Short-Term Urban Link Travel Time Prediction using Dynamic Time Warping with Disaggregate Probe Data

Ruotian Tang\(^1\), Ryo Kanamori\(^2\), and Toshiyuki Yamamoto\(^3\)

\(^1\)Department of Civil Engineering, Nagoya University, Nagoya, 464-8603, Japan
\(^2\)Institute of Innovation for Future Society, Nagoya University, Nagoya, 464-8603, Japan
\(^3\)Institute of Materials and Systems for Sustainability, Nagoya University, Nagoya, 464-8603, Japan

Corresponding author: Ruotian Tang (e-mail: tang.ruotian@a.mbox.nagoya-u.ac.jp).

ABSTRACT There is increasing demand for short-term urban link travel time prediction to build an advanced intelligent transportation system (ITS). With the development of data collection technology, probe data are receiving more attention but the penetration rate of probe vehicles capable of sending probe data is still limited. Most research pertaining to short-term travel time prediction tends to aggregate probe data to obtain useful samples when the penetration rate is low. However, as a result, the prediction can only provide a general description of the travel time and changes in travel time during a short time interval are neglected. To overcome this limitation, a non-parametric model using disaggregate probe data based on dynamic time warping (DTW) was developed in this study. Data from the crossing direction are introduced to separate the data into different signal phases instead of identifying the exact signal pattern. A classical k-nearest neighbor (KNN) model and a naive model were compared with the proposed model. The models were tested in three scenarios: a computer simulation and two real cases from Nagoya, Japan. The results showed that the proposed model outperforms the other two models under different data penetration rates because it can reflect changes in travel time during a traffic signal cycle. Moreover, the proposed model has wider applicability than the KNN model because it is free from the equal time interval constraint.

INDEX TERMS travel time prediction; disaggregate probe data; short term; dynamic time warping; traffic signal cycle; penetration rate; urban link

I. INTRODUCTION

Due to the rapid process of urbanization and motorization, traffic-related problems have become major social problems in many cities. For example, traffic congestion not only causes inconvenience in daily life but also causes large workforce productivity time loss and excess fuel consumption [1]. Many governments previously attempted to relieve traffic congestion by constructing more roads, but this practice stimulated additional traffic demand [2].

Recently, the intelligent transportation system (ITS) has been widely recognized as an effective solution for traffic congestion and other traffic-related problems because of the development of information technologies such as big data [3]. Within ITS, many pieces of traffic information, such as traffic flow and travel speed, are collected and analyzed to determine traffic control and management strategies for different traffic conditions [4]-[6]. Among all of this information, travel time plays the most important role because it is the most sensitive information to individual users and it can be used in both transportation planning and traffic management. In ITS, there is always a time lag between system and user, especially for systems applying pull-based strategies [3]. Therefore, short-term travel time prediction, the prediction horizon of which ranges from a few seconds to a few hours, has become increasingly crucial for different applications in ITS [7].

To satisfy the increasing requirement for urbanization and motorization, the research interest of short-term travel time forecasting is switching from freeways and arterials to urban roads. Urban traffic conditions change so frequently that large amounts of data are required to update the real-time traffic condition. With the development of data collection technology, researchers have tended to use probe data instead.
of conventional data because of their low cost and wide coverage. The amount of probe data depends on the sampling frequency and the penetration rate of the probe vehicles. If the penetration rate is low, the sampling frequency has to be high to ensure that abundant data can be collected, and vice versa [8]-[9]. Probe data are usually collected from the onboard unit of a vehicle called a probe vehicle. Currently, only a portion of the vehicles (e.g., taxis and buses) on the road are equipped with the onboard unit, so the penetration rate of probe vehicles is low. Although every vehicle could function as a probe vehicle in the future, most research has pointed out that the penetration rate of vehicles capable of actually sending probe data might still be limited because of the cost of the data processing and storage and the privacy issue [3]-[10]-[13]. Moreover, probe data usually distribute unevenly over time, so many researchers tend to aggregate the raw data to obtain useful samples. Fusco and Gori [14] pointed out that a 5-min interval is required in the advanced traveler information system. Therefore, the data are usually aggregated in 5 min in many aspects of research [15]-[18]. However, some studies have adopted a longer time interval, such as 15 min [19]-[21].

Although the aggregation of data is necessary, especially for multistep prediction, there are two constraints for application on urban networks. First, forecasting results are also aggregated. Because the urban traffic condition changes quickly, forecasting results that are aggregated by minutes cannot react to changes in traffic conditions that occur in seconds. Second, the influence of the traffic signal cannot be reflected properly. The variation of travel time in urban networks comes mainly from the traffic signal. Thus, because the length of most aggregation time intervals that have been adopted is longer than the length of a signal cycle, the aggregation might reduce the variation of travel time. The reduction of this variation is good for the model, but the temporal change of the traffic condition could be missed. Although some aggregate models can provide travel time distributions for signalized sections of roadway, their assumptions about distributions might vary according to the study site [22]. Such distributions can provide a general description of travel time during the time interval, but they cannot reflect temporal changes in a short period, which play an important role in applications such as online personal car navigation. To overcome these constraints, this paper proposes a model based on dynamic time warping (DTW) to predict the short-term link travel time in urban networks using disaggregate probe data. Moreover, to reflect temporal changes in travel time, the forecasting step is based on the traffic signal cycle instead of a fixed time interval.

In conclusion, the main contribution of this paper is using the raw probe data (i.e., disaggregate data) directly without aggregation to predict the urban link travel time in the short term so that changes in travel time during a short time interval can be reflected. The rest of this paper is organized as follows. In Section 2, a literature review of models pertaining to short-term travel time prediction is given. Section 3 explains the details of the proposed model. Section 4 describes the datasets used in this paper, which include simulation data and two real cases from Nagoya, Japan. In Section 5, the proposed model is analyzed and compared with a naïve model and a classical k-nearest neighbor (KNN) model under different levels of data penetration rates. The last section provides the conclusions of this paper and discusses future work.

II. Literature review

To date, many models have been developed to predict short-term travel time. They can be categorized into four types, namely, the naïve model, traffic flow-based model, data-driven model, and hybrid model [23]. Naïve models use the average value of the instantaneous or historical data as the prediction. Despite their low accuracy, they are employed widely due to their low computation cost and convenient implementation. They are also used as a benchmark for comparison with other complex models in academic research.

Traffic flow-based models include macroscopic, microscopic, and mesoscopic simulation models [24]. They can provide detailed information, such as the delay and the queue length, for decision-making based on traffic theory. However, the real traffic condition is so complex that it is difficult to make a general simulation. Unlike traffic flow-based models, data-driven models perform better under complex and dynamic traffic conditions because there is no restricted theory. Although there is a lack of interpretability, these models can make considerable predictions as long as there are enough data available at the study site. Hybrid models refer to combinations of traffic flow-based models and data-driven models [15]-[25]. Widespread application of hybrid models in the future could be promising because they have the advantages of both the traffic-flow model and the data-driven model. For more details, readers can refer to reviews of up-to-date methods concerning traffic forecasting [7]-[23]-[26].

Currently, data-driven models attract the most interest because they can deal with the complex traffic condition of urban networks and the computation cost is relatively low. There are two main types of data-driven models: the parametric model and the non-parametric model. Among the different types of parametric models, time series models, which take advantage of the temporal relationship between travel time and traffic condition, have received the most attention. For example, a Kalman filter (KF) is commonly used to estimate the traffic condition and can be associated with an autoregressive integrated moving average (ARIMA) model, which can project the recurrent pattern of the traffic condition into the future [18]-[19]. Other methods such as the particle filtering model [17] and the Bayesian dynamic linear model [27] are also used to make predictions under both recurrent and non-recurrent traffic conditions. Most parametric models can provide accurate prediction under
linear situations, such as freeways and arterials, but they have difficulty addressing non-linear situations, such as urban roads. In the case of a non-linear situation, non-parametric models are preferred.

Among the different types of non-parametric models, neural network (NN) models and pattern recognition models are most popular. Researchers have devoted substantial efforts to developing different types of NN models. Jiang and Zhang [28] proposed a generalized regression neural network that can convert the average speed into a travel time and tested the model on an arterial road in China. Lint et al. [29] proposed a state-space neural network (SSNN) to predict travel time on freeways with missing data. Lint [30] further combined SSNN and KF to solve the inherently delayed travel time prediction problem. Fusco et al. [31] applied a NN model to an urban network, and the results showed that the NN model outperforms the ARIMA model under recurrent congestion conditions. Although NN models are capable of extracting the complex relationships between different traffic variables, there are three main drawbacks that constrain them from practical applications on large networks. First, NN models require long training processes and have to train at each study site. Second, NN models require large storage capacity to store a huge amount of data, but the storage capacity of a vehicle’s onboard unit is limited because of other smart functions. Third, NN models lack interpretability due to their ‘black box’ nature [3] [15] [23].

On the contrary, pattern recognition models are easy to implement and transfer to different sites without data training. For example, the KNN model, which is more popular than other pattern recognition models such as support vector machine [32] and cluster analysis [33], is used widely to predict different traffic variables. Smith and Demetsky [34] showed that KNN models have the potential to serve as accurate and portable prediction models and have advantages over NN models. Habtemichael and Cetin [20] compared the KNN model with the adaptive KF model and the seasonal ARIMA model when predicting the traffic flow on freeways. The results demonstrated that the KNN model outperforms the adaptive KF model and the seasonal ARIMA model in real-time traffic control and management for freeways. Robinson and Polak [35] predicted the urban link travel time in London and tested their model using different KNN design parameters. Cai et al. [16] focused on improving the search algorithm and result integration when predicting travel speeds on urban networks. They used a spatiotemporal state matrix to describe the traffic state instead of only a time series as in the classic KNN model. The results showed that both the classic KNN model and the improved KNN model outperform the Elman-NN model.

However, most pattern recognition models have to use aggregated data because they compare the alignment between two time-sequential groups of data, which must be sorted in exactly the same time interval and have exactly the same length. To avoid the aforementioned constraints and reflect real-time changes in travel time, a new pattern recognition model using DTW has been developed and is compared with the classic KNN model in this paper. DTW was developed to classify nonlinear time-sequential sequences with different lengths and has long been used in speech pattern recognition [36]. DTW has also been applied successfully to various fields such as data mining [37] and bioinformatics analysis [38]. In the field of transportation, DTW has been employed for out-of-sequence traffic classification by comparing two traffic flows with different lengths [39]. Another application was to estimate the vehicle speed with a distorted magnetic signature due to behaviors such as acceleration and deceleration within the monitoring distance [40].

III. METHODOLOGY

Unlike estimation, which has the goal of providing a general description of the traffic condition on a certain road, the objective of prediction is to forecast the travel time for a trajectory. Because there might be a large number of possible trajectories for one combination of origin and destination, most researchers pay attention to predicting the link travel times separately and then summing the link travel times to obtain the travel time of the trajectory [23]. Because the uncertainty of urban travel time usually comes from the traffic signal, our attention was focused on the urban link defined by the segment of a roadway between two adjacent signalized intersections.

Although information about signal timing may be accessible to researchers or even to the public in the future, it is considered confidential at present. Several methods have been developed to estimate signal timing using the completed trajectory data obtained from probe vehicles [41]-[44]. However, the delay caused by the traffic signal cannot be estimated separately when making a prediction because the trajectory is incomplete. In this paper, the link travel time consists of the running time on the link and the stopping time at the downstream signalized intersection. The time stamp is defined as the time when vehicles leave the link at the downstream signalized intersection.

It has been demonstrated that different turning movements experience different delays at signalized intersections. Feng et al. [45] pointed out that vehicles turning left or right at an intersection need to yield to vehicles going straight, which make up the largest portion of the traffic in most cases. Therefore, this paper focuses on the link travel time of vehicles going straight through the intersection. These vehicles are defined as object vehicles. To reflect the influence of traffic signals, vehicles in the crossing direction are also recorded. These vehicles are defined as crossing vehicles.
A. SAMPLE SELECTION
Similar to the KNN model, the proposed model compares samples (i.e., link travel time sequences) from the historical database with the real-time sample. Figure 1 shows an example of a sample in the proposed model, along with the corresponding example in the KNN model.

![Figure 1. Examples of samples with two signal cycles.](image)

One sample consists of two sections: a reference section and a prediction section. The reference section is used to make a comparison between the historical sample and the real-time sample, whereas the prediction section is used to produce a prediction. In this paper, the length of each section is measured by the number of signal cycles. A signal cycle is divided into two phases: a green phase when the object vehicle can move through the intersection and a red phase when crossing vehicles can move through the intersection. The continuous object vehicle data stream is used to represent the green phase, whereas the continuous crossing vehicle data stream is used to represent the red phase. Because the value of the crossing vehicle is not of interest, it is set as $L$, which is significantly greater than any possible value of the object vehicle data.

In the proposed model, the travel times from both object vehicle and crossing vehicle are directly recorded in the same link travel time sequence to represent the sample, whereas average travel times of the object vehicles are used in the KNN model. For both sections, the start point is the first observation in the continuous object vehicle data stream, and the end point is the last observation in the continuous object vehicle data stream. A reference section must contain more than two object vehicle observations to avoid a crossing vehicle data stream. A reference section must be used to calculate the local cost.

2. If the current sampling time is in the green phase, one crossing vehicle observation is added at the end of the sample, and the observations before it are recorded in the same way as in condition 1.

For the historical sample, the prediction section contains only one signal cycle, whereas the reference section can contain successive signal cycles. For comparison with the historical sample, the reference section in the real-time sample should have the same length as in the historical sample. On the other hand, to simulate the multi-step prediction, the prediction section in the real-time sample contains successive signal cycles.

B. DTW ALGORITHM IN THE PROPOSED MODEL
In the proposed model, the classic DTW algorithm is modified to compare reference sections of the historical sample and the real-time sample, e.g., $P = (p_1, p_2, ..., p_N)$ and $Q = (q_1, q_2, ..., q_M)$. To compare the different travel times between the two sections, a local cost measure is needed. There are various methods for calculating the local cost [36], such as the Manhattan distance and the Euclidean distance. Equation (1) is used in the proposed model to calculate the local cost.

$$c_{ij} = (p_i - q_j)^2, \quad i \in N, j \in M$$

A local cost matrix, in which each element represents the local cost $c_{ij}$ between the points $p_i$ and $q_j$, is built, and several optional alignments between $P$ and $Q$ can be found. An optional alignment is represented by a warping path $A = (a_i, a_2, ..., a_H)$ with $a_h = (i, j)$ for $h \in [1, H]$, which is shown in Figure 2. There are two basic constraints on the warping path selection [36]:

1. Boundary constraint: $a_1 = (1, 1)$ and $a_H = (N, M)$.
2. Unit step-size constraint: For any $h \in [1, H - 1]$, $a_{h+1} - a_h \in \{(0, 1), (1, 0), (1, 1)\}^T$.

![Figure 2. Illustrations of a warping path.](image)
called monotonicity. In other words, for any \( h \in [1, H - 1] \), if \( a_{h+1} = (i, j) \) and \( a_h = (i', j') \), then \( i \geq i' \) and \( j \geq j' \). The monotonicity constraint ensures that there is no step back while searching for the warping path.

The purpose of the DTW is to ascertain the shortest warping path between the sections \( P \) and \( Q \). Each path has its own accumulated cost, which is the summation of each local cost corresponding to each path obtained by using (2).

\[
C_a(P,Q) = \sum_{h=0}^{H} \alpha_h c_h(a_h), \quad a_h \in A
\]

\[
\alpha_h = \begin{cases} 
1, & a_h - a_{h-1} \in \{(1,0),(0,1)\} \\
2, & a_h - a_{h-1} = (1,1) \\
0, & a_h = (0,0) 
\end{cases}
\]

The coefficient \( \alpha_h \) is calculated by (3) to avoid the preference of choosing the diagonal path. The shortest warping path is the one with the minimum accumulated cost, i.e., \( DTW(P,Q) = \min[C_a(P,Q) | A \text{ is a warping path between sequences } P \text{ and } Q] \), which represents the similarity between the two sections.

If the accumulated cost of each possible warping path is calculated, the shortest path can be found. However, it is computationally expensive to do so, particularly when the local cost matrix is large. However, there is an existing method to find the shortest warping path without calculating the costs of all the possible paths [36] [40]. The minimum accumulated cost can be calculated using the algorithm listed in Table 1. First, a local cost matrix between the sections \( P \) and \( Q \) is built (Lines 1–2). Then, an additional column and an additional row are added to the original local cost matrix (Lines 3–5). Finally, the cost of the shortest warping path between the sections \( P \) and \( Q \) is calculated using (4) (Lines 6–13).

**Table 1.** Modified classic DTW algorithm

1. Build an \( N \times M \) local cost matrix \( C \) between sequences \( P = (p_1, p_2, \ldots, p_N) \) and \( Q = (q_1, q_2, \ldots, q_M) \)
2. Add an additional row, where \( c_{0j} = \infty, i > 0 \)
3. Add an additional column, where \( c_{ij} = \infty, j > 0 \)
4. \( c_{00} = 0 \)
5. For \( i = 1 \) to \( N \)
6. For \( j = 1 \) to \( M \)
7. \( \text{DTW}(p_i, q_j) = \min \left( \text{DTW}(p_{i-1}, q_j) + c_{ij}, \text{DTW}(p_i, q_{j-1}) + c_{ij}, \text{DTW}(p_{i-1}, q_{j-1}) + 2c_{ij} \right) \)
8. \( \text{DTW}(P,Q) = \text{DTW}(p_N, q_M) \)

Table 2 presents an example of searching for the shortest warping path between the reference sections of two samples: \( (70, 20, 15, L, L, 90, 88, L) \) and \( (68, 60, 26, 10, L, L, 82, 77, 69, L, L) \). The reference sections of both samples contain two signal cycles. The shaded portion from the upper left to the lower right represents the shortest warping path. In Table 2, elements represent the local costs calculated using (1), except for those in the first column and the first row. The local cost matrix can be divided into several sub-matrices using the crossing vehicle data \( L \) as long as the value of \( L \) is sufficiently high. The sub-matrices, through which the shortest warping path traverses, represent the local cost matrices between the travel time sequences in the corresponding signal phases. When the penetration rate is high, it is necessary to record the crossing vehicle data so that the travel time sub-sequence in one sample can be compared with the one in the corresponding signal phase in another sample if there is more than one signal cycle in the reference section. When the penetration rate is low, the crossing vehicle might be misleading. Therefore, the shortest warping path between the reference sections of two samples without recording the crossing vehicle should also be found.

**Table 2.** Example of searching the shortest warping path in the local cost matrix

<table>
<thead>
<tr>
<th></th>
<th>70</th>
<th>20</th>
<th>15</th>
<th>L</th>
<th>L</th>
<th>90</th>
<th>88</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>68</td>
<td>4</td>
<td>2304</td>
<td>2809</td>
<td>L'</td>
<td>L'</td>
<td>484</td>
<td>400</td>
<td>L'</td>
</tr>
<tr>
<td>60</td>
<td>100</td>
<td>1600</td>
<td>2025</td>
<td>L'</td>
<td>L'</td>
<td>700</td>
<td>784</td>
<td>L'</td>
</tr>
<tr>
<td>26</td>
<td>1936</td>
<td>36</td>
<td>121</td>
<td>L'</td>
<td>L'</td>
<td>4096</td>
<td>3844</td>
<td>L'</td>
</tr>
<tr>
<td>10</td>
<td>3600</td>
<td>100</td>
<td>25</td>
<td>L'</td>
<td>L'</td>
<td>6400</td>
<td>6084</td>
<td>L'</td>
</tr>
<tr>
<td>1</td>
<td>144</td>
<td>3844</td>
<td>4489</td>
<td>L'</td>
<td>L'</td>
<td>64</td>
<td>36</td>
<td>L'</td>
</tr>
<tr>
<td>77</td>
<td>49</td>
<td>3249</td>
<td>3844</td>
<td>L'</td>
<td>L'</td>
<td>169</td>
<td>121</td>
<td>L'</td>
</tr>
<tr>
<td>69</td>
<td>1</td>
<td>2401</td>
<td>2916</td>
<td>L'</td>
<td>L'</td>
<td>441</td>
<td>361</td>
<td>L'</td>
</tr>
<tr>
<td>82</td>
<td>144</td>
<td>3844</td>
<td>4489</td>
<td>L'</td>
<td>L'</td>
<td>64</td>
<td>36</td>
<td>L'</td>
</tr>
<tr>
<td>77</td>
<td>49</td>
<td>3249</td>
<td>3844</td>
<td>L'</td>
<td>L'</td>
<td>169</td>
<td>121</td>
<td>L'</td>
</tr>
</tbody>
</table>

**C. PREDICTION**

In the classical KNN model, the average travel time for object vehicles in each signal cycle is used. The prediction \( \hat{\tau} \) is given by the weighted summation of \( \vec{\tau}_i \) from the first \( K \) historical samples with the highest weight, as shown in (5) and (6):

\[
W_i = 1 / (\sum_i W_i \cdot (\vec{\tau}_i - \vec{\tau}_r)^2)^{1/2}
\]

\[
\vec{\tau} = \sum_i W_i \cdot \vec{\tau}_i
\]

where \( W_i \) is the weight, \( \vec{\tau}_r \) is the average travel time for the \( r \)th signal cycle in the reference section of the historical sample, \( \vec{\tau}_i \) is the average travel time for the \( r \)th signal cycle of the real-time sample, \( N_i \) is the number of signal cycles in the reference section, and \( \vec{\tau}_i \) is the average travel time for the prediction section of the historical sample.

Some pieces of information are lost when using the average travel time in a signal cycle. For example, the link travel time tends to decrease after the traffic signal turns...
green because vehicles leaving in the subsequent green phase might pass through the intersection without a stop. Therefore, in the proposed model, the prediction is based on the link travel time sequence instead of one single value. Because the exact length and start time point of a signal phase are unknown, it is unreasonable to merge the travel time sequences from the prediction section of different historical samples. Therefore, the prediction section of the historical sample with the highest similarity is used directly for the prediction. The similarity is calculated by (7):\

$$s_{i} = 1/[(\min(DTW(P, Q)/l, DTW(P', Q)/l'))^{1/2}]$$ (7)\

where $P$ is the reference section of the historical sample with the crossing vehicle data, $Q$ is the reference section of the real-time sample with the crossing vehicle data, $l$ is the length of the shortest warping path between $P$ and $Q$, $P'$ is the reference section of the historical sample without the crossing vehicle data, $Q'$ is the reference section of the real-time sample without the crossing vehicle data, and $l'$ is the length of the shortest warping path between $P'$ and $Q'$. As mentioned above, the prediction section of the historical data contains only one signal cycle, implying that predictions can be made only in the next signal cycle. For the multi-step prediction, a new reference section of the real-time sample is made at each step by removing the first signal cycle of the reference section and adding the current prediction to the end of the reference section. Then, the prediction at the next step can be made by comparing the historical samples and the real-time sample with the new reference section.

**IV. DATA PROCESS**

**A. SIMULATION DATA**

Currently, probe data are usually collected from special vehicles (e.g., taxis and buses), so the penetration rate of vehicles that can send probe data is usually low. This study used probe data collected from taxis, which means that it also suffers from the problem of low penetration rate. Therefore, computer simulation was employed to evaluate the proposed model under different penetration rates. The traffic simulation was constructed using VISSIM (version 7.0) at a normal cross intersection with four 200-m-long links. One link was selected as the target link, the travel time of which was predicted. For simplicity, vehicles were made to only move straight through the intersection, and their initial speeds varied from 10 to 50 km/h. The traffic volumes of the target link were 200, 600, and 100 vehicles/h in the first, second, and third 10 min, respectively, and this pattern was repeated every 30 min. For the other three links, the traffic volume was fixed at 200 vehicles/h. Changes in the traffic volume were considered to simulate a situation in which congestion appears and then disappears in the target link. All of the vehicles were moving according to the default setting. The signal pattern was set to have two phases without the all-

red phase, and the length of each phase was 60 s.

The simulation was repeated 30 times, and vehicles were delivered randomly for each time according to the default setting. Five trials were used as the testing data. The remaining 25 trials made up the historical database. The amount of data in the testing database was 1,475, whereas the amount of data in the historical database was 7,388. The average link travel time in the test database was 82.0 s, whereas that in the historical database was 86.6 s. Figure 3 shows the distribution of travel time and the changes in travel time in the target link. Fusco et al. [31] showed that the connection between the traffic signals and the individual vehicle speeds can be inferred by the distribution of individual vehicle speed. For the same reason, the connection between the traffic signals and the travel times can be inferred by the distribution of travel time. For example, the peak at approximately 60 s represents data from vehicles having a stop at the intersection, whereas the peak at approximately 110 s represents data from vehicles having at least one stop at the intersection. In this manner, the traffic condition was defined as congested when the average travel time was greater than 80 s in the simulation. Congestion was observed from the 14th min to the 24th min and from the 44th min to the 54th min.

![FIGURE 3. Distribution and changes in travel time on the target link.](image-url)
B. REAL DATA

The probe data used in the real cases were obtained from taxis in Nagoya, Japan, in February and June 2015 (58 days in total). Each data record represented a taxi that passed through a link and included information such as the times the taxi entered and exited the link, the ID and length of the link, the ID of the next link, and the latitude and longitude of the endpoints for each link. Data pertaining to the stopping of the taxi on the link to pick up or deliver passengers were discarded. Because taxis only accounted for a small portion of all vehicles on the road, some of the links had insufficient data. Consequently, two links with a relatively large amount of data were selected from the entire network as target links. One link (Link 1) was a 209-m-long west–east link on Hirokoji Street, which connects the city center and a suburban area. The other one (Link 2) was a 223-m-long north–south link on Ootsu Street at the city center. Both of the links connect signalized intersections.

The amount of data for Link 1 was 23,815, among which the number of data of taxis moving straight was 17,478. The amount of data for Link 2 was 47,001, and the number for taxis going straight was 26,586. The number of data of taxis going straight for both links was overwhelming compared to all possible turning choices, so it was reasonable to focus on taxis going straight. Figure 4 shows the changes in the average number and travel time of taxis going straight per hour during one day for both links in February and June.

V. EXPERIMENTS

As mentioned previously, the prediction in the proposed model is based on the travel time sequence instead of one single value. This makes it possible to reflect the changes in travel time during a signal cycle. Because the time interval and sample length of the forecasting result differ from the target, DTW is also used to compare the forecasting result and the target. As shown in Figure 2, the travel time in one sequence can be considered as the travel time in another according to the warping path. For example, in Figure 2, if $P$ represents the forecasting result and $Q$ represents the target, then $p_1$, $p_2$ refers to $q_1$ and $p_2$ refers to $q_2$. Therefore, according to the shortest warping path, the travel time in the target can be represented by the average value of the corresponding elements in the forecasting result. For example, if the forecasting result is $(68, 60, 26, 10)$ and the target is $(70, 20, 15)$, the forecasting result can be transformed into $(64, 26, 10)$. Aside from the KNN model, the naïve model, which uses the average value of the historical data as the forecasting result, was also used to compare with the proposed model. The average value used in the KNN model or in the naïve model was considered as a forecasting result containing only one element. Two indices were used to measure the accuracy of the proposed model. One is the mean absolute percentage error (MAPE) and the other is the root mean squared error (RMSE), which are defined as follows:

$$\text{MAPE} = \frac{\sum_{i}^{N} |\hat{t}_i - t_i|}{N} \times \frac{1}{\hat{t}_i}$$  \hspace{1cm} (8)

$$\text{RMSE} = \left( \frac{\sum_{i}^{N} (\hat{t}_i - t_i)^2}{N} \right)^{1/2}$$  \hspace{1cm} (9)

where $N$ is the total amount of object vehicle data in the target, $t_i$ is the true value of the travel time, and $\hat{t}_i$ is its corresponding forecasting result.

A. PARAMETER CALIBRATION

The simulation data were used to calibrate the parameters. The proposed model and the KNN model were tested under different penetration rates in the simulation. If more than 75% of the data were reduced, the number of effective samples would be insufficient for parameter calibration. Therefore, the data were randomly reduced by 50% and 75% to simulate conditions of 50% and 25% penetration rates, respectively. When calibrating the parameters, the prediction horizon was fixed as one signal cycle.
For the KNN model, two parameters required calibration: the length of the reference section, which influences the sample size and the information provided by the sample, and the value of $K$, which is the number of historical samples used for prediction. Figure 5 shows the MAPE of the KNN model with different combinations of reference section length and $K$ under different penetration rates.

![Figure 5: MAPE of the KNN model with different combinations of reference section length and $K$ under different penetration rates.](image)

The MAPE fluctuates slightly when the penetration rate is 100%. The MAPE decreases at first and then increases under the other two penetration rates. The MAPE decreases because the additional length is expected to help provide more information. The increase in the MAPE is due to the lack of crossing vehicle data, which can separate the signal cycles under low penetration rates. If the data are not separated into corresponding signal cycles, more data lead to more redundancy. This can also explain why the increase in the MAPE under the 25% penetration rate is faster than that under the 50% penetration rate. Therefore, to ensure sufficient samples and that the proposed model can attain a stable accuracy, the length was set as 2.

B. RESULTS AND DISCUSSION

The proposed model, along with the KNN and naïve models, was tested against the simulation and real cases. Because the length of the reference section of the real-time samples used in the proposed model was shorter than that in the KNN model, the number of real-time samples in the proposed model was larger. Therefore, when testing the proposed model, the same real-time samples as in the KNN model were used to make sure that the real-time sample size was the same, but only the last two signal cycles were used when comparing the reference sections. In the simulation, multi-step predictions were made using the models under different penetration rates. As mentioned previously, there must be more than two object vehicle observations in each signal cycle of a sample. However, when the penetration rate is 10%, each signal cycle of the samples contains at least two object vehicle observations to make sure there are enough samples. Despite this, there is no qualified sample with a prediction horizon of five when the penetration rate is 10%. Figure 7 shows the accuracy of the three models under different penetration rates in terms of MAPE and RMSE.

![Figure 6: MAPE of the proposed model with different reference section lengths under different penetration rates.](image)
The proposed model outperforms the other two models irrespective of the penetration rate because it can reflect the changes in travel time during a signal cycle. When the prediction horizon increases, the accuracy of the proposed model deteriorates. This occurs because it cannot catch up with the change in traffic time in the long term if there are no new data to update the current traffic condition. However, when the penetration rate is 10%, the accuracy fluctuates. The accuracy of the naïve model is relatively stable under different prediction horizons, except for the case of the 10% penetration rate. Therefore, it is reasonable to conclude that the fluctuation comes from the bias of samples under the low penetration rate.

When the penetration rate decreases, the MAPE values of all three models tend to decrease because the sample diversity decreases and the predictable range shrinks. We now consider samples with a prediction horizon of one as an example, as listed in Table 3.
Because each sample must contain four successive signal cycles and each signal cycle must contain at least two object vehicle observations according to the previous setting, if the penetration rate decreases, the sample size also decreases. For the same reason, successive signal cycles that contain sufficient object data tend to appear congested under low penetration rates. Therefore, both the historical and real-time samples tend to be collected when the traffic is congested under low penetration rates. Moreover, because the samples are mainly obtained from congested conditions in which the travel time is considerable, the MAPE tends to be low. Concerning the computation cost, the program was built by C++ and run by a computer with a 2.5 GHz Intel Core i7 processor and 16 GB of 1600 MHz DDR3 memory. The running time for one prediction in the proposed model was approximately 0.45 s. In the KNN model it was approximately 0.02 s.

As for the real cases, the parameter setting was the same as that in the simulation. Because the penetration rate was extremely low, the prediction horizon was fixed as one signal cycle and the length of the reference section was set as 2 to collect sufficient real-time samples. For the proposed model, the sample sizes in Links 1 and 2 were 246 and 115, respectively, whereas for the KNN model, the sample sizes in Links 1 and 2 were 66 and 9, respectively. Therefore, data were used more effectively in the proposed model than in the KNN model. Although the amount of data in Link 1 was less than that in Link 2, the sample size in Link 1 was greater than that in Link 2. This occurred because most of the data in Link 1 were distributed during the nighttime, whereas in Link 2, the data were distributed more evenly, as shown in Figure 4. Leave-one-out cross-validation was performed to test the different models in Links 1 and 2, as listed in Tables 4 and 5. The proposed model still outperformed the KNN model and the naïve model under extremely low penetration rates. The accuracy of the KNN model is lower than that of the naïve model because the length of the signal cycle changes for the real cases, implying that the time interval when collecting samples in the KNN model is not equal. However, the proposed model is free from the equal time interval constraint, thus exhibiting a higher accuracy.

To compare the proposed model and the KNN model further, residual analysis was performed and the results are shown in Figure 8. The residual is defined as the difference between the observed travel time and its prediction. The distribution of residuals in the KNN model tends to be linear because it uses the weighted average value as the prediction. The distribution of residuals in the proposed model shows more randomness than that in the KNN model because its prediction is in the form of a travel time sequence, which can reflect changes in travel time. However, when the observed travel time is too short or too long, the residuals of the proposed model also have a bias. This occurs because the penetration rate is so low that the proposed model cannot find similar patterns in the past to make predictions. On the whole, it is reasonable to conclude that the residual has a more desirable random distribution in the proposed model than in the KNN model.

VI. CONCLUSION AND FUTURE WORK

In this study, a pattern recognition model using the DTW is developed to predict the short-term urban link travel time with disaggregate probe data based on traffic signals. The DTW is a method used to compare two sets of the time

<table>
<thead>
<tr>
<th>TABLE 3. Sample size and composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penetration rate (%)</td>
</tr>
<tr>
<td>Sample size</td>
</tr>
<tr>
<td>Ratio of samples in congestion (%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 5. Accuracy of different models in Link 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed model</td>
</tr>
<tr>
<td>MAPE (%)</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 4. Accuracy of different models in Link 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed model</td>
</tr>
<tr>
<td>MAPE (%)</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
</tbody>
</table>
sequential data and is free from constraints of the equal time interval and equal sample length. In the proposed model, crossing vehicle data are introduced to divide the object data into different traffic signal cycles instead of identifying the exact signal patterns. The proposed model predicts the travel time sequence in a signal cycle instead of the average travel time which can reflect the changes in travel time caused by the traffic signal.

A classical KNN model and a naïve model were compared with the proposed model. The models under different data penetration rates were tested in three scenarios. One was a computer simulation and the other two were real cases from Nagoya, Japan. In the simulation, the models were employed for multi-step predictions under different penetration rates. The longest prediction horizon was five signal cycles, which corresponds to ten minutes. The proposed model outperformed the other two models because it could reflect the changes in travel time during a signal cycle. However, it failed to reflect the changes in travel time in the long term. For lower penetration rates, all three models tended to have smaller MAPE values because the effective samples tended to be congested and it was difficult to collect samples when the traffic condition was uncongested, implying that the travel time in the samples was relatively high. In the real cases, the prediction horizon was fixed as one signal cycle to collect sufficient real-time samples. The results showed that the proposed model is better than the other two models even under extremely low penetration rates and that it can be applied more widely because it is free from the constraint of the equal time interval and can reflect changes in travel time.

Generally, because the proposed model takes advantage of disaggregate data and because it is not required to estimate the exact signal pattern, the method can use data more effectively and has wider applicability. However, although the method can reflect changes in travel time in a traffic signal, it has some drawbacks. For example, it is computationally expensive to make the prediction, especially when the sample length is considerable; it cannot deal with the uncongested traffic condition when the penetration rate is low; and it cannot catch up with the changes in travel time in the long term. Hence, there is still room to improve the proposed model. Some future plans include: (i) considering vehicles that turn at the intersection, (ii) reducing the computation cost, and (iii) predicting travel time under the uncongested traffic condition when the penetration rate is low.

ACKNOWLEDGMENT
This research was partially supported by the Center of Innovation Program from the Japan Science and Technology Agency, JST and JSPS KAKENHI grant number JP16H02367.

REFERENCES
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2929791, IEEE Access


R. Kanamori received a Doctor of Engineering from Nagoya University, Japan, in 2007 where he works as a research associate professor. His research interests include evaluation of transport policies with travel demand forecasting models and travel behavior analysis. He is currently analyzing driving-stress with multiple biological indexes in driving experiments and also evaluating the impact of mobility services such as ride-sharing systems.

T. Yamamoto received a Doctor of Engineering from Kyoto University, Japan, in 2000. He is a professor of the Institute of Materials and Systems for Sustainability at Nagoya University, Japan. His research interests include ITS technologies for environmentally sustainable transportation, household vehicle ownership, activity analysis, and traffic safety.