Research Article

Applying Clustered KNN Algorithm for Short-Term Travel Speed Prediction and Reduced Speed Detection on Urban Arterial Road Work Zones

Hyun Su Park ¹, Yong Woo Park ², Oh Hoon Kwon ², and Shin Hyoung Park ³

¹The Institute of Urban Science, University of Seoul, Seoul 02504, Republic of Korea

²Department of Transportation Engineering, Keimyung University, Daegu 42601, Republic of Korea

³Department of Transportation Engineering, University of Seoul, Seoul 02504, Republic of Korea

Correspondence should be addressed to Shin Hyoung Park; shinhpark@uos.ac.kr

Received 18 June 2021; Revised 20 December 2021; Accepted 5 January 2022; Published 4 February 2022

Academic Editor: Inhi Kim

Copyright © 2022 Hyun Su Park et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study developed and verified a travel speed prediction model based on the travel speed and work zone statistics collected from the advanced traffic management system (ATMS) real-time data in Daegu, South Korea. A clustered K-nearest neighbors (CKNN) algorithm was used to predict travel speed, resulting in a 6.9% average mean absolute percentage error (MAPE) using the data from 1,815 work zones. Furthermore, road network impact due to road work was calculated by comparing the travel speed prediction results obtained from the historical speed data. The predicted travel speed data in a work zone generated from this study is expected to allow drivers to select optimized paths and use them for traffic management strategies to operate in a work zone efficiently.

1. Introduction

A downside involved in road works is the reduction in the capacity of a road, leading to traffic congestion and inconveniencing drivers. In Daegu Metropolitan City, South Korea, an average of 20 road works are carried out per day with limited information provided in advance, such as construction schedules. Other essential details, including expected traffic congestion and estimated travel time based on road works, are not disclosed. Therefore, drivers navigating or near the work zone may experience significantly longer travel times than expected due to the restricted notices. Predicting the network impact of road constructions can provide drivers with opportunities to choose detours [1, 2] and allow road managers to use it as data for establishing traffic operation strategies in case of traffic congestion. It is necessary to have a system that predicts the impact of work on traffic flow to reduce congestion caused by frequent road works while providing information to drivers or road managers ahead of time. In addition, an algorithm

for predicting the speed of neighboring road links after road work and a method for understanding the effect of road construction on the network should be developed.

Daegu Metropolitan City operates the advanced traffic management system (ATMS) to provide traffic condition information on urban roads. However, the data generated by the system does not reflect real-time traffic information, so inconsistencies in the actual road conditions arise. It is necessary to provide the system's users with estimated travel speed or time to change their travel plans appropriately based on the provided traffic information. Meanwhile, numerous studies have been conducted to predict traffic conditions in urban networks [3], but studies on predicting traffic conditions, particularly in urban construction sections, are relatively insufficient [4]. This study aims to develop an algorithm to predict traffic speed after road work in an urban area and present a method for determining whether the road network is affected. The travel speed prediction model for the work area was developed through the clustered K-nearest neighbors (CKNN) algorithm using

traffic statistics collected from Daegu ATMS and work data gathered from the urban traffic information system (UTIS). Furthermore, road network impact due to road construction was calculated by comparing the driving speed prediction results obtained from the historical speed data.

This study is composed of five sections. Section 2 reviews previous studies related to this study, while Section 3 describes data used for this study and its preprocessing. Moreover, Section 4 describes how to design and verify a CKNN model, and Section 5 discusses the research result and the follow-up study.

1.1. Literature Review. Predicting travel speed or travel time has been an active research topic for decades, and as a result, various predictive models have been developed [5]. In early studies, parametric methods such as autoregressive moving average (ARMA), Kalman filter [6], autoregressive integrated moving average (ARIMA) [7, 8], and seasonal autoregressive integrated moving average (SARIMA) [9] were utilized [10, 11]. However, parametric methods were difficult to implement in real-time traffic systems due to some problems such as model calibration, validation, and computational challenges [5]. In addition, they have been proven to encounter poor performance compared to nonparametric methods in unstable traffic conditions and complex road settings [12, 13]. Neural network (NN), K-nearest neighbors (KNN) [14], Bayesian network (BN) [15], and support vector machine (SVM) [10, 16] are the representatives of nonparametric algorithms [17]. Such approaches were advantageous as they are free of assumptions regarding the underlying model formulation and the uncertainty in estimating the model parameters [18]. Recently, studies using deep learning techniques have been conducted to improve the prediction accuracy of traffic conditions [17, 19]. These include long short-term memory (LSTM) [20, 21], deep belief network (DBN) [22], stacked autoencoder (SAE) [23, 24], and convolutional neural network (CNN) [25], which were widely used and had achieved good results in predicting traffic conditions [26]. Nevertheless, a large amount of traffic data was required to utilize nonparametric methods and deep learning strategies, which increased algorithm execution times, making it difficult to present prediction results in real-time [27, 28].

This study considered two factors when selecting the travel speed prediction algorithm. First, the travel speed prediction algorithm must be implemented in the traffic information and management systems. Second, the algorithm should be easily understandable through a traffic manager's level of knowledge and experience. The parametric process was unsuitable for this study because of its complications in implementing it in the system based on these two points. Conversely, a nonparametric method was easier to apply and provided superior prediction performance than a parametric method in unstable traffic conditions. In particular, due to their excellent prediction results, neural networks and KNN algorithms have been used in related studies for a long time. However, it was difficult for analysts and managers to understand neural

network models since they use numerous neurons, complex structures, and nonlinear functions [29–31]. Therefore, a KNN model implemented in a real-time traffic system that traffic managers can easily understand would be more appropriate.

KNN algorithms could perform relatively accurate predictions as the data increases, but computation time becomes longer. Liu et al. [32] recognized this problem and used a clustering method to improve it. Clustering is a procedure for grouping data with similar characteristics, and when used in the KNN algorithm, it shortens prediction times and maintains good performance [33]. Hence, a clustering method was applied to the KNN algorithm to compress prediction times and improve the accuracy.

Most previous studies on travel speed prediction in work zones were conducted on highways [34–38]. Prior studies focused on work zones of urban arterial roads, but these were limited to specific links or routes [4, 39]. Based on the cited studies emphasizing the efficacy and potential of the KNN algorithm, this study aims to develop an algorithm that predicts traffic speeds based on changing traffic conditions caused by work zones on urban roads. It also aims to present a method to understand its effect on the road network. This study is not limited to a specific link or route but conducts a travel speed prediction targeting all arterial roads in Daegu Metropolitan City. In addition, this study presents a difference in that few studies suggest a method for judging the effect on networks due to construction on urban arterial roads.

2. Data Description and Preparation

2.1. Standard Node Link Data. Standard node link is Korea's standard transportation network database with a unified identification (ID) system. Among them, link data includes various road information (link ID, number of lanes, road name, speed limit, etc.), as shown in Table 1. In Daegu, the two major systems that collect, process, and provide traffic information are UTIS and ATMS, which efficiently use standard node link-based link IDs to match data between systems. The calculation time of the travel speed prediction algorithm in a work zone was shortened, and the accuracy of the prediction result was improved by clustering 1,672 links according to their attributes' similarities.

2.2. ATMS Data. Daegu Metropolitan City provides real-time traffic information to road users by building ATMS, a part of intelligent transportation systems (ITS). ATMS collects individual vehicle travel information such as vehicle IDs and detection time through dedicated short-range communication (DSRC) when a vehicle equipped with an onboard unit (OBU) passes through roadside equipment (RSE) installation points. The collected traffic speed data was generated using the vehicle detection times. Meanwhile, the distance between the roadside devices was then processed into traffic speed data in units of five minutes for each road link. As shown in Table 2, ATMS data fields include standard node link-based ID (STD_LINK_ID), aggregated time

TABLE 1: Field items of standard link data.

Field name	Description	Field name	Description
LINK_ID	Link ID	ROAD_NAME	Road name
F_NODE	Start node ID	ROAD_TYPE	Road type
T_NODE	End node ID	MAX_SPD	Max. speed limit
LANES	Number of lanes	REST_VEH	Restricted vehicle
ROAD_RANK	Road grade	REST_W	Restricted width
ROAD_NO	Road number	REST_H	Restricted height

TABLE 2: Field items of advanced traffic management system (ATMS) data.

Field name	Description
STD_LINK_ID	Link ID
GENERATEDDATE	Collected time
SPEED	Speed

(GENERATEDDATE), and speed (SPEED) calculated by aggregating data collected over five minutes. In this study, the traffic speed of the work zone was predicted using ATMS data collected for a total of eight months, from November 2018 to June 2019.

2.3. UTIS Data. UTIS contains event information such as traffic accidents, road construction, events, and weather conditions that happened on the road. Table 3 shows the UTIS data, including various information such as event ID, link ID based on standard node link, event start and end date, event information, and location of occurrence. After collecting the UTIS data from November 2018 to June 2019, the results were used to extract (1,815 cases) ATMS data at the time of road work through link ID matching.

2.4. Data Preprocessing. This study applied travel speed data on arterial roads (1,672 links) collected through ATMS and road works statistics (1,815 cases) for eight months (November 2018 to June 2019). For the same links, ATMS speed data was classified into days with or without road works, and some links showed that road work was performed twice or more within eight months. Since the day of the week was one of the many factors that affect the travel speed, details were constructed by classifying the days from Sunday to Saturday so that the characteristics of the day can be reflected in the travel speed prediction model for the work zones. The network statistics with road works in progress were extracted and used as training data, and the statistics on the networks under the normal condition without road works were utilized to analyze the network impact caused by these construction or maintenance activities.

Because the traffic data generates random noise of measured values from its stochastic characteristics, it is required to remove the noise in historical speed data through smoothing [40]. Thus, moving average, which is considered a smoothing method, was implemented using five-minute time intervals (travel speed for ten minutes before and after including the speed of the i -th time) as in equation (1). The historical speed data was transformed into a smoother form

of the time series data with outliers removed, as shown in Figure 1.

$$V_t = \frac{1}{5} \sum_{i=-2}^{i=2} V_{t+i}^h, \text{ for all samples.} \quad (1)$$

In equation (1), V_t is the speed in time t at which the smoothing operation was performed, and V^h is the historical speed data.

3. Methodology

3.1. Cluster Analysis. Cluster analysis refers to grouping data having a similar pattern [33]. In this study, cluster analysis was performed to improve the accuracy of prediction results and the computation time required for prediction. Travel speed, which is affected by various factors such as road environment (e.g., speed limit, number of lanes, etc.), can be predicted more accurately because the noise from inconsistent data can be removed when clustered by links with similar road environments [33]. Moreover, it is possible to improve the prediction speed of the KNN algorithm by grouping data, which deteriorates as the number of samples increases [32, 41].

As a partitional clustering method, the k-means clustering algorithm was applied because the concept is relatively simple, making it easier for traffic managers to understand. The calculation time is short, making it effortless to use in a real-time information system [42]. The number of lanes and speed limit were used as input variables for k-means clustering. As seen in Table 1, the link contains numerous pieces of information, but the input variables used for k-means clustering analysis are limited. For example, since the road grade or road type means the hierarchy of roads (expressway/general road, highway/urban road/rural road, etc.) rather than link information, it is challenging to use them to cluster similar links. Conversely, the number of lanes and speed limit affect the capacity of the construction section network [43, 44] since network capacity is related to the traffic speed [45]. Therefore, similar links were classified as k-means clustering input variables using the number of lanes and speed limit.

TABLE 3: Field items of urban traffic information system (UTIS) data.

Field name	Description	Field name	Description
INCIDENTID	Event ID	TROUBLEGRADE	Event grade
LINKID	Link ID	INCIDENTTITLE	Event title
STARTDATE	Start date	INCIDENTINFO	Event information
ENDDATE	End date	INCIDENTCODE	Event type
COORDX	Longitude	INCIDENTSUBCODE	Event subtype
COORDY	Latitude	LOCATION	Address

FIGURE 1: Smoothed historical speed data. (a) Raw data. (b) Smoothed data.

Next, the optimal number of clusters (k) was determined. Various methods for determining k include elbow method, gap statistic, silhouette coefficient, and canopy [42]. The elbow method is utilized in this study, which is the most frequently used k determination method. It is used to select k as the point at which cluster variability (within-cluster sum of squares) becomes smooth with an increase in the number of clusters [46]. For that reason, it was appropriate when the value of k is 3, which is the inflection point of the graph, as illustrated in Figure 2. The three clusters classified through this can be characterized as follows: Cluster 1 is a road with a speed limit of 50 km/h or slower and three lanes or less. Meanwhile, Cluster 2 is a road with a speed limit of 60 km/h or higher and four lanes or more, and Cluster 3 is a road with a speed limit of 60 km/h or higher and three lanes or less. The cluster analysis result was used to find the travel speed pattern most similar to the past when predicting the travel speed of the work zone through the KNN algorithm.

3.2. CKNN Algorithm. The KNN algorithm is a nonparametric method used for classification or regression. It predicts situations by referring to the K training data, which is most similar to the input data [33]. The measure of similarity usually uses the Euclidean distance, which is preferred in predicting a short-term traffic condition because its basic model and calculation time are short, with data matching based on simple similarity. In particular, its prediction is excellent for complex nonlinear problems and can reflect traffic conditions with incidents or traffic jams.

Moreover, the KNN algorithm used in this study predicted the speed up to the forecast duration by referring to the training data of K numbers. This was most similar to the travel speed pattern data during the lag duration before the road work's starting time (t). The detailed analysis procedure of the KNN algorithm is presented in Figure 3. When designing the KNN algorithm, 1,453 out of 1,815 units of data were used as training data for predictive model design, and 362 units of data were utilized as test data for finding the K and algorithm verification.

$$d = \sqrt{(V_t^c - V_t^h)^2 + (V_{t-1}^c - V_{t-1}^h)^2 + (V_{t-2}^c - V_{t-2}^h)^2 + \dots + (V_{t-l}^c - V_{t-l}^h)^2}. \quad (2)$$

Data with the same day of the week and link cluster number as the input data is filtered from historical travel speed statistics to find the most similar travel speed pattern to the past. This study referred to the CKNN algorithm because the cluster results were used to find similar travel speed patterns [33]. Then, equation (2) was used to calculate the Euclidean distance between travel speeds of lag duration, and K training data was selected in order of the smallest Euclidean distance.

Here, d is Euclidean distance, V_t is speed data at the road work start time t , V^c is real-time speed data, V^h is historical speed data, and l is lag duration.

Finally, the travel speed for each period (in five-minute intervals) since the start of road work of the most similar K training data was reflected in equation (3) to predict travel speed after the current road work commenced. The method for pattern matching using Euclidean distance is shown in Figure 4.

FIGURE 2: Determination of k using the elbow method.

FIGURE 3: Framework of clustered K-nearest neighbors (CKNN) algorithm.

FIGURE 4: Speed prediction method based on Euclidean distance.

$$V_{t+j}^p = \frac{1}{K} \sum_{i=1}^{i=K} V_{t+j}^h(i). \quad (3)$$

In equation (3), V^p is the predicted travel speed, V^h is the historical speed data, $V(i)$ is i -th nearest neighbor speed, K is the number of nearest neighbors, and j is the prediction horizon (every five minutes).

3.3. Selection of Optimal K and Appropriate Lag Duration (in CKNN Algorithm). When designing the CKNN algorithm, it is imperative to determine the lag duration required for Euclidean distance calculation and the optimal K and forecast duration. The lag duration and K can be selected using mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE), which are methods for evaluating the predictive power of a model. In this study, predictions were performed by varying the lag duration and K value using test data, and the results were compared to select the most suitable lag duration and K for the model. The lag duration and the optimal K were chosen based on the prediction accuracy of the CKNN algorithm using equations (4)–(6) or the three error criteria.

$$\text{MAPE}(\%) = \frac{1}{n} \frac{1}{m} \sum_{i=1}^{i=n} \left(\sum_{j=1}^{j=m} \left| \frac{V_{t+j}^a(i) - V_{t+j}^p(i)}{V_{t+j}^a(i)} \right| \right) \times 100, \quad (4)$$

$$\text{MAE} = \frac{1}{n} \frac{1}{m} \sum_{i=1}^{i=n} \left(\sum_{j=1}^{j=m} |V_{t+j}^a(i) - V_{t+j}^p(i)| \right), \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \frac{1}{m} \sum_{i=1}^{i=n} \left(\sum_{j=1}^{j=m} (V_{t+j}^a(i) - V_{t+j}^p(i))^2 \right)}. \quad (6)$$

Here, V^a is actual speed data, V^p is the predicted travel speed, $V(i)$ is i -th work zone link speed, n is the number of data on the road works, and m is the number of time intervals for forecast duration.

3.4. Determination of the Impact of Road Work on the Network and Travel Speed Degradation. Road work in an urban area may or may not reduce the network traffic speed depending on the scale or type of work. Thus, a probability distribution model was applied to determine whether the actual road work causes the decrease in the travel speed under road work. The impact of road work on the network was determined by comparing the speed under normal network conditions with speed predicted through the CKNN algorithm by checking if the confidence level at 95% is met. Assuming that speeds under normal conditions were the standard normal distribution when the predicted value satisfied the 95% confidence level of the average speed under normal conditions, the road work had no impact on the network.

$$z = \frac{|V_i^p - \bar{v}_i^h|}{\sigma}. \quad (7)$$

In equation (7), V_i^p is the predicted travel speed in time i , $\bar{v}_i^h$ is the average speed in time i , and σ is the standard deviation in normal speed in time i .

FIGURE 5: Impact of lag duration and number of candidates on prediction error. (a) Forecast duration: 1 hour. (b) Forecast duration: 2 hours. (c) Forecast duration: 3 hours.

FIGURE 6: Optimal K given the lag duration is 20 min.

If the z-score calculated through equation (7) is higher than $z_{0.025}$, or the critical value of the 95% confidence interval of the average travel speed at normal conditions, the road work affected the network. When calculating the predicted network degradation caused by road work, equation (8) must be applied to determine the speed degradation against the average speed.

$$\text{speed degradation (\%)} = \frac{V_i^p - \bar{v}_i^h}{\bar{v}_i^h} \times 100. \quad (8)$$

3.5. Case Study. In this section, the forecast accuracy of the prediction algorithm was evaluated by selecting the optimal K and lag duration when predicting the travel speed of the work zone. The predicted speed during road work and the normal speed without road work were determined, including the work zone's impact on network and travel speed degradation.

First, the CKNN algorithm and 362 test data were used to select the optimal K and lag duration. Based on the three error criteria presented above (equations (4)–(6)), the accuracy of the CKNN algorithm was analyzed according to the increase in K and lag duration for each prediction time (one hour, two hours, three hours), as illustrated in Figure 5.

Second, the prediction accuracy based on the three error criteria was most appropriate when the lag duration was 20 minutes. Thus, the speed pattern for the previous 20 minutes should be used when designing the CKNN algorithm. Forecast duration was less accurate when forecasting for a long time, so one hour was identified as the most suitable. Lastly, to find the optimal K, a predictive power evaluation was performed according to the change of the K value when the lag duration was 20 minutes and the forecast duration was 1 hour (Figure 6).

Table 4 shows the values of MAPE, MAE, and RMSE when K has values from 1 to 10. Figure 6 shows that when the K value is ten or more, the value of each criterion continues to increase upward; hence it is unable to find the minimum value. As a result, MAPE obtained the minimum value when K was 5, MAE reached the minimum value when K was 2, and RMSE acquired the minimum value when K was 2 and 5. Since the difference between the MAE values when K was 2 and 5 was detected as small and with

TABLE 4: Mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE) according to K values when lag duration is 20 minutes.

K	MAPE (%)	MAE	RMSE
1	7.54	2.35	2.66
2	7.02	2.20	2.48
3	7.02	2.21	2.49
4	6.97	2.23	2.50
5	6.90	2.22	2.48
6	6.97	2.26	2.52
7	6.98	2.28	2.53
8	7.05	2.29	2.54
9	7.12	2.32	2.57
10	7.26	2.37	2.61

scale-dependent errors, the K value was then identified as 5, which minimized MAPE.

The test details were used as the input data to verify the model for predicting travel speed in a work zone and the travel speed for an hour after the road work was predicted. The result is presented in Figure 7.

As a result of predicting the test set using the CKNN algorithm, the average MAPE, 6.9%, exhibited excellent predictive power, as indicated in Table 5. In some cases, the MAPE for the predicted value exceeded 15%, but most of them were predictable within 10%. Thus, the model accuracy was considered high [47].

Table 6 specifies the results of analyzing the network impact due to the road work carried out at 10:25 am on Friday, February 15, 2019. The network was classified as Cluster 1. Using the CKNN algorithm, the travel speed one hour after road work was predicted and compared with the network under normal conditions.

Consequently, there was no difference between the speed predicted by CKNN and the normal speed at the beginning of the road work. It was found that the network effect occurred from about 25 minutes after the start of the work, and the speed decreased by about 11%–17% compared to normal conditions, suggesting that road or traffic managers need to establish a strategy to reduce congestion about 30 minutes after starting. Predicting the speed and judging the network impact can also forecast congestion intensity by time, enabling more active and preemptive traffic and congestion management.

FIGURE 7: Mean absolute percentage error (MAPE) results at 1-hour prediction ($k = 5$, lag duration = 20 min).

TABLE 5: Prediction reliability based on mean absolute percentage error (MAPE).

MAPE	Prediction reliability
$0\% \leq \text{MAPE} < 10\%$	Very accurate prediction
$10\% \leq \text{MAPE} < 20\%$	Accurate prediction
$20\% \leq \text{MAPE} < 50\%$	Reasonable prediction
$50\% > \text{MAPE}$	Not accurate prediction

TABLE 6: Impact analysis result of road work based on actual data.

Time	Predicted travel speed (km/h)	Average travel speed at normal conditions (km/h)	Standard deviation of travel speed at normal conditions (km/h)	z-score	Whether the construction affects the speed	Speed degradation (%)
$t + 1$ 10:30	19.36	21.29	1.51	1.28	No	—
$t + 2$ 10:35	19.48	21.65	1.50	1.45	No	—
$t + 3$ 10:40	19.52	21.65	1.45	1.47	No	—
$t + 4$ 10:45	19.48	21.77	0.92	2.49	No	—
$t + 5$ 10:50	19.44	21.97	1.35	1.87	Yes	-11
$t + 6$ 10:55	19.28	21.58	1.12	2.06	No	—
$t + 7$ 11:00	18.84	21.84	1.21	2.47	Yes	-11
$t + 8$ 11:05	18.64	22.35	1.60	2.32	Yes	-14
$t + 9$ 11:10	18.60	22.39	1.60	2.36	Yes	-17
$t + 10$ 11:15	18.56	22.45	1.43	2.71	Yes	-17
$t + 11$ 11:20	18.72	21.94	0.89	3.60	Yes	-15
$t + 12$ 11:25	19.04	22.10	0.79	3.87	Yes	-14

4. Conclusion

It is crucial to prepare an appropriate traffic management strategy for the expected congestion level by predicting the travel speed after road work to prevent congestion caused by road works. This study developed a model that predicts the travel speed of the work zone using the CKNN algorithm. Furthermore, a method to grasp how much the traffic speed decreases due to road work was compared with the normal speed pattern.

Most proposed methodologies for short-term speed prediction presented by several existing studies were methods for predicting speed in normal road conditions. Since roads in the work zone were entirely or partially blocked, a speed pattern differed from normal road conditions. Applying the proposed methodology through a case study can accurately predict the speed from the start of road construction up to an hour later. Furthermore, it was likewise feasible to provide useful information for preemptive traffic congestion management by detecting the timing of link speed degradation caused by capacity reduction due to road work.

However, this study had limitations that need improvement through future studies. First, the established model for predicting travel speeds in a work zone filtered the data using the day of the week and link clusters classified according to road characteristics. Still, it is necessary to use work type as a filter. For example, road works that block or occupy roads largely affect traffic conditions. However, work conducted on drains or sidewalks will only slightly influence traffic conditions. Therefore, better results can be achieved if the travel speed of the work zone can be predicted by considering the work type.

Second, a prediction model was developed using eight-month data for major arterial roads installed with traffic information collection devices. Although the amount of data was not small, it was still insufficient for securing details similar to the input data.

Third and last, this study used only the CKNN algorithm for speed prediction. However, evaluating the appropriateness of the methodology proposed in this study compared to results predicted by other clustering methods such as support vector machines, random forests, and neural networks is required.

Data Availability

The data used to support the findings of this study are not publicly made available according to the data security policy of Daegu Metropolitan City.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the 2021 Research Fund of the University of Seoul.

References

- [1] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.
- [2] E. J. Kim, H. C. Park, S. Y. Kho, and D. K. Kim, "A hybrid approach based on variational mode decomposition for analyzing and predicting urban travel speed," *Journal of Advanced Transportation*, vol. 2019, Article ID 3958127, 12 pages, 2019.
- [3] H. C. Park, S. Kang, S. Y. Kho, and D. K. Kim, "Investigation of effects of inherent variation and spatiotemporal dependency on urban travel-speed prediction," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 5, 2020.
- [4] Y. Hou, P. Edara, and C. Sun, "Traffic flow forecasting for urban work zones," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1761–1770, 2014.
- [5] Z. Zheng and D. Su, "Short-term traffic volume forecasting: a K-Nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 143–157, 2014.
- [6] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.
- [7] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1678, no. 1, pp. 179–188, 1999.
- [8] D. Billings and J. S. Yang, "Application of the ARIMA models to urban roadway travel time prediction—a case study," in *2006 IEEE International Conference on Systems, Man and Cybernetics*, vol. 3, pp. 2529–2534, 2006.
- [9] A. M. Khoei, A. Bhaskar, and E. Chung, "Travel time prediction on signalised urban arterials by applying SARIMA modelling on Bluetooth data," in *Proceedings of the 36th Australasian Transport Research Forum (ATRF)*, Brisbane, Australia, 2013.
- [10] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6164–6173, 2009.
- [11] L. Kang, G. Hu, H. Huang, W. Lu, and L. Liu, "Urban traffic travel time short-term prediction model based on spatio-temporal feature extraction," *Journal of Advanced Transportation*, vol. 2020, Article ID 3247847, 16 pages, 2020.
- [12] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.
- [13] L. Li, S. He, J. Zhang, and B. Ran, "Short-term highway traffic flow prediction based on a hybrid strategy considering temporal-spatial information," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 2029–2040, 2016.
- [14] B. Sun, W. Cheng, P. Goswami, and G. Bai, "Short-term traffic forecasting using self-adjusting k-nearest neighbours," *IET Intelligent Transport Systems*, vol. 12, no. 1, pp. 41–48, 2017.
- [15] H.-C. Park, D.-K. Kim, and S.-Y. Kho, "Bayesian network for freeway traffic state prediction," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 45, pp. 124–135, 2018.
- [16] Y. Cong, J. Wang, and X. Li, "Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm," *Procedia Engineering*, vol. 137, pp. 59–68, 2016.
- [17] P. Wu, Z. Huang, Y. Pian, L. Xu, J. Li, and K. Chen, "A combined deep learning method with attention-based LSTM model for short-term traffic speed forecasting," *Journal of Advanced Transportation*, vol. 2020, Article ID 8863724, 15 pages, 2020.
- [18] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 61–78, 2016.
- [19] H. F. Yang, T. S. Dillon, and Y. P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2371–2381, 2016.
- [20] Y. J. Lee and O. Min, "Long short-term memory recurrent neural network for urban traffic prediction: a case study of Seoul," in *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1279–1284, Maui, HA, USA, November, 2018.
- [21] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.
- [22] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.
- [23] Y. Jia, J. Wu, and Y. Du, "Traffic speed prediction using deep learning method," in *Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1217–1222, Rio de Janeiro, Brazil, December, 2016.
- [24] Y. Lv, Y. Duan, and W. Kang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2014.
- [25] C. Song, H. Lee, C. Kang, and W. Lee, "Traffic speed prediction under weekday using convolutional neural networks concepts," in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1293–1298, Los Angeles, CA, USA, June, 2017.
- [26] C. Zhang, J. J. Q. Yu, and Y. Liu, "Spatial-temporal graph attention networks: a deep learning approach for traffic forecasting," *IEEE Access*, vol. 7, pp. 166246–166256, 2019.

- [27] X. Luo, D. Li, Y. Yang, and S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *Journal of Advanced Transportation*, vol. 2019, Article ID 4145353, 10 pages, 2019.
- [28] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin, "Deep learning on traffic prediction: methods, analysis and future directions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, 2021.
- [29] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 871–882, 2013.
- [30] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 65–78, 2014.
- [31] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 166–180, 2018.
- [32] T. Liu, J. Ma, W. Guan, Y. Song, and H. Niu, "Bus arrival time prediction based on the K-Nearest neighbor method," in *Proceedings of the 2012 Fifth International Joint Conference on Computational Sciences and Optimization*, pp. 480–483, Heilongjiang, China, June, 2012.
- [33] M. Akbari, P. J. v. Overloop, and A. Afshar, "Clustered K nearest neighbor algorithm for daily inflow forecasting," *Water Resources Management*, vol. 25, no. 5, pp. 1341–1357, 2011.
- [34] M. Kamyab, S. Remias, E. Najmi, S. Rabinia, and J. M. Waddell, "Machine learning approach to forecast work zone mobility using probe vehicle data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 9, pp. 157–167, 2020.
- [35] H. Wang, L. Liu, S. Dong, Z. Qian, and H. Wei, "A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework," *Transportation Business: Transport Dynamics*, vol. 4, no. 3, pp. 159–186, 2016.
- [36] D. R. Taylor, S. Muthiah, B. T. Kulakowski, K. M. Mahoney, and R. J. Porter, "Artificial neural network speed profile model for construction work zones on high-speed highways," *Journal of Transportation Engineering*, vol. 133, no. 3, pp. 198–204, 2007.
- [37] X. Jiang and H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," *Journal of Transportation Engineering*, vol. 131, no. 10, pp. 771–779, 2005.
- [38] H. T. Zwahlen and A. Russ, "Evaluation of the accuracy of a real-time travel time prediction system in a freeway construction work zone," *Transportation Research Record*, vol. 1803, no. 1, pp. 87–93, 2002.
- [39] Q. Meng and J. Weng, "Cellular automata model for work zone traffic," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2188, no. 1, pp. 131–139, 2010.
- [40] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, "A spatiotemporal correlative K-Nearest neighbor model for short-term traffic multistep forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 21–34, 2016.
- [41] H. Xiaoyu, W. Yisheng, and H. Siyu, "Short-term traffic flow forecasting based on two-tier K-Nearest neighbor algorithm," *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 2529–2536, 2013.
- [42] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *D-J Series*, vol. 2, no. 2, pp. 226–235, 2019.
- [43] Transportation Research Board, *Highway Capacity Manual 2010*, TRB National Research Council, Washington, DC, USA, 2010.
- [44] A. H. Mashhadi, M. Farhadmanesh, A. Rashidi, and N. Marković, "Review of Methods for Estimating Construction Work Zone Capacity," *Transportation Research Record*, vol. 2675, 2021.
- [45] A. Dhamaniya and S. Chandra, "Influence of operating speed on capacity of urban arterial midblock sections," *International Journal of Civil Engineering*, vol. 15, no. 7, pp. 1053–1062, 2017.
- [46] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-means clustering," *International Journal*, vol. 1, no. 6, pp. 90–95, 2013.
- [47] C. D. Lewis, *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*, Butterworth Scientific, London, UK, 1982.