The Open Transportation Journal

Content list available at: https://opentransportationjournal.com

CrossMark

RESEARCH ARTICLE

Data Processing Techniques for Real-Time Traveler Information: Use of Dedicated Short-Range Communications Probes on Suburban Arterial

Jinhwan Jang[1,*]

[1]*Next Generation Infrastructure Research Center, Korea Institute of Civil Engineering and Building Technology, 283 Goyangdae-Ro, Illsanseo-Gu, Goyang-Si, Gyeonggi-Do10223, Republic of Korea*

Abstract:

Background:

As wireless communication technologies evolve, probe-based travel-time collection systems are becoming popular around the globe. However, two problems generally arise in probe-based systems: one is the outlier and the other is time lag. To resolve the problems, methods for outlier removal and travel-time prediction need to be applied.

Methods:

In this study, data processing methods for addressing the two issues are proposed. After investigating the characteristic of the travel times on the test section, the modified z-score was suggested for censoring outliers contained in probe travel times. To mitigate the time-lag phenomenon, a recurrent neural network, a class of deep learning where temporal sequence data are normally treated, was applied to predict travel times.

Results:

As a result of evaluation with ground-truth data obtained through test-car runs, the proposed methods showed enhanced performances with prediction errors lower than 13% on average compared to current practices.

Conclusion:

The suggested methods can make drivers to better arrange their trip schedules with real-time travel-time information with improved accuracy.

Keywords: Travel time, Probe, Outlier, Prediction, DSRC, Data processing techniques.

1. INTRODUCTION

Real-time travel-time information is an essential element of modern traffic management systems. It enables drivers to make detours in their routes to less congested ones or to adjust their trip schedules to avoid traffic jams. With the development of wireless communication technologies coupled with the increased market penetration of relevant on-board units, probe-based travel-time systems are being deployed worldwide due to their ability to directly obtain link travel times. In Korea, freeways and major arterials are equipped with Dedicated Short-Range Communications (DSRC) scanners (Fig. **1**). DSRC is originally used for the electronic toll collection

* Address correspondence to the author at the Next Generation Infrastructure Research Center, Korea Institute of Civil Engineering and Building Technology, 283 Goyangdae-Ro, Illsanseo-Gu, Goyang-Si, Gyeonggi-Do10223, Republic of Korea; E-mail: jhjang@kict.re.kr

systems in Korea and around 80% of vehicles are equipped with DSRC on-board units, indicating that probe sample size is not a critical issue.

However, the DSRC systems installed on suburban arterials on which many intersections and roadside stores exist inevitably generate substantial outliers due to intermediate stops at stores and/or gas stations, exit/entry maneuvers on the route, U-turns, driving illegally on the shoulder lane during congestion, and so on. Moreover, DSRC scanners cannot identify the direction of a detected probe, so if a probe makes a round trip on the roadway section, the probe is detected twice by the scanner. In this case, the second detection generates an abnormally long travel time and should be classified as an outlier. The severity of the outlier problem of the study site is exemplified in Fig. (**2**), where substantial outliers are observed. If these outliers are not properly treated, the travel-time

information could be useless.

Another issue to be addressed is a time-lag phenomenon. Probe travel times obtained using DSRC scanners inevitably include time lags equivalent to the travel times experienced by the probes on the segment because they, which normally called

arrival time-based travel times, are to be computed after probes complete the trip. However, the drivers who receive the arrival time-based travel-time information experience the departure time-based travel time (Fig. **3**). Therefore, applying a prediction technique is more emphasized in probe-based real-time systems than conventional detector-based systems.
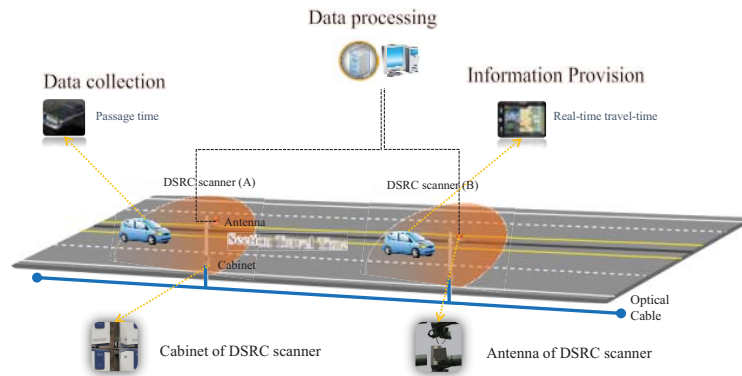
Fig. (1). Schematic of the DSRC traffic information system.
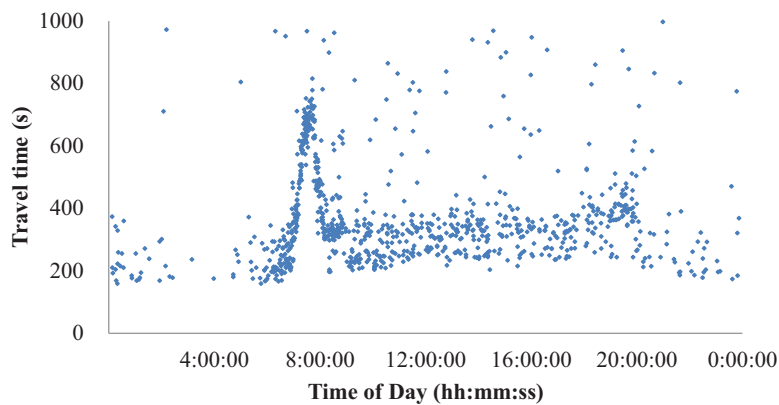
Fig. (2). The plot of raw probe travel times collected at the study site.
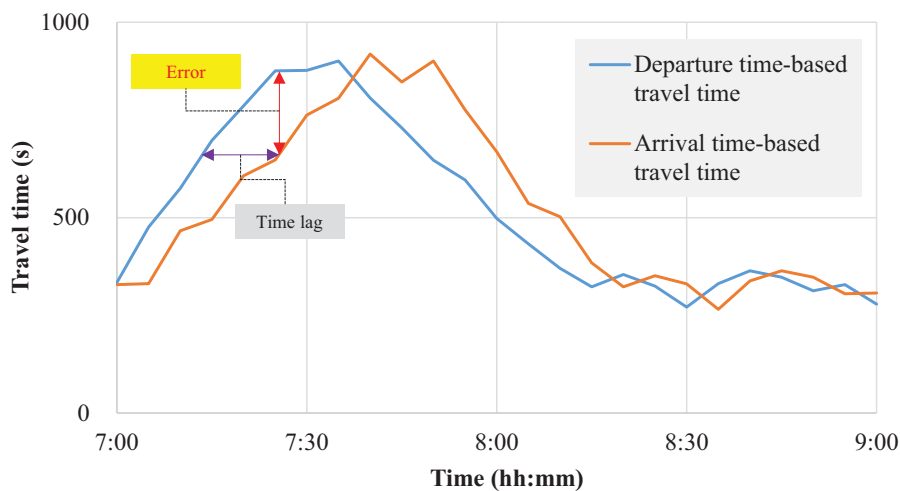
Fig. (3). The plot of aggregated probe travel times collected at the study site.

To resolve the above-mentioned problems, data processing methods for outlier removal and travel-time prediction were proposed in this study. After identifying the problems in the current practices, new methods were developed to resolve the problems. Investigation on the distribution of the probe travel times revealed that the current outlier removal technique based on the z-score is inappropriate for the test section. Therefore, an alternative method is suggested that takes the travel-time characteristic into account. Also, a widely recognized deep learning model is employed to enhance the travel-time prediction performance. The details will be described in the subsequent chapters.

2. LITERATURE REVIEW

Studies have been performed to deal with outlier problems in probe-based travel-time systems. The Southwest Research Institute [1] developed the TransGuide algorithm, which uses a simple validity window calculated from the observations in the previous aggregation interval. Clark *et al.* [2] suggested a statistical technique to filter outliers contained in probe travel times collected using devices that recognize vehicle registration numbers. Dion and Rakha [3] suggested an algorithm to estimate travel times with an assumption that they follow a lognormal distribution. Ma and Koutsopoulos [4] proposed a median filter that uses the median, instead of the mean, as a measure of location. Boxel *et al.* [5] developed a method that filters out outlying observations using a confidence interval predefined through Greenshield's traffic flow theory. Soroush and Bruce [6] invented an adaptive outlier detection algorithm that uses historical and current interval data gathered on signalized arterials to determine a validity window. Park and Kim [7] proposed a model to filter outliers in travel time data obtained using DSRC on the interrupted traffic flow section.

Several methods for predicting real-life travel times have been adopted for the past several decades. Among the most reputed methods are the Kalman Filter (KF), Artificial Neural Network (ANN), Time Series Analysis (TSA), and the k-Nearest Neighbor (k-NN) technique. KF, firstly proposed by R. E. Kalman [8], is an algorithm that utilizes observations in a time sequence and generates estimates of unknown variables [9]. ANN, created with inspiration from the biological neural networks, is a computing algorithm that learns from archived historical measurements to conduct a certain mission [10, 11]. Among various ANN algorithms, recurrent neural network algorithms that mostly treats with measurements in a time sequence are recently receiving much attention from developers on the globe [12 - 14]. TSA is a method for analyzing time-series data to forecast unknown variables based on the values observed earlier [15, 16]. k-NN, a sort of nonparametric technique mostly adopted for classification or regression, has widely exploited to forecast real-time travel times [17 - 26]. In k-NN methods, the input variables are the k closest training instances in the feature domain.

3. TEST SITE DESCRIPTION

Probe travel times collected from DSRC scanners installed on a signalized arterial in Pyeong Taek region, South Korea (Fig. **4**) were used for evaluating the data processing techniques proposed in this study. Traffic on the 4 km stretch of the highway with two lanes in each direction is controlled by six traffic signals. Due to intermediate stops by the signals, the average travel speed on the section under low volume situations is around 50 km/h even though drivers can drive up to 80 km/h by the regulation.

Individual probe travel times, as shown in Fig. (**5**), without outlying observations, exhibited a normal commuter corridor with apparent morning peak hours from 7:00 to 9:00 a.m. during weekdays. Travel times during peak hours are 3-4 times higher than the remaining hours, indicating that real-time traveler information is highly stressed during peak hours.

4. OUTLIER TREATMENT

4.1. Current Practice for Outlier Removal

Presently, an outlier removal technique based on the z-score is employed that uses the z-score of observations to censor outlying observations (refer to equations 1 - 3). Previous 30 probe samples, which is the minimum sample size for the application of the Gaussian distribution are used to compute the relevant statistics. However, the technique has been intuitively applied to the system since it was initially deployed; that is, no detailed analysis of the characteristics of probe travel times on the roadway has been performed to justify using it.
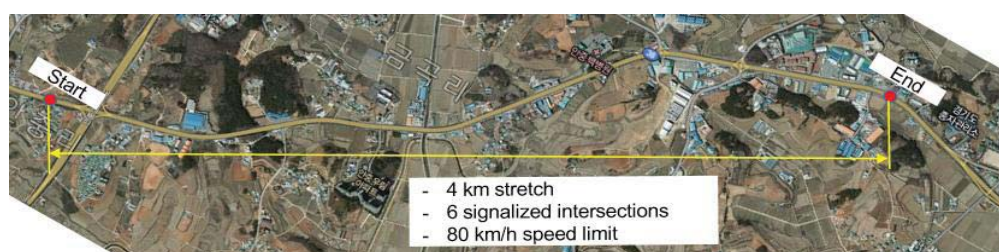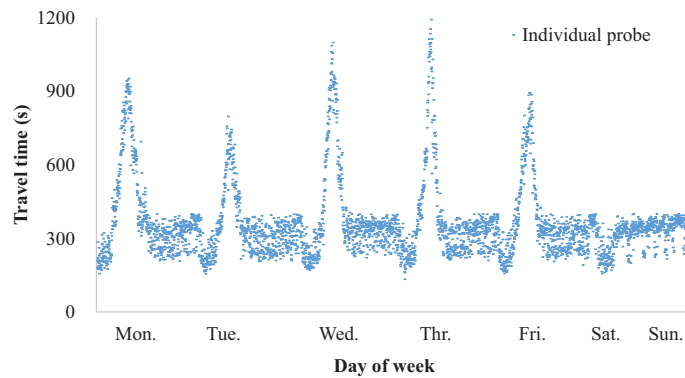

Fig. (4). Study site.

Fig. (5). Individual probe travel times on the test section.

$$tt_{valid} = \{tt_i | (\mu_t - z_{99.9\%}\sigma_t) \leq tt_i \leq (\mu_t + z_{99.9\%}\sigma_t)\} \quad \textbf{(1)}$$

$$\mu_t = mean\{tt_i, \cdots, tt_{i-29}\} \quad \textbf{(2)}$$

$$\sigma_t = standard\ deviation\{|tt_i - \mu_t|, \cdots, |tt_{i-29} - \mu_t|\} \quad \textbf{(3)}$$

where

tt_{valid} = valid travel times,

tt_i = travel time of *i*-th probe vehicle, and

$z_{99.9\%}$ = z-score at 99.9% confidence interval

4.2. Problems in Current Practice

As shown in Fig. (**6**), the current method showed poor performance with substantial outliers unfiltered. The premise behind the current method is that the probe travel times follow the normal distribution. To verify the assumption, a normality test on probe travel times on the test section was performed.

For the test, the Maximum Likelihood Estimation (MLE) for estimating parameters and the Kolmogorov-Smirnov (K-S) statistic at a 5% significance level for goodness-of-fit tests were employed. The MLE technique is generally recognized to be the best-unbiased estimator for the single distribution models that this study aims to investigate [27]. The K-S statistic, which fits a cumulative distribution to observations point by point, is considered to be the most conservative test compared to χ^2 and Anderson Darling tests; in other words, it has the lowest possibility of falsely rejecting a correct fit, and for that reason, it is frequently employed by many researchers [28].

The results of the fit tests are presented in Table **1** and the relevant travel-time histogram superimposed by the distribution curves is shown in Fig. (**7**). The results show that the travel times do not follow the normal distribution; rather, they were correctly fitted by the lognormal distribution. This indicates that the use of the current z-score cannot be theoretically viable.
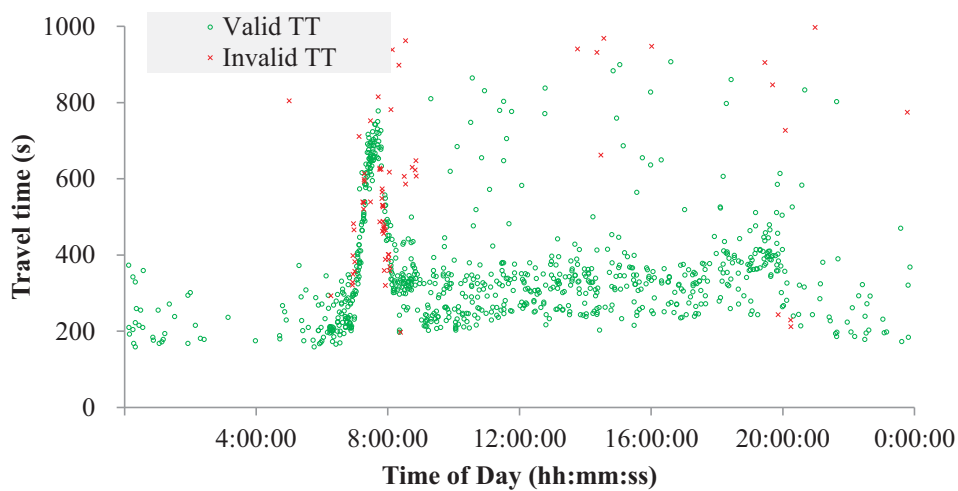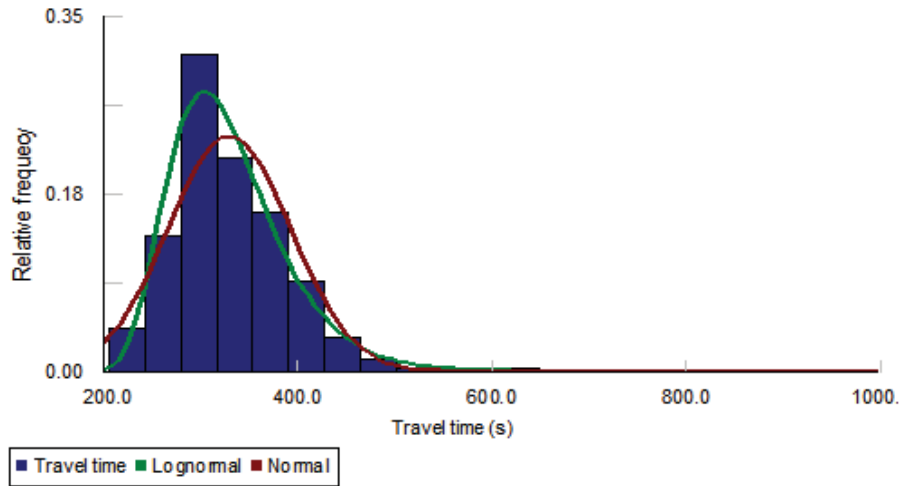

Fig. (6). Result of application of the current outlier filtering method.

Table 1. Goodness-of-fit tests of normal and lognormal distributions.

Distribution	Number of Samples	μ	σ	K-S Statistic	Critical K-S Statistic at sig.ᵃ = 0.05	p-value
Normal	1169	330	64	8.85e-2	3.93e-2	1.47e-8
Lognormal	1169	5.19	0.31	3.66e-2	3.93e-2	0.08

ᵃsig. = significance level.

Fig. (7). Histogram of probe travel times superimposed by distribution curves.

Instead, the z-score calculated from the logarithm (base e) of the travel times is justifiable. Taking these findings into consideration, a new outlier removal method is developed and will be described in detail in the section that follows.

4.3. New Technique for Outlier Removal

The proposed method expressed in Equations 4 to 7 is based on the modified z-score where the median instead of the mean as a measure of location is used. Also, the natural logarithm is applied to the raw travel times to take the characteristic of the travel-time distribution into account. The modified median filter is recognized to be resistant to abnormally deviated values compared to the standard z-score, because the median is less susceptible to the impact of abnormal values than is the mean [29, 30].

$$tt_{valid} = exp(ln(tt_i)_{valid}) \tag{4}$$

$$ln(tt_i)_{valid} = \{tt_i | (M_t - z_{mod,99.9\%} M_{AD}) \le ln(tt_i) \le (M_t + z_{mod,99.9\%} M_{AD})\} \tag{5}$$
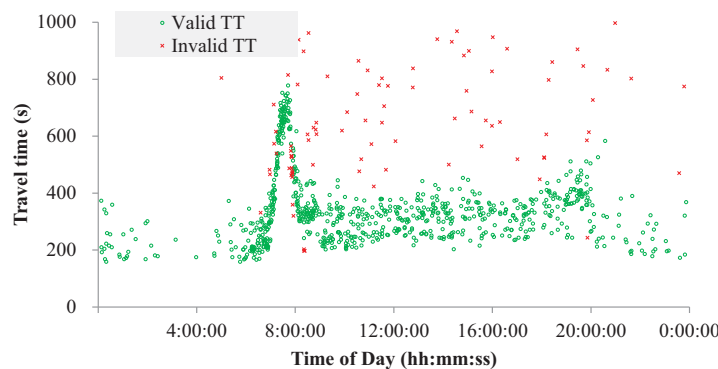
$$M_t = median\{ln(tt_i), \cdots, ln(tt_{i-29})\} \tag{6}$$

$$M_{AD} = median\{|ln(tt_i) - M_t|, \cdots, |ln(tt_{i-29}) - M_t|\} \tag{7}$$

where

tt_{valid} = valid travel times,

$ln(tt_i)$ = logarithm of travel time of i-th probe vehicle, and

$z_{mod\ 99.9\%}$ = modified z-score at 99.9% confidence interval ($4.45 = 3*0.6745^{-1}$).

Fig. (8). Result of application of the proposed outlier removal method.

Fig. (**8**) shows the result of the application of the proposed method to the same data presented in (Fig. **6**). Most of the outliers mainly caused by intermittent stops and round trips were properly filtered. Although some valid data were categorized into outliers, they seemingly have little influence on the travel-time information to be generated. The cleaned individual probe travel times are then aggregated into 5 min bins to be used for the prediction to reduce the time-lag in the arrival time-based travel times.

5. METHODS

5.1. Current Prediction Method

The k-NN technique is initially applied for predicting travel times on a freeway by Davis and Nihan [31]. In their model, the continuously archived probe travel times constituted the state vector as an input variable. The established model finds out k-nearest neighbors in the predefined archived travel times combined in a group (n collection intervals). The model basically assumes that the future travel times to be forecast are analogous to the past travel times in archived records. Considering repetitive human activities that produce traffic on the road, this assumption can be regarded to be reasonable. The k-NN algorithm, expressed in equations 8 to 10, constitutes five phases as described below:

1. Prepare archived travel times. For this study, month-long travel times in a 5-min bin were utilized.

2. Choose n seamlessly combined previous intervals in a group. For this study, a 30 minute-long (or six 5-min bins) travel times that exhibited the lowest prediction error for the k-NN algorithm according to an earlier study [32] were adopted.

3. Calculate the distance metric defined as the Euclidean distance and choose the number of k. The selected number of k was set at 4 based on an earlier study [25] that argued that 4 is the optimal k number considering both computing efficiency and accuracy of the model. Their rule was also proven to be valid in this study.

4. Determine k-nearest neighbors based on the calculated distance metric.

5. Predict the future travel times (the departure time-based travel times) averaged by the k-nearest neighbors weighted by the computed distance values.

$$TT(D_t) = \frac{\sum_{k=1}^{k}[TT_h^k(D_t) \times w^k(A_t)]}{\sum_{k=1}^{k} w^k(A_t)} \quad (8)$$

$$w^k(A_t) = d^k(A_t)^{-1} \quad (9)$$

$$d^k(A_t) = \left[\sum_{i=0}^{n}\{TT(A_t - i) - TT_h^k(A_t - i)\}^2\right]^{\frac{1}{2}} \quad (10)$$

where

$TT(D_t)$ = departure time-based travel time at time t,

$TT_h^k(D_t)$ = k-th departure time-based travel time at historical time t,

$w^k(A_t)$ = k-th weight of arrival time-based travel time at time t,

$d^k(A_t)$ = k-th distance between current and historical arrival time-based travel time at time t,

$TT(A_t)$ = current arrival time-based travel time at time t,

$TT_h^k(A_t)$ = k-th historical arrival time-based travel time at time t, and ,

i= aggregation interval (*e.g.* 5 min).

5.2. Alternative Prediction Method

The current k-NN method has been proven to be inefficient for real-time applications due to its relatively long searching time: around 2 min for 6 month-long historical data using the server system currently deployed. To improve prediction performance as well as real-time capabilities, a Recurrent Neural Network (RNN), a class of deep neural networks where connections between nodes form a directed graph along a temporal sequence, is adopted for an alternative method. Unlike the conventional feedforward neural networks, recently developed RNN models use their internal states to process sequences of inputs. This characteristic allows them to be widely applicable to tasks including handwriting recognition, speech recognition, and time-series data prediction [33, 34].

Among RNN models, a Lng Short-Term Memory (LSTM) model which can avoid the vanishing gradient problem was adopted in this study. As LSTM models are normally augmented by recurrent gates, it prevents backpropagated errors from vanishing or exploding [35]. In other words, LSTM has the merit of learning tasks that need memories of events that happened discrete time steps earlier. Also, a mix of low and high-frequency components can be handled by LSTM models [36]. All these characteristics make LSTM-based RNN models be the optimal solution for forecasting travel times in a time sequence. They can, moreover, forecast travel times on a real-time basis with less than several seconds.

The input data format, as represented in Table **2**, was arranged to have a sequence length of one, a data dimension of six, and an output dimension of one. The constructed model with 10 hidden nodes was optimized by the widely-recognized adaptive moment estimation. One month-long historical data were divided into training data of 80% and testing data of 20%. The performance of the applied model is plotted in Fig. (**9**). For comparison with the k-NN model, six 5-min intervals data were used as independent variables where it revealed the lowest error.

Table 2. Input data format (before normalization).

Independent Variables (Previous 30 min)						Dependent Variable
A-25 min	A-20 min	A-15 min	A-10 min	A-5 min	A-0 min	D-0 min
219	197	300	223	283	329	334

(Table 2) cont.....

Independent Variables (Previous 30 min)						Dependent Variable
197	300	223	283	329	331	334
300	223	283	329	331	466	476

Note: A indicates arrival time-based travel times and D indicates departure time-based travel times.
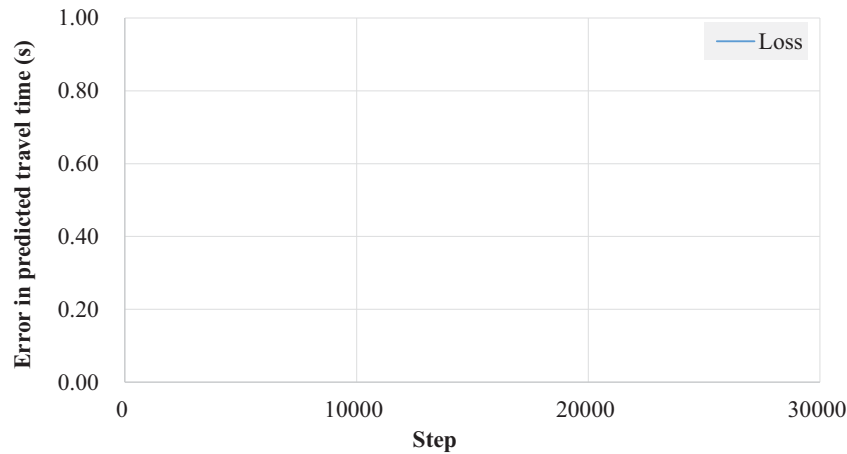

Fig. (9). Training process of the proposed LSTM-based RNN model.

6. RESULTS

The significance of travel-time prediction is highlighted during peak hours because, under noncongested conditions, predicted *vs.* current travel times do not show notable differences. The peak hour (7-9 a.m.) probe travel times on three weekdays (Wed. – Fri.) were therefore considered for evaluation of the two prediction methods. A month-long weekday travel times on the same stretch were included in the historical database. As shown in Table **3**, the average travel times on the two datasets were similar due to the recurrent travel-time patterns observed in the section during weekdays (Fig. **5**).

Table 3. Data description.

Item	Historical Data Set	Prediction Data Set
Period[a]	Jan. 1st – 31st, 2019 (Weekdays)	Feb. 6th (Wed.) – 8th (Fri.), 2019
Average nonpeak hour travel time	298 s	308 s
Average peak hour travel time[b]	654 s	667 s

[a]Weekends are not included.
[b]From 7 to 9 a.m.

Ten experiment cars were employed to collect ground truth data. The drivers were instructed to operate the cars with the floating car method where the number of cars overtaking and overtaken was equivalent. A total of 10 experiment cars with roughly 2 min headway were driven along the bi-directional roadway repeatedly during the peak hours for the three days. A total of 220 test-car runs in 75 5-min collection bins (2-4 ground-truth data in each bin) were obtained. The mean absolute error (Equation 11) was utilized to quantitatively measure the prediction errors.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|x(t) - \hat{x}(t)| \times 100(\%) \qquad \textbf{(11)}$$

where MAE= mean absolute error,

n= number of samples,

$x(t)$= actual (observed) travel time, and

$\hat{x}$ = predicted travel time.

Table **4** and Fig. (**10**) show comparisons between the ground truth and predicted travel times. In all cases, the proposed LSTM-based RNN model revealed higher performances than the current k-NN method. It also decreases the variance of prediction errors that can reduce uncertainty and therefore allows drivers to better plan activities who rely on arrival time. Further, quartile error comparisons (Fig. **11**) reinforces the superiority of the proposed method where lower dispersions were observed in the alternative method.

Table 4. Travel-time prediction error.

Statistics (s)	Wednesday		Thursday		Friday	
	Current	Alternative	Current	Alternative	Current	Alternative
Mean	48	37	87	77	41	39
Variance	48	29	105	69	48	32
Maximum	177	102	347	293	237	127

(Table 4) cont.....

Statistics (s)	Wednesday		Thursday		Friday	
75th percentile	65	45	115	132	52	50
Median	33	33	40	42	27	31
25th percentile	17	13	26	32	12	19
Minimum	3	2	2	0	0	0

(a) Wednesday

(b) Thirsday

(c) Friday

(d) Comparison
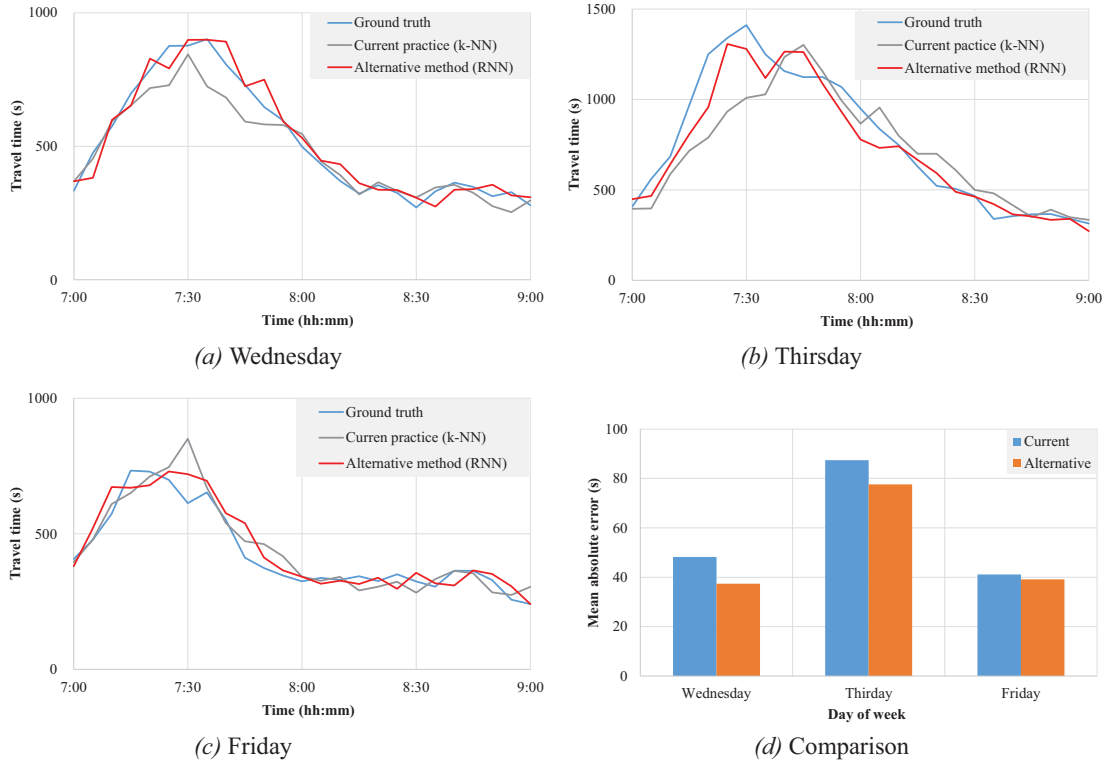
Fig. (10). Performances of the current (k-NN) and the proposed (RNN) predicton methods.

(a) Wednesday
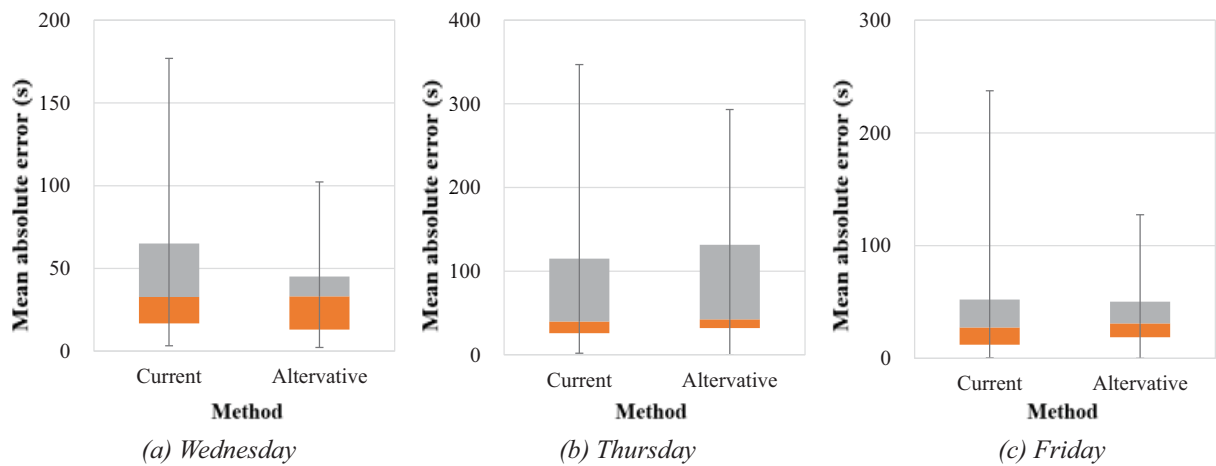
(b) Thursday

(c) Friday

Fig. (11). Boxplots of the errors in the current and the proposed prediction methods.

7. DISCUSSION

Although probe-based travel-time systems have merits in that they can directly collect link travel times and generally require a lower budget compared to the traditional vehicle detector-based systems, improper treatment of outlying observations and time-lag phenomenon could make them obsolete. After identifying shortcomings in the current methods, a new outlier treatment technique and an LSTM-based RNN model were proposed in this study.

To censor outliers in the probe travel times, the modified z-score that is less susceptible to aberrant values was proposed. Also, to take the characteristic of the probe travel times into consideration, the natural logarithm was applied to the raw travel times. The developed method was proven to outperform the current one. To mitigate the time-lag phenomenon, an LSTM-based RNN model, which is recently garnering much attention for forecasting time-series data is applied and its prediction performance was compared to the current k-NN method. Resultantly, it generated fewer prediction errors when evaluated by 3-day-long ground-truth data collected by test-car runs during the peak hours (7-9 a.m.) on the stretch.

CONCLUSION

Although the proposed methods were thoroughly tested on the segment, it would be preferable to be further verified through follow-up research into spatiotemporal universality by applying the proposed methods to other arterial and/or freeway systems.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Southwest Research Institute, *Automatic Vehicle Identification Model Deployment Initiative-System Design Document.,* Texas Department of Transportation, 1998.

[2]   S. D. Clark, S. Grant-Muller, and H. Chen, "Cleaning of Matched License Plate Data", *Transportation Res Rec,* no. 1804, 2002.

[3]   F. Dion, and H. Rakha, *Estimating Dynamic Roadway Travel Times Using Automatic Vehicle Identification Data. Trans Res Part B.,* Elsevier, 2006.

[4]   X. Ma, and H. Koutsopoulos, "Estimation of the automatic vehicle identification based spatial travel time information collected in stockholm", *IET Intell. Transp. Syst.,* vol. 4, no. 4, 2010.
[http://dx.doi.org/10.1049/iet-its.2009.0149]

[5]   D.V. Boxel, W.H. Schneider IV, and C. Bakula, "An innovative real-time methodology for detecting travel time outliers on interstate highways and urban arterials", *TRB 2011 Annual Meeting CD-ROM,* 2011 Washington D.C.

[6]   S. S. Moghaddam, and B. Hellinga, "Algorithm for detecting outliers in bluetooth data in real time", *Trans Res Rec,* no. 2442, 2014.

[7]   H. Park, and Y. Kim, "Model for filtering the outliers in dsrc travel time data on interrupted traffic flow sections", *KSCE J. Civ. Eng.,* vol. 22, no. 9, 2018.
[http://dx.doi.org/10.1007/s12205-017-1333-z]

[8]   R.E. Kalman, "A new approach to linear filtering and prediction problems", *J. Basic Eng.,* vol. 82, p. 35, 1960.
[http://dx.doi.org/10.1115/1.3662552]

[9]   D.M. Wolpert, and Z. Ghahramani, "Computational principles of movement neuroscience", *Nat. Neurosci.,* vol. 3, suppl. Suppl., pp. 1212-1217, 2000.
[http://dx.doi.org/10.1038/81497] [PMID: 11127840]

[10]   J. Weng, *Natural and Artificial Intelligence: Introduction to Computational Brain-Mind.,* BMI Press, 2012.

[11]   P. Lingras, S. Sharma, and M. Zhong, "Prediction of recreational travel using genetically designed regression and time-delay neural network models", *Trans Res Rec. J. Trans Res Board,* vol. 1805, 2003.

[12]   L. Li, and X. Jiang, "Predicting the travel time in using recurrent neural networks: A case study of fuzhou, international conference on smart vehicular technology, transportation, communication and applications", *Advances in Smart Vehicular Technology, Transportation, Communication and Applications,* 2017 Springer

[13]   Y. Chen, Y. Lv, Z. Li, and F. Wang, "Long short-term memory model for traffic congestion prediction with online open data", *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC).,* , 2016

[14]   H. Zhang, H. Wu, W. Sun, and B. Zheng, "DEEP TRAVEL: a Neural Network Based Travel Time Estimation Model", *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18),* 2018 Stockholm

[15]   H. Kantz, and S. Thomas, *Nonlinear Time Series Analysis.,* Cambridge University Press: London, 2004.

[16]   M. Voort, M. Dougherty, and S. Watson, "Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow", *Transp. Res., Part C Emerg. Technol.,* vol. 4, no. 5, 1996.

[17]   N. S. Altman, *An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression,* vol. 46, no. 3, p. 175, 1992.

[18]   P. Hall, B.U. Park, and R.J. Samworth, "Choice of neighbor order in nearest-neighbor", *Ann. Stat.,* vol. 36, no. 5, p. 2135, 2008.
[http://dx.doi.org/10.1214/07-AOS537]

[19]   J. Myung, D. Kim, and S. Kho, "Travel time prediction using k nearest neighbor method with combined data from vehicle detector system and automatic toll collection system", *Trans Res Rec: J Trans Res Board,* vol. 2256, 2011.

[20]   W. Qiao, A. Haghani, and M. Mamedi, "A nonparametric model for short-term travel time prediction using bluetooth data", *J Intelligent Trans Sys,* vol. 17, Taylor and Francis Group, no. 2, p. 165, 2013.
[http://dx.doi.org/10.1080/15472450.2012.748555]

[21]   W. Qiao, A. Haghani, C. Shao, and J. Liu, "Freeway path travel time prediction based on heterogeneous traffic data through nonparametric model", *J. Intell Trans Sys,* vol. 20, Taylor and Francis Group, no. 5, 2013.

[22]   S. Wu, Z. Yang, X. Zhu, and B. Yu, "Improved k-nn for short-term traffic forecasting using temporal and spatial information", *J. Trans Eng,* vol. 140, American Society of Civil Engineers, no. 7, 2014.
[http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000672]

[23]   S. Lim, H. Lee, S. Park, and T. Heo, "a study of travel time prediction using k-nearest neighborhood method", *Korean J. App Sci,* vol. 26, no. 5, p. 835, 2013. [in Korean].
[http://dx.doi.org/10.5351/KJAS.2013.26.5.835]

[24]   S. Tak, S. Kim, K. Jang, and H. Yeo, *Real-Time Travel Time Prediction Using Multi-level k-Nearest Neighbor Algorithm and Data Fusion Method. Computing in Civil and Building Engineering.,* American Society of Civil Engineers: Orlando, Florida, 2014.

[25]   B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "k-Nearest Neighbor Model for Multi-Time-Step Prediction of Short-Term Traffic Condition", *J. Transp. Eng.,* vol. 142, no. 6, 2016.

[26]   J. Zhong, and S. Ling, "Key factors of k-nearest neighbor nonparametric regression in short-time traffic flow forecasting",

Proceedings of the 21st International Conference on Industrial Engineering and Engineering Management 2014, , 2015

[27] G. Zhang, Y. Wang, H. Wei, and Y. Chen, "Examining headway distribution models using urban freeway loop event data", *Trans Res Rec: J Trans Res Board,* no. 1999, 2007.

[28] J. Jang, "Analysis of time headway distribution on suburban arterial", *KSCE J. Civ. Eng.,* vol. 16, no. 4, 2012.
[http://dx.doi.org/10.1007/s12205-012-1214-4]

[29] B. Iglewicz, and D. Hoaglin, *How to Detect and Handle Outliers, Vol. 16 of The ASQC Basic References in Quality Control: Statistical Techniques,* 1993.E. F. Mykytka,

[30] J.T. Behrens, "Principles and procedures of exploratory data analysis", *Psychol. Methods,* vol. 2, no. 2, 1997.
[http://dx.doi.org/10.1037/1082-989X.2.2.131]

[31] G Davis, and N. Nihan, "Nonparametric regression and short-term freeway traffic forecasting", *J. Trans Eng,* vol. 117, American Society

of Civil Engineers, no. 2, 1991.
[http://dx.doi.org/10.1061/(ASCE)0733-947X(1991)117:2(178)]

[32] J. Arroyo, and C. Mate, "Forecasting histogram time series with k-nearest neighbors methods", *Int J Forecasting,* vol. 25, Elsevier, no. 1, 2009.

[33] H. Sak, A. Senior, and F. Beaufays, *Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling.,* Interspeech, 2014.

[34] F. Gers, N. Schraudolph, and J. Schmidhuber, "Learning Precise Timing with LSTM Recurrent Network", *J. Mach. Learn. Res.,* p. 3, 2002.

[35] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, *Long Short Term Memory Networks for Anomaly Detection in Time Series.,* Computational Intelligence and Machine Learning, 2015.

[36] J. Schmidhuber, "Deep learning in neural networks: An overview", *Technical Report IDSIA-03-14,,* 2014.