IET Intelligent Transport Systems

Special Section: Innovative Data, Methodology and Technology for Future Transportation Systems: Selected Papers from the 19th COTA International Conference of Transportation Professionals

Time-aware gated recurrent unit networks for forecasting road surface friction using historical data with missing values

ISSN 1751-956X
Received on 3rd July 2019
Revised 8th December 2019
Accepted on 5th February 2020
E-First on 13th March 2020
doi: 10.1049/iet-its.2019.0428
www.ietdl.org

Ziyuan Pu[1], Zhiyong Cui[1], Shuo Wang[2], Qianmu Li[2], Yinhai Wang[1] ✉

[1]Smart Transportation Application and Research (STAR) Laboratory, Department of Civil and Environmental Engineering, University of Washington, 101 More Hall, Seattle, WA, USA

[2]School of Computer Science and Engineering, Nanjing University of Science and Technology, 200 Xiaolingwei Street, Nanjing, Jiangsu, People's Republic of China

✉ *E-mail: yinhai@uw.edu*

Abstract: An accurate road surface friction forecasting algorithm can allow travelers and managers to schedule trips and maintenance activities based on the road weather condition to enhance traffic safety and efficiency in advance. Previously, scholars developed multiple forecasting models to predict road surface conditions using historical data. However, historical dataset used for model training may have missing values caused by multiple issues, e.g. the data collected by on-vehicle sensors may be influenced when vehicles cannot travel due to high economic and labor cost or weather-related issues. The missing values in the road surface condition dataset can damage the effectiveness and accuracy of the existing prediction methods. This study proposed a road surface friction forecasting algorithm by employing a time-aware Gated Recurrent Unit (GRU-D) networks that integrate a decay mechanism as extra gates of the GRU to handle the missing values and forecast the road surface friction in future periods simultaneously. The evaluation results present that the proposed GRU-D networks outperform all selected baseline algorithms. The impacts of missing rate on predictive accuracy, learning efficiency, and learned decay rates are investigated as well. The findings can help improve the forecasting accuracy and efficiency of road surface friction prediction using historical data with missing values, therefore mitigating the negative impact of wet or icy road conditions on traffic safety and efficiency.

1 Introduction

Road surface friction is defined as the resistance to the motion between vehicle and road surface, which has a strong influence on the distance, which is required for a vehicle to decelerate and the minimum required reaction time when a vehicle requires to break for avoiding collisions [1]. In the winter season, road surface friction reduces substantially caused by dramatically decreasing air temperature, which is one of the major causal factors of car accidents [2, 3]. FHWA reports that, in the United States, the majority of traffic accidents were reported to happen during wet or icy road surface conditions, where 73% of accidents occurred on wet pavements, and 17% on snow or sleet [4]. Besides, existing studies indicated that the intelligent systems which have the capacity for sharing the timely road condition-related information could potentially mitigate the negative influence of road surface condition on traffic safety [5]. Thus, considering the road surface friction is a direct quantitative measurement of road surface conditions, an efficient and cost-effective road surface friction prediction methodology is needed for improving traffic safety.

Previously, several sensing technologies were developed for monitoring road surface condition parameters, e.g. DSC-111 and DST-111 sensor sets [6–8], Road Condition Monitor (RCM) 411 sensor [9], and image-based sensing technology [10]. Such sensing technologies can sense the road surface state (dry, moist, wet, icy, snowy/frosty, or slushy), pavement surface temperature, air temperature, relative humidity, etc. Previous studies evaluated the performance of existing sensing technologies, e.g. RCM-411 is accurate in friction level, and road surface status detection [9, 11, 12], DSC-111 can provide accurate surface state measurement. Still, the friction detection of DST-111 is not precise [7], etc. Some of these technologies have already been employed for real-time road monitoring implementations, e.g. Road Weather Information Station (RWIS) in the US and other countries [11, 13–16].

By having such affluent data resources, scholars have developed plenty of data-driven estimation and prediction models for road surface condition forecasting. Liu developed a road surface temperature prediction model based on the gradient extreme learning machine boosting algorithm [17]. Solol developed a road surface temperature prediction model based on energy balance and heat conduction models [18]. Also, some researchers developed road surface condition recognition algorithms based on computer vision technologies [19–23]. However, these studies can only estimate the road surface condition in the current period based on environmental measurements, e.g. air temperature, humidity, etc. They are not able to predict road surface conditions in the future. By considering the time-series features of road surface conditions [24], multiple studies established Artificial Intelligence-based prediction models for forecasting short-term road surface condition parameters [25–27]. The inputs to these models are the historical road surface condition measurements with a fixed temporal resolution for each time step. The output of the forecasting model is the same road surface condition parameter in the next time step. For example, if the target of the model is to predict the road surface condition one-day after, the time interval of each timestamp should be 1-day in the input dataset. However, for road surface friction data collection, it cannot guarantee to collect data at every consecutive timestamp due to weather or cost issues, e.g. the data collected by on-vehicle sensors may have missing values when the vehicle cannot travel due to weather or vehicle maintenance issues. Such missing values can damage the accuracy and effectiveness of the existing models.

Based on the above considerations, the primary objective of this study is to develop a road surface friction forecasting algorithm based on time-aware Gated Recurrent Unit (GRU-D) networks to handle missing values. The road surface friction data sensed by the RCM-411 sensor were employed as the algorithm input in this research. To evaluate the predictive performance of the proposed algorithm, several baseline predictions and imputation models were

selected for the comparison purpose. Besides the prediction performance evaluation, the impact of missing rates, the learning efficiency, and the learned decay rates was also investigated. The findings of this study can help to improve the prediction model effectiveness by handling missing values, so that mitigate the negative impact of road surface conditions on road traffic safety and mobility, especially in the winter season.

2 Methodology

2.1 Forecasting road surface friction using GRU-D networks

2.1.1 GRU structure: A multivariate time series with D variables of length T is denoted as $X = (x_1, x_2, \ldots, x_T)^D \in \mathbb{R}^{T \times D}$, where for each time step $t \in \{1, 2, \ldots, T\}, x_t \in \mathbb{R}^D$ represents the measurements of all variables at a time step t and x_t^d denotes the measurements of the dth variable of x_t. In this study, only road surface friction at a time step t is considered as a variable. Thus, D equals 1 and a time series measurement is a $T \times 1$ vector. Then the T-length sequence vector is used as the input of each GRU. Fig. 1*a* shows the original GRU structure. As shown in the figure, each jth GRU utilise a reset gate r_t^j and an update gate z_t^j to control the hidden state h_t^j at the tth time step. The following equations show how the parameters are updated in each GRU [28].

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + \boldsymbol{b}_r) \tag{1}$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + \boldsymbol{b}_z) \tag{2}$$

$$\widetilde{h_t} = \tanh W x_t + U(r_t \odot h_{t-1} + \boldsymbol{b}) \tag{3}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \widetilde{h_t} \tag{4}$$

where W_r, W_z, W, U_r, U_z, U and vectors $\boldsymbol{b}_r, \boldsymbol{b}_z, \boldsymbol{b}$ are model parameters. σ is used for element-wise sigmoid function, and $\odot$ is the element-wise multiplication. In the GRU update formulation, all measurements are assumed as the observed values with zero missing rate. For handling missing values, the GRU-D unit will be introduced in the next section.

2.1.2 GRU-D unit structure: For the time series X with missing values, there are two representations of the missing pattern handled in the GRU-D unit, which is masking and time interval [29]. Masking conveys the information of which inputs are observed or absent to the model, while time interval informs GRU-D unit the time-series patterns of the input observations. A masking vector $\boldsymbol{m}_t \in \{0, 1\}^D$ and the time interval vector $\boldsymbol{\delta}_t^d \in \mathbb{R}$, which are shown in (5) and (6) for capturing the masking and time interval of the input observations.

$$m_t^d = \begin{cases} 1, & \text{if } x_t^d \text{ is observed} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$\boldsymbol{\delta}_t^d \begin{cases} s_t - s_{t-1} + \delta_{t-1}^d, & t > 1, m_{t-1}^d = 0 \\ s_t - s_{t-1}, & t < 1, m_{t-1}^d = 1 \\ 0, & t = 1 \end{cases} \tag{6}$$

where $s_t \in \mathbb{R}$ denotes the time stamp when the tth measurement is observed and the first measurement is observed at time stamp 0. Equation (7) provides an example of the masking and time interval vectors

$$\begin{cases} \boldsymbol{X} = [x_1 \quad x_2 \quad \text{missing} \quad \text{missing} \quad x_5 \quad x_6 \quad \text{missing} \quad x_8] \\ \boldsymbol{S} = [0 \quad 2 \quad 5 \quad 6 \quad 10 \quad 12 \quad 13 \quad 18] \\ \boldsymbol{M} = [1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 1] \\ \boldsymbol{\Delta} = [0 \quad 2 \quad 3 \quad 1 \quad 8 \quad 2 \quad 1 \quad 6] \end{cases} \tag{7}$$

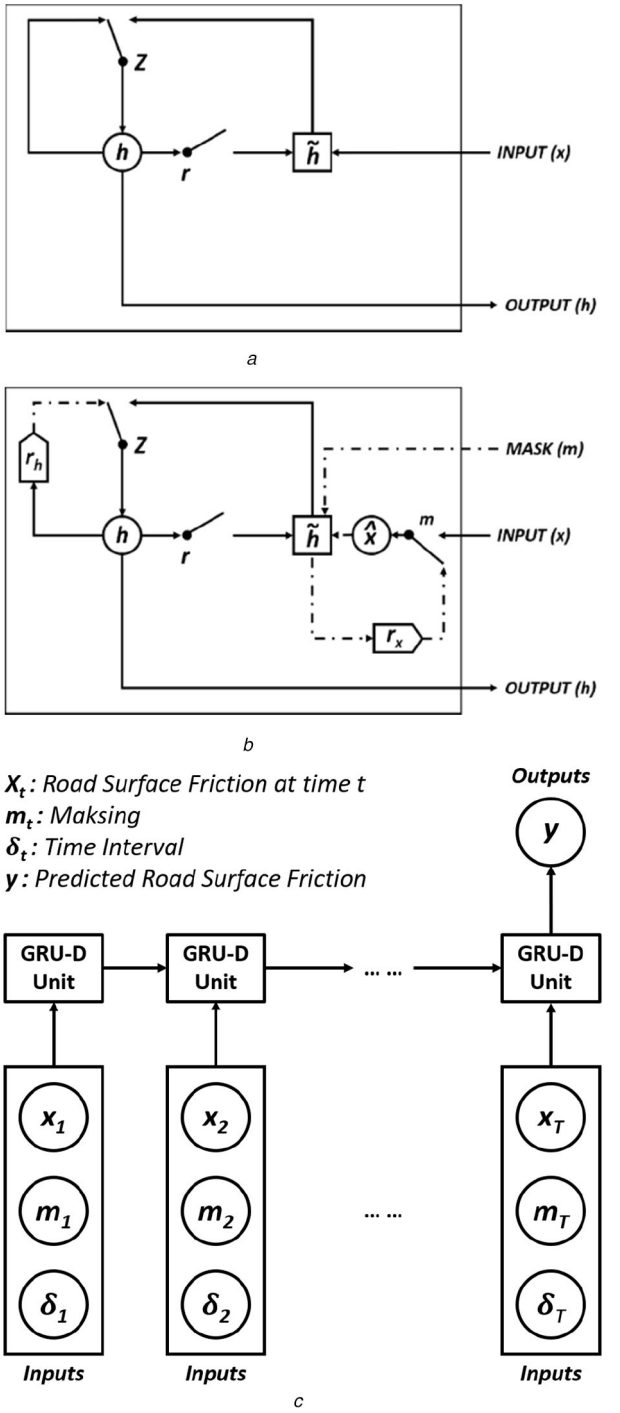


Fig. 1 *Graphical illustrations of the algorithm structure*
(a) Original GRU structure, ***(b)*** GRU-D unit structure, ***(c)*** GRU-D networks structure

where $\boldsymbol{X}$ is an 8-length time-series measurement with three missing values at time steps 3, 4 and 7. $\boldsymbol{S}$ is the vector denote the time stamp when the measurement x_t is observed. $\boldsymbol{M}$ is the vector representing the missing status of x_t. Then, the input data of each GRU-D unit is formed as $D = \{(X_n, s_n, M_n)_{n=1}^N\}$, where N is the total number of data points in input dataset.

As shown in Fig. 1*b*, GRU-D unit deploys two decay rates γ_{x_t} and γ_{h_t} to handle the missing values. The following equation calculates the decay rate:

$$\gamma_t = \exp\{-\max(0, W_\gamma \delta_t + b_\gamma)\} \tag{8}$$

where W_γ and b_γ are model parameters that are trained jointly with all the other parameters of the GRU-D unit. The decay rates are learned from the training data, and the exponentiated negative

rectifier is deployed to keep the decay rate in a range between 0 and 1. Then, the input and hidden states are calculated based on the following equations:

$$\hat{x}_t^d = m_t^d x_t^d + \left(1 - m_t^d\right)\left(\gamma_{x_t}^d x_{t'}^d + \left(1 - \gamma_{x_t}^d\right)\tilde{x}^d\right) \tag{9}$$

$$\hat{h}_{t-1} = \gamma_{h_t} \odot h_{t-1} \tag{10}$$

where $x_{t'}^d$ is the last observation of the current time step, and $\tilde{x}^d$ is the empirical mean of the training dataset. h_{t-1} is the hidden state of the previous GRU-D unit. Then, the update functions of GRU-D unit would be shown in the following equations:

$$r_t = \sigma\left(W_r \hat{x}_t + U_r \hat{h}_{t-1} + V_r \boldsymbol{m}_t + \boldsymbol{b}_r\right) \tag{11}$$

$$z_t = \sigma\left(W_z \hat{x}_t + U_z \hat{h}_{t-1} + V_z \boldsymbol{m}_t + \boldsymbol{b}_z\right) \tag{12}$$

$$\widetilde{h_t} = \tanh\left(W \hat{x}_t + U\left(r_t \odot \hat{h}_{t-1}\right) + V \boldsymbol{m}_t + \boldsymbol{b}\right) \tag{13}$$

$$h_t = (1 - z_t) \odot \hat{h}_{t-1} + z_t \odot \widetilde{h_t} \tag{14}$$

where x_t and h_{t-1} are substituted by $\hat{x}_t$ and $\hat{h}_{t-1}$ from (9) and (10). The masking vector $\boldsymbol{m}_t$ is integrated with the model, and V_r, V_z, and V are new parameters for the masking vector. In this paper, a GRU-D network is formed by t-length GRU-D unit for the road surface friction prediction task, which is shown in Fig. 1*c*.

2.2 Predictive performance evaluation

2.2.1 Baseline missing data imputation methods: To evaluate the performance of the decay mechanism of the GRU-D unit for handling missing values, three simple methods are used to handle the missing values, which are referred as average, last, and simple. Equations (15)–(17) present how they handle missing values

$$x_t^d \leftarrow m_t^d x_t^d + \left(1 - m_t^d\right)\tilde{x}^d \tag{15}$$

$$x_t^d \leftarrow m_t^d x_t^d + \left(1 - m_t^d\right)x_{t'}^d \tag{16}$$

$$x_t^d \leftarrow m_t^d x_t^d + \left(1 - m_t^d\right)\left(\hat{\delta}_t^d x_{t'}^d + \left(1 - \hat{\delta}_t^d\right)\tilde{x}^d\right) \tag{17}$$

where

$$\tilde{x}^d = \sum_{n=1}^{N}\sum_{t=1}^{T_n} m_{t,n}^d x_{t,n}^d / \sum_{n=1}^{N}\sum_{t=1}^{T_n} m_{t,n}^d,$$

which is the historical average value of the road surface friction, and it is calculated using the training dataset and used for both training and testing datasets. $x_{t'}^d$ is the last road surface friction measurements of the current time step t, and $\hat{\delta}_t^d$ is the normalised time interval at a time step t.

2.2.2 Baseline prediction models: The performance of GRU-D networks in road surface friction forecasting is compared to that of many classical baseline short-term prediction models. Typically, ARIMA, support vector regression (SVR), random forest (RF), Kalman filter, tree-based model, feed-forward NN, and LSTM NN were used for addressing short-term prediction problems [30–34], e.g. traffic speed, travel time and driving behaviour prediction [35–39]. However, several time-series prediction models were demonstrated that the predictive performance is not as accurate as others, e.g. ARIMA and Kalman filter. Therefore, based on the previous research results, SVR, RF, feed-forward NN, and LSTM NN were selected for comparing the performance of road surface friction prediction with the proposed GRU-D NN model in this study. Among these models, feed-forward NN, which is also called multilayer perceptron, and LSTM NN are popular for precise performance in short-term prediction [33, 40]. RF and SVR are also well-known models for efficient predictive performance [30, 32]. For the parameters of model development, the radial basis function (RBF) kernel is deployed in the SVR model. Ten trees were built, and there was no predefined limitation for the maximum depth of the trees for the RF model. The feed-forward NN and LSTM NN were composed of two hidden layers.

2.2.3 Evaluation measurements: Mean absolute error (MAE), mean square error (MSE), and mean absolute percentage error (MAPE) are used as the measurements of predictive performance. The following equations present the measurement formulation:

$$\text{MAE} = \frac{\sum_{i=1}^{N}\left|Y_i - \widehat{Y_i}\right|}{N} \tag{18}$$

$$\text{MSE} = \frac{\sum_{i=1}^{N}\left(Y_i - \widehat{Y_i}\right)^2}{N} \tag{19}$$

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \tag{20}$$

where N is the total number of samples in the testing data set, Y_i is the ground truth of the road surface friction, which is detected by the RCM 411 sensor in this study, and $\hat{Y}$ is the predicted road surface friction of the proposed prediction model. Typically, the MAE is used to measure the absolute error associated with a prediction, the MAPE presents a measure of the percentage of average misprediction of the model and the MSE measures the relative error for a prediction. The prediction model with the smaller values of MAE, MSE, and MAPE performs better.

3 Data

3.1 Testing field

The data used in this study were collected by an on-vehicle RCM 411 sensor on the road segments of the E75 route from Sodankylä to Kemi in Finland. The total length of the testing field is 186 miles. In the winter season, from October to next April, the air temperature is historically relatively low in the testing field. It could be as low as negative 40°C, and the average minimum air temperature is about negative 15°C. The average maximum air temperature is still under the water ice point for most of the time. In the other seasons, the air temperature is not as high as the temperature in common areas. The historical average maximum air temperature in July is about 20°C when is the month with the highest temperature in this area. Therefore, the testing field has road weather issues, e.g. icing and snow happened with high frequency. There was a detection vehicle equipped with an RCM 411 sensor to keep collecting the road surface friction data since February 2017. In addition, the sensor system also estimates the road surface status based on road surface friction and other related measurements. Six statuses are used to label the road surface, including dry, wet, moist, slush, ice, and snow or hoar frost. Fig. 2 shows the spatial distribution of the calculated road surface status of two randomly selected days. One is in winter and another one is in summer. In the figure, calculated road surface status was variated along the road. Most of the road was covered by snow or hoar frost in winter, while most of the road was dried in summer.

3.2 Data description

3.2.1 Road segmentation: As road surface friction on adjacent road segments has no spatial correlation, we proposed a spatial clustering method based on the K-means clustering algorithm to segment the study site into 1487 road segments. The detailed road segmentation method was introduced in the previous work [27]. Fig. 2 visualises the examples of road segmentation results in the small image patches. There is only one road surface status exists within the road segments. After segmenting the testing field based

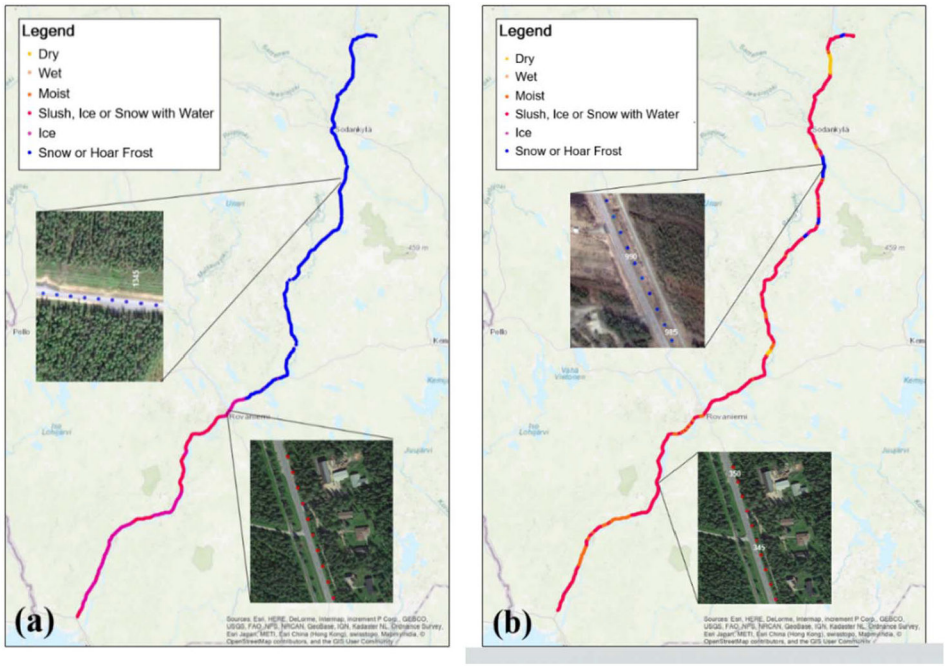


Fig. 2 *Road surface status spatial distribution and examples of road segmentation results*
(a) Winter season, February 14th, 2018, ***(b)*** Summer season, August 21st, 2018

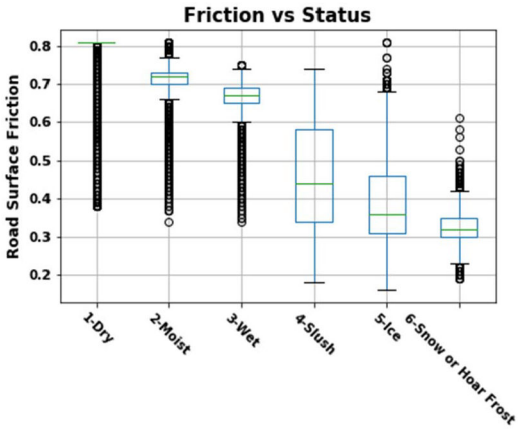


Fig. 3 *Boxplot of road surface friction with different surface status in the study site*

on the proposed criteria, the temporally average road surface friction could be used to represent the measurements of each road segment at a specific timestamp. Otherwise, the road surface friction measurements could spatially vary a lot at one timestamp.

3.2.2 Statistical description: The data used in this study was collected by the Road Condition Monitor (RCM) 411 sensor, which is an optical sensing-based on-vehicle road surface condition sensor [9]. The dataset covers 446 days from February 17th, 2017 to May 9th, 2018. During this period, the vehicle equipped with an RCM 411 sensor traversed the testing field at least once per day, so that every road segment has at least one friction measurement for every single day.

As mentioned before, besides the road surface friction, RCM 411 sensor also provides the calculated road surface status with high accuracy [12]. In most cases, road surface status is defined or estimated based on the friction information. In addition, the road surface friction coefficient is the most important indicator to characterise the anti-sliding performance of a road segment. Fig. 3 shows the road surface friction distribution with different road surface status in the testing field. As seen in the plot, the road surface friction decreased when the road surface status got worse. When the road surface was not covered by any type of ice or snow, the average road surface friction ranged from 0.75 to 0.8. Once any kind of ice or snow existed on the road surface, the average road surface friction dropped under 0.5. According to the literature, the stopping distance is almost doubled when road surface friction drops from 0.8 to 0.5 [41].

In addition, the road surface friction also has time-series features. Fig. 4 presents the road surface friction of five randomly selected road segments in the period from February 17th, 2017 to May 9th, 2018. The road surface friction of the five selected road segments shared a similar time-series trend. During the period from March to October 2017, the road surface friction was around 0.7 for most of the time. In other months, the road surface friction fluctuated a lot, and low friction value happened more frequently due to the winter weather influence.

3.3 Data preparation for GRU-D networks and baseline models

Since it is assumed no spatial correlation between road segments, each road will be modelled independently. Thus, the spatial dimension of the input data of each road is set as $P = 1$.

The time interval between time-steps is set as 1-day, then, the dataset has 446 time-steps for each road segment. Suppose the number of the time-lag is set as $T = t$ with 1 days between each time-step, which means the model used the data in previous t consecutive time-steps to predict the road surface friction in the following $t + 1$ time-step. Based on the previous work, $T = 7$ was demonstrated to generate the best predictive performance in this application [27]. Then the dataset is separated into samples with $T = 7$, and the sample size is $N = 446 - 7 = 439$. Thus, each sample of the input data, X_n, is a two-dimensional vector with the dimension of $[T, P] = [7, 1]$, and each sample of the output data is a one-dimensional vector with one component. The input of the model for each road segment is a three-dimensional vector which dimension is $[N, T, P] = [439, 7, 1]$. Before feeding into the model, all samples are randomly divided into the training, validation, and testing dataset with the ratio 7:2:1. All prediction models share the same data structure. For training GRU-D networks, the model was trained by minimising the mean squared error with the batch size of 16 and a learning rate of 10^{-5}. The RMSProp was employed as the gradient descent optimiser with the alpha of 0.99 and the early stopping mechanism was utilised to avoid the overfitting.

The original dataset has no missing values. For testing the performance of GRU-D networks, the masking vectors were randomly generated by given a ratio of missing values. Then, the

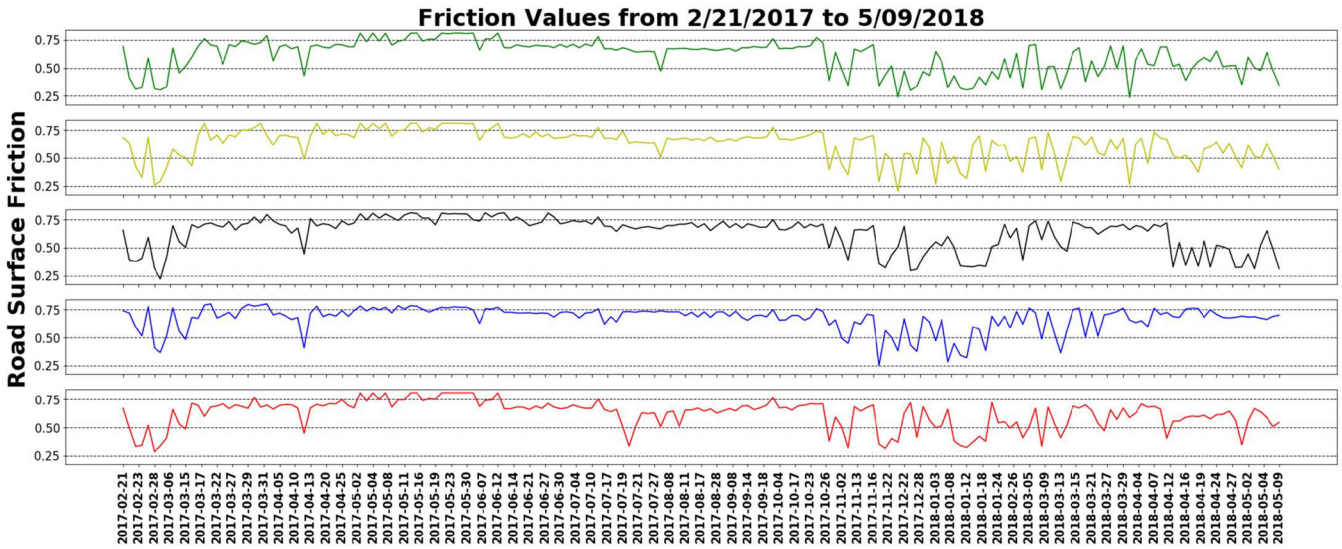


Fig. 4 *Temporal distribution of road surface friction of five randomly selected road segments*

Table 1 Predictive performance of GRU networks and baseline prediction models without missing values

Prediction models	MSE	MAE	MAPE, %
feed-forward NN	0.0281	0.1482	26.81
SVR	0.0161	0.1111	19.99
RF	0.0113	0.0792	15.00
LSTM	0.0085	0.0697	12.85
GRU	0.0080	0.0640	11.92

Table 2 Predictive performance of non-RNN baseline models with 20% missing values

Non-RNN models-imputation methods	MSE	MAE	MAPE, %
feed-forward NN-average	0.0301	0.1539	28.52
feed-forward NN-last	0.0301	0.1539	28.53
feed-forward NN-simple	0.0301	0.1538	28.50
SVR-average	0.0183	0.1152	21.01
SVR-last	0.0175	0.1139	20.94
SVR-simple	0.0192	0.1195	21.97
RF-average	0.0128	0.0860	16.20
RF-last	0.0119	0.0809	15.32
RF-simple	0.0114	0.0820	15.71

time interval vectors were calculated based on the corresponding masking vectors. As introduced in Section 2.2.1, three simple methods were employed to evaluate the performance of the decay mechanism. The randomly generated missing values were imputed by these three baseline imputation methods before input to the prediction model. All prediction models were combined with baseline imputation methods for testing. The performance was compared with GRU-D networks and will be introduced in the next section.

4 Numerical results

4.1 Predictive performance comparison of GRU and baseline prediction models

The prediction models were tested using the dataset without any missing values first. All prediction models were trained based on the same training dataset for each road segment independently, and the predictive performance for each model was calculated using the testing dataset. Table 1 shows the predictive performance of all prediction models. The presented measurements were the average value of all road segments. In general, GRU networks outperformed all prediction models in terms of all three measurements. Among non-RNN prediction models, RF performed much better than SVR and feed-forward NN in terms of all three metrics, which is reasonable due to the majority votes mechanism of the RF model. The feed-forward NN had the worst predictive performance, which is caused by the sparsity of the data. The vanilla LSTM model performed better than all non-RNN models due to beneficial of long and short-term memory. The GRU networks achieved better performance than the LSTM model due to the sparse data features and less training parameters.

Table 3 Predictive performance of and GRU-D and RNN baseline models with 20% missing values

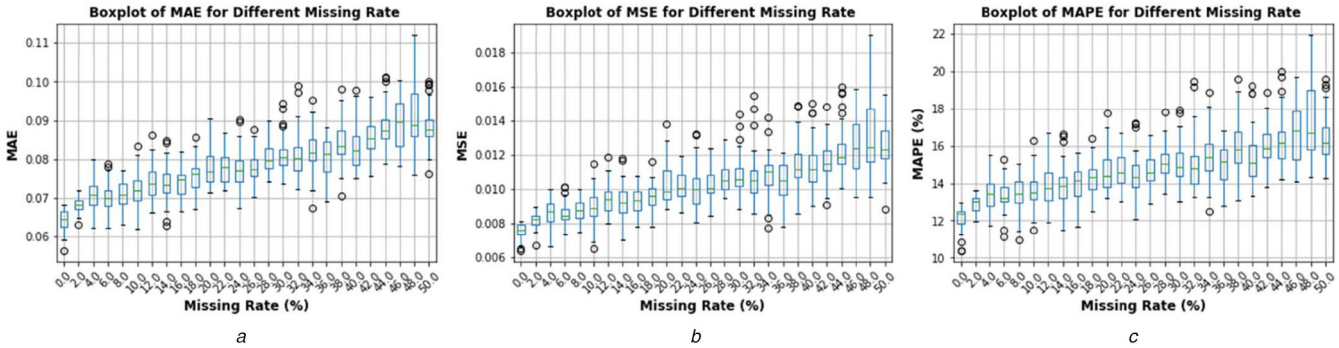
RNN models-imputation methods	MSE	MAE	MAPE, %
LSTM-average	0.0106	0.0794	15.84
LSTM-last	0.0108	0.0808	15.46
LSTM-simple	0.0105	0.0802	15.51
GRU-average	0.0107	0.0811	15.80
GRU-last	0.0108	0.0808	15.46
GRU-simple	0.0109	0.0799	15.50
GRU-D	0.0102	0.0783	14.61

4.2 Predictive performance comparison of GRU-D and baseline imputation methods

In this section, the dataset with 20% missing values was utilised as the input of the prediction models. For baseline missing data imputation methods, the missing values were imputed by three simple methods before modelling. Tables 2 and 3, respectively, show the predictive performance of non-RNN models and RNN models with three simple missing value imputation methods. In general, the predictive performance kept the same order with the predictive performance using dataset without missing values, but all prediction models had a slight decrease in prediction accuracy due to the impact of missing values. The last data imputation method outperformed common and simple methods except for the RF models, and the simple method achieved better performance than the average method. For the GRU-D networks, since both short-term features and empirical average value were considered in the decay mechanism. The GRU-D networks outperformed all other baseline imputation methods in terms of all three predictive performance measurements.

4.3 Impact of missing rates on prediction accuracy

To investigate the impact of missing rates on the predictive performance of the proposed GRU-D networks, the input datasets with the missing rates from 0 to 50% were used for testing. The



a *b* *c*

Fig. 5 *Predictive performance of GRU-D networks with different missing rates*
(a) MAE, ***(b)*** MSE, ***(c)*** MAPE

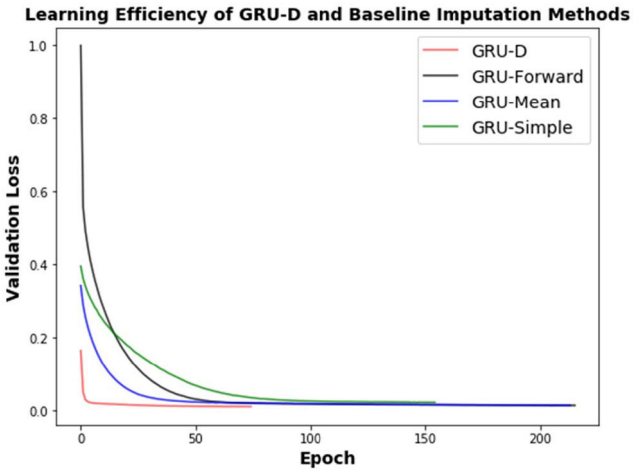


Fig. 6 *Learning efficiency of GRU networks with different missing data imputation methods*

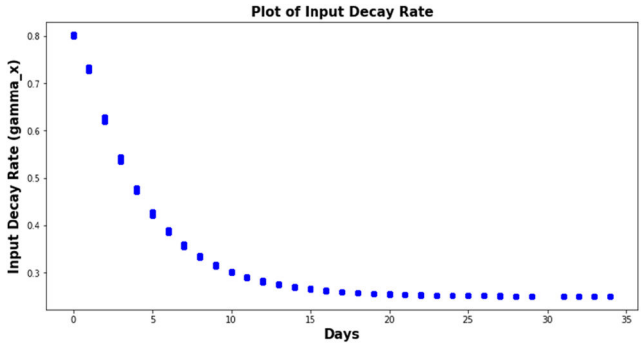


Fig. 7 *Learned input decay rate (the x-axis is the time interval δ_t^d from 0 35 days, the y-axis is the value of learned input decay rates $\gamma_{x_t}^d$)*

data of all 1487 road segments were tested. Fig. 5 shows the boxplots of three predictive performance measurements.

When the missing rate equals 0%, the predictive performance of GRU-D networks has the same prediction accuracy with GRU networks. As the missing rate increased, the prediction precision gradually dropped. When the missing rate equals 50%, the GRU-D generated averaged 0.089 MAE with a 0.0128 standard error. It is noticed that stopping distance and driver reaction time have a non-linearly negative correlation curve with road surface friction. When the road surface friction lower than 0.5, the incremental stopping distance with one unit decrease in road surface friction is rapidly increased. 0.1 error in road surface friction could cause tens of metres of error in stopping distance estimation [41]. Thus, even the proposed GRU-D networks are capable of handling the missing values; the impact became more considerable in this implementation when the missing rate is getting larger.

4.4 Learning efficiency and learned decays interpretation

In this section, the learning efficiency and the learned input decay rate are investigated. Fig. 6 shows the validation loss curves versus the training epochs. Due to the early stopping mechanism is used during the training process, the number of training epochs is different. The GRU-D networks need fewer epochs to converge than the GRU networks with other baseline missing data imputation methods. Also, the loss of the GRU-D networks decreased the fastest and achieved lower loss among the compared models.

Fig. 7 shows the learned input decay rate γ_{x_t}, which can demonstrate how the informative missing pattern was utilised for missing value imputation. According to the figure, the learned input decay rate ranged from 0.8 to around 0.2. The decay rate dropped with a large slope when the time interval is smaller than ten days. Then, the declining rate of the input decay rate decreased. Finally, it kept a constant when the time interval is larger than 20 days. This indicates that the value of the observation at the current and nearby time step is significant for the road surface friction prediction and the model relies less on the previous observations with the time interval larger than 20 days.

5 Conclusion and future work

This study proposed GRU-D networks for forecasting road surface friction with missing values. The GRU-D unit handled the missing values by integrating a decay mechanism with the original GRU. Masking and time interval missing patterns are both considered in the GRU-D networks. To evaluate the predictive performance of the proposed method, feed-forward NN, SVR, RF, and LSTM were selected as baseline prediction models, and three simple methods for data imputing were selected as baseline imputation methods. The evaluation results indicate the proposed model outperformed all baseline models with the most efficient learning efficiency. The analysis of learned decay rates presents the value of the observation at the current and nearby time step is more important for the road surface friction prediction at the testing field and the model relies less on the previous observations with the time interval larger than 20 days. For future work, the model will be improved to increase the predictive accuracy and interpretability of the model parameters.

6 Acknowledgments

This research was supported by the multi-institutional project (Exploring Weather-related Connected Vehicle Applications for Improved Winter Travel in Pacific Northwest) of Pacific Northwest Transportation Consortium (PacTrans) USDOT University Transportation Center for Federal Region 10.

7 References

[1] Mayora, J.M.P., Piña, R.J.: 'An assessment of the skid resistance effect on traffic safety under wet-pavement conditions', *Accid. Anal. Prev.*, 2009, **41**, (4), pp. 881–886

[2] Chen, F., Chen, S., Ma, X.: 'Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data', *J. Safety Res.*, 2018, **65**, pp. 153–159

[3] El Esawey, M., Walker, S., Sowers, C., *et al.*: 'Safety assessment of the integration of road weather information systems and variable message signs in British Columbia', *Transp. Res. Rec.*, 2019, **2673**, pp. 305–313

[4] (FHWA), FHA: 'How do weather events impact roads?', 2005

[5] Panahandeh, G., Ek, E., Mohammadiha, N.: 'Road friction estimation for connected vehicles using supervised machine learning'. 2017 IEEE Intelligent Vehicles Symp. (IV), Redondo Beach, CA, USA, 2017, pp. 1262–1267

[6] Ye, Z., Shi, X., Strong, C.K., *et al.*: 'Vehicle-based sensor technologies for winter highway operations', *IET Intell. Transp. Syst.*, 2012, **6**, (3), p. 336

[7] Feng, F., Fu, L.: 'Evaluation of two new Vaisala sensors for road surface conditions monitoring final report', August 2008

[8] Ewan, L., Al-Kaisy, A., Veneziano, D.: 'Remote sensing of weather and road surface conditions', *Transp. Res. Rec., J. Transp. Res. Board*, 2013, **2329**, (2329), pp. 8–16

[9] Haavasoja, T., Nylander, J., Nylander, P.: 'Experiences of mobile road condition monitoring'. Proc. 16th Int. Road Weather Conf., Helsinki, Finland, May 2012, pp. 23–25

[10] Roychowdhury, S., Zhao, M., Wallin, A., *et al.*: 'Machine learning models for road surface and friction estimation using front-camera images'. 2018 Int. Joint Conf. on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1–8

[11] Maenpaa, K., Sukuvaara, T., Ylitalo, R., *et al.*: 'Road weather station acting as a wireless service hotspot for vehicles'. Proc. - 2013 IEEE 9th Int. Conf. Intelligent Computer Communication Processing ICCP 2013, Cluj-Napoca, Romania, 2013, pp. 159–162

[12] Fay, L., Akin, M., Muthumani, A.: 'Quantifying salt concentration on pavement – phase II', 2018

[13] Karsisto, V., Nurmi, P.: 'Using car observations in road weather forecasting'. Proc. 18th Standing Int. Road Weather Commission Conf., Fort Collins, CO, USA, 2016, http://sirwec.org/wp-content/uploads/Papers/2016-FtCollins/D-017.pdf

[14] Singh, G., Bansal, D., Sofat, S., *et al.*: 'Smart patrolling: an efficient road surface monitoring using smartphone sensors and crowdsourcing', *Pervasive Mob. Comput.*, 2017, **40**, pp. 71–88

[15] Saarikivi, P.: 'Development of mobile optical remote road condition monitoring in Finland', May 2012, pp. 23–25

[16] Boselly, S.E., Thornes, J.E., Ulburg, C., *et al.*: 'Road weather information systems volume 1: research report', SHRP-H-351, Strategic Highway Research Program, National Research Council, Washington, DC, 1993

[17] Liu, B., Yan, S., You, H., *et al.*: 'Road surface temperature prediction based on gradient extreme learning machine boosting', *Comput. Ind.*, 2018, **99**, (March), pp. 294–302

[18] Sokol, Z., Bližňák, V., Sedlák, P., *et al.*: 'Ensemble forecasts of road surface temperatures', *Atmos. Res.*, 2017, **187**, pp. 33–41

[19] Linton, M.A., Fu, L.: 'Connected vehicle solution for winter road surface condition monitoring', *Transp. Res. Rec., J. Transp. Res. Board*, 2016, **2551**, pp. 62–72

[20] Sukuvaara, T., Nurmi, P.: 'ID: 001 connected vehicle safety network and road weather forecasting – the WiSafeCar project', May 2012, pp. 23–25

[21] Pan, G., Fu, L., Yu, R., *et al.*: 'Winter road surface condition recognition using a pretrained deep convolutional network', *Proc. Transp. Res. Board 97th Annu. Meet.*, 2018, https://arxiv.org/abs/1812.06858v1

[22] Jonsson, P., Edblad, J., Thörnberg, B.: 'Developing a cost effective multi pixel NIR camera for road surface status classification in 2D', 2014, pp. 1–6

[23] Chen, X., Wang, S., Shi, C., *et al.*: 'Robust ship tracking via multi-view learning and sparse representation', *J. Navig.*, 2018, pp. 1–7, http://dx.doi.org/10.1017/S0373463318000504, 00.9

[24] Kangas, M., Heikinheimo, M., Hippi, M.: 'RoadSurf: a modelling system for predicting road weather and road surface conditions', *Meteorol. Appl.*, 2015, **22**, (3), pp. 544–553

[25] Smith, B.A., Hoogenboom, G., McClendon, R.W.: 'Artificial neural networks for automated year-round temperature prediction', *Comput. Electron. Agric.*, 2009, **68**, (1), pp. 52–61

[26] Venkadesh, S., Hoogenboom, G., Potter, W., *et al.*: 'A genetic algorithm to refine input data selection for air temperature prediction using artificial neural networks', *Appl. Soft Comput.*, 2013, **13**, (5), pp. 2253–2260

[27] Pu, Z., Liu, C., Wang, Y., *et al.*: 'Road surface condition prediction using long short-term memory neural network based on historical data', 2019

[28] Chung, J., Gulcehre, C., Cho, K., *et al.*: 'Empirical evaluation of gated recurrent neural networks on sequence modeling', 2014, available Online, http://arxiv.org/abs/1412.3555

[29] Che, Z., Purushotham, S., Cho, K., *et al.*: 'Recurrent neural networks for multivariate time series with missing values', *Sci. Rep.*, 2018, **8**, (1), p. 6085

[30] Chen, Y., Lv, Y., Li, Z., *et al.*: 'Long short-term memory model for traffic congestion prediction with online open data'. 2016 IEEE 19th Int. Conf. on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 2016, pp. 132–137

[31] Guo, J., Huang, W., Williams, B.M.: 'Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification', *Transp. Res. C, Emerg. Technol.*, 2014, **43**, pp. 50–64

[32] Wu, C.-H., Ho, J.-M., Lee, D.-T.: 'Travel-time prediction with support vector regression', *IEEE Trans. Intell. Transp. Syst.*, 2004, **5**, (4), pp. 276–281

[33] Gers, F.A., Schmidhuber, J., Cummins, F.: 'Learning to forget: continual prediction with LSTM', 1999

[34] Tang, J., Chen, X., Hu, Z., *et al.*: 'Traffic flow prediction based on combination of support vector machine and data denoising schemes', *Phys. A, Stat. Mech. Appl.*, 2019, **534**, p. 120642

[35] Tang, J., Li, L., Hu, Z., *et al.*: 'Short-term traffic flow prediction considering spatio-temporal correlation: a hybrid model combing type-2 fuzzy C-means and artificial neural network', *IEEE Access*, 2019, **7**, pp. 101009–101018

[36] Ma, X., Tao, Z., Wang, Y., *et al.*: 'Long short-term memory neural network for traffic speed prediction using remote microwave sensor data', *Transp. Res. C, Emerg. Technol.*, 2015, **54**, pp. 187–197

[37] Cui, Z., Ke, R., Wang, Y.: 'Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction', 2018, pp. 1–12

[38] Tang, J., Yu, S., Liu, F., *et al.*: 'A hierarchical prediction model for lane-changes based on combination of fuzzy C-means and adaptive neural network', *Expert Syst. Appl.*, 2019, **130**, pp. 265–275

[39] Cui, Z., Henrickson, K., Ke, R., *et al.*: 'High-order graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting', 2018, pp. 1–9

[40] Lv, Y., Duan, Y., Kang, W., *et al.*: 'Traffic flow prediction with big data: a deep learning approach', *IEEE Trans. Intell. Transp. Syst.*, 2014, **16**, (2), pp. 865–873

[41] Wallman, C.-G., Åström, H.: '*Friction measurement methods and the correlation between road friction and traffic safety: a literature review*' (Statens väg-och transportforskningsinstitut, Sweden, 2001)