Travel time forecasting on a freeway corridor: a dynamic information fusion model based on the random forests approach

Bo Qiu and Wei Fan

*USDOT Center for Advanced Multimodal Mobility Solutions and Education
(CAMMSE), Department of Civil and Environmental Engineering,
University of North Carolina at Charlotte, Charlotte, North Carolina, USA*

Abstract

Purpose – Metropolitan areas suffer from frequent road traffic congestion not only during peak hours but also during off-peak periods. Different machine learning methods have been used in travel time prediction, however, such machine learning methods practically face the problem of overfitting. Tree-based ensembles have been applied in various prediction fields, and such approaches usually produce high prediction accuracy by aggregating and averaging individual decision trees. The inherent advantages of these approaches not only get better prediction results but also have a good bias-variance trade-off which can help to avoid overfitting. However, the reality is that the application of tree-based integration algorithms in traffic prediction is still limited. This study aims to improve the accuracy and interpretability of the models by using random forest (RF) to analyze and model the travel time on freeways.

Design/methodology/approach – As the traffic conditions often greatly change, the prediction results are often unsatisfactory. To improve the accuracy of short-term travel time prediction in the freeway network, a practically feasible and computationally efficient RF prediction method for real-world freeways by using probe traffic data was generated. In addition, the variables' relative importance was ranked, which provides an investigation platform to gain a better understanding of how different contributing factors might affect travel time on freeways.

Findings – The parameters of the RF model were estimated by using the training sample set. After the parameter tuning process was completed, the proposed RF model was developed. The features' relative importance showed that the variables (travel time 15 min before) and time of day (TOD) contribute the most to the predicted travel time result. The model performance was also evaluated and compared against the extreme gradient boosting method and the results indicated that the RF always produces more accurate travel time predictions.

Originality/value – This research developed an RF method to predict the freeway travel time by using the probe vehicle-based traffic data and weather data. Detailed information about the input variables and data pre-processing were presented. To measure the effectiveness of proposed travel time prediction algorithms,

the mean absolute percentage errors were computed for different observation segments combined with different prediction horizons ranging from 15 to 60 min.

1. Introduction
Nowadays, travel time prediction plays a significant role as it can greatly help route planning and also the development of countermeasures to reduce traffic congestion. Metropolitan areas are adversely affected by frequent road traffic congestion not only in peak hours but also in off-peak periods. Therefore, the capability to forecast traffic conditions, particularly travel times, is of utmost importance in traffic management applications aimed at relieving negative social, environmental and economic impacts for people. The definition of travel time is the total time for a vehicle to travel from one point to another over a specified route (Zhu *et al.*, 2009). Travel time has been widely used to measure the effectiveness of transportation systems and increasingly becomes one of the most popular traffic information that travelers are interested in gathering. The ability to accurately predict travel time in transportation networks is a critical component of the traveler information system. Accurate travel time prediction can enhance the performance of the traffic management systems, in which travelers are given the opportunities to react to the traffic proactively (Oh *et al.*, 2015). Furthermore, as an important performance indicator, accurate predicted travel times can be used for quantitatively comparing different traffic management systems. Nowadays, with the explosive availability of abundant data collected by sensors and monitors, the big data storage and processing issues have become more and more relevant (Šemanjski, 2015).

In travel time prediction, a reliable prediction method needs to achieve the following three objectives: accuracy, robustness and adaptability (Van Lint, 2006). Traditional data-based (e.g. linear regression and time series) models have been widely applied to predict travel times based on the historical data. However, with the consideration of effectiveness, accuracy and feasibility, these models may have become outdated and replaceable. Recently, different machine approaches (such as neural networks, ensemble learning and support vector machines) have been used by different researchers and the results indicate that such approaches to prediction are adaptable and can give better performances than traditional models. Therefore, the machine learning-based approaches are selected for the travel time prediction in this study. The purpose of this study is to propose an approach to systematically analyze the relationship between travel time and various traffic features. In that regard, a machine learning-based approach (e.g. the random forest [RF] model) is used to predict the freeway travel time. The proposed approach is also tested using a freeway corridor in Charlotte, North Carolina using the probe vehicle-based traffic data. The advantages and disadvantages of the proposed model are also identified and compared. Finally, the effectiveness and efficiency of the proposed model are also evaluated.

2. Literature review
Transportation researchers and data scientists have developed various techniques in the past three decades to provide more reliable future travel time estimation methods (Oh *et al.*, 2015). Generally speaking, such techniques can be classified into three groups: naive methods, traffic theory-based methods and data-driven methods. As the name indicates, the naive prediction models are very simple methods, which typically do not involve the estimation of model parameters. As the model assumptions are usually restrictive, they are

not actually fulfilled in many situations (Wunderlich *et al.*, 2000). As one of the traffic theory-based methods, traffic flow simulation and user-optimal dynamic traffic assignment have been widely used in freeway travel time prediction. Examples include Papageorgiou *et al.* (2010) and Dion *et al.* (2004). In data-based traffic time prediction models, the function that relates traffic factors with the prediction result (dependent variable) is not obtained from predetermined traffic theory, as the relationships of variables come from the sample data itself by using statistical data mining methods. This approach greatly expands the pool of researchers who can participate in travel time prediction because they no longer have to become experts in traffic theory. However, such data-based methods usually need a lot of data, which is not always available. The data-based models are strongly subjected to data availability and accessibility (Van Lint, 2006).

In general, the data-based models can be divided into two categories, which are parametric and non-parametric models. In the parametric models, the parameters can be estimated to define the function, which are predefined and set in a finite-dimensional space. The most widely applied parametric model is linear regression, where the dependent variable is always a linear function of the explanatory input variables. Generally, the independent temporal variables are traffic observations in several past time intervals. The second type of parametric model is the Bayesian net, which assumes that the explanatory variables are always conditionally independent given the dependent variable. The third group of the parametric models in modeling travel time is time series models, of which the most widely used one is the autoregressive integrated moving average model.

In the non-parametric models, the structure of the model is not predefined and the intrinsically complex relationships cannot be expressed by simple functions. Furthermore, the term non-parametric does not mean that there are no parameters to be estimated, but on the contrary, it means that the number and typology of the parameters are unknown a priori and possibly infinite depending on the sample data set (Mori *et al.*, 2015). With the rapid development of data science, the methodologies for non-parametric estimation are also being quickly updated. Along this line, the most widely seen in the literature of travel time prediction is the artificial neural networks (ANN). ANN models are widely used in transportation because of their ability to capture complex relationships in large data sets (Dharia and Adeli, 2003). Unlike multivariable models, ANN models are developed without a predetermined form of function, whereas they can overcome multicollinearity problems. Different types of neural networks have been applied in travel time prediction, from regular multilayer feedforward neural networks (Yildirimoglu and Ozbay, 2012) to more complex spectral basis neural networks (Park *et al.*, 1999). Another choice for travel time prediction is using support vector machine (SVM) methods. This advanced algorithm consists of decision function, the application of the kernel functions and the sparsity of solutions. The SVM models have a good performance on travel time prediction with historical travel time data. Some researchers (Yildirimoglu and Geroliminis, 2013; Wu *et al.*, 2004) used SVM methods to estimate travel times. In the calculation process, the kernel function can map the input data into a higher-dimensional space. In the model generating process, the flattest linear function is identified which relates to the transferred input vectors into the target variables. Travel time prediction will be based on the function, which can be mapped into the initial space by the flattest linear function. Both the ANN and SVM models tend to be overfitting due to their complicated structures and the large number of parameters that need to be calibrated, which is a serious problem that commonly existed in the non-parameter machine learning algorithm.

The local regression approach is another non-parametric approach that always produces accurate and reliable results. The main idea of local regression is to generate a method to choose a set of historical data points which have similar properties to the current situation

and predict the travel time using a constructed model with these chosen data points. Various local regression models can be used depending on the type of methods used to select the set of similar historical points and depending on the methodology chosen to fit the model (Mori *et al.*, 2015).

There have also been some semi-parametric models developed, as a combination of parametric and non-parametric methods, in travel time prediction. Some of the strict assumptions of the parametric model are loosened to obtain a more flexible structure (Ruppert *et al.*, 2003). In the application of travel time prediction, semi-parametric models are presented as varying coefficient regression models. The prediction result (travel time) was defined as a linear function of the naive historical and instantaneous predictors; however, the parameters vary depending on the departure time interval and prediction horizon (Schmitt and Jula, 2007).

In summary, with the wide applications of big data in the field of transportation, different machine learning approaches have been deployed in the travel time prediction area. The methodologies include, but are not limited to, the following: SVM regression, neural network approaches (e.g. state-and-space neural network, long short-term memory neural network), nearest neighbor (e.g. *k*-nearest neighbor) and ensemble learning (e.g. RF and gradient boosting), etc. Table 1 provides a summary of the studies reviewed in chronological order.

3. Data collection and processing

3.1 Data collection

3.1.1 Travel time data. In this study, the raw travel time data are gathered from the regional integrated transportation information system (RITIS), an advanced traffic system that includes segment analysis, probe data analytics and signal analytics. A series of major freeway segments in Charlotte, North Carolina are selected for the case study: as one of the most heavily traveled interstate freeways in the City of Charlotte area, I-485 is an interstate highway loop encircling the city, which completed the last segment on June 5, 2015. Charlotte metropolitan area has been growing and in the past 25 years, the Charlotte area population has increased from 688,000 to 1.4 million and more than 500,000 more residents are anticipated over the next 20 years. In 2018 alone, there was over \$1bn in capital investment in the region. One result of this growth is increased traffic congestion. I-485 freeway segments in the vicinity of the southern Charlotte area experiences massive traffic congestion during weekdays due to heavy commuter and interstate traffic. As the recurrent congestion seriously affects the travel and further economic development in this area, the I-485 Express Lanes project will add one express lane in each direction along I-485 between I-77 and US 74 (Independence Boulevard), resulting in a seamless network of express lanes in southern Mecklenburg county that could improve travel time reliability and traffic flow in this critical transportation corridor. The project will also add one general-purpose lane in each direction along I-485 between Rea Road and Providence Road. The estimated cost is \$346m and the construction began in summer 2019 and the completion date is 2022.

In the RITIS system, the selected section of I-485 Southern loop starts from the interchange with I-77 (Exit 67) and ends at the interchange with US-74 (Exit 51). The directions include clockwise and counter-clockwise and 37 miles of roadways and 32 traffic message channel (TMC) code segments are selected in this study. All the selected segments have uninterrupted coverage in the RITIS data 24 h per day and 365 days a year. The data set is collected from January 1, 2019 to December 1, 2019, and the interval is 15 min. An example of the raw time data used in this study is shown in Table 2 below.

3.1.2 Weather data collection. The historical weather data are also collected at locations that are close to the Charlotte Douglas International airport. The raw weather data includes information on different categories such as temperature, dew point, humidity, pressure, visibility, wind direction, wind speed, gust speed, precipitation and conditions. The raw

Author (year)	Location	Roadway category	Data source	Method category	Data type	Prediction method
Wunderlich et al. (2000)	N/A	N/A	Simulated data from integration	Navie model	Travel time	Exponential filtering
Dion et al. (2004)	Virginia, USA	N/A	Simulated data from integration	Traffic theory-based model	Travel time	Delay models
Van Lint et al. (2002)	N/A	Freeway	Freeway operations simulation (FOSIM)	Non-parametric	Travel time, travel speed	State-space neural network
Wu et al. (2005)	Taiwan	Highway	Loop detector	Non-parametric	Travel speed	SVR
Schmitt and Jula (2007)	California, USA	Urban road	Loop detector	Navie model	Travel time	Switch model
Zou et al. (2008)	Maryland, USA	Highway	Roadside detector	Hybrid non-parametric	Travel time	Combined clustering neural networks
Li et al. (2009)	Atlanta, USA	N/A	Simulated data from VISSIM	Hybrid non-parametric	Travel time, travel speed	Combined boosting and neural network
Papageorgiou et al. (2010)	N/A	N/A	Simulated data from MATANET	Traffic theory-based model	Travel time	Macroscopic simulation
Hammer (2010)	N/A	N/A	GPS	Non-parametric	Travel speed	RF
Myung et al. (2011)	Korea	N/A	ATC system	Non-parametric	Travel time	KNN
Wisitpongphan (2012)	Bangkok, Thailand	Highway	GPS	Non-parametric	Travel time, GPS	BP neural network
Zheng et al. (2013)	Delft, The Netherlands	Urban road	GPS data	Non-parametric	Vehicle position, travel speed	State-space neural network
Yildirimoglu and Geroliminis's (2013)	California, USA	Freeway	Loop detector	Hybrid non-parametric	Travel time	Combined Gaussian mixture, PCA and clustering
Zhang and Haghani (2015)	Maryland, USA	Interstate highway	INRIX	Non-parametric	Travel time	Gradient boosting
Joao et al. (2015)	Porto, Portugal	Urban road	STCP system	Hybrid non-parametric	Travel time	Combined RF, projection pursuit regression and SVM
Duan et al. (2016)	England	Highway	Cameras, GPS and loop detectors	Non-parametric	Travel time	LSTM neural network
Li and Bai (2016)	Ningbo, China	N/A	N/A	Non-parametric	Truck trajectory, travel time, travel speed	Gradient boosting

(continued)

Table 1.
Summary of travel time prediction using machine learning approaches

Table 1.

Author (year)	Location	Roadway category	Data source	Method category	Data type	Prediction method
Liu et al. (2017)	California, USA	Interstate highway	PeMS	Non-parametric	Travel time	LSTM neural network
Fan et al. (2018)	Taiwan	Highway	Electric toll	Non-parametric	Travel time, vehicle information	RF method
Yu et al. (2017)	Shenyang, China	bus route	AVL system	Non-parametric	Bus travel time	RF and KNN
Wang et al. (2018)	Beijing, China	Urban road	Floating ar data	Non-parametric	Taxi travel time, vehicle trajectory data	LSTM neural network
Wei et al. (2018)	China	Urban road	Vehicle passage records	Non-parametric	Travel time	LSTM neural network
Wang et al. (2018)	Beijing and Chengdu, China	Urban road	GPS	Non-parametric	Vehicle trajectory data	LSTM neural network
Gupta et al. (2018)	Porto, Portugal	Urban road	GPS	Non-parametric	Taxi travel speed	RF and gradient boosting
Moonam et al. (2019)	Madison, Wisconsin, USA	Freeway	Bluetooth detector	Non-parametric	Travel speed	KNN, KF
Kumar et al. (2019)	Chennai, India	Urban road	GPS	Non-parametric	Travel time	KNN
Cristóbal et al. (2019)	Gran Canaria, Spain	Urban road	Public transport network	Non-parametric	Travel time	K-medoid clustering technique
Kwak and Geroliminis (2020)	California, USA	Freeway	PeMS	Parametric	Travel time	Dynamic linear model
Fu et al. (2020)	Beijing, Suzhou, Shenyang, China	Urban road	Ride-hailing platform	Non-parametric	Travel time	Graph attention network
Chiabaut and Faitout (2021)	Lyon, French	Highway	Loop detector	Non-parametric	Travel time	PCA and Clustering
Wu et al. (2021)	Houston, USA	Urban	AWAM	Parametric	Travel time	Autoregressive with exogenous inputs (NARX) model and feed-forward neural network

Notes: SVR = support vector regression; VISSIM = Verkehr In Städten - SIMulationsmodell (German for "Traffic in cities - simulation model"); PeMS = performance measurement system; GPS = global positioning system; ATC = air traffic control; KNN = k-nearest neighbours; AVL = automatic vehicle location; KF = K filter; AWAM = anonymous wireless address matching; NARX = nonlinear autoregressive exogenous model

TMC code	Timestamp	Speed (mile/h)	Travel time (second)
125N04784	1/1/2019 0:00	62.91	53.58
125N04783	1/1/2019 0:00	61.17	12.82
125N04786	1/1/2019 0:00	60.43	47.56
125N04785	1/1/2019 0:00	61.3	11.85
125N04780	1/1/2019 0:00	63.97	14.59
125N04782	1/1/2019 0:00	63.04	21.73
125N04781	1/1/2019 0:00	62.79	12.42
125N04788	1/1/2019 0:00	65.03	29.6
125N04787	1/1/2019 0:00	63.5	53.76
125N04789	1/1/2019 0:00	64.79	54.5
125–04783	1/1/2019 0:00	62.98	33.22
125–04782	1/1/2019 0:00	62.75	35.68
125–04785	1/1/2019 0:00	60.54	5.16
125N04784	1/1/2019 0:00	62.91	53.58

Notes: The table includes the following information: TMC code: the RITIS system uses the TMC and assigns a unique identifier code to each road segment. In other words, the TMC code is a road segment ID. Timestamp: this shows the timestamp of the record. Speed: this presents the current estimated harmonic mean speed on the roadway segment in miles per hour. Travel time: this field indicates the time that it takes to drive along the roadway segment

Table 2.
Sample raw travel
time data

weather data were recorded on a per hour basis, and as such, the discrepancy in the time intervals was treated by a mapping methodology to combine the traffic data with the weather data. An example of the raw weather data used in this study is shown in Table 3 below.

3.2 Data processing
Based on previous studies, it was revealed that travel speed is much more sensitive to severe weather events. The weather conditions in the Charlotte area were originally classified into 30 detailed weather conditions. However, in this study, the weather conditions are further categorized into only three groups including normal, rain and snow/fog/ice. Table 4 presents the detailed classification of the newly grouped weather conditions. To keep the sample size to the extent that is acceptable, "snow," "fog," "ice pellet" and other similar conditions are combined because of their rates of occurrence.

Date	Time (EDT)	Visibility	Conditions
Saturday, Oct 5, 2019	7:55 a.m.	2.0 mi	Rain
Saturday, Oct 5, 2019	8:55 a.m.	2.0 mi	Rain
Saturday, Oct 5, 2019	9:55 a.m.	2.0 mi	Light rain
Saturday, Oct 5, 2019	10:55 a.m.	2.0 mi	Light rain
Saturday, Oct 5, 2019	11:55 a.m.	3.0 mi	Light rain
Saturday, Oct 5, 2019	12:55 a.m.	2.0 mi	Light rain
Saturday, Oct 5, 2019	13:55 a.m.	3.0 mi	Light rain
Saturday, Oct 5, 2019	14:55 a.m.	7.0 mi	Light rain
Saturday, Oct 5, 2019	15:55 a.m.	6.0 mi	Light rain
Saturday, Oct 5, 2019	16:55 a.m.	7.0 mi	Light rain
Saturday, Oct 5, 2019	17:55 a.m.	4.0 mi	Rain

Note: EDT = Eastern daylight time

Table 3.
Sample raw weather
data

Snow/fog/ice	Rain	Normal
Haze	Light rain	Clear
Fog	Rain	Partly cloudy
Smoke	Heavy rain	Mostly cloudy
Patches of fog	Light drizzle	Scattered clouds
Mist	Heavy thunderstorm	Overcast
Shallow fog	Thunderstorms an	Unknown
Light freezing R	Light thunderstorm	Squalls
Light ice pellet	Thunderstorm	
Light freezing D	Drizzle	
Light freezing F		
Ice pellets		
Light snow		
Snow		
Heavy snow		

Table 4.
Classification of the
weather conditions

To merge the link travel times data set with the historical weather data set, the issue of different intervals of two data sets should be resolved first. The RITIS data sets are aggregated into 15 min intervals, while the weather data set is aggregated into 1 h intervals. Therefore, the weather conditions are distributed evenly with the RITIS data set based on the timestamp.

4. Travel time prediction methodology
4.1 Ensemble learning methodology
An ensemble itself is a supervised learning algorithm, which can be trained and used to make predictions. The ensemble learning-based algorithms consist of multiple base models (e.g. decision tree model), each of which provides an alternative solution to the problem. The prediction results tend to be more accurate when there is a strong diversity among the models (Kuncheva and Whitaker, 2003). Decision trees always suffer from high variance which causes the instability of the prediction results. Bootstrap aggregating (bagging) is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms. In the bagging process, the algorithm builds multiple models from the same original samples data set to reduce the variance. However, the bagging can make the trees highly correlated. RF is an extension of bagging in that in addition to building trees based on multiple samples of the original training data, it also constrains the features that can be used to build the trees, forcing trees to be different. To date, the RF models have been widely applied to various research fields (Greenhalgh and Mirmehdi, 2012; Jia *et al.*, 2016). For classification tasks, RF typically gives high accuracy while also having a faster classification time. An RF classifier requires training with large data sets, which in our study are obviously available because of the nature of the travel record data collected. Furthermore, the RF computational process runs efficiently on large data sets, which can reduce model complexity, overcome the overfitting to some extent and improve the efficiency. As known, overfitting means that the estimated model fits the training data too well. Generally, this is caused by the fact that the model function is too complicated to consider each data point and even outliers. The RF method can build a large number of random trees and then combine the results from each individual tree. The benefit of using the RF methods is that through averaging, the variance can be reduced.

4.2 Random forest algorithm
RF is an algorithm that can compete with gradient boosted trees in ensemble learning, especially because of its convenient parallel training, which is very attractive in the era of big data and large samples. RF is an ensemble tool which takes a subset of features to build multiple decision trees. Before the explanation of RF, one needs to mention decision trees. A decision tree is a very simple algorithm, which is a supervised learning algorithm based on the if-then-else rules. Its explanation is strong, and it is in line with human intuitive thinking. For each separate decision tree, the feature selection is conducted randomly, which means there is no correlation between different decision trees. The low correlation between models is the key in which uncorrelated models can produce ensemble predictions that are more accurate than any of the individual predictions. RF is an integrated algorithm that is composed of decision trees, which can get the final prediction better than any best separate judge. The RF algorithm procedure consists of the following main steps:

- Step 1: Randomly draw the samples from a given data set.
- Step 2: Construct a decision tree for each sample and predict the result.
- Step 3: Voting will be performed from the independent prediction results.
- Step 4: Select the popularly voted result for the classification problem or average the results for regression.

Figure 1 shows the prediction process of the RF algorithm, which is described as follows:

- The number of training data points is N and the number of variables in the classifier is M;
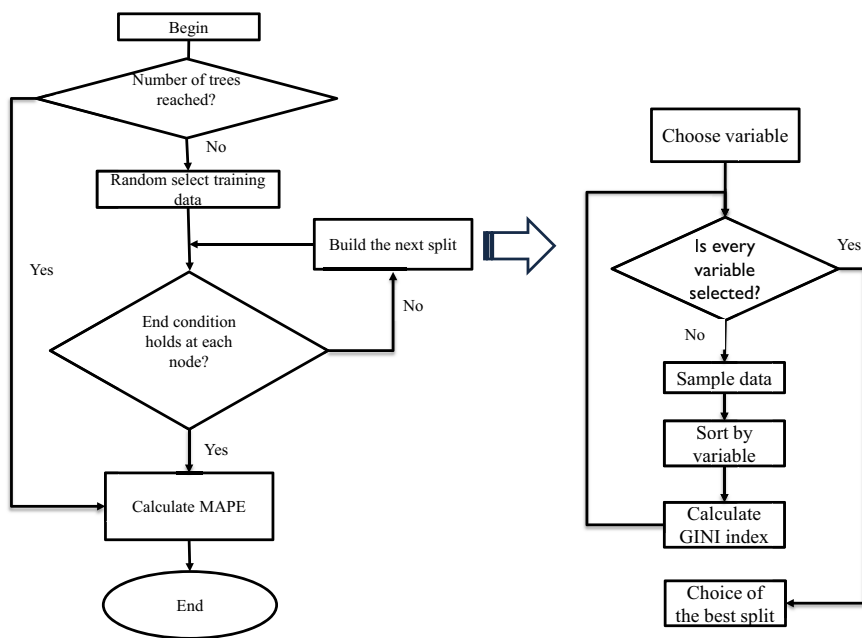
Qiu and Fan

Figure 1.
RF algorithm
processing flow

- Select the m variables in the whole variable set M to determine the decision at a node of the tree (Note that m is always considerably smaller than M);
- To construct the forest by trees, choose a training set k times with replacement from all N training data set. Each of these data sets is called a bootstrap data set. The number k is the number of the trees to be trained;
- For each tree node, randomly choose m variables on which to make the decision at that node. Calculate and get the best split based on these m variables in the training set; and
- The *Gini index* is used for calculating the Gini value to determine the best split point, which can be used to describe the purity after the split. The *Gini index* will fall between 0 to 1 and the smaller the value, the better the split. If a data set contains elements from two classes, the *Gini index* is defined as follows:

$$Gini(T) = 1 - \sum_{j=1}^{n} \left(p_j^2 \right) \tag{1}$$

where p^j is the relative proportion of class j in the original data set T and n is the number of classes in data set T.

$$Gini_{split}(T) = \frac{N_1}{N} Gini(T_1) + \frac{N_2}{N} Gini(T_2) \tag{2}$$

5. Proposed travel time prediction approaches

5.1 Feature selection and pre-processing steps

In the prediction model, the southern part of the I-485 freeway is divided into 32 sections by the recorded sensor segment in this study. Traffic data on each segment (from sensor to sensor) contains information on the subject segment and adjacent segment travel times, day of week (DOW), time of day (TOD), segment length and space mean speed. The RITIS real-world travel time data used for this study has a less than 0.5% missing rate (i.e. 4,246 out of 981,083). Note that in this study, the missing values are simply replaced with the mean of its closest surrounding values. From the previous studies (Wang *et al.*, 2018), the variables that have a significant impact on the travel time prediction included the basic variables (such as TOD, DOW, month and weather) and the spatial and temporal characteristics of the adjacent road segments. Furthermore, in this study, the travel times (which are collected several steps ahead of the travel time to be predicted) are also accounted for in the model estimation. The prediction model is developed under normal traffic conditions and does not consider unexpected conditions (e.g. special events). The data on each segment will be used to train one forest which consists of decision trees. The RF model prediction includes two major steps: training and prediction. The forests are constructed by using randomly selected parameter combinations and different numbers of trees during the training step. The selected variables include the temporal features, such as travel time at prediction segment 15, 30 and 45 min before, which is defined as T_{t-1}, T_{t-2} and T_{t-3}, respectively. The travel time at prediction segment exactly 1 week before, which are defined as T_{t-w}; TOD and DOW as important temporal features are also included. The spatial features include road segment identification (ID), segment length. In the data preparation, the temporal-spatial features are also generated, including the travel time of the nearest downstream and upstream road segment 15 min before, which are defined as

T_{t-1}^{i+1} and T_{t-1}^{i-1}, respectively. The detailed information and definition of the selected variables can be seen in Table 7.

5.2 Model development

To achieve the best modeling results, it is important to explore the effect of different combinations of parameters on the RF model prediction performance. Based on previous studies, there are primarily three features that can be tuned to optimize the predictive power of the model: the maximum number of features (*Max_features*), the number of trees (*N_estimators*) and minimum leaf size (*Min_sample-leaf*). They are presented as follows:

5.2.1 Max_features. This is the maximum number of features in the RF model that is allowed to try in each tree. There are multiple options available in Python to assign maximum features. "Auto/None" is a command that simply takes all the features that make sense in every tree, which simply does not put any restrictions on the individual tree. The "SQRT" option takes the square root of the total number of features in each individual run. For example, if the total number of variables is 100, under this option the system can only take 10 of them in each individual tree. The "log2" option is another similar type of option used for *max_features*. In this study, after several tests, the random subspace method is applied. The number of features considered at each internal node of RF is m, which is randomly chosen to be $m = INT(log_2 M + 1)$, where m is the total number of features, as suggested by Breiman (2001a, 2001b).

5.2.2 n_Estimators. This is the number of trees that one wants to build before taking the maximum voting or averages of predictions. A larger number of trees will give one better performance with a compromise of computing efficiency. As such, one should choose a value as high as what the processor can handle because this makes the predictions stronger and more stable.

5.2.3 Min_sample_leaf. This is the minimum leaf size. The leaf is the end node of a decision tree, which is the number of cases or observations in that leaf. A smaller leaf makes the model more prone to capture noise in the train data. To optimize the RF model, it is important to estimate the effect of different combinations of parameters on the model's performance. Based on this information, in this study, the tool RandomSearch is applied to optimize the tuning process to achieve a lower prediction error. In this study, after several trials of different *min_sample_leaf*, a minimum leaf size of 30 is chosen. When the parameters select 50 as the number of trees and 30 as the minimum leaf size, the mean absolute percentage error (MAPE) reaches the lowest 5.97%. This process is shown in Figure 2 and Table 5 demonstrates how the performance varies with different combinations of parameters (i.e. the number of trees and the minimum leaf size). RF models are not sensitive if the features are independent or dependent, though many will perform better if the data are preprocessed. A simple way to identify dependence among features is to calculate a correlation coefficient between each feature and all other features. From Figure 2 and Table 5, the results clearly show that when the number of trees reaches 50, the value of MAPE becomes nearly the same. In statistics, overfitting is the co-product of an analysis that corresponds perfectly to the sample set of data, and therefore, may fail to fit additional data or predict future observations reliably, which is a general problem of traditional ensemble learning methods. For example, the prediction error usually increases when the number of trees increases after it reaches the optimized point in the tree base model (Zhang and Haghani, 2015). There is also a need to consider the tradeoff between prediction accuracy and computational time. As when a large number of trees are being fitted, model complexity also increases and requires more computational time. The "randomness" in an RF means two things: n training samples are randomly extracted from the training set and

PARAMETER TUNING

Figure 2.
RF travel time
prediction model
performance
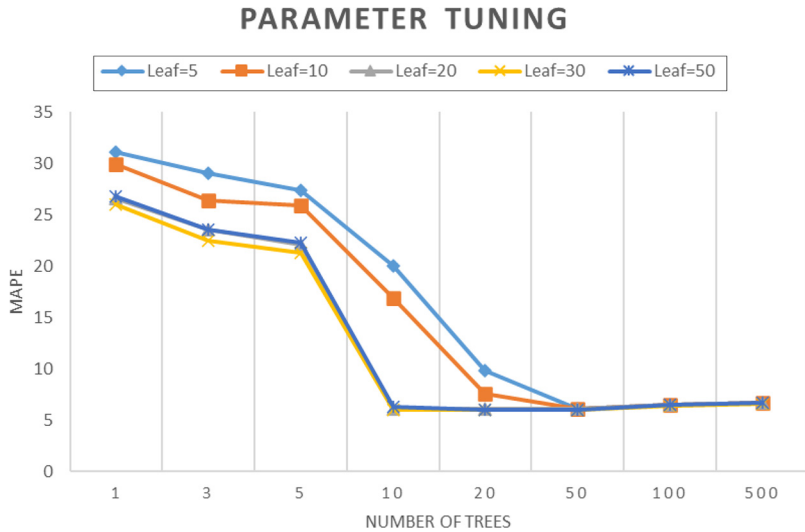
the m feature subsets are randomly drawn from M features. The introduction of such randomness is very important to the performance of RF. Due to their introduction, the RF is not prone to overfitting and is very noise-resistant (i.e. insensitive to default values).

It is also important to note that the performance measure used in this study is the MAPE. The MAPE statistic usually expresses accuracy as a percentage that is calculated as follows:

$$MAPE = \frac{1}{m}\sum_{i=1}^{m} |y_i - \hat{y}_i|$$

where

m = the total number of the data points;

$\hat{y}_i$ = the predicted travel time value in the test data set of record i; and

y_i = the actual travel time value in the test data set of record i.

To measure the effectiveness of different travel time prediction algorithms, the MAPEs are computed for three different observation segments (where A, B, C are three observation

No. of trees	Leaf = 5	Leaf = 10	Leaf = 20	Leaf = 30	Leaf = 50
1	31.11	29.87	26.56	26.01	26.74
3	29.05	26.34	23.52	22.46	23.59
5	27.38	25.9	22.09	21.28	22.24
10	19.98	16.87	6.13	6.01	6.26
20	9.78	7.56	6.1	5.99	6.05
50	6.13	6.14	6.12	5.97	5.99
100	6.46	6.48	6.51	6.42	6.54
500	6.7	6.72	6.73	6.6	6.72
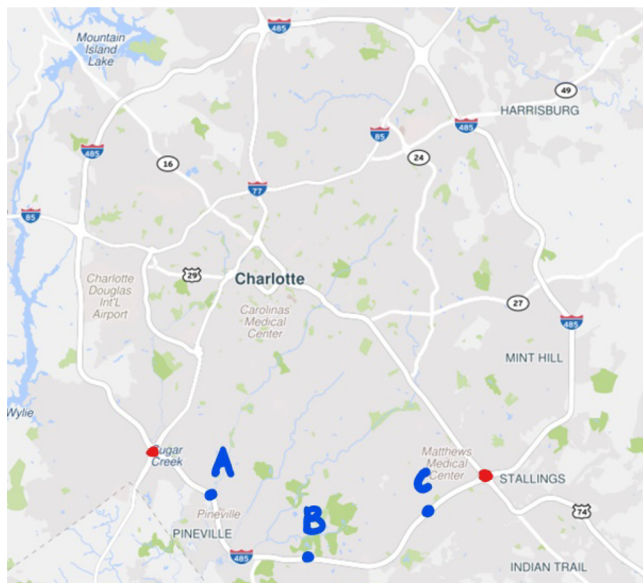
Table 5.
The MAPEs based
on the combination of
parameters (unit: %)

segments along the selected freeway for study, shows in Figure 3) with different prediction horizons from 15 min to 60 min. According to the comparison shown in Table 6 and Figure 4, the performance of the proposed RF is better than the eXtreme Gradient Boosting (i.e. [XGBoost], another widely used tree-based ensemble method), especially when the horizon of prediction time is long. The MAPEs of the RF model is significantly smaller than XGBoost when the horizon is long enough (i.e. longer than 45 min).

In the machine learning area, usually, only part of the predictor variables have significant impacts on the prediction results. Exploring the impact on the individual feature can help researchers and policymakers better understand contributing variables. Higher relative importance indicates a higher influence on travel time. Table 7 presents the relative importance of each variable and its ranks in the optimized RF model. From Table 7, each predictor variable has significant and different degrees of impact on the predicted travel time. The model result shows that the variable T_{t-1} (travel time 15 min before) contributes the most (34.85%) to the predicted travel time result. This result is expected and consistent with a previous study (Zhang and Haghani, 2015), which demonstrates that the immediate previous traffic condition will directly influence the traffic condition in the future. TOD is the second-highest ranked variable with the relative importance value of 30.12% and this result is also under expectation. T_{t-w} is the fourth-highest ranked variable with an importance value of 9.87%, which can be interpreted as a highly similar pattern of traffic times between weeks.

| Models | MAPE (%) of different road segments with a different prediction horizon | | | | | | | | | | | | |
| | 15 min | | | 30 min | | | 45 min | | | 60 min | | | |
	A	B	C	A	B	C	A	B	C	A	B	C	
RF	6.49	6.15	6.39	9.69	9.97	10.67	15.29	16.19	17.37	24.59	25.66	26.76	
XGBoost	6.57	6.14	6.39	10.58	9.98	10.89	15.35	15.98	17.90	25.90	26.06	28.09	

MAPE of different observation point with different
prediction time range
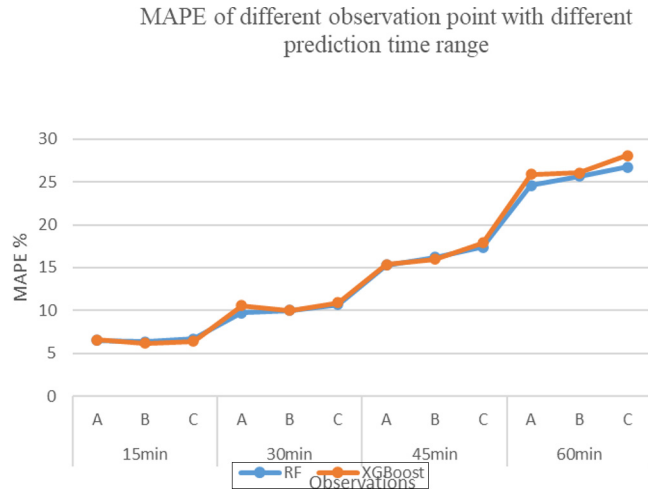
The result in Table 7 also shows that the spatial impact is less than the temporal impact, as, except for the variable road ID with a relative importance value of 2.28%, all the relative importance values of other spatial variables are less than 1%. Several variables such as the travel time of the two upstream segments (with the relative importance value of 0.31% and 0.42%, respectively) and the travel time of the two downstream segments (with the relative importance value of 0.35% and 0.61%, respectively) one time-step ahead are considered in the model. With respect to the travel time change value, the relative importance values of the two upstream segments are both 0.29% and the relative importance values of the two upstream segments are 0.79% and 0.37%, respectively. Based on these results, it could be explained that the relative importance values of the downstream segments are higher than those of upstream segments. The reason is caused by the spatial characteristics of the roadway. When a bottleneck occurs at the downstream segment, the upstream will be impacted very shortly.

6. Conclusions and recommendations

The tree-based ensemble methods are widely used in the field of prediction. By combining a simple tree with a forest, RF always produces high prediction accuracy (Zhang and Haghani, 2015). In this study, the authors applied an RF method to analyze and model freeway travel time to improve the prediction accuracy and model interpretability. Most existing machine learning models can capture the nonlinear pattern of travel time but suffer from over-fitting. Study results indicated that the RF model has its considerable advantages in freeway travel time prediction, the performance evaluation result showed that the RF-based model can have better predictions in terms of prediction accuracy. RF model showed a reasonable performance compared with other approaches. When the prediction horizon is no more than 15 min, the RF algorithm is relatively accurate. However, when the prediction horizon is longer than 30 min, the prediction error increases dramatically like other methods. Different from other machine learning methods, RF methods provide interpretable results with different types of predictor variables. RF can also handle data with very high dimensions (many features) without specific feature selection (because feature subsets are randomly selected) and identifies which features are more important after the training

process. Furthermore, it has an effective way of estimating missing data and maintaining accuracy when a significant proportion of the data are missing. The relative importance of the features shows that the travel time one step ahead (15 min before) contributes the most to the predicted travel time. Features such as the TOD, DOW and the travel time at prediction segment one week before and weather also have higher relative importance values in the model than other features. Adding up the most important eight variables' relative importance values (T_{t-1}, TOD, Speed, T_{t-w}, DOW, Weather, Road ID, Month) in the Table 7 will be as high as 94.77%, which means that these eight selected variables include most of the information needed in the travel time prediction. The proposed RF travel time prediction method has considerable advantages over the other tree-based approach.

However, the practice of RF algorithm and other tree-based ensemble methods in the travel time prediction area is still very limited. The future focus of the research would be hybrid models (combination models) which can combine several models of the same or

Variable	Definition	Relative importance (%)	Rank
ID	Road segment ID	2.28	7
L	Length of the road segment	0.17	23
Speed	Space mean speed	10.59	3
TOD	TOD is indexed from 1 to 96, which represent the time from 0:00–24:00 by every 15 min timestep	30.12	2
DOW	DOW is indexed from 1 to 7, which represent from Monday through Sunday	2.84	5
Month	The month is indexed 1 to 12, which represent from January to December	1.59	8
Weather	Weather is indexed from 1 to 3, which represent normal, rain and snow/ice/fog	2.63	6
T_{t-1}	The travel time at prediction segment 15 min before	34.85	1
T_{t-2}	The travel time at prediction segment 30 min before	0.57	11
T_{t-3}	The travel time at prediction segment 45 min before	0.28	18
T_{t-w}	The travel time at prediction segment one week before	9.87	4
ΔT_{t-1}	The ravel time change value at T_{t-1}	0.24	19
ΔT_{t-2}	The ravel time change value at T_{t-2}	0.20	21
ΔT_{t-3}	The travel time change value at T_{t-3}	0.18	22
ΔT_{t-w}	The travel time change value at T_{t-w}	0.22	20
T_{t-1}^{i-1}	The travel time of the nearest upstream road segment 15 min before	0.31	15
T_{t-1}^{i-2}	The travel time of the second nearest upstream road segment 15 min before	0.42	12
ΔT_{t-1}^{i-1}	The travel time change value at the nearest upstream road segment 15 min before	0.29	16
ΔT_{t-1}^{i-2}	The travel time change value at the second nearest upstream road segment 15 min before	0.29	16
T_{t-1}^{i+1}	The travel time of the nearest downstream road segment 15 min before	0.35	14
T_{t-1}^{i+2}	The travel time of the second nearest downstream road segment 15 min before	0.61	10
ΔT_{t-1}^{i+1}	The travel time change value at the nearest downstream road segment 15 min before	0.79	9
ΔT_{t-1}^{i+2}	The travel time change value at the second nearest downstream road segment 15 min before	0.37	13

Table 7.
Relative importance of each variable and their ranks in the RF model

different types of prediction models to enhance the model performance and prediction. The RF method can be combined with other tree-based methods or another type of machine learning method in the preprocessing step or prediction step. Experimental results showed the combination methods have a better prediction result than using a method alone (Li *et al.*, 2009). As the combination model method has been proved superior in terms of prediction accuracy, this should be given careful consideration in the future.

References

Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5-32.

Chiabaut, N. and Faitout, R. (2021), "Traffic congestion and travel time prediction based on historical congestion maps and identification of consensual days", *Transportation Research Part C: Emerging Technologies*, Vol. 124, p. 102920.

Cristóbal, T., Padrón, G., Quesada-Arencibia, A., Alayón, F., de Blasio, G. and García, C.R. (2019), "Bus travel time prediction model based on profile similarity", *Sensors*, Vol. 19 No. 13, p. 2869.

Dharia, A. and Adeli, H. (2003), "Neural network model for rapid forecasting of freeway link travel time", *Engineering Applications of Artificial Intelligence*, Vol. 16 Nos 7/8, pp. 607-613.

Dion, F., Rakha, H. and Kang, Y.S. (2004), "Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections", *Transportation Research Part B: Methodological*, Vol. 38 No. 2, pp. 99-122.

Fan, S.K.S., Su, C.J., Nien, H.T., Tsai, P.F. and Cheng, C.Y. (2018), "Using machine learning and big data approaches to predict travel time based on historical and real-time data from Taiwan electronic toll collection", *Soft Computing*, Vol. 22 No. 17, pp. 5707-5718.

Fu, K., Meng, F., Ye, J. and Wang, Z. (2020), "CompactETA: a fast inference system for travel time prediction", *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 3337-3345.

Greenhalgh, J. and Mirmehdi, M. (2012), "Real-time detection and recognition of road traffic signs", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13 No. 4, pp. 1498-1506.

Gupta, B., Awasthi, S., Gupta, R., Ram, L., Kumar, P., Prasad, B.R. and Agarwal, S. (2018), "Taxi travel time prediction using ensemble-based random Forest and gradient boosting model", *Advances in Big Data and Cloud Computing*, Springer, pp. 63-78.

Hamner, B. (2010), "Predicting travel times with context-dependent random forests by modeling local and aggregate traffic flow", *2010 IEEE International Conference on Data Mining Workshops*, IEEE, pp. 1357-1359.

Jia, J., Xu, Y., Zhang, S. and Xue, X. (2016), "The facial expression recognition method of random forest based on improved PCA extracting feature", *In 2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, IEEE, pp. 1-5.

Kumar, B.A., Jairam, R., Arkatkar, S.S. and Vanajakshi, L. (2019), "Real time bus travel time prediction using k-NN classifier", *Transportation Letters*, Vol. 11 No. 7, pp. 362-372.

Kuncheva, L.I. and Whitaker, C.J. (2003), "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy", *Machine Learning*, Vol. 51 No. 2, pp. 181-207.

Kwak, S. and Geroliminis, N. (2020), "Travel time prediction for congested freeways with a dynamic linear model", IEEE Transactions on Intelligent Transportation Systems.

Li, Y., Fujimoto, R.M. and Hunter, M.P. (2009), "Online travel time prediction based on boosting", *2009 12th International IEEE Conference on Intelligent Transportation Systems*, IEEE, pp. 1-6.

Mori, U., Mendiburu, A., Álvarez, M. and Lozano, J.A. (2015), "A review of travel time estimation and forecasting for advanced traveller information systems", *Transportmetrica A: Transport Science*, Vol. 11 No. 2, pp. 119-157.

Oh, S., Byon, Y.J., Jang, K. and Yeo, H. (2015), "Short-term travel-time prediction on highway: a review of the data-driven approach", *Transport Reviews*, Vol. 35 No. 1, pp. 4-32.

Papageorgiou, M., Papamichail, I., Messmer, A. and Wang, Y. (2010), "Traffic simulation with METANET", *Fundamentals of Traffic Simulation*, Springer, New York, NY, pp. 399-430.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003), *Semiparametric Regression*, Cambridge university press, No. 12.

Schmitt, E.J. and Jula, H. (2007), "On the limitations of linear models in predicting travel times", *2007 IEEE Intelligent Transportation Systems Conference*, IEEE, pp. 830-835.

Šemanjski, I. (2015), "Analysed potential of big data and supervised machine learning techniques in effectively forecasting travel times from fused data", *PROMET – Traffic&Transportation*, Vol. 27 No. 6, pp. 515-528.

Van Lint, J.W. (2006), "Reliable real-time framework for short-term freeway travel time prediction", *Journal of Transportation Engineering*, Vol. 132 No. 12, pp. 921-932.

Wu, Z., Rilett, L.R. and Ren, W. (2021), "New methodologies for predicting corridor travel time mean and reliability", *International Journal of Urban Sciences*, pp. 1-24.

Wunderlich, K.E., Kaufman, D.E. and Smith, R.L. (2000), "Link travel time prediction for decentralized route guidance architectures", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 1 No. 1, pp. 4-14.

Yildirimoglu, M. and Geroliminis, N. (2013), "Experienced travel time prediction for congested freeways", *Transportation Research Part B: Methodological*, Vol. 53, pp. 45-63.

Zhang, Y. and Haghani, A. (2015), "A gradient boosting method to improve travel time prediction", *Transportation Research Part C: Emerging Technologies*, Vol. 58, pp. 308-324.

Zhu, T., Kong, X. and Lv, W. (2009), "Large-scale travel time prediction for urban arterial roads based on kalman filter", *2009 International Conference on Computational Intelligence and Software Engineering*, IEEE, pp. 1-5.

Zou, N., Wang, J. and Chang, G.L. (2008), "A reliable hybrid prediction model for real-time travel time prediction with widely spaced detectors", *2008 11th International IEEE Conference on Intelligent Transportation Systems*, IEEE, pp. 91-96.

Further reading

Du, L., Peeta, S. and Kim, Y.H. (2012), "An adaptive information fusion model to predict the short-term link travel time distribution in dynamic traffic networks", *Transportation Research Part B: Methodological*, Vol. 46 No. 1, pp. 235-252.

He, H.S., Keane, R.E. and Iverson, L.R. (2008), "Forest landscape models, a tool for understanding the effect of the large-scale and long-term landscape processes", *Forest Ecology and Management*, Vol. 254 No. 3, pp. 371-374.254.

Jiang, X. and Adeli, H. (2004), "Wavelet packet-autocorrelation function method for traffic flow pattern analysis", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 19 No. 5, pp. 324-337.

Leshem, G. and Ritov, Y. (2007), "Traffic flow prediction using adaboost algorithm with random forests as a weak learner", *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 19, pp. 193-198.

Mendes-Moreira, J., Jorge, A.M., de Sousa, J.F. and Soares, C. (2015), "Improving the accuracy of long-term travel time prediction using heterogeneous ensembles", *Neurocomputing*, Vol. 150, pp. 428-439.

Ramezani, M. and Geroliminis, N. (2012), "On the estimation of arterial route travel time distribution with Markov chains", *Transportation Research Part B: Methodological*, Vol. 46 No. 10, pp. 1576-1590.

Sun, H., Liu, H.X., Xiao, H. and Ran, B. (2002), "Short term traffic forecasting using the local linear regression model".

Vlahogianni, E.I., Golias, J.C. and Karlaftis, M.G. (2004), "Short-term traffic forecasting: overview of objectives and methods", *Transport Reviews*, Vol. 24 No. 5, pp. 533-557.

Xu, B. and Qiu, G. (2016), "Crowd density estimation based on rich features and random projection forest", *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 1-8.

Yu, B., Wang, H., Shan, W. and Yao, B. (2018), "Prediction of bus travel time using random forests based on near neighbors", *Computer-Aided Civil and Infrastructure Engineering*, Vol. 33 No. 4, pp. 333-350.

Corresponding author

Wei Fan can be contacted at: wfan7@uncc.edu.