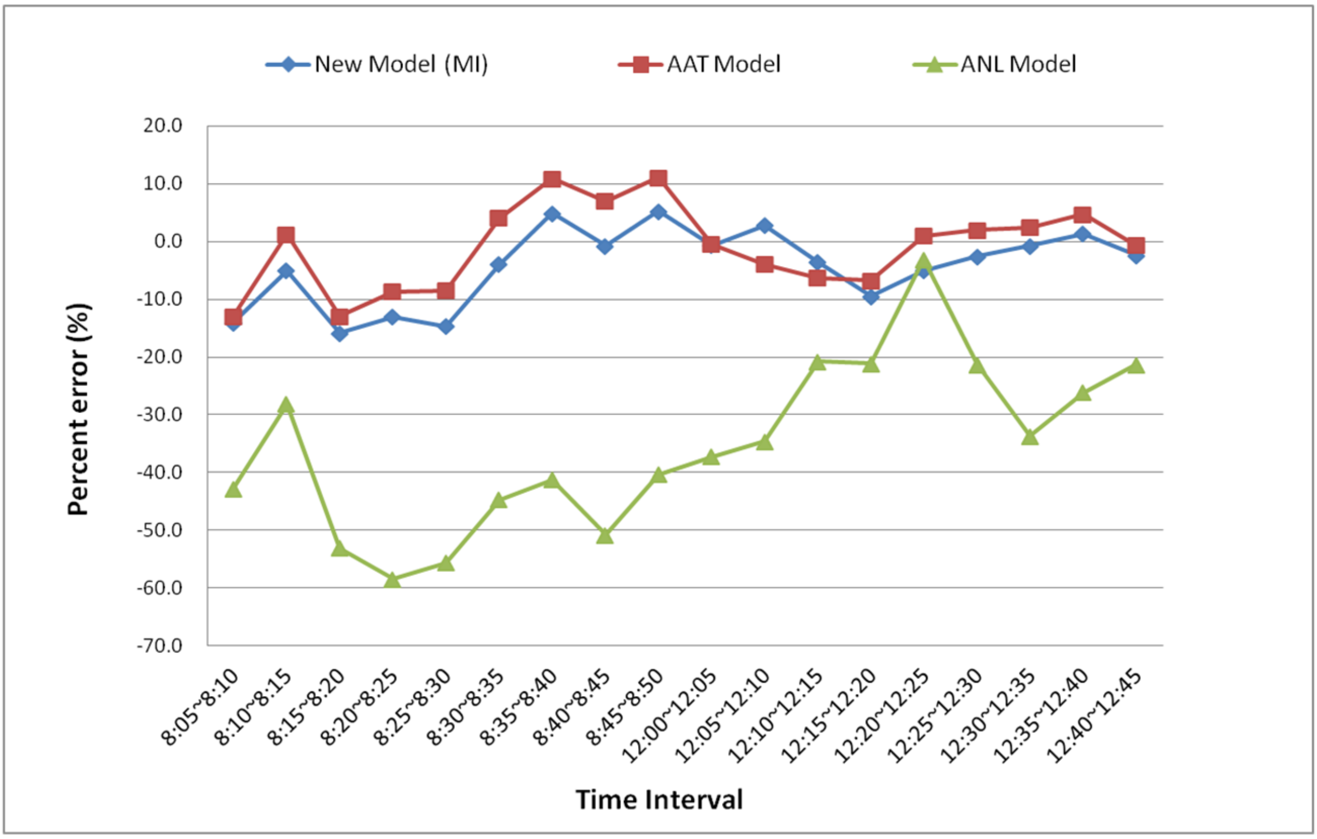

Fig. 6. Field test results for National Highway 39

VI. CONCLUSIONS

Predicting travel speeds on roads is a difficult task, but it provides crucial information in today's ITS applications. Loop detectors and AVI systems have been used in the past to obtain travel speed estimates, but the performance of these methods is not adequate. We have initiated the KTIS in South Korea, which automatically collects and provides average link travel speed information to drivers. The KTIS is installed in Seoul city, but there are still missing data caused by construction, detector failure, and other factors.

This paper presented an imputation model for link travel speed measurement in the traffic information center. The main concept underlying the proposed model is that of an imputing algorithm that uses multiple regression to improve the accuracy of travel speed measurement. The key feature of the new model is its integration of the missing data imputation with travel speed prediction, and its direct estimation of the missing travel speed using available information of the neighbor links and the adjacent time periods.

We performed an experimental test to evaluate the performance of the new model on urban roads in Incheon and Bucheon. The experiment was conducted on August 4 and 11. In the field test, the new model successfully imputed missing data and significantly reduced the variance of percentage errors as compared with two conventional models (AAT and ANL models). We can therefore conclude that the proposed model significantly improves the accuracy of link travel speed measurement. Because the proposed imputation algorithm was applied to urban arterial streets, future studies will be needed in order to explore its applicability to longer links and other types of environments.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education. (No. NRF-2017R1D1A1B03034065).", "A Study on Providing Incident Information for Automated Vehicle.".

REFERENCES

[1] Turner, S. M., W. L. Eisele, B. J. Gajewski, L. P. Algert, and R. J. Benz. ITS Data Archiving: Case Study Analyses of San Antonio TransGuide Data. Texas Transportation Institute, College Station, 1999.

[2] Pu, W., J. Lin, and L. Long, Real-Time Estimation of Urban Street Segment Travel Time Using Buses as Speed Probes. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2129*, TRB, National Research Council, Washington, D.C., 2009, pp. 81-89.

[3] Wang, J., N. Sou, and G.L. Chang, Travel Time Prediction. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2049*, TRB, National Research Council, Washington, D.C., 2008, pp. 81-91.

[4] Xu, H., and M. Barth, Travel Time Estimation Techniques for Traffic Information Systems Based on Inter vehicle Communications. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1944*, TRB, National Research Council, Washington, D.C., 2006, pp. 72-81

[5] Turner, S., L. Albert, B. Gajewski and W. Eisele. Archived Intelligent Transportation System Data Quality. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1719*, TRB, National Research Council, Washington, D.C., 2000, pp. 77-84.

[6] Ni, D., and J.D. Leonard Ⅱ. Markov Chain Monte Carlo Multiple Imputation using Bayesian Networks for Incomplete Intelligent Transportation System Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1935*, TRB, National Research Council, Washington, D.C., 2005, pp. 57-67.

[7] Nguyen, L., and W. T. Scherer. *Imputation Techniques to Account for Missing Data in Support of Intelligent Transportation Systems*. Center for Transportation Studies, university of Virginia Report No. UVACTS-13-0-78. Office of University Programs, Research and Special Programs Administration, U.S. Department of Transportation, Washington, D.C., 2003.

[8] Ni, D., J. Leonard Ⅱ, A. Guin, and C. Feng. Multiple Imputation Scheme for Overcoming the Missing Values and Variability Issues in ITS Data. *Journal of Transportation Engineering*, Vol. 95, 2000, pp. 144-154.

[9] Zhong, M., S. Sharma, and P. Lingras. Genetically Designed Models for Accurate Imputation of Missing Traffic Counts. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1879*, TRB, National Research Council, Washington, D.C., 2004, pp. 71-79.

[10] Chen, M., J. Xia, R. Liu. Developing a Strategy for Imputing Missing Traffic Volume Data. *Journal of the Transportation Research Forum*. Vol. 45, No. 3, 2006, pp. 59-77.

[11] Ki, Y. K., and W. T. Jeong, H. J. Kwon, M. R. Kim, “An Algorithm for Incident Detection Using Artificial Neural Networks.” *In Proc, 25th FRUCT*, 2019.