



The Open Transportation Journal

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



RESEARCH ARTICLE

Travel-time Prediction Using K-nearest Neighbor Method with Distance Metric of Correlation Coefficient

Jinhwan Jang^{1,*}

¹Highway Research Division, Korea Institute of Civil Engineering and Building Technology, 283 Goyangdae-Ro, Ilsanseo-Gu, Goyang-Si, Gyeonggi-Do 411-712, Republic of Korea

Abstract:

Background:

Real-time Travel Time (TT) information has become an essential component of daily life in modern society. With reliable TT information, road users can increase their productivity by choosing less congested routes or adjusting their trip schedules. Drivers normally prefer departure time-based TT, but most agencies in Korea still provide arrival time-based TT with probe data from Dedicated Short-Range Communications (DSRC) scanners due to a lack of robust prediction techniques. Recently, interest has focused on the conventional k-nearest neighbor (k-NN) method that uses the Euclidean distance for real-time TT prediction. However, conventional k-NN still shows some deficiencies under certain conditions.

Methods:

This article identifies the cases where conventional k-NN has shortcomings and proposes an improved k-NN method that employs a correlation coefficient as a measure of distance and applies a regression equation to compensate for the difference between current and historical TT.

Results:

The superiority of the suggested method over conventional k-NN was verified using DSRC probe data gathered on a signalized suburban arterial in Korea, resulting in a decrease in TT prediction error of 3.7 percent points on average. Performance during transition periods where TTs are falling immediately after rising exhibited statistically significant differences by paired t-tests at a significance level of 0.05, yielding p-values of 0.03 and 0.003 for two-day data.

Conclusion:

The method presented in this study can enhance the accuracy of real-time TT information and consequently improve the productivity of road users.

Keywords: : Travel time, Prediction, k-nearest neighbor, Correlation coefficient, Regression equation, DSRC.

Article History

Received: June 09, 2019

Revised: August 21, 2019

Accepted: September 11, 2019

1. INTRODUCTION

Real-time Travel Time (TT) information has become an essential part of modern society. It can reduce road users' travel times by choosing a less congested route or rescheduling their trip plans to nonpeak hours. Traditionally, TTs were collected using vehicle detectors that use different sensors, such as loop, video image, microwave, ultrasonic. However, as wireless communication technologies evolve, probe-base systems such as GPS, Dedicated Short-Range Communications (DSRC), and Bluetooth/Wi-Fi are attracting more interest in

developed countries. These wireless technologies have two main merits: efficiency for deployment and maintenance, and their ability to obtain direct link TT. Compared to vehicle detector-based TT collection systems, DSRC-based systems are known to be more affordable so DSRC scanners are broadly installed on many roadways by many road agencies in Korea. Using traffic data from vehicle detectors, link TTs should be estimated with the assumption of homogeneity of TTs in the segment covered by a detector. However, this assumption is frequently violated on the signalized arterials that this study focuses on. In Korea, an Electronic Toll Collection System (ETCS) is already based on 5.8 GHz DSRC technology, indicating that agencies could simply install DSRC scanners on the road to collect DSRC-equipped probe data. As

* Address correspondence to this author at the Highway Research Division, Korea Institute of Civil Engineering and Building Technology, 283 Goyangdae-Ro, Ilsanseo-Gu, Goyang-Si, Gyeonggi-Do 411-712, Republic of Korea; E-mail: jhjang@kict.re.kr

of 2017, more than 20% of vehicles on the road could be used as probes with DSRC scanners. The penetration rate was similar in 2013 when the data were gathered for this study. Of course, standard measures for protecting drivers' privacy are applied to DSRC TT systems not only by encrypting user identifications, but also by discarding individual probe data after generating link TTs. However, probe TTs obtained using DSRC scanners inevitably include time lags equivalent to the TTs experienced by the probes on the segment because TTs, which normally called arrival time-base travel times, are calculated after probe vehicles complete the trip. Therefore, TT prediction is more emphasized in probe-based real-time TT provision systems than detector-based TT provision systems.

Several prediction techniques have been applied to real-time TT information systems. Among the most popular techniques are the Kalman Filter (KF), Artificial Neural Network (ANN), time series analysis (TSA), and the k-Nearest Neighbor (k-NN) method. KF, named after R. E. Kalman, one of the primary developers of its theory, is an algorithm that uses time series measurements and produces estimates of unknown variables [1]. The algorithm is recursive, using a two-step process. In the prediction step, it generates estimates of the current state variables, which include uncertainties. In the second step, when the next measurement is obtained, these estimates are updated using a weighted average. In this case, less weight is assigned to the estimates with higher uncertainty [2]. Since it only uses the current input measurements and previously estimated state along with its uncertainty matrix, its implementation is quite simple. In contrast, other techniques listed above require a large amount of high quality historical TTs. However, KF can only predict TTs one step ahead even though road users normally want to know TTs several steps ahead (or departure time-based TTs: the travel times to be experienced by vehicles that are departing the start point after receiving TT information). ANN is a computing system inspired by the biological neural networks that organize animal brains, learning from archived data to perform a certain task [3]. Among many ANN algorithms, time-delay neural networks that recognize features in sequential data are regarded as an optimal solution for predicting traffic data, which, reflecting people's daily activities, normally shows repetitive day-to-day patterns [4]. TSA is a methodology for interpreting data in a time sequence to predict future values using previously observed values [5]. Autoregressive, integrated, and moving average models are three general classes for modeling variations in time sequential observations. Combinations of the three models produce Autoregressive Integrated Moving Average (ARIMA) models that were proven appropriate for predicting traffic data [6]. However, both ANN and TSA are notorious for their lack of universality (or transferability), indicating that if new TT patterns deviate from TT patterns used for training, performance could deteriorate significantly. In reality, changes in TT patterns frequently occur due to local events, traffic accidents, weather, roadwork, and other difficult to predict phenomena. Hence, although based on a sophisticated theoretical background, real-world systems using ANN and TSA to predict TTs are hardly observed. k-NN is a nonparametric method applied for classification or regression. In both cases, the input variables are the k closest learning

examples in the feature space [7]. The method is useful because it assigns more weight to nearer neighbors by measuring the distance between current and historical values [8]. A commonly used distance metric is the Euclidean distance. From the perspective of TT prediction, k-NN has advantages over KF, ANN, and TSA. It can not only predict TTs multiple steps ahead (or departure time-based TTs), but it can also recursively include newly collected TTs as a training set in order to continuously adjust to new TT patterns over time. For this reason, k-NN is drawing substantial interest from road agencies in Korea. Table 1 summarizes the pros and cons of each technique described above.

Table 1. Comparison of travel-time prediction methods.

Method	Advantage	Disadvantage
KF	Do not require large amount of historical data	Impossible to predict multiple-step-ahead travel times
ANN and TSA	Sophisticated and accurate	Do not consider changes in travel-time pattern
k-NN	Flexible to changes in travel-time pattern	Long time to execute

Increasing interest in k-NN for real-time TT prediction is not limited to Korea. Studies have also been performed to apply k-NN to predict traffic data worldwide. Myung *et al.* proposed a TT prediction system based on k-NN using data from vehicle detector systems and automatic toll collection systems deployed on freeways, and demonstrated the superiority of the proposed model over six other models including KF [9]. Qiao *et al.* applied a set of four prediction models including historical average, ARIMA, KF, and two types of k-NN algorithms to probe TT data from Bluetooth scanners installed on I-66 in Virginia. k-NN models outperformed other techniques by decreasing errors in prediction accuracy by more than 10% throughout the day and by 20% for peak hours [10]. They also applied a new k-NN model for predicting multi-step ahead TT and verified the superiority of the suggested model over ARIMA using both historical and real-time heterogeneous traffic and weather data collected on I-95 in Maryland [11]. Wu *et al.* compared several TT prediction techniques including k-NN, historical average, and ANN, and demonstrated that the proposed k-NN model that considered both temporal and spatial information significantly improved forecasting accuracy. This was a result of numerical tests using probe data obtained from 180 taxis running in Guiyang, China [12]. Lim *et al.* predicted departure time-based real-time TT with k-NN, KF, and historical average techniques using arrival time-based probe TT data from DSRC scanners installed on suburban arterials, and demonstrated the superiority of the applied k-NN technique over other techniques [13]. Tak *et al.* applied a multi-level k-NN algorithm designed for forecasting TT with increased computational efficiency and prediction accuracy. The proposed algorithm consisted of three parts: classification, global matching, and local matching. It showed 5% error range even in congested traffic situations, indicating satisfactory performance compared to existing methods [14]. Yu *et al.* proposed a short-term traffic condition prediction model based on k-NN. They proposed a multi-step prediction model with consideration of the time-varying and continuous characteris-

tics of traffic flow. Evaluation using GPS data of taxis in Foshan city, China, revealed better performance compared with the support vector machine, ANN, real-time-data model, and history-data model [15]. Zhong *et al.* investigated the key factors of k-NN through four tests and presented two findings: no linear improvement with increases in data size and enhancement of prediction accuracy with more information included in the state vector [16].

All the previous studies described above used Euclidean distance to find the predefined (k in number) nearest neighbors. However, since the Euclidean distance can only consider the proximity between current and historical values, it potentially shows some deficiencies in TT transition situations, such as increasing TTs followed by decreasing TTs. During the TT transition period, the TT trend along with nearness play an important role in accurate TT prediction. This will be discussed in detail with examples in the following section.

In this study, a correlation coefficient was employed to measure the similarity between current and historical TTs to resolve deficiencies of previous studies. In addition, to compensate for the differences between current and historical TTs, a regression equation was applied to historical TT to obtain the optimal value for a predicted TT. Since correlation coefficients can consider the trend in the current and historical TTs and regression equations can adjust the difference between current and historical TTs, the proposed k-NN method is expected to overcome shortcomings of previous methods. The remainder of this article is structured as follows. In the next section, the traditional k-NN is described in detail and its problems in TT transition situations are described, followed by the description of a new k-NN method that surmounts the problems of the conventional k-NN method. Subsequently, the proposed method is evaluated using 10 experiment cars driven on the segment with 5 minutes headway at the maximum. Finally, the significance of this study will be discussed with a summary and suggestions for future studies.

2. METHODS

2.1. Traditional k-NN Method

Davis and Nihan firstly applied the traditional k-NN method to short-term freeway traffic forecasting [17]. The input variables to constitute the state vector for their model are TTs in the continuously archived previous time intervals. The model searches k-nearest neighbors in the predefined historical records using a combined group (n aggregation intervals) of continuous TTs. The basic assumption behind the method is that the future TTs to be predicted are closely related to the TTs in previous time intervals. This assumption seems to be reasonable in that traffic on the road is generated from human activities that are recursive in nature. The process of the conventional k-NN algorithm can be expressed by equations 1 - 3 and operates in five steps as follow:

- [1] Construct a historical database. For this study, a month of weekday TTs in 5-min interval were used.
- [2] Select n continuously combined previous intervals. For this study, a 30 minute (or six 5-min intervals) series

of data that showed the optimal performance for the conventional k-NN according to a previous study (13) were employed.

- [3] Compute the distance metric defined as the Euclidean distance and select the number of k. For this study, the selected number of k was 4 according to a previous study [15] that demonstrated that k values greater than 4 show little improvement in prediction accuracy. This rule was found to hold for the TT data in this study.
- [4] Determine k-nearest neighbors according to the computed distance metric.
- [5] Forecast the target TT (the departure time-based TT at the same time with the arrival time-based TT) by taking the average of the k-nearest neighbors weighted by the calculated Euclidean distance.

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

$$w^k(A_t) = d^k(A_t)^{-1} \tag{2}$$

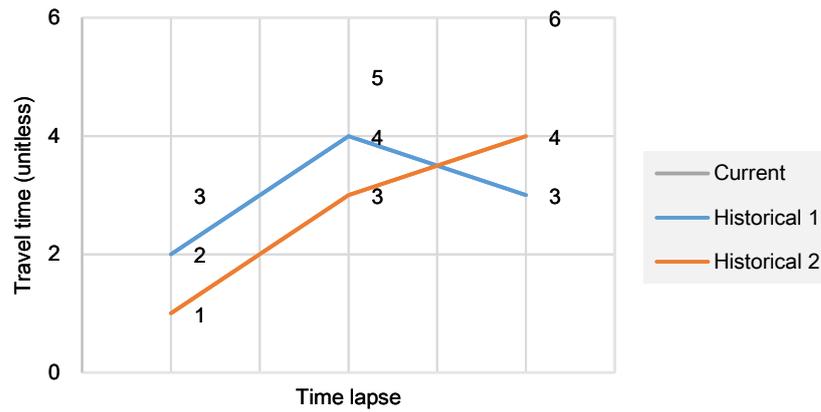
$$d^k(A_t) = \left[\sum_{i=0}^n \{TT(A_t - i) - TT_h^k(A_t - i)\}^2 \right]^{\frac{1}{2}} \tag{3}$$

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).

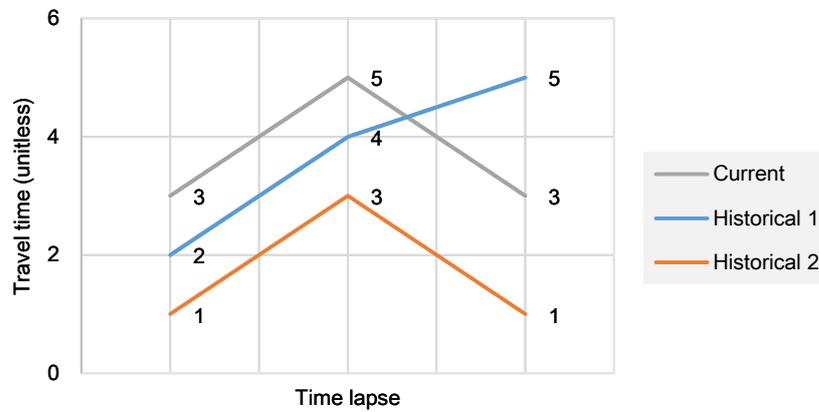
2.2. Problems in Traditional k-NN Method

The traditional k-NN method is one of the most widely used nonparametric pattern matching techniques and its robustness has been proven [18 - 20]. The essence of the nonparametric regression is to find the optimal forecasting values for current observations in the neighborhood in a historical database with similar state vectors. However, as stated earlier, since the traditional k-NN method that uses the Euclidean distance only considers the closeness between the present and the historical observations to find the nearest neighbors, it shows shortcomings in certain situations as exemplified in Fig. (1).

In Fig. (1), one current series and two historical candidate series for the nearest neighbors are plotted. One historical series in (a) and the current series in (b) show trend-changing conditions where values are falling immediately after rising. From the perspective of probe TTs, these changing conditions are normally observed during peak hours. When employing Euclidean distance as a measure of similarity, the predicted values could deviate from the actual observations due to failing to consider the trends. To overcome this problem, a correlation coefficient was used to find historical TTs with the similar trend. A regression equation derived by the ordinary least squares method was then applied to adjust the difference between the current and the historical series. The following section describes the details of the proposed k-NN method.



(a) Transition situation of historical travel time



(b) Transition situation of current travel time

Fig. (1). Cases showing shortcomings in the traditional k-NN method using Euclidean distance.

2.3. Proposed k-NN Method

The process of the proposed k-NN in this study is essentially the same with the conventional k-NN method. The only difference is that a correlation coefficient combined with a regression equation is included. To find the optimum value for prediction, the trend as well as the proximity between current and historical series is taken into account. Due to its inability to account for the trend, the traditional k-NN method showed some deficiencies (Fig. 1). On the other hand, since the proposed k-NN method can account for the trend using a correlation coefficient as well as the proximity using a regression equation, it is expected to resolve the shortcomings in traditional k-NN. TT prediction values are obtained by taking the average of the adjusted historical TTs weighted by the correlation coefficients. The proposed k-NN can be expressed in equations 4-6.

$$TT(D_t) = \frac{\sum_{k=1}^k [ATT_h^k(D_t) \times w^k(A_t)]}{\sum_{k=1}^k w^k(A_t)} \tag{4}$$

$$ATT(D_t) = \alpha + \beta TT_k^h(D_t) \tag{5}$$

$$w^k(A_t) = r^k \{ TT(A_t - i), TT_h^k(A_t - i) \}, i = 0, \dots, n \tag{6}$$

where $ATT_h^k(D_t)$ = k-th adjusted departure time-based travel time at historical time t, α and β = intercept and slope of regression line for a set of $TT(A_t - i)$ and $TT_h^k(A_t - i)$, and $r^k(A_t)$ = correlation coefficient for each pair of current and k-th historical arrival time-based travel times.

3. EVALUATION

3.1. Data for Evaluation

TT data for this study were obtained from the DSRC traffic information system deployed on a signalized arterial in Pyeong Taek region, South Korea (Fig. 2). Traffic on the 4 km stretch of the 4-lane roadway with an 80 km/h speed limit is controlled by six traffic signals. Due to interruptions by the traffic signals, the average travel speed on the stretch under free-flow conditions is approximately 50 km/h.

As illustrated in Fig. (3), the DSRC traffic information system consists of a roadside unit (RSE) and an On-Board Unit (OBU). The RSE scans the unique identifications (IDs) of OBU mounted on passing cars. As mentioned above, the original purpose of installing the OBU was for ETCS. Once a car equipped with DSRC OBU passes two consecutive RSEs, a probe TT is generated by matching the unique ID of the OBU.

Through the OBUs, real-time TT and incident (accident, lane closure, etc.) information can also be received from the RSEs.

Many outlying observations that deviated significantly from valid observations were gathered. The main reason for outlier occurrence was entry/exit maneuvers and intermittent stops at roadside stores on the corridor. Hence, a robust outlier removal technique developed by Jang [21] for probe TT data on signalized arterials was applied to clean the probe TTs. The applied technique was proven to be superior to other techniques when evaluated with the TT data on the subject segment. Individual probe TTs filtered using this technique for one week (Jan. 21-27, 2013) showed a typical commuter route where morning peak hours (7:00-9:00 a.m.) during weekdays are obvious and no congestion is observed during weekends (Fig. 4). TTs during peak hours on weekdays are 3-4 times longer than those during nonpeak hours. Also, the range of TTs under free-flow conditions is relatively high, possibly due to traffic signals. The significance of real-time TT prediction is highlighted during peak hours because, under noncongested conditions, predicted vs. current TTs do not generally show notable differences. The peak hour (7:00-9:00 a.m.) probe TTs on two weekdays (Feb. 5th-6th, 2013) were therefore considered for the application of the conventional and the proposed k-NN methods. A month-long weekday TTs (January in 2013) on the same stretch were included in the historical database. As shown in Table 2, the average TTs on the two data sets were similar due to the recurrent TT patterns observed on the section during weekdays.

Table 2. Data description.

Item	Historical data set	Prediction data set
Period ^a	Jan. 1st – 30th, 2013	Feb. 5th – 6th, 2013
Average nonpeak hour travel time	298 s	308 s
Average peak hour travel time ^b	654 s	667 s

^aWeekends are not included. ^bFrom 7 to 9 a.m.

Ten experiment cars were employed to collect ground truth data. The drivers instructed to drive the cars with floating car method which makes the number of cars overtaking the experiment car and overtaken by the experiment car to be equivalent. At least one experiment car was driven within the aggregation interval of 5 minutes. The ground truth TTs were obtained based on the departure time of each experiment car. Fig. (5) shows the comparison between the ground truth TTs and arrival time-based TTs obtained by the DSRC system. The time lag phenomenon is explicitly observed when real-time TT information is provided with arrival time-based TTs, which is the problem with the current practice. As mentioned earlier, in probe-based TT systems, TTs on the corridor can be obtained only after the probes pass the end point. Hence, drivers provided with arrival time-based TT information are frustrated due to the time lags. For this reason, real-time TT predictions should be emphasized to maximize the utility of probe-based TT systems.

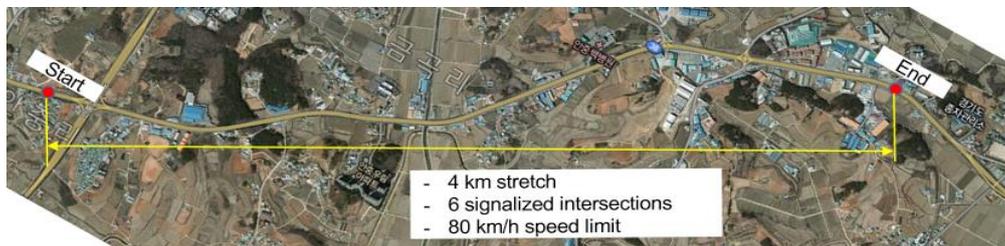


Fig. (2). Test site.

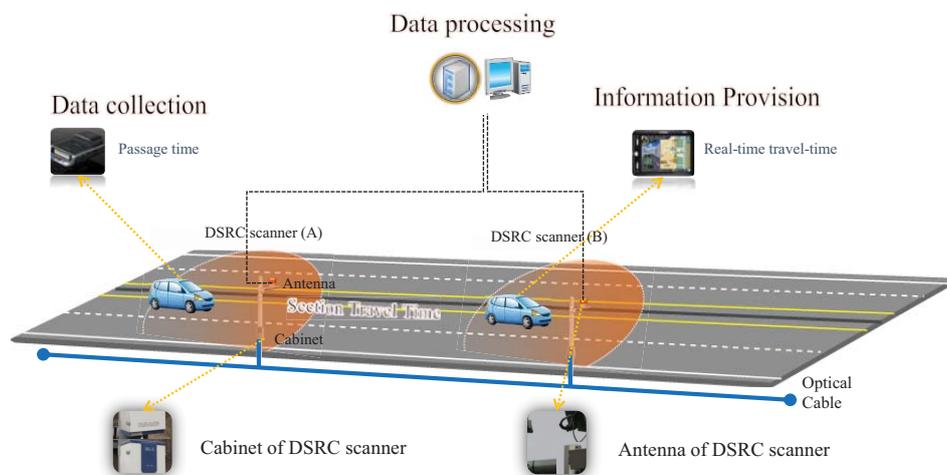


Fig. (3). Schematic of DSRC traffic information system in Korea.

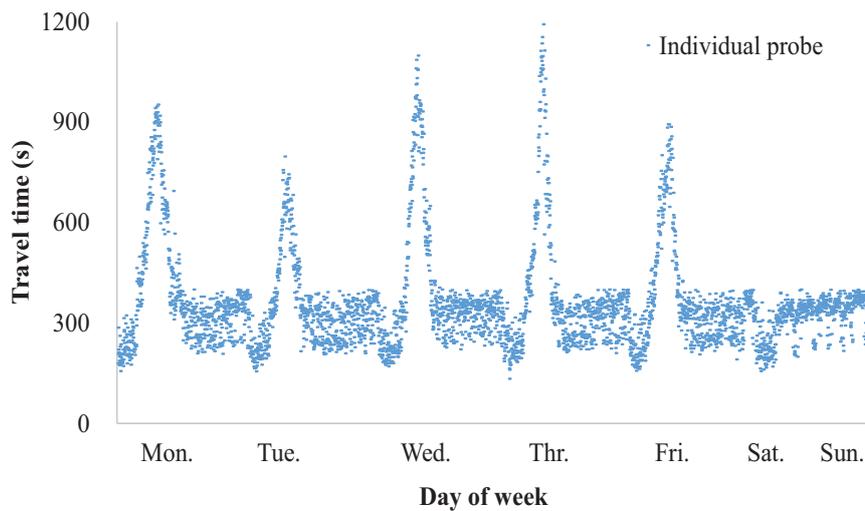


Fig. (4). Daily travel time pattern of the test section.

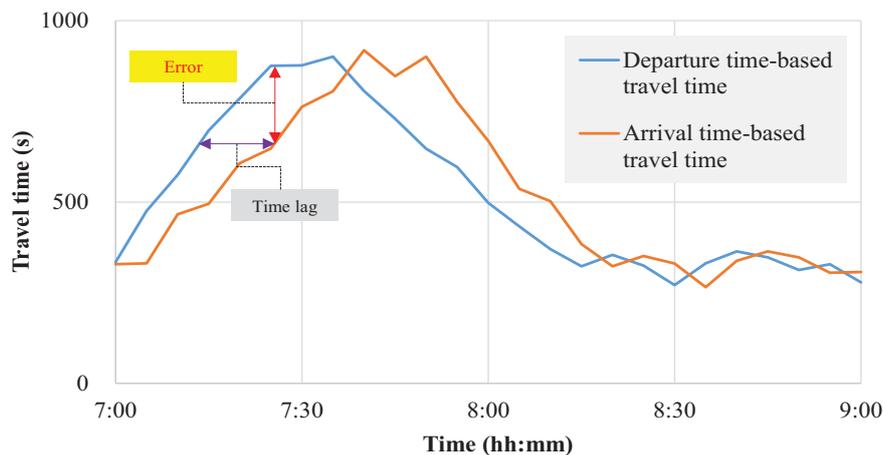


Fig. (5). Time lag between departure time-based and arrival time-based travel times of the test section.

3.2. Numerical Analysis

The conventional and proposed k-NN methods were applied to the data described in Table 1 to predict probe TTs on two days. Predicted TTs along with the ground truth data (departure time-based TTs) are plotted in Fig. (6). Compared to Fig. (5), the two methods apparently reduced the time lag and therefore increased the accuracy of TT information. However, the prediction methods cannot be easily differentiated. Therefore, predicted TTs were evaluated with an evaluation index of mean absolute percent error (MAPE) as expressed in equation 7, and the results were analyzed for comparison.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|x(t) - \hat{x}(t)|}{x(t)} \times 100(\%) \quad (7)$$

where MAPE= mean absolute percent error, n = number of samples, $x(t)$ = actual (observed) travel time, and $\hat{x}(t)$ = predicted travel time.

Table 3 shows a summary of the statistical analyses. Since

the proposed k-NN was developed to improve the performance of conventional k-NN in TT transition situations, the results were categorized into two parts: the entire period and the transition period. Considering the daily TT patterns in Fig. (6), the time gap between 7:20 and 8:00 a.m. was arbitrarily selected as the transition period. As investigated in Fig. (1), the traditional k-NN shows deficiencies in trend-changing (or transition) situation so the evaluation period was categorized into two parts-whole and transition period.

Average errors in predicted TTs on the two days were decreased 3.7 percentage point for the whole period and 5.3 percentage point for the transition period when the proposed k-NN method was applied. The differences in prediction errors were also significant according to t-statistics at the 95% confidence interval: the t-statistics for one-tailed tests were higher than the critical t-statistics for all cases except the whole period on Tuesday. The correlation coefficients computed four times for every single TT prediction ranged from 0.61 to 0.99. The slopes and y-intercepts of the regression models ranged

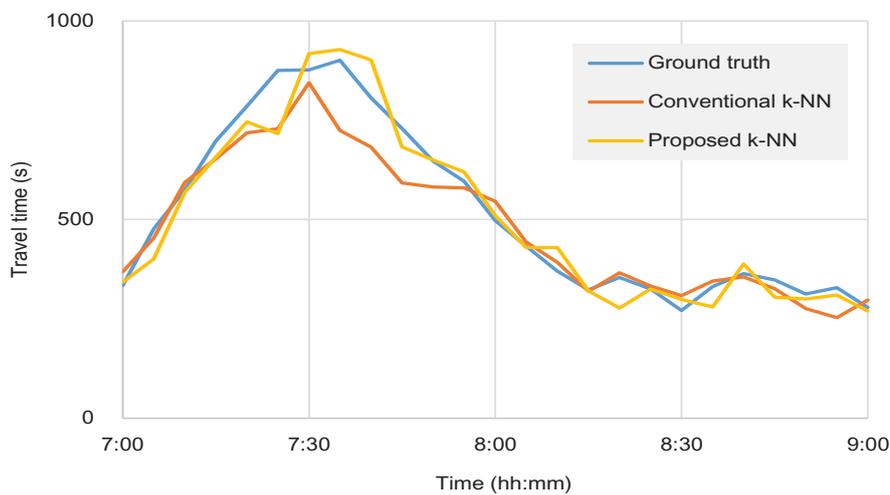
from 0.51 to 1.53 and 0 to 388, respectively. All the statistics was statistically significant at 0.05 significance level.

Due to difficulties in gathering ground truth data using experiment cars, the evaluation period was limited to two days. However, considering the similarity of the daily TT patterns plotted in Fig. (4), the consequences in Table 2 can be reasonably accepted. Fig. (7) shows the predicted TTs by the traditional and proposed methods against departure time-based TTs. Although the departure time-based TTs are not ground-truth data, it can reinforce the superiority of the suggested method in that the TTs predicted by the suggested method are closer to the departure time-based TTs during the transition periods where the proposed method assumed to outperform.

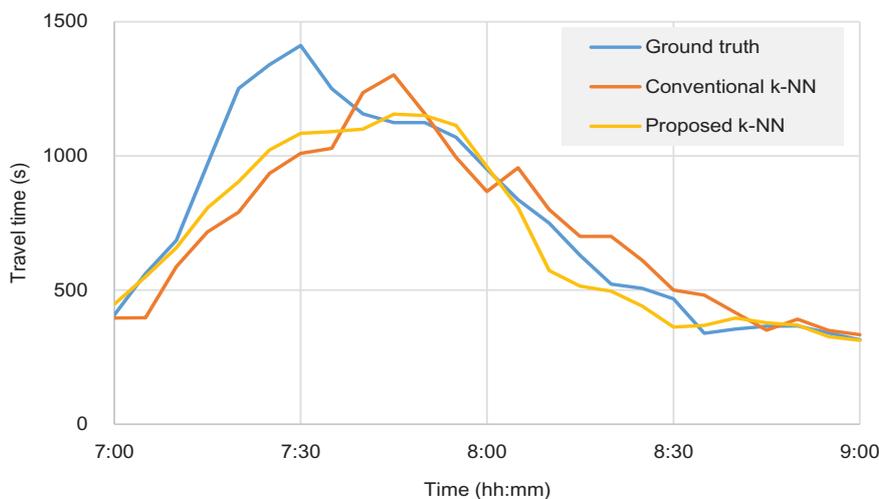
4. DISCUSSION

As described above, the newly proposed k-NN method mitigates the time lag phenomenon in the DSRC probe-based

TT provision system deployed in signalized arterials in Korea. Error reductions were more prominent in TT transition states where the conventional k-NN showed relatively low performance; this phenomena can be attributed to the fact that, whereas the conventional k-NN only takes closeness into account, the proposed method considers both closeness and trend simultaneously. This outcome satisfies the intention of this study. For Tuesday data, the improvement in the proposed k-NN was minimal over the whole period, but the TT prediction accuracy was significantly enhanced during the transition period. For Wednesday data, improvements during the transition period as well as during the whole period were statistically significant. Furthermore, comparison of the quartiles of the conventional and proposed method where lower values were obtained by the proposed method reinforces the superiority of the proposed method to the convention k-NN (Fig. 8).

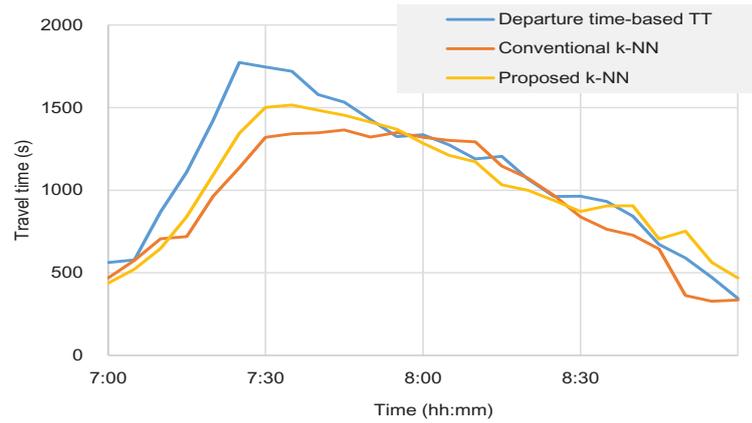


(a) Tuesday (Feb. 5th)

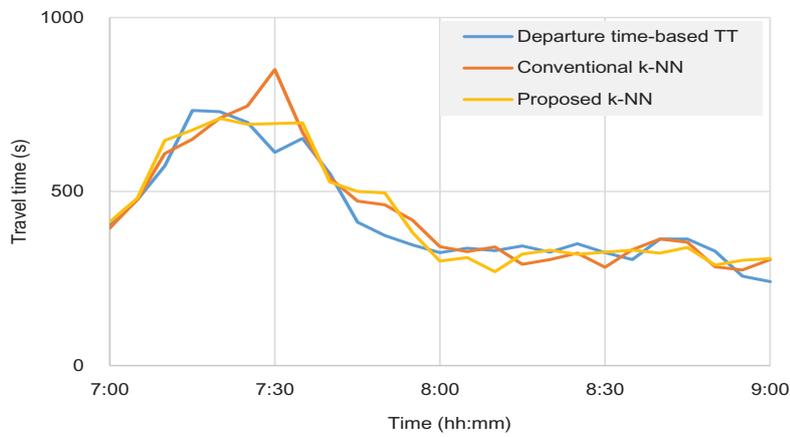


(b) Wednesday (Feb. 6th)

Fig. (6). Distribution of predicted travel times using conventional and proposed k-NN method against ground-truth data for experiment cars.

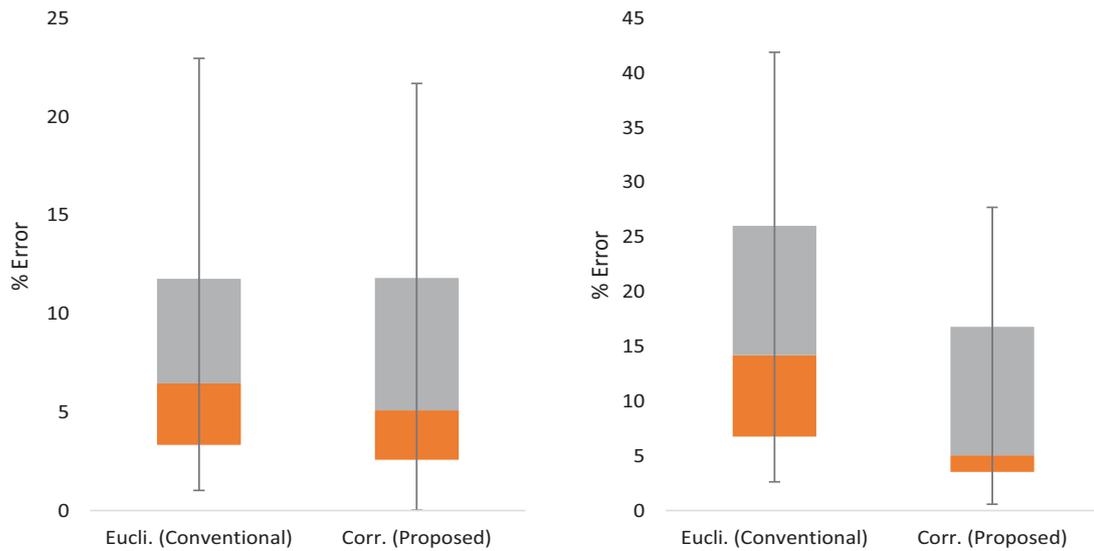


(a) Monday (Feb. 4th)



(b) Friday (Feb. 8th)

Fig. (7). Distribution of predicted travel times using conventional and proposed k-NN method against departure time-based travel times.



(a) Tuesday

(b) Wednesday

Fig. (8). Box plot of errors.

Table 3. Analysis of percent errors in travel times predicted by the conventional and the proposed k-NN method.

Classification	Statistics	Tuesday (Feb. 5th)		Wednesday (Feb. 6th)	
		Con. k-NN	Pro. k-NN	Con. k-NN	Pro. k-NN
Whole period (7:00–9:00 a.m.)	Mean	8.5	7.1	16.0	10.0
	Variance	38.8	39.4	139.8	75.7
	Sample	25		25	
	t-stat.	0.82		2.54	
	Cri. t-stat. at 95% con. int.	1.71		1.71	
	p-value	0.21		0.009	
Transition period (7:20–8:00 a.m.)	Mean	11.8	6.7	18.2	12.7
	Variance	44.0	32.3	156.7	113.6
	Sample	8		8	
	t-stat.	2.18		3.84	
	Cri. t-stat. at 95% con. int.	1.89		1.89	
	p-value	0.03		0.003	

Note: con. = conventional; pro. = proposed; Cri.=critical; stat. = statistic; con. int.: confidence interval.

CONCLUSION AND FUTURE STUDIES

Real-time TT prediction has been one of the most important research topics of the last several decades in the transportation field of study [22]. Nonetheless, few traffic Management Centers (TMCs) employ TT prediction Techniques due to their inability to generate reliable TT information. Among several widely recognized TT prediction methods, k-NN is drawing considerable interest from many TMCs in Korea. However, conventional k-NN that uses the Euclidean distance to find the nearest neighbors showed deficiencies in TT transition situations. The deficiency was due to its inability to consider the TT trend when searching neighbors. Therefore, a new k-NN method was suggested to resolve this shortcoming. The proposed k-NN employs a correlation coefficient to find the closest neighbors and applies a regression equation to adjust the difference between the current and historical observations. Based on its ability to take both the proximity and the trend into consideration, the suggested method was proven to overcome the deficiency in the conventional k-NN.

The next step research would be verification of the robustness, or the time-space transferability, of the proposed method by applying it to link TT data from other time periods and roadway sections. Also, as in the previous study [23], unusual factors such as weather, road works, social events, where available, need to be considered to enhance prediction accuracy.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

This work was supported by a grant (No. 18TLRP-B148386-01) from the Korea Agency for Infrastructure Technology Advancement (KAIA).

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

DISCLAIMER

This paper is a revised version of the paper submitted to 2019 TRB Annual Meeting in Washington D.C.

REFERENCES

- [1] 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]
- [2] D.M. Wolpert, and Z. Ghahramani, *Nat. Neurosci.*, vol. 3, suppl. Suppl., pp. 1212-1217, 2000. [http://dx.doi.org/10.1038/81497] [PMID: 11127840]
- [3] J. Weng, "Natural and artificial intelligence: introduction to computational brain-mind", ISBN 978-0985875725
- [4] 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.
- [5] H. Kantz, and S. Thomas, *Nonlinear Time Series Analysis.*, Cambridge University Press: London, 2004.
- [6] 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.
- [7] N. S. Altman, *An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, the American Statistician.*, vol. 46, no. 3, p. 175, 1992.
- [8] 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]
- [9] 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, pp. 51-59, 2011.
- [10] W. Qiao, A. Haghani, and M. Mamedi, "Nonparametric model for short-term travel time prediction using bluetooth data", *J Intel Trans Sys.*, vol. 17, Taylor and Francis Group, no. 2, pp. 165-175, 2013. [http://dx.doi.org/10.1080/15472450.2012.748555]
- [11] W. Qiao, A. Haghani, C. Shao, and J. Liu, "Freeway Path travel time prediction based on heterogeneous traffic data through nonparametric model", *J Intel Trans Sys.*, vol. 20, Taylor and Francis Group, no. 5, 2013.
- [12] S. Wu, Z. Yang, X. Zhu, and B. Yu, "Improved k-NN for short-term

- traffic forecasting using temporal and spatial information", *J Trans Engineering.*, vol. 140, American Society of Civil Eng, no. 7, 2014.
[[http://dx.doi.org/10.1061/\(ASCE\)TE.1943-5436.0000672](http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000672)]
- [13] S. Lim, H. Lee, S. Park, and T. Heo, "A Study of travel time prediction using k-nearest neighborhood method", *Korean J Appl Stat.*, vol. 26, no. 5, p. 835, 2013. [in Korean].
[<http://dx.doi.org/10.5351/KJAS.2013.26.5.835>]
- [14] 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.
- [15] "Song, X., Guan, F., Yang, Z., and Yao, B. k-nearest neighbor model for multi-time-step prediction of short-Term traffic condition", *J. Transp. Eng.*, vol. 142, no. 6, 2016.
- [16] J. Zhong, and S. Ling, "Key factors of k-nearest neighbor nonparametric regression in short-time traffic flow forecasting", *Pro 21st Int Conf on Ind Eng Engineering Manag 2014.*, 2015
- [17] 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\(178\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(1991)117:2(178))]
- [18] J. Arroyo, and C. Mate, "Forecasting histogram time series with k-nearest neighbors methods", *Int J Fore.*, vol. 25, Elsevier, no. 1, 2009.
- [19] B. Smith, and M. Demetsky, "Traffic flow forecasting: comparison of modeling approaches", *J of Trans Eng.*, vol. 123:4, American Society of Civil Engineers, no. 261, 1997.
- [20] B. Smith, B. Williams, and R. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting", *Transport Research Part C: Emerging Technologies.*, vol. 10, Elsevier, no. 4, 2002.
[[http://dx.doi.org/10.1016/S0968-090X\(02\)00009-8](http://dx.doi.org/10.1016/S0968-090X(02)00009-8)]
- [21] J. Jang, "Data-cleaning technique for reliable real-life travel time estimation: use of dedicated short-range communications probes on rural highway", *Trans Res Rec: J Trans Res Board.*, vol. 2593, 2016.
- [22] B. I. Bustillos, and Y. C. Chiu, "Real-time freeway-experienced travel time prediction using N-curve and k nearest neighbor methods", *Trans Res Rec: J Trans Res Board.*, vol. 2243, Elsevier, 2011.
- [23] 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, conference paper",
[<http://dx.doi.org/10.1061/9780784413616.231>]

© 2019 Jinhwan Jang.

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: (<https://creativecommons.org/licenses/by/4.0/legalcode>). This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.