Travel Time Prediction of Bus Journey with Multiple Bus Trips

Peilan He, Guiyuan Jiang, Siew-Kei Lam, Dehua Tang

***Abstract*—Accurate travel time prediction of public transport is essential for reliable journey planning in urban transportation systems. However, existing studies on bus travel/arrival time prediction often focus only on improving the prediction accuracy of a single bus trip. This is inadequate in modern public transportation systems where a bus journey usually consists of multiple bus trips. In this paper, we investigate the problem of travel time prediction for bus journey that takes into account the passenger's riding time on multiple bus trips, and also his/her waiting time at transfer points (interchange stations or bus stops). A novel framework is proposed to separately predict the riding and waiting time of a given journey from multiple datasets (i.e. historical bus trajectories, bus route, and road network), and combining the results to form the final travel time prediction. We empirically determine the impact factors of bus riding times and develop a Long Short-Term Memory (LSTM) model that can accurately predict the riding time of each segment of the bus lines/routes. We also demonstrate that the waiting time at transfer points significantly impacts the total journey travel time, and estimating the waiting time is non-trivial as we cannot assume a fixed distribution waiting time. In order to accurately predict the waiting time, we introduce a novel Interval-based Historical Average (IHA) method that can efficiently address the correlation and sensitivity issues in waiting time prediction. Experiments on real-world data show that the proposed method notably outperforms six baseline approaches for all the scenarios considered.**

***Index Terms*—Travel time prediction, bus journey, waiting time prediction, LSTM, interval-based historical average.**

I. Introduction

Efficient and easy-to-use public transportation system is an important element in sustainable cities as it can boost reduction in traffic congestion and lower carbon emissions from vehicles [1]. A key enabler to the success of public transportation system lies in the provision of accurate travel time information for travelers to make reliable journey planning. This is especially vital for bus services which typically account for the majority ridership among all public transport journeys [2]. Unfortunately current bus services have difficulties in maintaining deterministic travel times as the bus networks are inherently unstable due to heterogeneous traffic conditions and passenger demands.

One of the main issues that contribute to the decrease in quality of bus service is the perennial problem of long and unknown waiting time during transfer, which discourages passengers from continuing to use public transport [3], [4]. This will become more pronounced in connective networks [5] or high performance bus (HPB) network [6], [7] that are designed to increase the number of travel options through the use of more frequent services and transfers. This problem can be effectively circumvented by providing accurate travel time prediction to passengers so that they can plan a preferred route and departure time. Travel time prediction is also elementary to dynamic route guidance systems that provide intermodal transport options and routes to travellers based on real-time data. The availability of accurate and timely information such as bus travel/arrival time will become indispensable not only in current public transportation networks that are constantly expanding, but also in emerging agile transportation systems where buses are scheduled and routed based on travel demand rather than adhering to fixed time-tables and routes [8].

Many researchers have proposed algorithms to predict travel time for taxis or private cars between an origin-destination (OD) pair [9], [10], [11], [12], but research on bus journey time prediction is rather limited. The former approaches cannot be directly applied to bus travel time prediction because bus networks are not only affected by traffic conditions (vehicle speed, traffic flow, traffic signal etc.), but also by factors such as travel demands, dynamic bus load, bus schedules, etc. As such, compared to travel time prediction for cars, travel time prediction for buses is more complex and pose significant challenges in achieving optimal solutions. Moreover, existing approaches for bus travel time prediction fail to consider waiting time and transfer time at transfer points (interchange stations and bus stops), which is a non-negligible part of a passengers' total journey time. This challenge is compounded by the lack of sufficient historical travel records of individual passengers to enable proper validation of the algorithms.

In this paper, we focus on predicting the bus journey travel time for passengers. We make a distinction between journey and trip, where a *trip* is traveled on a single route using one mode (e.g. single bus line/route using one bus service without transferring), and a *journey* consists of one or more trips where transfers are made between services occurring within a certain time frame (e.g. 60 minutes). Unlike existing works on travel/arrival time prediction for a single bus trip [6-32], we address the problem of travel time prediction for a bus journey that involves riding times of multiple bus trips and passenger waiting times at transfer points. We conduct extensive experiments using real bus travel data to evaluate the prediction accuracy of the proposed method for the bus riding time, waiting time, and overall journey time. Experimental results demonstrated the superiority of the proposed approach over several state-of-the-art methods. The main contributions

P. He, G. Jiang, S.K. Lam are with the School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore. (e-mail: phe002@e.ntu.edu.sg, {gyjiang, assklam}@ntu.edu.sg).

D. Tang is with the School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin, 300387, China (e-mail: de_hua_tang@126.com).

of this work are summarized as follows:

1) To the best of our knowledge, our work is the first to predict the travel time of a bus journey consisting of riding times on multiple bus trips, and waiting times during transfers. We propose a Partitioning and Combination Framework (PCF) that addresses the heterogeneous distribution of waiting time at each bus stop, as well as the riding time of bus trips with different frequency and travel speed. A given journey is first partitioned into multiple components that comprises of waiting times at each transfer points (waiting time components) and bus riding times on each line segments (riding time components). Prediction is performed on the waiting time and the riding time components separately, and the results are combined to obtain the travel time of the full journey. We adopt a data driven approach (instead of using real-time traffic data), since traffic information pertaining to future bus trips may not be available at the time of prediction.

2) In order to accurately predict the riding time components, we studied the impact of roadway characteristics (e.g. distance, number of bus stops, traffic signals, etc.) and traffic conditions (characterized by trip start time, day of week, route spatial distribution, etc.) on the bus travel time. Our studies revealed a close relationship between bus riding time and roadway characteristics as well as traffic conditions. We propose a deep learning approach based on Long Short-Term Memory (LSTM) to predict the travel time of each segment of a given bus line. The LSTM model relies on features extracted from multiple data sources that not only characterizes the bus route but also the traffic conditions and travel demands.

3) We demonstrate that the waiting times at a bus stop are very sensitive to the arrival time of the passenger at that bus stop (sensitivity problem), and the waiting times at different bus stops are correlated to the bus travel times (correlation problem). With this, we developed a novel Interval-based Historical Average (IHA) method to estimate the waiting time at a bus stop, which does not assume a fixed distribution waiting time. The proposed method effectively handles the sensitivity and correlation problems.

The remainder of the paper is organized as follows. Related works are discussed in Section II. Section III introduces important definitions and the problem formulation. Section IV discusses the proposed PCF approach for travel time prediction of bus journey. We evaluate the performance of the proposed approach in Section V, and Section VI concludes the paper.

II. Related Works

Over the past decade, the problem of predicting travel time of vehicles (such as buses, taxis and private cars) has received wide attention. However, we are not aware of any reported works that address the same problem considered in this paper, i.e. estimating the total travel time of a bus journey, which involves both riding times on multiple bus trips and waiting times at transfer points. In this section, we review related works and highlight their differences with ours. These related works can be categorized into prediction of bus travel time, bus arrival time, and journey travel time.

A. Bus Travel Time Prediction

In general, efforts to predict bus travel times can be categorized into four approaches:

1) *Analytical model* based approaches explore the physical relationship between travel time and traffic variables (e.g. traffic flow, occupancy, signal phase plans, etc.) [35], [13].

For example, speed-based models split the bus route into segments and estimate the average speed on each segment separately. Traffic conditions such as traffic flow, travel speed, signal phase plans as well as road capacities are relied upon for the estimation. Song et al. [14] proposed a method to predict bus travel time based on real-time GPS (Global Positioning System) and RFID (Radio-Frequency IDentification) data. A self-adaptive exponential smoothing algorithm is first proposed to predict the bus speed based on the short-term speeds of taxis and buses. A bus travel time prediction model is then proposed that takes into account the delay caused by the signal control and acceleration/deceleration. The work in [15] developed a prediction method that considers both temporal and spatial variations in travel time. In their work, the conservation of vehicles equation in terms of flow and density was first re-written in terms of speed in the form of a partial differential equation using traffic stream models, and discretized using the Godunov scheme. A Kalman Filter based prediction approach is then proposed using the speed based equation.

In general, approaches that are based on analytical models require real-time traffic data of high density and frequency. As such, these approaches are usually not applicable to large scale transportation networks due to the substantial cost involved in obtaining such data.

2) *Historical average* approaches predict the travel time of a trip by relying on the historical average travel time for the same daily period over different days. This approach builds a non-parametric model that does not make any assumptions on the underlying data, and does not use any explicit training data.

The work in [16] builds a prediction model based on k-nearest method using data from Vehicle Detector System (VDS) and Automatic Tool Collection (ATC) system. However, this work is devoted to predicting the travel time on expressways. Lee et al. [17] proposed a historical trajectory based travel/arrival time prediction (HTTP) framework for real-time prediction of travel time of an on-going bus journey. HTTP first samples a collection of historical trajectories "similar" to the current on-going bus trajectory as the basis for prediction, explores different prediction schemes to prune the sample set of similar trajectories, and then return the average travel time of the pruned set as the prediction. However, the challenge of the historical average approach lies in collecting sufficient historical travel records for prediction. While it is possible to collect travel trajectories from mobile phones of passengers, it is difficult to accurately estimate the journey time using mobile phone trajectories as there is a need to infer the users' travel modes (walking, bus, taxi, metro) from sparse trajectories due to existence of noise and high spatial inaccuracy. Furthermore, obtaining dense trajectories from mobile phones is not feasible as this will incur high power consumption leading to poor user experience.

3) *Kalman Filters* (KF) approaches.

The first work using Kalman Filter to predict bus arrival time is reported in [18], where a combination of GPS data and historical data is used. The KF model is applied to track a vehicle location and a statistical estimation technique is used to predict travel time. A KF algorithm is proposed in [19] to predict travel times under heterogeneous traffic conditions on urban roadways in the city of Chennai, India using GPS data collected from buses. A hybrid method is developed to predict the travel time of Bus Rapid Transit (BRT) vehicle using the GPS data of BRT line 2 in Chaoyang district, Beijing. In this work, a Support Vector Machine (SVM) approach is used to predict an initial travel time and then the KF algorithm is applied to dynamically adjust the results of the initial prediction [20]. The results show that the prediction accuracy is higher for off-peak hours. Existing work has shown that the KF method can be applied to predict bus travel time using Automatic Vehicle Location (AVL) data [21]. However, a KF model typically cannot be generalized to the prediction of different time series as discussed in [22]. Moreover, the KF approach is sensitive to anomalies which are common in bus journeys due to uncertainties caused by bunching, delays at intersections, etc. As a result, the KF method becomes unreliable if there is a huge difference in travel time between two consecutive time steps.

4) *Time Series Analysis*. Time series forecasting uses models to predict future values using previously observed values by taking into account possible internal structure in the data.

The work in [23] applied Autoregressive Integrated Moving Average (ARIMA) model to forecast the short-term travel time along a corridor by incorporating traffic information from neighboring links. The authors reported that the travel times for consecutive segments are highly correlated, and upstream segments have a higher effect on travel time than downstream segments. The work in [24] apply the time series model ARIMA to predict bus travel time, using dataset collected from the bus service operated on a divided 4-lane 2-way highway in Ipoh-Lumut corridor, Perak, Malaysia. However, it has been shown that the time series based approaches cannot produce predictions with very high accuracy for complex scenarios such as urban bus travel time prediction. This is because bus travel time is affected by not only traffic and route conditions, such as travel distance, traffic congestion, number of bus stops, intersections and traffic signals etc., but also the dynamic travel demands such as the number of alighting and boarding passengers, dynamic bus loads, etc.

5) *Machine learning* models based on regression, SVM and neural network have been proposed for travel time prediction.

Classic regression models build relationships between travel times and related factors. Previous studies created regression models using travel distance, traffic flow, transit bus frequency, heavy vehicles proportion, bus-stopping time, number of stops, and time period, which resulted in promising predictions under specific conditions. Linear regression models have been widely applied in traffic prediction to capture the linear relationship between travel times and the related factors [25], [26]. This model is computationally efficient but usually produce undesirable results for nonlinear systems. As SVM have greater generalization capacity and can guarantee global minima for given training data, the works in [9], [27] have applied Support Vector Regression (SVR) for predicting travel-time on highways. However, this model suffers from high computation overhead. Many Neural Network (NN) approaches have beed developed to predict bus travel time using both historical and real-time data [26], [29], [30], [31], [32]. The features, such as the travel distance, number of stops, number of passengers boarding and alighting at each stop, average non-stop trip time, dwell time, bus schedule, have been considered as inputs for the existing NN prediction models. Approaches using the NN model has demonstrated advantage over the Kalman Filter model, historical average model, ARIMA and classic regression models.

Some of the above mentioned methods are used in a hybrid manner. For example, the KF method is combined with the theory of traffic flow in [15] to predict bus travel time. There also exist works that combine the neural network method with the KF method [33], [34].

B. Bus Arrival Time Prediction

Existing works on bus arrival time (BAT) prediction also use similar methods, such as historical average, linear regression, SVR [39], NN [36], [37], [38], Kalman Filter [33], and hybrid models [40]. In particular, the BAT problem in [37] considers the arrival time prediction of multiple buses at a transfer point (including the origin stop). The term 'multi-line' in their work refers to multiple choice of bus services at a transfer point. This differs from the notation used in our work where multiple consecutive bus lines constitute to a bus *journey*. In our problem, we investigate the total journey time of a passenger where only one bus service is used at each transfer point. Moreover, existing works typically focus on short term prediction which predicts the arrival time of a bus that is currently in operation.

C. Journey Travel Time Prediction

Existing research on journey travel time prediction mainly targets at estimating the time for vehicles traveling on a path of road segments. Most of this works employ a route-based strategy for travel time estimation, i.e. they first identify a route and partition the route into multiple road segments. The journey time for this route is then estimated by aggregating the travel time spent on each segment using historical trajectories. Existing approaches estimate the travel time using real freeway traffic data [9], [41], rural highway data [10], and urban roadway data [42], [43], [11]. The spatial or temporal correlations of link travel time have also been taken into consideration for estimating the probability distribution of trip travel times [11], [49]. Different from the traditional route-based approaches, the work in [12] proposed to use a large amount of taxi trips without relying on intermediate trajectory points to estimate the travel time between source and destination. This approach can achieve efficient prediction with average Mean Relative Error (MRE) of 21%. The work in [44] investigated the problem of online travel time prediction in the context of a bus journey, using both historical data and real-time data streams. It partitioned each bus line into road segments based on bus stops, and the travel time over each segment is estimated using data from multiple bus lines that travel through the same segment. However, the approach requires real-time information to predict the travel time of an ongoing journey. Moreover, the waiting time and transfer time at interchange stations along the journey have not been taken into consideration.

Our work differs from the existing methods for bus travel/arrival time prediction as they only target a single bus trip. These methods cannot be directly applied to our problem where the waiting times during transfer need to be considered. Also, unlike the BAT problem, our work aims to develop a more general method that not only predicts short term bus travel time, but also long term travel time of bus journey that may not have started yet. Note there exist some works on waiting time prediction [45], [46], but they do not consider the problem of bus journey time prediction for passengers. Moreover, they mainly focus on analyzing anomalies in bus arrival records and calculate the headways to derive waiting time. The existing approaches to estimate the journey travel time for taxis and private cars from link level travel time distributions also cannot be applied to the problem considered in this paper due to the following reasons: 1) Unlike the car, the travel time of buses is not only affected by the traffic conditions but also by other factors such as travel demand, the dwell time at each bus stop, bus schedule, etc. 2) Multiple bus lines of a passenger's journey usually have different travel time distributions due to variance on the traffic conditions and frequencies of buses. 3) Waiting times of the passengers at the transfer points must be considered for travel time prediction.

III. Preliminaries

A. Terminology

In this section, we define some terms that will be used throughout the remainder of this paper.

TABLE I
List of important notations.

L	bus line L
n	number of bus stops in L
$p_i^L = (x_i, y_i)$	a bus stop in line L
$R_{i,j}^L$	line segment from stop p_i^L to p_j^L
$dist(R_{i,j}^L)$	the length of segment $R_{i,j}^L$
$R_{i,i+1}^L$	unit segment locates at stop p_i^L in L
$\#.stops$	number of bus stops
$\#.intersections$	number of intersections
$\#.signals$	number of traffic signals
$HA(t_0)$	predicted waiting time with start time t_0
BAT_p	dataset of bus arrival records at stop p
$Pr(A)$	probability of event A
$\mathbf{E}[t_a]$	expectation of random variable t_a
$\mathbf{Var}[e_i]$	variance of random variable e_i
ε	mean absolute error, i.e. MAE
N	number of instances in test set
y	predicted journey time
$\hat{y}$	groundtruth of journey time
e_i	prediction error $e_i = \lvert y_i - \hat{y}_i \rvert$.

A bus line is a fixed route that is regularly traveled by the bus, and it includes a sequence of points $L =< p_1^L, p_2^L, \ldots, p_n^L >$ where $p_i^L = (x_i, y_i)$, for $i = 1 \ldots n$, is the GPS location of the i-th bus stop along the bus line L. We use bus stops as the route points to represent a bus route as it is of high interest to predict the arrival time at a bus stop. Bus passengers tend to be only interested in the arrival time at a bus stop rather than a random point along the route, since anyone can only board or alight at a bus stop. In this paper, we use the notation bus line and bus route interchangeably.

Along a bus line, consecutive points are connected by *unit segments*. In particular, the i-th unit segment of bus line L is denoted as $R_i^L = < p_i^L, p_{i+1}^L > (i = 1 \ldots (n-1))$, and the entire bus line L consists of unit segments $R_1^L, R_2^L, \ldots, R_{n-1}^L$, where n is the number of bus stops in bus line L. A *bus line segment* is a set of connected unit segments, e.g. $R_{i,j}^L =< p_i^L, p_{i+1}^L, \ldots, p_j^L > (i < j)$ indicating the segment from stop p_i^L to stop p_j^L of the bus line L.

We define a *trip* as a traversal on a single bus line without any transfers. In contrast, a *journey* indicates a complete travel from the passenger's origin to the destination, which may involve multiple trips using different bus lines/services. In practice, the end point of a previous bus trip and the start point of the new bus trip should be close to each other since passengers typically avoid walking long distances during transfer. The transfer points between two consecutive bus trips consist of bus stops or interchange stations. Without loss of generality, the bus stop/interchange station where the passenger waits for the first bus service is also called a transfer point.

A *bus trajectory* is a sequence of points in a bus trip, where each point represents the GPS location of a bus stop/interchange station and the arrival time of the bus at the bus stop/interchange station.

B. Problem Formulation

In general, travel time prediction provides an estimate of the journey's duration from an origin to the destination. Typically, the travel times vary greatly over different periods of the day/week/month, e.g., due to different levels of traffic load. Consequently, prediction is inherently time-dependent, and hence predictors are usually a function of the origin, the destination, and the time at which the journey is made. We assume that for a given journey, the bus line segments and transfer points are fixed.

Problem description: *Given a bus journey (may cover portions of multiple bus lines), with the origin, destination, transfer points and the journey start time, predict the total journey time (including the riding time in each bus line as well as any initial waiting and transfer times) based on historical bus trajectories, bus network and road network information.*

Existing works show that the waiting time is perceived to be heavier than in-vehicle riding time, and different individual passengers typically have different perceptions [47], [48]. In this paper, we predict the absolute travel time and waiting time, instead of relying on perceived waiting/travel time, as it is difficult to obtain a unified perception for all passengers.

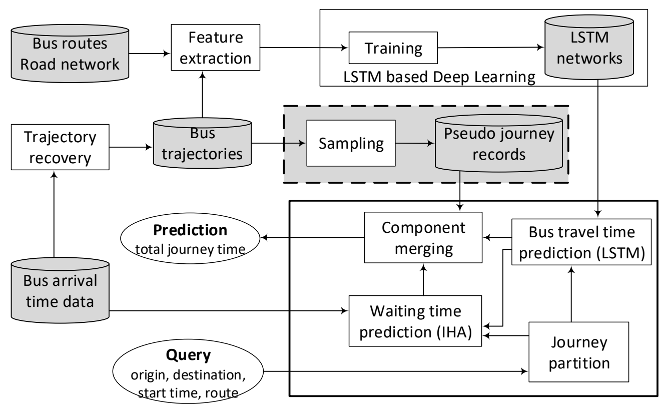


Fig. 1. Framework for journey travel time prediction.

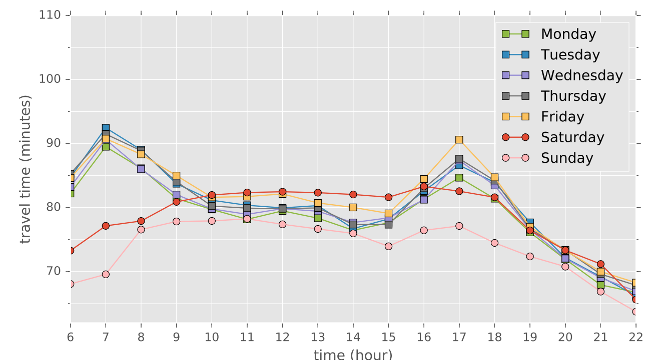


Fig. 2. Comparison of total travel time of bus line 855 in Singapore on different days.

IV. Journey Travel Time Prediction

We propose a novel Partitioning and Combination Framework (PCF) (shown in Fig. 1) to address the problem of bus journey travel time prediction as described in the previous section. As shown in the figure, bus trajectories are first recovered from bus arrival data. We employ LSTM based deep learning for bus travel time prediction, using features extracted from multiple datasets including bus trajectories, bus route information (e.g. travel distance, number of bus stops between origin and destination, etc.) and roadway characteristics (e.g. number of intersections and traffic signals along the travel route etc.). Using the PCF approach, an input journey is first partitioned into multiple components based on the transfer points along the journey route. Specifically, the components consist of the waiting times at the transfer points, and the riding times on the bus line segments. The riding time components are predicted using the LSTM networks and the waiting time components are predicted with the proposed IHA approach. The predicted waiting time and riding time components are then merged to obtain the total travel time of the entire journey.

The component merging step to obtain the final travel time can be implemented in two ways: 1) The first way is denoted as `PCF-sum`, which calculates the direct sum of all components as the final prediction, 2) The second way uses machine learning based methods such as linear regression (denoted as `PCF-LR`) to combine all components to form the final result, etc. The latter method requires extra journey records to train machine learning models. This can be achieved by generating sufficient (pseudo) journey records from historical bus trajectories. In our experiments, we have implemented both `PCF-sum` and `PCF-LR` to evaluate the performance.

It is noteworthy that our work is significantly different from the work in [44], that partitions a journey using all bus stops along the journey, which results in large number of bus line segments. Using more segments for travel time prediction will lead to the higher travel time uncertainty for the entire journey. Our method, which partitions the journey at transfer points only, has the following advantages: 1) The entire journey is partitioned into fewer components, thus leading to lesser accumulated errors when merging all the component prediction results. 2) Since there are lesser components in our method, we require much lower computation overhead to process each component and merging all the components. 3) The bus dwelling time at each bus stop that is not a transfer point has considerable impact on the bus travel time. In our method, this has already been accounted for in the travel time of each riding component. Thus, even though we do not specifically consider the bus dwelling time at those stops, we can still capture their impacts on the total travel time.

A. Riding Time Prediction

1) Feature Analysis and Extraction: In this section, we analyze the features that affect the bus riding time component, i.e. bus travel time on a given bus line segment.

Fig. 2 illustrates the total travel time of the bus line 855 in Singapore. The bus trajectories are first grouped based on day of week, such that each group contain trajectories covering 9 days. For example, the group of Monday data contains bus trajectories collected on 9 Mondays. For each group, we partition the time of a day into intervals with length of one hour, and for each time interval, we calculate the travel time as the average of all trajectories whose trip start time fall into the time interval. The results for each group are shown in Fig. 2. The x-axis indicates the journey start time, and the y-axis indicates the total travel time along the entire bus line. It can be observed that working days (Monday to Friday) have similar travel time patterns, and the same can be observed for weekends (Saturday and Sunday). In the same day, journeys with different start time have significant variance in the travel time. This shows that the journey start time has an impact on the travel time and should be used as a feature for travel time prediction. Moreover, weekdays and weekends have different travel time distribution, and Sunday exhibits the best travel condition than other days on the bus line 855. There is no morning peak on weekends, and the average travel time on Saturday during 10:00am to 15:00pm is evidently longer than that on workdays. This observation indicates that the day-of-week should also be considered as a feature to characterize the traffic condition for travel time prediction. Note that the bus line 855 has fixed services (same route and number of stops) throughout the week.

Fig. 3 shows the average travel time between all OD pairs on bus line 855, for the first 20 bus stops. The y-axis indicates the

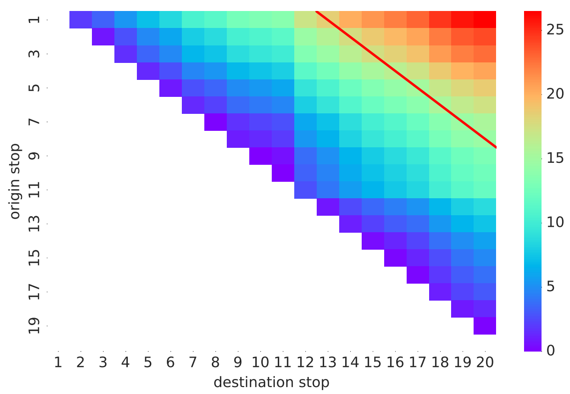


Fig. 3. The average travel time (minutes) between any origin-destination pairs on bus line 855, for the first 20 bus stops.

index of the origin stop, the x-axis indicates the index of the destination stop, and the color indicates the travel time needed for traveling between the origin and destination. It shows that the number of bus stops between the origin and the destination have direct impacts on the total travel time. The reason that more bus stops leads to longer travel time is not only due to longer travel distance, but also because of the increase in bus dwelling time (more number of bus stoppings) and bus deceleration/acceleration at the bus stops.

The red line in Fig. 3 covers cells corresponding to bus line segments of the same number of bus stops, i.e. 12 stops. It can be observed that even for trips of the same number of bus stops, their spatial distribution also have considerable impact on the travel time. This is due to the fact that traffic conditions at different bus line segments can differ significantly. In addition, the travel demands among different stops are not the same, leading to unbalanced bus dwelling time distribution along the trip. In our work, we characterize the spatial location of a trip by the origin stop location and the segment length in terms of the number of bus stops. On the other hand, more bus stops generally correspond to longer travel distance, more intersections and more traffic signals. Similar figures can be obtained by replacing number of bus stops with travel distance, number of intersections and number of traffic signals.

Existing work [21] also takes into account the effect of passenger demand on bus journey using Automated Passenger Count (APC) data. Since the APC data is not available to us, we have relied on time features (e.g. time of day and day of week) and location of the stops in the bus line (e.g. journey original stop, number of stops in the journey), which are highly correlated with the temporal patterns (e.g. periodical and seasonal patterns) of the passenger demand. Using these features, we are able to incorporate the travel demand patterns with our LSTM based deep neural network.

2) Proposed LSTM: Long Short-Term Memory (LSTM) [50] is a specific recurrent neural network (RNN) architecture that is well-suited for learning from experience to classify, process and predict time series with time lags of unknown size. It has been successfully used in many real-world problems for processing sequential data, including link-level travel time estimation. However, the existing approaches only consider travel time in different time slots and make predictions based on the historical travel time data. They also need to train a separate model/network for each of the road links [52]. In contrast, we train one LSTM network for each bus line such that the travel time of any segments of the bus line can be accurately predicted by the same model. The training process takes into consideration the travel time and the impact factors/features discussed in the previous section.

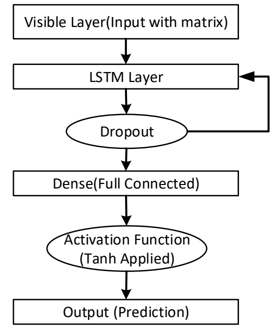


Fig. 4. The structure of the deep learning network.

① *Network Structure:* Our model is illustrated by the flow diagram in Fig. 4. The input matrix is fed into two LSTM layers, where the first LSTM layer has 128 neurons and the second has 64 neurons. In order to prevent overfitting, a dropout mechanism is applied to each LSTM layer as shown in Fig 4. The rate of dropout is set to 0.5. The output of the second layer goes into a dense layer, which is a fully connected NN. The dense layer makes prediction based on the feature information at the output of LSTM layer. Finally, we apply a tanh activation function and obtain the prediction results. The Mean Absolute Error (MAE) is applied as the loss function for training the model and the Adam optimizer is utilized as the gradient descent optimization algorithm. The training process repeats for 50 epochs.

② *LSTM Structure:* The basic structure of a single LSTM memory cell is shown in Fig. 5, which can be described with the following equations:

$$\begin{aligned}
\mathbf{i}_t &= \sigma(\mathbf{W}_{ix}\mathbf{x}_t + \mathbf{W}_{ih}\mathbf{h}_{t-1} + \mathbf{b}_i)\\
\mathbf{f}_t &= \sigma(\mathbf{W}_{fx}\mathbf{x}_t + \mathbf{W}_{fh}\mathbf{h}_{t-1} + \mathbf{b}_f)\\
\mathbf{o}_t &= \sigma(\mathbf{W}_{ox}\mathbf{x}_t + \mathbf{W}_{oh}\mathbf{h}_{t-1} + \mathbf{b}_o)\\
\widetilde{\mathbf{C}_t} &= \tanh(\mathbf{W}_{Cx}\mathbf{x}_t + \mathbf{W}_{Ch}\mathbf{h}_{t-1} + \mathbf{b}_C)\\
\mathbf{C}_t &= \mathbf{i}_t * \widetilde{\mathbf{C}_t} + \mathbf{f}_t * \mathbf{C}_{t-1}\\
\mathbf{h}_t &= \mathbf{o}_t * \tanh(\mathbf{C}_t)
\end{aligned}$$

where t stands for the t-th timestamp, $\mathbf{i}_t$, $\mathbf{f}_t$, $\mathbf{o}_t$ refer to the output of the input gate, forget gate and output gate respectively. $\mathbf{x}_t$, $\mathbf{c}_t$, $\mathbf{h}_t$ are the input vector, state vector and hidden vector respectively, and $\mathbf{h}_{t-1}$ is the former output of $\mathbf{h}_t$. $\widetilde{\mathbf{C}_t}$ and $\mathbf{C}_t$ are the input state and output state of the memory cell, and $\mathbf{C}_{t-1}$ is the former state of $\mathbf{C}_t$. σ is a sigmoid function. $\mathbf{W}_{ix}, \mathbf{W}_{fx}, \mathbf{W}_{ox}, \mathbf{W}_{Cx}$ are the weight matrices connecting $\mathbf{x}_t$ to the three gates and the cell input, $\mathbf{W}_{ih}, \mathbf{W}_{fh}, \mathbf{W}_{oh}, \mathbf{W}_{Ch}$ are the weight matrices connecting $\mathbf{x}_{t-1}$ to the three gates and the cell input, $\mathbf{b}_i$, $\mathbf{b}_f$, $\mathbf{b}_o$, $\mathbf{b}_C$ are

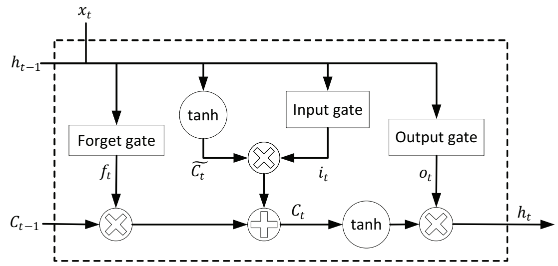


Fig. 5. Structure of the LSTM memory cell.

the bias terms of the three gates and the cell gates.

③ *Input Matrix:* In our model, the input record x_t for time step t is a vector representing the information of a trip, that consists of two parts: travel time and the impact factors of the travel time, i.e. the features as shown in Table II. For a given bus line segment, we collected data record x_t at a 30-min time interval for a period of 63 days, to obtain a data sequence of 36×63 records, where 36 is the number of time intervals in each day. Note that there are only 36 records instead of 48 records for each day because there are no buses in operation during 12:00pm to 06:00am. A bus line with n bus stops has $\frac{n\cdot(n-1)}{2}$ different bus line segments in total, thus the number of data records of the input matrix is $36 \times 60 \cdot \frac{n\cdot(n-1)}{2}$ for a bus line with n bus stops. In other words, the input matrix contains multiple data sequences, such that each of the $\frac{n\cdot(n-1)}{2}$ bus line segments is associated with a sequence. The input vectors are scaled to the range of (0, 1) based on the min-max normalization and is used to train a LSTM model such that it can be used to make accurate travel time prediction for any segment. The sequence of records contains the temporal trend of the travel time of each bus line segment, while the $\frac{n\cdot(n-1)}{2}$ sequences include information of segments that are spatially adjacent to the segment to be predicted. Therefore, we can make travel time predictions for any bus line segment based on temporal data dependencies and spatial data correlations.

TABLE II
List of extracted features for each trip record.

Extracted features	from dataset
day of week	bus trajectories
time of day	bus trajectories
travel distance	bus route data
number of bus stops	bus route data
location of the origin stop in the bus line	bus route data
number of intersections	road networks
number of traffic signals	road networks

B. Waiting Time Prediction

The challenges in waiting time prediction are as follows: 1) Distribution problem: To date, it is not clear which models can best characterize the waiting time distribution in the context of bus journey; 2) Sensitivity problem: The waiting time at a transfer point is very sensitive to the arrival time of the passenger, making it more challenging for prediction because the exact travel/arrival time of the bus prior to the transfer point is not known in advance; 3) Correlation problem: The waiting times at different transfer points as well as the bus travel time between the consecutive transfer points are correlated, such that the errors of different components (e.g. riding time, waiting time) will propagate and lead to extremely large error for components at the end of the journey.

Case study. Route 3 of Table IV utilizes bus lines 154 and 179, starts at bus stop 81119 using bus line 154, and switches to bus line 179 at the bus stop 22009 (22009 is common to both bus lines 154 and 179). It finally reaches its destination stop 27251. Thus a passenger needs to wait for bus 154 at stop 81119, and then waits for another bus 179 at stop 22009. The waiting times at the two bus stops with varying journey start times are illustrated in Fig. 6, where the x-axis indicates the journey start time. From the figure it can be observed that the two curves have steep slopes, which implies that the waiting times are very sensitive to the journey start time. For example, as shown in the figure, bus 154 arrived at the stop 81119 at 07:40am. So if the passenger arrives at the stop 81119 at 07:40am, then the waiting time is 0 minute, however, the waiting time becomes 16 minutes if the passenger arrives at 07:41am. Estimating the waiting time at the second bus stop 22009 is more challenging than that at stop 81119. This is not only due to the fact that waiting times are very sensitive to the journey start time, but also because the exact bus travel time between stops 81119 and 22009 (or the bus arrival time at stop 22009) cannot be predicted without errors. This shows that the travel times are correlated with the waiting times.

We take the example of predicting the waiting time for bus 179 at the stop 22009 to describe our method. For simplicity, we call the bus 154 between stop 81119 to stop 22009 as the 'prior bus', and it is expected to arrive at bus stop 22009 at time t_0, meaning that the passenger is expected to wait for the bus 179 from time t_0. A simple approach for waiting time prediction is Historical Average (HA). Assuming the dataset BAT_{22009} of historical bus arrival times at bus stop 22009 containing data of d days, then the historical average waiting time can be calculated as

$$HA(t_0) = \frac{\sum_{i=1}^{d}(t_i - t_0)}{d},$$

where t_i is the first time bus 179 arrives after time t_0 in the i-th day of the dataset. Thus $t_i - t_0$ is the historical waiting time on the i-th day.

In our work, the HA approach is further optimized in the following ways: 1) The historical dataset of bus arrival time is partitioned into two groups based on weekday and weekend, as bus frequencies on weekends are much lower than weekdays; 2) The existence of noises and missing values in the dataset of bus arrival time results in many incorrect records of historical waiting times. These anomalies (extremely large values of waiting time) will affect the prediction accuracy. To mitigate the impact of these anomalies on our prediction model, we rank the waiting time records and remove those that ranked above the 90th percentile, as these records are likely to be anomalies caused by missing values or abnormal traffic conditions. The remaining records of historical waiting times will be used in the HA method. 3) The average time interval between two consecutive bus arrivals (during a period of 1

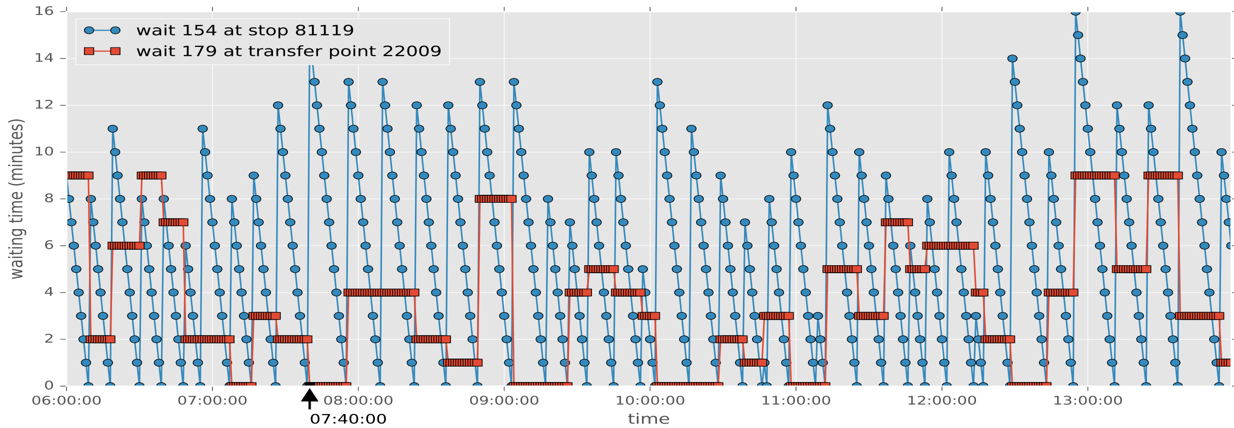


Fig. 6. Observed waiting times at two transfer points along journey route 3 of Table IV, with varying journey start time.

hour) are calculated, then all records that have waiting time larger than twice of the average time interval are removed. Even passengers can wait more than twice the average waiting time in some extreme cases, this constitutes to a small percentage of occurrences in Singapore [53]. As such, we have removed records from the dataset pertaining to waiting times larger than twice the average waiting time as they are likely to be anomalies caused by data collection.

Since the HA method is a data-driven approach, it avoids the assumption of a fixed distribution of waiting times at a bus stop. However, it cannot address the problems of correlation and sensitivity, and hence it tends to be biased significantly when the bus arrival rates are time-varying. Based on the above mentioned optimizations, we propose a novel approach, denoted as Interval-based Historical Average (IHA), to handle the problems of correlation and sensitivity by integrating a set of results obtained using the optimized HA approach. The major challenge is that the waiting time is sensitive to the arrival time t_0 of the prior bus 154 at the stop 22009, which cannot be previously predicted. To mitigate the influence of the sensitivity problem, we utilize a time interval $[\mathbf{E}[t_a]-\varepsilon, \mathbf{E}[t_a]+\varepsilon]$ to characterize the arrival time of the prior bus instead of using a single time point t_0, where $\mathbf{E}[t_a]$ is the expectation of arrival time based on historical bus trajectories, and ε is set to be the mean absolute error of the LSTM network presented in the previous section. We next show that the exact arrival time of the prior bus will fall into this interval with probability above $1 - \frac{\mathbf{Var}[e_i]}{\varepsilon^2}$.

Specifically, for each instance in the test set of bus travel times, we can calculate a prediction error $e_i = |y_i - \widehat{y_i}|$, as well as its expectation $\mathbf{E}[e_i]$ and variance $\mathbf{Var}[e_i]$. Based on the Chebyshev's Inequality, it can be calculated that the probability where prediction errors fall out of the interval $[-\varepsilon, \varepsilon]$ is bounded by

$$Pr(|e - \mathbf{E}[e_i]| \geq \varepsilon) \leq \frac{\mathbf{Var}[e_i]}{\varepsilon^2}.$$

Therefore, the prediction errors will fall into $[-\varepsilon, \varepsilon]$ with probability at least $1 - \frac{\mathbf{Var}[e_i]}{\varepsilon^2}$, which indicates that the exact bus arrival time will fall in interval $[\mathbf{E}[t_a] - \varepsilon, \mathbf{E}[t_a] + \varepsilon]$ with probability of at least $1-\frac{\mathbf{Var}[e_i]}{\varepsilon^2}$, where $\mathbf{Var}[e_i]$ is the variance of prediction error and ε is the mean absolute error of our prediction model (LSTM network).

Based on the above analysis, we predict the waiting time as

$$IHA(\mathbf{E}[t_a], \varepsilon) = \frac{\sum_{i=-\varepsilon}^{\varepsilon}(HA(\mathbf{E}[t_a] + i))}{2\varepsilon + 1},$$

where $\mathbf{E}[t_a]$ is the expectation of arrival time based on historical bus trajectories, and ε is the mean absolute error of the LSTM network for bus travel time prediction. $HA(\mathbf{E}[t_a] + i)$ indicates the estimated waiting time using the optimized HA approach if the passenger arrives at the bus stop at time $\mathbf{E}[t_a] + i$. For estimating the waiting time at the first transfer point (i.e. 81119), t_a is set to journey start time t_0 and ε is set to 0 as t_0 is the exact arrival time of the passenger.

V. Results and Analysis

A. Datasets and Preprocessing

Road Networks: Our experiments are based on Singapore's road network, which comprised of 41,732 nodes and 98,539 road segments. The road network is utilized to derive the information of intersections ($\#.intsections$) as well the number of traffic signals ($\#.signals$) for any journey routes.

Bus Lines: The bus route information includes the ID (a five digit number) of each bus stop in sequential order, the GPS location (latitude and longitude) of each bus stop, the travel distance $dist(R_{i,j}^L)$ between any 2 bus stops p_i^L and p_j^L (i.e. the i-th and j-th stops of bus line L), and the number of bus stops between p_i^L and p_j^L. We also map the bus routes to the road network using the GPS locations of bus stops to find out the road segments of the bus line. The results are verified by comparing with Google Map via visualization. Based on the map-matched bus line routes, the number of intersections ($\#.intersections$) as well the number of traffic signals ($\#.signal$) for any journey routes can be easily calculated. Table III shows the statistical information of the 5 bus lines considered in our experiments. The first three bus lines connect the western part of Singapore to the eastern part, while the last two bus lines connect the northern part of Singapore to the southern part of the city as shown in Fig. 7.

Bus Trajectories: A bus trajectory dataset is derived based on the real-world Bus Arrival Time dataset (the arrival time of

TABLE III
Bus line routes used in the experiments.

busline	origin stop ID	dest. stop ID	total distance	$\#.stops$	avg. travel time	frequency	$\#.intsections$	$\#.signals$
179	22009	22009	9.8 km	24	32 min	4-8 min	65	19
154_2	82009	22009	33.8 km	74	83 min	7-15 min	214	32
24	54009	54009	49.5 km	100	139 min	7-15 min	298	44
980_1	58009	80009	25.4 km	64	61 min	10-16 min	178	22
855_1	59009	14009	26.7 km	59	70 min	9-15 min	195	21

TABLE IV
Journey routes used in the experiments.

	origin ID	destination ID	used buslines	distance	#.stops	transfers	travel time (minutes)
route 1	95019	27251	24-154-179	52.1 km	109	82049,22009	min: 127, avg: 172, max: 234
route 2	95019	22009	24-154	45.8km	94	82049	min: 108, avg: 147, max: 207
route 3	81119	27251	154-179	28.4km	52	22009	min: 90, avg: 125, max: 181
route 4	28251	14141	980-855	27.6km	64	53029	min: 63, avg: 96, max: 136
route 5	59039	14141	855	23.8km	52	NA	min: 46, avg: 74, max: 111
route 6	57039	01341	980	16.7km	43	NA	min: 34, avg: 64, max: 95
route 7	82049	41011	154	12.4km	24	NA	min: 27, avg: 45, max: 78
route 8	22521	27251	179	4.9km	12	NA	min: 8, avg: 19, max: 40

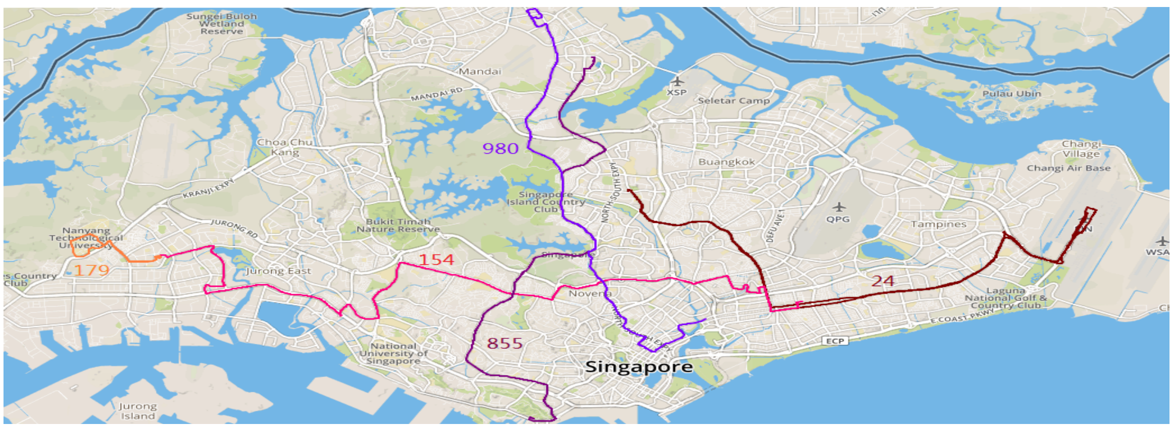


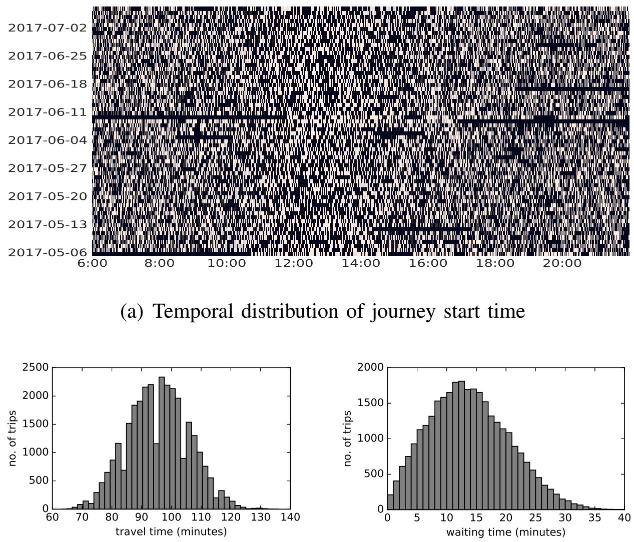
Fig. 7. The spatial distribution of 5 bus lines.

the next bus for each bus stop, at every minute) provided by the Land Transport Authority, Singapore. The dataset contains bus trajectory data of 5 bus lines from May 06 to July 07, 2017 as illustrated in Table III. Each bus trajectory consists of a sequence of points, where each point contains the information of the stop ID, the GPS location of the stop, the timestamp (arrival time of the bus at the stop), and the bus line ID. Based on the historical bus trajectories, the following features are extracted for each trip traveling along the bus line segments $R^L_{i,j}$ (with origin p^L_i and destination p^L_j of bus line L): the day-of-week, the trip start time, as well as the trip duration (i.e. the total travel time for the bus line segment).

Pseudo Journey Travel Records: Since it is not possible to obtain sufficient travel records of individual passengers, a dataset of journey travel records is generated based on the available bus trajectories, via random sampling. The involved journey routes are summarized in Table IV. The investigated journey routes differ in terms of geometric locations, travel distance (stops), total travel time, transfers, bus frequency, etc.

To explain how the pseudo journey records are generated, we use journey route 4 in Table IV as an example. A passenger will first wait at the bus stop 28251 and board the first bus 980 and alight at the bus stop 53029. He will then board the first bus 855 that arrives and alight at bus stop 14141. If the date and start time of the trip are given, we can calculate the exact waiting times at bus stops 28251 and 53029, as well as the exact bus riding times of bus 980 between 28251 and 53029, and bus 179 between 53029 and 14141, because the historical bus travel trajectories are known. Therefore, the travel information of the journey can be obtained. By repeating this procedure, for each of the investigated trip routes, we generate 500 journey records for each observation day, by randomly selecting the journey start times.

Fig. 8(a) illustrates the temporal distribution of the journey records for trip route 4 of Table IV, where the y-axis indicates the dates from May 06 to Jul. 07, 2017, and the x-axis indicates the journey start time of the day. It can be seen that the generated journey records (denoted as white cells) are evenly distributed in the temporal space. We employ uniform sampling for generating the journey records to ensure that the training process of our prediction model does not discriminate between peak and off-peak hours. For example, if more records



(a) Temporal distribution of journey start time

(b) journey travel time distribution (c) journey waiting time distribution

Fig. 8. Statistics of the generated journey travel records for route 4.

are generated during peak hours, then the training process will pay more attention to those journeys, leading to higher prediction accuracy for peak-hour journeys than off-peak hour journeys. There exist some horizontal black row-segments indicating failure in generating journey records during those periods due to missing data on the corresponding days (e.g. May 06, 10, Jun. 06, 10, 17 etc.). The journey travel times range from 63 to 136 minutes following the distribution shown in Fig. 8(b), and the waiting time span follows the distribution depicted in Fig. 8(c). In the experiments, we do not estimate the travel times of journeys with start time earlier than 6:00am and after 10:00pm, since there are almost no buses during those times. Since the day-of-week has been considered as a feature of the journeys, we use the same prediction model for a journey regardless of whether it is taken during weekday or weekend. For the journey routes listed in Table IV, a total number of 31,500 journey records are generated. The dataset is partitioned into two sets for the baseline methods: a training dataset consisting of the 28,000 journeys generated from May 06, 2017 to Jun. 30, 2017, and a test dataset consisting of the 3,500 journeys from the last week of the entire time period. The features as illustrated in Table II are extracted for each of the generated journey records.

During the generation of journey records, we assume that a passenger can always board the first bus that arrives. It is noteworthy that we can also extend this to the case where some passengers have to wait for the second bus due to overcrowding. To achieve this, we can first observe the rate, say $\rho\%$, of passengers that failed in boarding the first bus at each transfer point. During the journey generation procedure, we can allow $\rho\%$ passengers who need to transfer to use the second bus trajectory.

As shown Table IV, 8 journey routes are utilized for performance evaluation, where routes 1-4 are long range travel routes that traverses residential, commercial, and institutional areas, as shown in Fig. 7. The routes 5-8 are short and middle range routes: Route 5 is a middle range route that has light traffic congestions along its entire trip; Route 6 is a middle range route where the first half of the route goes through light congested areas while the second half goes through heavy congested areas; Route 7 is a short range route that passes through heavy congested areas; Route 8 is a short range route with good traffic conditions and the involved bus line (bus 179) is of high frequency.

B. Baseline Methods for Bus Travel Time Prediction

Since there is no existing approach for the same problem considered in this paper, we compare our method with the following 6 well known techniques for journey travel time prediction, which are briefly described below.

1) Historical average (HA): HA is a naive prediction method and is commonly used as a baseline for travel time prediction. Given the origin, destination, journey start time (interval), and historical journey records, the predicted travel time is the average of all historical travel records of the same period that have the same origin and destination.

2) k-Nearest Neighbor (kNN): Similar to the HA method, kNN approach matches current input variables with historical observations that have similar input variables. The difference is that kNN only use the k most similar records instead of all the previous travel records that have the same origin and destination. The similarity of historical records is evaluated as the closeness on the journey start time. The parameter k is set to 10, which is selected to maximize the performance of the KNN method on our dataset, by trying different values, i.e. 5-30 stepped by 5.

3) TensorFlow Time Series (TFTS): We use the open source tool TFTS as one of our baseline[1]. For testing each OD pair, the travel times with journey start time at different time intervals are modeled as a time series, where each value of the series corresponds to the average travel time of all historical records whose journey start time fall into the same 30-minute interval.

4) Linear Regression (LR): LR is utilized to model the relationship between journey travel time and all the impact factors/features discussed in Section IV-A1.

5) Fully Connected Neural Network (FCN): A three layer fully connected neural network is implemented where the first layer contains 128 neurons, the second layer contains 64 neurons, and the third layer is a dense layer. Tanh is applied as the active function, the MAE is applied as the loss function for training the model and Adam optimizer is utilized as the gradient descent optimization algorithm. The training process repeats for 50 epochs.

6) Support Vector Regression (SVR): As a variant of the classic classification technique Support Vector Machines (SVM), SVR is targeted towards regression problems based on a straightforward idea: construct a hyperplane that sets apart the classes of data. Due to its high accuracy when trained with a sufficiently large dataset, SVR has been used for travel time prediction in some existing works such as [9], [27].

[1] https://github.com/tgjeon/TensorFlow-Tutorials-for-Time-Series

C. Baseline Methods for Waiting Time Prediction

Since the above mentioned baseline approaches for predicting journey travel time do not estimate the waiting times separately, the following baseline methods are utilized for evaluating the proposed IHA approach to estimate the waiting times at transfer points. 1) Bus Frequency based Approach (FA): This approach partitions a day into several time intervals of 30 minutes each, and calculates the mean and variance of the time headways of each interval, where the headway is the time interval between two consecutive bus arrivals of the same services. The average waiting time, corresponding to an interval with mean μ and variance σ^2, is calculated as $\frac{\mu}{2} \cdot (1 + \frac{\sigma^2}{\mu^2})$ [28]. 2) Historical Average approach (HA): HA method has been introduced in section IV-B, 3) Machine Learning approaches including Linear Regression (LR), Support Vector Regression (SVR) as well as Fully Connected Neural Networks (FCN).

D. Overall Performance

The performance measures used are the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE). Besides MAE and MAPE, we also compute the average error of travel time per km (MAE/distance).

The overall performance is evaluated in this section, while the evaluations on bus riding times and waiting times are presented in the following two sections. During the training process, 30% of the training set is used for validation. We evaluate the performance for a very long prediction period of seven days, i.e. from July 01 to July 07, 2017. In other words, the ratios of training set, validation set and test set are 62.3%, 26.7% and 11%, respectively [51], [52]. All algorithms are implemented using TensorFlow 1.3 running on E5 2.0 GHZ CPU with 64 GB memory.

Table V presents the overall performance. In particular, for the PCF approach, we evaluate the two different methods for merging the waiting time components and riding time components. PCF-sum indicates the results obtained as a direct summation of the waiting time components and riding time components, while the PCF-LR indicates the results obtained by using linear regression. It is evident that the PCF approach outperforms all the baseline methods for all the three metrics. In addition, PCF-LR slightly outperforms the PCF-sum approach with an average improvement of 0.18 minutes in terms of MAE. This is due to the fact that the journey record information are utilized in PCF-LR during the solution merging procedure. Note that, even without using the journey records, the PCF-sum method achieves better results than PCF-LR in some cases (e.g. route 7 and route 8) in terms of MAPE.

The results also reveal that longer journeys tend to result in smaller MAPE (the journey distance decreases from route 1 to route 8). This is because the uncertainty in overall travel time is predominantly caused by traffic congestions at the road segments, delays at intersections, the bus dwelling times at bus stops, and the waiting/transferring time at transfer points. Since journeys with short travel distance have fewer road segments, intersections, bus stops, and transfer points, large variances in the congestion condition, bus dwell times, intersection delays, waiting times and bus stopping times are expected. However, with the increase in journey distance, the average values of the above mentioned components tend to be more stable, hence better prediction accuracy is obtained. The value of MAPE on route 8 is generally high (13% for the PCF approach). The reason for this is that the length of route 8 is very short, i.e. 4.9km. Since the journey time is very short, even small MAEs, i.e. 2-3 minutes, will lead to high MAPEs.

E. Results on Prediction of Bus Riding Times

Table VII illustrates the results of the proposed LSTM network for bus travel time prediction of a single bus line in comparison with the baseline approaches, where the waiting time at the first bus stop is not considered. The journey routes 5-8 utilized for the evaluation in this experiment, as shown in Table VI, are the same as those in Table IV with the exception that the waiting times at the first stop are not considered.

Since existing works on bus travel time prediction only predict the bus travel time of a fixed travel route, we train a separate prediction model for each of the 4 journey routes when implementing the baseline approaches. It is worth noting that the proposed PCF differs from the baseline approaches in that it trains a prediction model for each bus line. This means that the number of required models of our approach is not influenced by the number of journey routes, but only depends on the number of involved bus lines. In other words, if any two or more journey routes use the same bus line, their travel time on that bus line can be predicted using the same model, regardless of their origins and destinations.

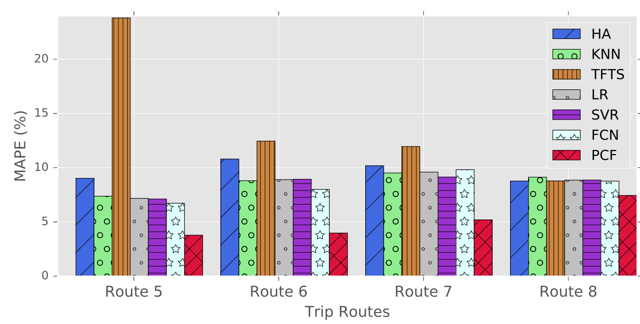


Fig. 9. Comparison of PCF and the baseline approaches in terms of MAPE on routes 5-8.

The comparisons on MAE, MAPE and MAE/distance are shown in Table VII. It is evident that the PCF achieves much better results than all the baselines, with average improvements of 55.2%, 48.4% and 47.3% in terms of MAE, MAPE and MAE/distance, respectively. This is because, our proposed LSTM networks make travel time predictions for any bus line segment based on temporal data dependencies and spatial data correlations, i.e., it considers not only the temporal trend of the travel time of each bus line segment, but also the location and spatially adjacent relationship of different segment for training the prediction model.

Fig. 9 shows the comparison of the MAPEs on each of the journey routes considered, i.e. route 5-8, respectively. For all the approaches, the MAPEs of different routes differ

TABLE V
Comparison of overall performance.

Evaluation metrics	Methods	route 1	route 2	route 3	route 4	route 5	route 6	route 7	route 8	average
MAE (min)	HA	10.32	9.96	9.11	8.14	6.63	6.94	5.03	3.12	7.41
	KNN	9.70	9.41	8.67	8.05	5.89	6.44	5.15	3.15	7.06
	TFTS	11.65	10.13	9.27	10.60	11.67	7.07	5.01	3.20	8.58
	LR	10.12	9.55	8.79	7.34	5.84	6.36	4.92	3.09	7.00
	SVR	10.13	9.56	8.69	7.33	5.82	6.36	4.90	3.09	6.99
	FCN	9.55	9.01	9.03	7.09	5.63	5.96	4.99	3.09	6.79
	PCF-sum	7.59	6.32	5.40	5.50	3.66	3.54	3.12	2.71	4.73
	PCF-LR	**7.07**	**6.15**	**5.22**	**5.27**	**3.59**	**3.41**	**3.06**	**2.64**	4.55
MAPE (%)	HA	5.10	6.69	7.07	8.58	9.28	11.04	11.60	15.68	9.38
	KNN	5.56	6.31	6.81	8.40	8.03	9.92	11.55	16.02	9.08
	TFTS	6.58	6.87	7.12	11.44	15.95	11.79	11.32	16.58	10.96
	LR	5.79	6.39	6.81	7.69	8.10	10.00	11.35	15.53	8.96
	SVR	5.79	6.39	6.84	7.68	8.07	9.98	11.27	15.43	8.93
	FCN	5.45	6.01	7.06	7.40	7.79	9.43	11.43	15.25	8.73
	PCF-sum	4.32	4.24	4.20	5.65	4.99	5.44	**6.83**	**13.08**	6.09
	PCF-LR	**4.08**	**4.17**	**4.10**	**5.46**	**4.90**	**5.22**	6.85	13.14	5.99
MAE/distance (minutes/km)	HA	0.198	0.217	0.321	0.295	0.279	0.416	0.406	0.637	0.346
	KNN	0.186	0.205	0.305	0.292	0.247	0.386	0.415	0.643	0.335
	TFTS	0.224	0.221	0.326	0.384	0.490	0.423	0.404	0.653	0.391
	LR	0.194	0.209	0.310	0.266	0.245	0.381	0.397	0.631	0.329
	SVR	0.194	0.209	0.306	0.266	0.245	0.381	0.395	0.631	0.328
	FCN	0.183	0.197	0.318	0.257	0.237	0.357	0.402	0.631	0.323
	PCF-sum	0.146	0.138	0.190	0.199	0.154	0.212	0.252	0.553	0.230
	PCF-LR	**0.136**	**0.134**	**0.184**	**0.191**	**0.151**	**0.204**	**0.247**	**0.539**	0.223

TABLE VI
Dataset description for bus travel time prediction.

Routes	route 5	route 6	route 7	route 8
affiliated busline	855	980	154	179
distance	23.8km	16.7km	12.4km	4.9km
Training Trajectories	4361	3906	4805	8568
Testing Trajectories	547	492	622	1056

TABLE VII
Evaluation on bus travel time prediction.

Methods	MAE (min)	MAPE (%)	MAE/distance (min/km)
HA	4.220	9.694	0.293
KNN	3.702	8.702	0.267
TFTS	7.230	14.248	0.430
LR	3.648	8.630	0.263
SVR	3.601	8.514	0.260
FCN	4.093	9.479	0.286
PCF	**1.978**	**5.098**	**0.158**

significantly, as the journey routes are associated with different degrees of traffic uncertainties (caused by traffic congestions, traffic signals and travel demands at the bus stops, etc). In particular, since route 8 is a short range route with good traffic conditions, all the approaches can make good predictions with MAEs no more than 2 minutes. However, because the total travel time of route 8 is short, ranging from 8 to 40 minutes, the MAPEs of all approaches are still high. Comparing the 4 travel routes, route 7 is the most challenging where most approaches exhibited the worst results due to the fact that route 7 passes through heavy congested areas.

F. Results on Prediction of Waiting Time

In addition to the bus travel time, we also evaluated the performance of the IHA method for waiting time prediction. Using journey route 3 as an example, Fig. 10 illustrates the performance comparison on the waiting time predictions obtained by IHA as well as the baseline approaches (including frequency based approach Freq, historical average HA, linear regression LR, support vector regression SVR and fully connected neural networks FCN) in terms of MAE. The figure shows the results of the total waiting time along route 3 as well as the waiting time at stop 81119 (first stop) and 22009 (transfer stop).

In general, the proposed IHA achieves better performance than all the baseline methods in terms of MAE. In addition, it can be observed that the machine learning based algorithms (e.g. LR, SVR and FCN) do not consistently achieve better results than the historical average HA. For example, the three machine learning based approaches produce smaller MAEs than HA in the overall waiting times of route 3, but obtained larger MAEs for the transfer point 81119. Moreover, the SVR method obtains larger MAEs for both of the individual interchange stations than HA. This is due to the fact that the bus arrival time at a bus stop has some degree of uncertainty due to the traffic conditions and travel demands. Moreover, there is a lack of effective features that the machine learning algorithms can use to train a prediction model.

VI. Conclusions

This paper investigates the problem of predicting bus journey travel time for individual passengers, that takes into account the bus riding times along the travel routes and the waiting times at transfer points. We proposed a PCF approach

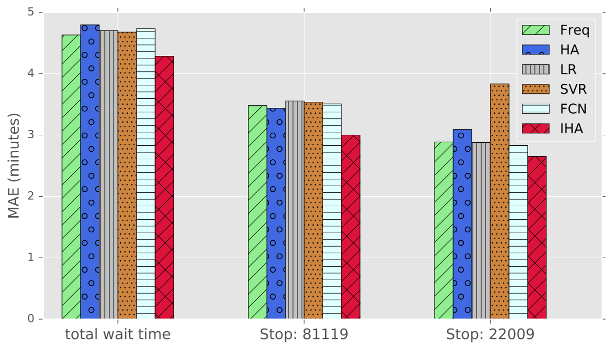


Fig. 10. Comparison of waiting times for journeys on route 3: the overall waiting time, waiting time for bus 154 at stop 81119, and waiting time for bus 179 at stop 22009.

to solve this problem by partitioning the entire journey into bus riding components and waiting components. The riding and waiting time components are predicted separately and the results are merged to obtain the final travel time. We have shown that features obtained from datasets of historical bus trajectories, bus routes, and the road network can well characterize traffic conditions, delays at intersections and dwelling time at the bus stops. These characteristics have significant influences on the bus travel time. Using a combination of these features, we developed a LSTM based approach that can accurately predict the bus travel time over any segment of the bus line. We also show that the waiting times at transfer points play a critical role in predicting the total travel time of the entire journey by demonstrating the challenges caused by sensitivity and correlation problems. To address those challenges, we proposed a novel IHA approach which can effectively address the correlation and sensitivity problems, without assuming a fixed distribution of waiting times. The experimental results demonstrated that the proposed approach significantly outperforms the baseline approaches.

In future, we plan to consider more features, such as weather condition, to improve the journey time prediction accuracy of our machine learning model. In addition, we plan to develop techniques for predicting waiting time that takes into consideration the extreme cases where buses are delayed due to operational service problems. Moreover, the prediction model will need to be continuously retrained using updated bus trajectory data to deal with evolving traffic conditions, transport infrastructures, government policies, etc.

Acknowledgment

We would like to thank the Land Transport Authority, Singapore for providing us with bus arrival time data. This research project is funded by the National Research Foundation Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme with the Technical University of Munich at TUMCREATE.

References

[1] Beaudoin, J., Farzin, Y.H., Lawell, C.Y.C.L., "Public transit investment and sustainable transportation: A review of studies of transit's impact on traffic congestion and air quality," *Res. Transp. Econ.*, vol. 52, pp. 15–22, 2015.

[2] Everlyn Lee, Public Transportation in Singapore, http://worksingapore.com/articles/live_4.php.

[3] Carrel, A., Halvorsen, A., Walker, J.L., "Passengers' perception of and behavioral adaptation to unreliability in public transportation," *Transp. Res. Rec.: J. Transp. Res. Board*, vol. 2351, no. 1, pp. 153–162, 2013.

[4] Badia, H., Argote-Cabanero, J., Daganzo, C.F., "How network structure can boost and shape the demand for bus transit," *Transp. Res. Part A: Policy Pract.*, vol. 103, pp. 83-94, 2017.

[5] It's More Than Buses (IMTB), The Advantage of Connective Networks, https://morethanbuses.ca/2016/10/12/the-advantage-of-connective-networks/

[6] Daganzo, C.F. , "Structure of competitive transit networks," *Transp. Res. Part B: Meth.*, vol. 44, no. 4, pp. 434-446, 2010.

[7] Estrada, M., Roca-Riu, M., Badia, H., Robusté, F., Daganzo, C.F., "Design and implementation of efficient transit networks: procedure, case study and validity test," *Transp. Res. Part A: Policy Pract.*, vol. 45, no. 9, pp. 935-950, 2011.

[8] Durán-Hormazábal, E. and Tirachini, A., "Estimation of travel time variability for cars, buses, metro and door-to-door public transport trips in Santiago, Chile," *Res. Transp. Econ.*, vol. 59, pp. 26–39, 2016.

[9] Wu, C.H., Ho, J.M., Lee, D.T., "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276-281, 2004.

[10] Jang, J., "Outlier filtering algorithm for travel time estimation using dedicated short-range communications probes on rural highways," *IET Intell. Transport Syst.*, vol. 10, no. 6, pp. 453-460, 2016.

[11] Salamanis, A., Kehagias, D.D., Filelis-Papadopoulos, C.K., Tzovaras, D., Gravvanis, G.A., "Managing spatial graph dependencies in large volumes of traffic data for travel-time prediction," *IEEE Trans. Intell. Transport. Syst.*, vol. 17, no. 6, pp. 1678-1687, 2016.

[12] Wang, H., Kuo, Y.H., Kifer, D., Li, Z., "A simple baseline for travel time estimation using large-scale trip data," in *Proc. ACM SIGSPATIAL GIS 2016*, pp. 61-64, Burlingame, California, USA, Oct. 2016.

[13] Geroliminis, N., Skabardonis, A., "Identification and analysis of queue spillovers in city street networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1107-1115, 2011.

[14] Xinghao, S., Jing, T., Guojun, C., Qichong, S., "Predicting bus real-time travel time basing on both GPS and RFID data," *Procedia Soc. Behav. Sci.*, vol. 96, pp. 2287-2299, 2013.

[15] Kumar, B.A., Vanajakshi, L., Subramanian, S.C., "Bus travel time prediction using a time-space discretization approach," *Transp. Res. Part C: Emerg. Technol.*, vol. 79, pp. 308-332, 2017.

[16] Myung, J., Kim, D-K., Kho, S-Y, Park, C-H, "Travel Time Prediction Using the K-Nearest Neighborhood Method with combined VDS and ATC Data," *Transport. Res. Rec.*, vol. 2256, pp. 1-19, 2011.

[17] Lee, W.C., Si, W., Chen, L.J., Chen, M.C., "HTTP: a new framework for bus travel time prediction based on historical trajectories," in *Proc. ACM SIGSPATIAL GIS 2015*, pp. 279-288, California, US, Nov., 2012.

[18] Z. Wall, D. J. Dailey, "An Algorithm for Predicting the Arrival Time of Mass Transit Vehicles Using Automatic Vehicle Location Data," in *Proc. 78th Annu. Meeting Transport. Res. Board*, pp. 279-288, Washington D.C., Jan. 1999.

[19] Vanajakshi, L., Subramanian, S.C., Sivanandan, R., "Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses," *IET Intell. Transp. Syst.*, vol. 3, no. 1, pp. 1-9, 2009.

[20] Xumei, C., Huibo, G., Wang, J., "BRT vehicle travel time prediction based on SVM and Kalman filter," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 12, no. 4, pp. 29-34, 2012.

[21] Shalaby, A., Farhan, A., "Bus travel time prediction model for dynamic operations control and passenger information systems," in *Proc. 82nd Annu. Meeting Transport. Res. Board*, pp. 1-15, Washington D.C., 2003.

[22] Fan, W., Gurmu, Z., "Dynamic Travel Time Prediction Models for Buses Using Only GPS Data," *Int. J. Transp. Sci. Technol.*, vol. 4, no. 4, pp. 353-366, 2012.

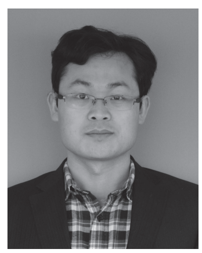
[23] Reza, R. M., Pulugurtha, S. S., Duddu, V. R., "Arima model for forecasting short-term travel time due to incidents in spatio-temporal context," in *Proc. 94th Annu. Meeting Transport. Res. Board*, pp. 1-18, paper no. 15-5553, Washington D.C., Jan. 2015.

[24] Suwardo, W., Napiah, M., Kamaruddin, I., "ARIMA models for bus travel time prediction," *J. Institut. Eng. Malaysia*, vol. 71, no. 1, pp. 49-58, 2010.

[25] Rice, J., Van Zwet, E., "A simple and effective method for predicting travel times on freeways," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 3, pp. 200-207, 2004.

[26] Padmanaban, R.P.S., Vanajakshi, L., Subramanian, S.C., "Estimation of bus travel time incorporating dwell time for APTS applications," in *Proc. IEEE IV 2009*, pp. 955-959, Xi'an, China, Jun. 2009.
[27] Vanajakshi, L., Rilett, L.R., "Support vector machine technique for the short term prediction of travel time," in *Proc. IEEE IV 2007*, pp. 600-605, Istanbul, Turkey, Jun. 2007.
[28] Ceder, A., "Public transit planning and operation: Modeling, practice and behavior," CRC press., 2007.
[29] Amita, J., Jain, S.S., Garg, P.K., "Prediction of Bus Travel Time Using ANN: A Case Study in Delhi," *Transport. Res. Procedia*, vol. 17, pp. 263-272, 2016.
[30] Mazloumi, E., Rose, G., Currie, G., Moridpour, S., "Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction," *Eng. Appl. Artif. Intel.*, vol. 24, no. 3, pp. 534-542, 2011.
[31] Wei, C. H., Lee, Y., "Development of freeway travel time forecasting models by integrating different sources of traffic data," *Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3682-3694, 2007.
[32] You, J., Kim, T. J., "Development and evaluation of a hybrid travel time forecasting model," *Transp. Res. Part C: Emerg. Technol.*, vol. 8, no. 1, pp. 231-256, 2000.
[33] Zaki, M., Ashour, I., Zorkany, M., Hesham, B., "Online bus arrival time prediction using hybrid neural network and Kalman filter techniques," *Int. J. of Mod. Eng. Res.*, vol. 3, no. 4, pp. 2035-2041, 2013.
[34] Liu, H., Van Zuylen, H., Van Lint, H., Salomons, M., "Predicting urban arterial travel time with state-space neural networks and Kalman filters," *Transp. Res. Record*, vol. 1968, pp. 99-108, 2006.
[35] Hofleitner, A., Herring, R., Bayen, A., "Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning," *Transp. Res. Part B: Meth.*, vol. 46, no. 9, pp. 1097-1122, 2012.
[36] Chien, S.I.J., Ding, Y., Wei, C., "Dynamic bus arrival time prediction with artificial neural networks," *J. Transport. Eng.*, vol. 128, no. 5, pp. 429-438, 2002.
[37] Yu, B., Lam, W.H.K., Tam, M.L., "Bus arrival time prediction at bus stop with multiple routes," *Transp. Res. Part C: Emerg. Technol.*, vol. 19, pp. 1157–1170, 2011.
[38] Jeong, R., Rilett, L.R., "Bus Arrival Time Prediction Using Artificial Neural Network Model," in *Proc. IEEE ITSC 2004*, pp. 988–993, Washington, D.C., USA, Oct. 2004.
[39] Bin, Y., Zhongzhen, Y. and Baozhen, Y., "Bus arrival time prediction using support vector machines," *J. Intell. Transport. Syst.*, vol. 10, no. 4, pp. 151-158, 2006.
[40] Yu, B., Yang, Z.Z., Chen, K., Yu, B., "Hybrid model for prediction of bus arrival times at next station," *J. Adv. Transport.*, vol. 44, no. 3, pp. 193-204, 2010.
[41] Innamaa, S., "Short-term prediction of travel time using neural networks on an interurban highway," *Transportation*, vol. 32, no. 6, pp. 649-669, 2005.
[42] Billings, D., Yang, J.S., "Application of the ARIMA models to urban roadway travel time prediction-a case study," in *Proc. IEEE SMC 2006*, pp. 2529-2534, Taipei, Taiwan, Oct. 2004.
[43] Jenelius, E., Koutsopoulos, H.N., "Travel time estimation for urban road networks using low frequency probe vehicle data," *Transp. Res. Part B: Meth.*, vol. 53, pp. 64-81, 2013.
[44] Gal, A., Mandelbaum, A., Schnitzler, F., Senderovich, A., Weidlich, M., "Traveling time prediction in scheduled transportation with journey segments," *Inf. Syst.*, vol. 64, pp. 266-280, 2017.
[45] McLeod, F., "Estimating bus passenger waiting times from incomplete bus arrivals data," *J. Oper. Res. Soc.*, vol. 58, no. 11, pp. 1518-1525, 2007.
[46] Barabino, B., Di Francesco, M., Mozzoni, S., "Time Reliability Measures in bus Transport Services from the Accurate use of Automatic Vehicle Location raw Data," *Qual. Reliab. Eng. Int.*, vol. 33, no. 5, pp. 969-978, 2007.
[47] Mishalani, R.G., McCord, M.M., Wirtz, J., "Passenger wait time perceptions at bus stops: Empirical results and impact on evaluating real-time bus arrival information," *J. Public Transport.*, vol. 9, no. 2, pp. 89-106, 2006.
[48] Fan, Y., Guthrie, A., Levinson, D., "Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security," *Transp. Res. Part A: Policy Pract.*, vol. 88, pp. 251-264, 2016.
[49] Ma, Z., Koutsopoulos, H.N., Ferreira, L., Mesbah, M., "Estimation of trip travel time distribution using a generalized Markov chain approach," *Transp. Res. Part C: Emerg. Technol.*, vol. 74, pp. 1-21, 2017.
[50] Hochreiter, S., Schmidhuber, J., "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997.
[51] Goodfellow I, Bengio Y, Courville A, Bengio Y., Deep learning, Cambridge: MIT Press, 2016, http://www.deeplearningbook.org.
[52] Duan, Y., Lv, Y., Wang, F.Y., "Travel time prediction with LSTM neural network," in *Proc. IEEE ITSC 2016*, pp. 1053-1058, Rio de Janeiro, Brazil, Nov. 2016.
[53] Yuen Sin, Bus services more reliable now: LTA. https://www.straitstimes.com/singapore/transport/bus-services-more-reliable-now-lta, Mar. 2017.

Peilan He received the Bachelor degree of Management from School of Economics and Management, Hainan University, Haikou, China, and the Master degree of Engineering from School of Computer Science and Technology, Tianjin University, China. She is currently working toward the PhD degree in the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her research interests include big data analytics in transportation, urban computing, multi-modal journey planning.



Guiyuan Jiang received the B.S. degree from Northwest University for Nationalities, China, the M.Eng. degree from Tianjin Polytechnic University, China, and the doctoral degree from Tianjin University (TJU), all in Computer Science and Technology. Currently, he works as a research fellow at School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include data analytics for citywide transportation modeling and optimization, design methodologies for reconfigurable computing systems, resource optimization for datacenter and sensor networks.

Lam Siew Kei received his BASc, MEng and PhD from School of Computer Science and Engineering (SCSE), NTU. He was a Visiting Research Fellow in the Imperial College of London, University of Warwick, and RWTH Aachen, Germany. He is currently an Assistant Professor in SCSE and his research focuses on devising custom computing solutions to meet the challenging demands of energy-efficiency, reliability and security in embedded systems. His current projects include architecture-aware algorithms for vision-enabled sensing, design methodologies for secure and reliable embedded systems, and transportation analytics.

Tang Dehua received his B.Eng. from School of Computer Science and Software Engineering, Tianjin Polytechnic University, China, where he is now pursuing the Master degree. Currently, he is working as an intern student at the School of Compute Science and Engineering, Nanyang Technological University, Singapore. His research interests include Machine Learning, Deep Learning in vision analytics.