# Predicting Travel Time of Bus Journeys with Alternative Bus Services

Peilan He, Yidan Sun, Guiyuan Jiang, Siew-Kei Lam
School of Computer Science and Engineering, Nanyang Technological University, Singapore
E-mail: {phe002, ysun014}@e.ntu.edu.sg, gyjiang@ntu.edu.sg, siewkei_lam@pmail.ntu.edu.sg

***Abstract*—Accurate travel time prediction of public transport services is essential for reliable journey planning. Existing methods for journey time prediction typically assume a fixed journey route with predefined bus services. However, there usually exist multiple alternative bus services that can serve the same journey route (or a segment of the route); thus the passengers could dynamically decide which bus service to take based on the dynamic bus arrivals. In this paper, we address the problem of travel time prediction of bus journeys with multiple alternative bus services (TP-BJMAS). We propose a novel framework to solve the TP-BJMAS problem by partitioning the journey route into several route segments based on the transfer points, such that each segment can be served by multiple bus services. The travel time of each segment is estimated using a segment prediction module based on neural network technique and the total journey time is obtained by aggregating the travel time of all segments. In the segment prediction module, the travel time using a specified bus service is obtained based on pre-trained riding time prediction model and waiting time prediction model. Since each route segment can be served by multiple alternative bus services, multiple estimations of segment travel time (ESTT) are calculated (each based on one bus service). The attention technique is utilized to fuse the ESTTs of all bus service considering the heterogeneous importance of different ESTTs. The effectiveness is evaluated using large scale real-world public transport networks and traffic data involving more than 30 bus services.**

***Index Terms*—Travel time prediction, bus journey, waiting time, alternative bus services.**

## I. Introduction

Efficient and easy-to-use public transportation system is an important element in sustainable cities as it can boost the reduction in traffic congestion and lower carbon emissions from vehicles [1], [2]. 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 [3]. Travel time prediction is also elementary to dynamic route guidance systems that provide intermodal transport options and recommended routes to travellers based on real-time data. Existing works on travel time prediction [4]–[6] typically assume a fixed journey route with predefined bus services. This could lead to unreliable prediction results as there usually exist multiple alternative bus services that can serve the same journey route (or a segment of the route). Thus, there is a need to provide accurate travel time prediction with consideration of multiple alternative bus services, such that the passengers could dynamically decide which bus service to take based on the dynamic bus arrivals.

In this paper, we address the problem of travel time prediction of bus journeys with multiple alternative bus services (TP-BJMAS). In general, the major works and contributions of this paper can be summarized as follows:

1) To the best of our knowledge, there exist no existing works that consider multiple alternative bus services during travel time prediction of bus journeys. We propose a novel framework to solve the TP-BJMAS problem by partitioning the input journey route into several route segments based on the transfer points, such that each segment can be served by multiple bus services. The travel time of each segment is estimated using a segment module based on neural network technique and the total journey time is obtained by aggregating the travel time of all segments.

2) We propose methods to predict the waiting time at the transfer points (including the origin stop), by considering two practical scenarios: online and offline scenario. For online scenario, a deep learning based model is developed to predict the waiting time at the origin stop relying on multiple spatiotemporal features (e.g. time-of-day, bus stop location, etc.) as well as previous bus arrival time records prior to the journey start time. For the offline scenario, since the previous bus arrival times are not available (at the transfer points), the method proposed in [7] is utilized.

3) Since each route segment can be served by multiple alternative bus services, multiple estimations of segment travel time are obtained (each based on one bus service). Different estimation contributes not equally as the bus service with earlier arrival time (i.e. shorter waiting time) is more likely to be taken. As such, we utilize the attention technique to calculate the segment travel time by fusing the estimations of all bus service. Specifically, we develop an attention network to dynamically determine the importance of different estimations, based on the predicted waiting time related to each bus service.

Through extensive experimental evaluations on real world public transport networks and traffic data involving more than 30 bus services, we show that our proposed method can efficiently predict the travel time of the TP-BJMAS problem.

This research project is partially 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.

The remainder of the paper is organized as follows. Section II reviews some related works and highlights the similarity and differences between this work and the existing ones. Section III introduces important definitions and problem description. Section IV discusses the proposed method for predicting the travel time of bus journeys with multiple alternative bus services. Section V evaluates the performance of the proposed approach, and Section VI concludes the paper.

## II. Related Works

### A. Bus Travel Time Prediction

In general, the existing methods for predicting bus travel times can be categorized into the following groups: historical average (HA) method [8], Kalman Filter (KF) approaches [9], [10], time series analysis (TS) [11], and machine learning (ML) based methods such as linear regression [12], support vector regression [4] and neural networks [13], [14]. HA approaches [8] predict the travel time of a journey by relying on the historical average travel time for the same daily period over different days. It constructs a non-parametric model that makes no assumptions on the underlying data. However, it is difficult to collect sufficient fine-grained travel records between OD pairs (e.g. using APPs running on mobile phones), not only because of high power consumption of mobile phones but also due to privacy issues. This negatively impacts the amount of data that can be collected and thus affect prediction accuracy. KF approaches use a series of travel time records observed over time to produces estimates of unknown travel times, by estimating a joint probability distribution over the travel time records for each time frame [9], [24]. Typically, a KF model cannot be generalized to the prediction of different time series [10] (e.g. a KF model that is used to predict the travel time of a bus line 15 minutes later may not work well for predicting the travel time of the same bus line 1 hour later). 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 is unreliable if there is a huge difference in travel time between two consecutive time steps. TS methods predict future values based on previously observed values by modeling possible internal structure in the data [11]. However, it was shown that the TS based approaches could not produce high-accuracy predictions for complex scenarios such as urban bus travel time prediction. It is because bus travel time is affected by not only by the dynamic travel demands (e.g. the number of alighting and boarding passengers, etc.) but also by the traffic and route conditions (e.g. route length, traffic congestion, number of bus stops, intersections and traffic signals etc.). ML models based on regression, SVM and neural network have been utilized in travel time prediction. Linear regression models have been widely used in traffic prediction to capture the linear relationship between travel times and the related impact factors [12]. This model is computationally efficient but usually, produce results of low accuracy for nonlinear systems. Since SVM have better generalization capacity and can guarantee global minima for a training data, existing works have applied Support Vector Regression (SVR) to travel time prediction of cars on highways [4] and buses in city road network [25]. SVR-based model suffers from high computation overhead. Many Neural Network (NN) based methods have been developed to predict bus travel time using both historical and real-time data [14], [26], [27]. Factors that affect the travel time, such as the travel distance, number of stops, number of passengers boarding and alighting at each stop, average non-stop journey time, dwell time, bus schedule, have been used as inputs for the existing NN prediction models. The NN based model has demonstrated advantages over the KF model, HA model, ARIMA and classic regression models. The ML approaches are often combined with other approaches to form hybrid methods [28]. Our work differs from the above methods of bus travel time prediction as they only target a single bus journey. These methods cannot be directly applied to our problem where the waiting times during transfer need to be considered.

### B. Journey Travel Time Prediction

Many works have been reported to estimate the travel time of vehicles (taxis or private cars) between an origin-destination (OD) pair [4]–[6], [29], [30]. They proposed route-based strategies, i.e. first partition the route into multiple route segments and then aggregate the travel time spent on each segment using historical trajectories. However, these approaches are not suitable for bus journey time prediction because: 1) Unlike the general autonomous vehicles (e.g. private cars, taxis), the travel time of buses is not only affected by the traffic conditions (e.g. traffic flow, vehicle speed, traffic signal) but also by other factors such as travel demand, the dwell time at each bus stop, bus service frequency, bus schedule timetable, etc. 2) Waiting times at the transfer points (including the origin stop) need to be considered as they are 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.

The work in [31] 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 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. Also, it is shown that simply summing up the travel time of each route segment does not result in high prediction accuracy [32]. The work in [7] investigated the problem of travel time prediction of bus journeys, where each journey may involve multiple bus services. They also considered the waiting time at the transfer points, including the origin stop. However, it assumes that the utilized bus services are fixed and given in advance, while neglecting the fact that journey route (segment) could be served by multiple bus services and the passengers may dynamically determine which bus to take based on dynamic bus arrival.

## III. Definitions and Problem Description

In this section, we first present several definitions and then provide our problem description.

*Bus Line*: A bus line is a fixed route regularly traveled by the bus service, and it can be represented by a sequence of points $BL_l =< p_1^l, p_2^l, \ldots, p_{n_l}^l >$ where $p_i = (x_i, y_i)$, for $i = 1 \ldots n$, is the GPS location of the $i$-th bus stop along the bus line $BL_l$, and $n_l$ is the number of bus stops in the bus line. We use bus stops as points to represent a bus line as predicting the arrival time at a bus stop is usually desired. Bus passengers tend to be only interested in the arrival time at a bus stop rather than a random point along the route. In this paper, the notation bus line, bus route, and bus service are used interchangeably. A *bus line segment* is a set of connected points, 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 $BL_l$. Particularly, the bus line segment $R_{i,j}^L$ is called a *unit segment* if $p_i^L$ and $p_j^L$ are consecutive bus stops of the bus line $BL_l$.

*Bus Trajectory:* A *bus trajectory* $Tr$ is a sequence of consecutive points that record a bus' travel information, i.e. $Tr = \{p_1, p_2, ..., p_{|Tr|}\}$. Each point $p_i$ contains the latitude information, longitude information and the timestamp information. A bus trajectory contains the arrival time of a bus at each of the bus stops along the bus line. Based on the bus trajectory, the actual bus travel time between any segment of the trajectory can be derived.

*Journey Route:* A *bus journey route* signifies a complete travel route from the passenger's origin to the destination, which may involve multiple bus line segments using different bus lines/services. Passengers typically need to wait for a bus service at the origin stop as well as the intermediate bus stops/interchange station. A journey route considered in this paper specifies the origin $p_o$, destination $p_d$, set of transfer stops $TP$.

*Alternative Bus Services:* A bus journey typically is achieved by using multiple bus services, and the passengers need to change to a different bus service at each transfer point. As such, a bus journey route typically can be partitioned into several segments based on the transfer points along the journey route. For each segment, e.g. starting from stop A and ending at stop B, there usually exist multiple alternative bus services that can take the passengers from A to B, as illustrated in Fig. 1. The alternative bus services may travel through either the same or different roadways.

The following describes our problem statement. *Given a journey start time $t_0$ and the journey route $R$, which may cover portions of multiple route segments, characterized by the origin $p_o$, the destination $p_d$ and a set of transfer points $TP$, our goal is to predict the total journey time including the riding time in each route segment and the time waiting for the bus services, with the consideration that each route segment can be served by multiple alternative services.*

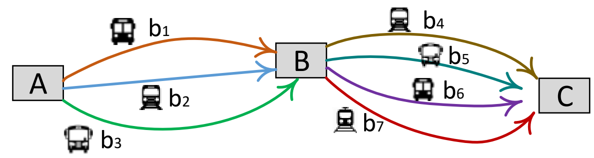



Fig. 1. An example of journey routes with multiple alternative bus services. There are 3 available bus services (i.e. $b_1$, $b_2$, $b_3$) to take passengers from stop A to stop B while 4 available services (i.e. $b_4$, $b_5$, $b_6$, $b_7$) between stop B and C. As such, there could be 12 different plans to travel from stop A to C. This paper predicts the journey time from A to C without fixing the utilized transport services in advance.

We make the following assumptions: 1) the journey route (specified by origin, transfer points and destination stop) is given, i.e. it has been specified by the user or generated by a route planner; 2) the alternative bus services for each route segment are known; 3) the passenger is notified with all the alternative services and will take the first arrival service.

## IV. Proposed Method

In this section, we first provide an overview of the proposed framework for the TP-BJMAS problem. Then, we present the details of each component of the framework in the following sections.

### A. Main Framework

Fig. 2 illustrates the framework for the TP-BJMAS problem. The input journey route is partitioned into several segments based on the transfer points, such that a journey route with $k$ different services is partitioned into $k$ segments. The time for travelling on each segment consists of both the riding time on the bus service and the time waiting for the bus service at the transfer point. In the proposed method, the travel time of each segment is predicted by a separate *Segment Module* (as shown in the figure) and the total travel time of the entire journey is obtained by aggregating the travel time of all segments. The Segment Module works in the following way.

1) We develop efficient machine learning methods to predict the waiting time at the transfer/origin stop, by considering two different scenarios, i.e. online and offline scenario. The model can predict the arrival time of the next three bus trips for each bus services (e.g. bus 241). For example, the expected arrival times of the next three bus 241 at stop A are 08:02, 08:09, 08:15, respectively. Based on the obtained bus arrival time, the expected waiting time for each alternative bus services can be easily calculated.
2) A deep learning approach based on Long Short-Term Memory (LSTM) is developed to predict the riding time on each segment of a given bus line. The LSTM based model relies on multiple features extracted from multiple data sources, which not only characterize 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.).

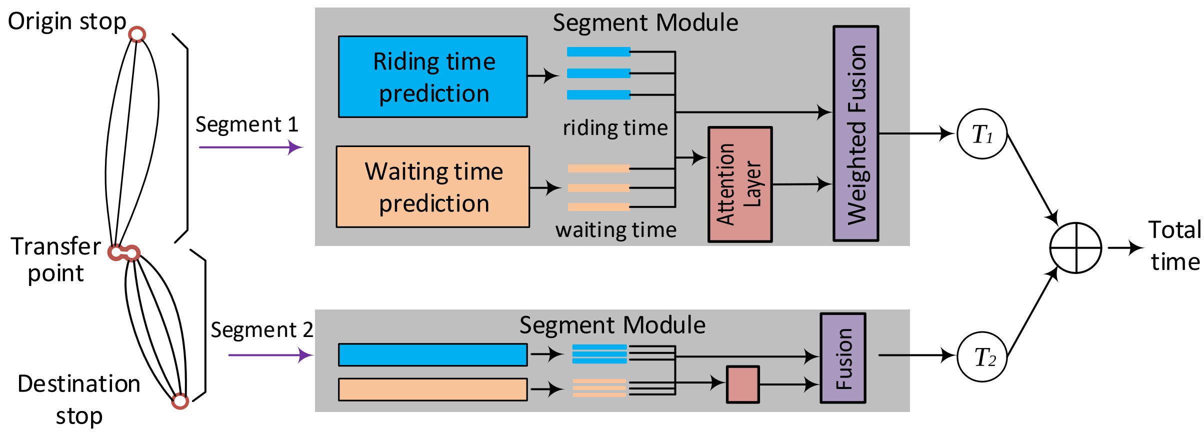



Fig. 2. Proposed framework for predicting travel time of bus journey with multiple alternative bus services.

3) Due to the uncertainty caused by the dynamic traffic conditions as well as the traffic signals, the actual bus arrival time might be different from the predicted. The bus service with smaller predicted waiting time is more likely to arrive earlier and be taken by the passenger. As such, the riding time regarding the bus service with smaller waiting time is more important than those with larger waiting time. We utilize the attention technique to integrate the journey times of using different bus services. The final total journey time can be calculated by aggregating the journey time over all segments of the given journey route.

In the following sections, we will present the details of each component, including 1) *Riding Time Prediction* component, which predicts the riding time on a route segment using a single bus service; 2) *Waiting Time Prediction* component, which predicts the waiting time for specific bus service at a bus stop; 3) *Attention based Fusion* component, which estimates the travel time on a route segment by fusing the results of multiple alternative bus services.

### B. Riding Time Prediction

This section describes the model for predicting the bus riding time using a single bus service without transfering. The Riding Time Prediction component is pretrained and will be directly used in the prediction model for the TP-BJMAS problem. We first discuss the training dataset and the features that are used to train the LSTM network, then we present the details of our LSTM network.

*1) Training data:* Given a bus trajectory, we can extract a training data consisting of a number of journey records, each journey record is a sub-trajectory of the complete bus trajectory. With a journey record, a vector representing the information of the trip record is extracted, which consists of two parts: travel time and a feature vector containing the following features that impact the journey travel time [7]:

- *Time of day*, i.e. journey start time. In the same day, journeys with different start time have significant variance in the travel time, which shows that the journey start time has an impact on the travel time and should be used as a feature for travel time prediction. The journey start time can be used as an indicator to characterize the variance of traffic conditions over a day.
- *Day of week*, i.e. day that journey will be made. Working days (Monday to Friday) have similar travel time patterns, and the same can be observed for weekends (Saturday and Sunday). On the other hand, weekdays and weekends have different travel time distribution. This indicates that the day of week should also be considered as a feature to differentiate the traffic conditions between working days and weekends for travel time prediction.
- *Travel distance*, i.e. total distance of the journey route. Longer travel distance typically leads to longer travel time. In addition, longer travel distance generally corresponds to more intersections and more traffic signals, which may cause unexpected delays.
- *Number of bus stops*, i.e. number of stops between the origin stop and the end stop along the journey route. It has direct impacts on the total travel time. The reason that more bus stops lead 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. This reflects the expected number of bus stopping and the bus dwelling time.
- *Number of intersections*, i.e. number of intersections along the journey route, including pedestrian crossings. Note that buses typically slow down at intersections.
- *Number of traffic signals*, i.e. number of traffic signals along the journey route. Buses often need to stop at the intersections with signals.
- *Weather condition*. This affects the bus moving speed.

The extracted features can characterize not only the temporal data dependencies but also the spatial data correlations. Formally, a sub-matrix $\mathbf{x}_i \in R^{F_i \times D}$ is extracted from each bus trajectory, where $F_i$ is the number of journey records obtained from trajectory $Tr_i$ and $D$ is the dimension of features. A bus trajectory $Tr_i$ has with $n_i+1$ bus stops has $\frac{n_i \cdot (n_i-1)}{2}$ different journey records in total. The sub-matrices obtained from all bus trajectories are combined together as $\mathbf{x} \in R^{F \times D}$ to train a LSTM model, which is able to make accurate travel time prediction for any segment of the bus service. The ground truth vector of the journey travel time is denoted as $\hat{\mathbf{y}} \in R^F$. In addition, we use $\mathbf{y} \in R^F$ to denote the target vector.

*2) LSTM Network Structure:* The input matrix is fed into two stacked LSTM layers, where each LSTM layer has 128 neurons. The LSTM memory cell can be described with the following equations:

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

where $t$ indicates 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 $\mathbf{C}_t$. $\sigma$ 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 the bias terms of the three gates and the cell gates.

The output of the second layer goes into several fully-connected layers, where each layer is of size 128. The fully-connected layers are connected with residual connections, which is shown to be efficient for training a very deep neural network [20]. For the first fully connected layer, its input is the output of the second LSTM layer. Let $\sigma_{f_i}$ be the $i$-th residual fully-connected layer, then the output of the first layer is $\sigma_{f_i}(\mathbf{o}_z)$, where $\mathbf{o}_z$ is the output of the LSTM layer. For the rest of the residual layers, let $\mathbf{o}_{f_i}$ be the output of the $i$-th layer, then the output of the $(i+1)$-th layer can be represented as $\mathbf{o}_{f_{i+1}}$=$\mathbf{o}_{f_i} \oplus \sigma_{f_{i+1}}(\mathbf{o}_{f_i})$, where $\oplus$ is an element-wise add operation.

Finally, we apply a tanh activation function and obtain the prediction results. In order to prevent overfitting, two widely used regularization techniques are employed: dropout [21] and $L_2$ regularization. The dropout mechanism is applied to each hidden layer, where the rate of dropout is set to 0.5. Moreover, we apply $L_2$ regularization on model weights to prevent possible overfitting. Formally, the loss function used for training the model is:

$$L_{loss} = \sum_{i=1}^{F} (\hat{y_i} - y_i)^2 + \lambda \parallel \mathbf{W} \parallel^2 \tag{2}$$

where $\lambda$ is a hyper-parameter to control the regularization strength and $\mathbf{W}$ denotes all weights in the network. The Adam optimizer is utilized as the gradient descent optimization algorithm. The training process repeats for 50 epochs.

### C. Waiting Time Prediction

Predicting the waiting time for bus service at the origin stop or at a transfer point is a challenging task. Assume that a passenger is expected to arrive at bus stop $p$ at time $t_0$ to wait for the bus (e.g. bus 179). We consider two different scenarios for waiting time prediction: 1) *Online scenario*, where the $\omega$ previous bus arrival time before time $t_0$ are known or can be preciously predicted; 2) *Offline scenario*, where the $\omega$ previous bus arrival time before time $t_0$ are not known and cannot be preciously predicted. For the online scenario, a neural network based prediction model is developed to predict the arrival time of the next 3 bus trips after time $t_0$, for a given bus service at a bus stop. For the offline scenario, since the previous bus arrival information is not available, we employ the method proposed in [7].

*1) Online scenario:* We rely on the large scale dataset of historical bus arrival times of buses at each bus stop to train a model to estimate the bus arrival times and thus the waiting times. Let $BAT_{p,i}$ be the dataset of historical bus arrival times at bus stop $p$ containing data of the $i$-th day. For each data sequence of $BAT_{p,i}$, we can extract a training data consisting of a number of training records, each record is a vector consisting of the following features:

- *Time of day*, i.e. the expected time that a passenger arrives at the bus stop, denoted as $t_0$. The $t_0$ affects the expected waiting time because a bus service typically has a higher frequency at peak hours and lower frequency at off-peak hours. As such, $t_0$ should be used as an indicator to characterize the variance of bus service frequency over a day.
- *Day of week*, i.e. day that journey will be made. Weekdays and weekends have different frequency of bus services and traffic conditions. This indicates that the day of week should also be considered as a feature to differentiate the traffic conditions between working days and weekends for bus arrival time prediction.
- *Previous $\omega$ records of bus arrival time*. The original sequence of $\omega$ bus arrival time records, as well as the sequence of time intervals between consecutive records are used.
- *Weather condition*. This affects the bus moving speed and travel demands.
- *The location of the bus stop*. It reflects the spatial information, as the bus services have uneven spatial distribution in both services frequency and coverage.

A sub-matrix is extracted from each $BAT_{p,i}$, and the combination of all sub-matrixes from all $BAT_{p,i}$ for all bus stops in all days form the entire training set. The ground truth for each record is the arrival times of the next three buses for the given bus service (e.g. bus 179) after time $t_0$ at the stop $p$. The utilized prediction model is the same as that used in section IV-B1.

*2) Offline scenario:* Since the previous $\omega$ records of bus arrival times may not be readily available for some cases (e.g. predicting the bus arrival times at the transfer point when the passenger has not started the journey), we employ the offline method presented in [7] for waiting time prediction. Specifically, $BAT_p$ is the sequence of the historical bus arrival times at stop $p$ consisting data of $d$ days, then the historical average waiting time is

$$HA(p) = HA(p, t_0) = \frac{\sum_{i=1}^{d}(t_i|_{BAT_p} - t_0)}{d} \tag{3}$$

where $t_i$ is the first bus arrival time after time $t_0$ in the $i$-th day of the dataset $BAT_p$. $t_i|_{BAT_p} - t_0$ is the historical waiting time on the $i$-th day.

For estimating the waiting time at an intermediate transfer point $p'$, the passenger's arrival time $t'_0$ at this stop can be estimated as $t_0$ plus the predicted journey time between the previous transfer (or origin) stop and $p'$. Then the waiting time $HA(p')$ at $p'$ is estimated as $HA(p', t'_0)$. However, the major challenge in waiting time prediction is that it is sensitive to the arrival time of the passenger at the transfer point, which cannot be accurately predicted at minute granularity. Thus, an enhanced method is developed based on the HA method. Specifically, 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 $\varepsilon$ is set to be the mean absolute error of the LSTM network presented in the previous section. The work in [7] proves that the exact arrival time at the transfer stop of the passenger will fall into this interval with a probability above $1 - \frac{\mathbf{Var}[e_i]}{\varepsilon^2}$. As such, the waiting time can be calculated as

$$Wait(\mathbf{E}[t_a], \varepsilon) = \frac{\sum_{i=-\varepsilon}^{\varepsilon}(HA(\mathbf{E}[t_a] + i))}{2\varepsilon + 1} \tag{4}$$

where $\mathbf{E}[t_a]$ is the expectation of arrival time based on historical bus trajectories, and $\varepsilon$ is the MAE (mean absolute error) of the bus riding time prediction model discussed in Section IV-B1. $HA(\mathbf{E}[t_a] + i)$ indicates the estimated waiting time using the HA approach if the passenger arrives at the bus stop at time $\mathbf{E}[t_a] + i$. For estimating the waiting time at the origin stop, $t_a$ is set to journey start time $t_0$ and $\varepsilon$ is set to 0 as $t_0$ is the exact arrival time of the passenger.

### *D. Attention based Journey Time Fusion*

As mentioned before, the proposed method partitioned the input journey route is partitioned into several segments based on the transfer points. Each segment can be served by multiple different bus services, e.g. $BL = \{b_1, b_2, \cdots, b_K\}$. This section introduces the method for estimating the journey time by fusing the journey time predicted based on each of the alternative bus services. For clarity, the journey time predicted based on single bus service is called a segment time component. The segment time component consists of both the bus riding time and the time waiting for the bus service at the transfer point. The bus riding time can be obtained using the prediction model presented in Section IV-B while the waiting time at the transfer point can be obtained using the prediction model presented in Section IV-C. Since different bus services will arrive at different time due to dynamic and uncertain traffic conditions, the bus services with shorter waiting time are more likely to be taken than others. This means that the segment time components obtained from all the alternative bus services do not contribute equally to the overall journey time prediction, the segment time component obtained based on the bus services with shorter waiting time is more important and should be given more attention. In our model, the attention mechanism [30] is employed to discriminate the importance of different bus services automatically. The key idea is to assign weights to different segment time component. Formally, the segment travel time is calculated as

$$T^i = \sum_{b_j \in BL^i} a_j \cdot (Wait_j^i + Ride_j^i) \tag{5}$$

where $Wait_j^i$ and $Ride_j^i$ are the waiting time and riding time of service $b_i$ ($b_i \in BL^i$), $BL^i$ is the set of all alternative bus services of the $i$-th segment, $a_j$ is the weight for term $Wait_j + Ride_j$, and $\sum a_j = 1$. The weight parameter $a_j$ is learned through the attention layer,

$$z_j = \mathbf{V}_{aj}^T \cdot relu(\mathbf{W}_{aj}(Wait_j + Ride_j) + \mathbf{b}_{aj}) \tag{6}$$

$$a_j = \frac{exp(z_j)}{\sum_j exp(z_j)} \tag{7}$$

where $\mathbf{W}_{aj}$ is weight matrices connecting neurons in attention layer and the input $Wait_j + Ride_j$, $\mathbf{V}_{aj}^T$ connect neurons in attention layer with $z_j$, $\mathbf{b}_{aj}$ is the bias terms.

After obtaining the segment travel times using the Segment Module for all the segments of the journey route, the total travel time of the entire journey can be obtained by aggregating the travel time of all segments.

## V. Results and Analysis

### *A. Dataset*

*Road Networks:* The road network (obtained from OpenStreetMap[1]) is utilized to derive the information of intersections ($\#.intsections$), the number of traffic signals ($\#.signals$) as well as the walking distance between transfer stop-pairs for any journey routes. Our experiment relies on the road network of Singapore, which comprised of 41,732 nodes and 98,539 road segments.

*Bus Route:* The bus route information[2] includes the ID (a five digit number) of each bus stop in sequential order, the GPS location (latitude and longitude) of each bus stop, and the travel distance between any two consecutive bus stops. We map the bus routes to the road network using the GPS locations of bus stops to determine the sequence of road segments

[1]https://www.openstreetmap.org/export

[2]https://www.mytransport.sg/content/mytransport/home/dataMall.html

traveled by 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 and the number of traffic signals for any journey routes can be calculated. More than 30 bus lines are used in the experiments, which are shown in Fig. 3.

*Bus Trajectories:* A bus trajectory dataset is derived based on the real-world Bus Arrival Time dataset (the arrival time of the next bus for each bus stop, at every minute) provided by the Land Transport Authority, Singapore. The dataset contains bus trajectory data of 30 bus lines from May 06 to July 07, 2017 (63 days in total). Each bus trajectory is a sequence of points, and each point contains the information of the stop ID, the GPS location of the bus stop, the timestamp (arrival time of the bus at the stop), and the bus line ID. With the trajectories, the following features are extracted for each trajectory segment: the day-of-week, the journey start time, as well as the trip duration (i.e. the bus riding time regarding the journey route segment).

*Pseudo Journey Records:* It is challenging to obtain sufficient journey records of individual passengers. Therefore, a dataset of journey records is generated based on the bus trajectories and bus arrival time records. The journey records are generated in the following way, which is similar to that used in [7]: We repeatedly select a journey route (may involve transfering among multiple bus services) on the bus network and then generate a certain amount of journey records by randomly selecting journey start time over a period of 63 days.

*Method to generate the journey routes*: a journey route contains the information of the origin $p_o$, the destination $p_d$ and a set of transfer points $TP$. The origin $p_o$ and destination $p_d$ are selected based on real-world bus/metro travel demand information[3]. We select the top 2000 OD-pairs (with highest travel demands) as the $p_o$ and $p_d$ to generate the journey routes to test the performance of the proposed prediction method and the baselines. Fig. 4(a) shows the distribution of journey volumes on the top 2000 OD-pairs. When a pair of $p_o$ and $p_d$ are determined, the set of transfer points $TP$ is constructed as follows: We first search all the feasible (reachable) path from $p_o$ to $p_d$ using at most 3 bus services. Then, all the journey paths are classified into three group: the paths of the first group need not transfer (denoted as $G_1$), each of the paths in the second group contains 1 transfer (denoted as $G_2$), and each of the paths in the third group contains 2 transfer points (denoted as $G_3$). For $G_1$, the set TP is empty. For $G_2$, the set TP contains a single transfer point, which is selected as the transfer point that is mostly used by the paths in the second group. For $G_3$, the set TP contains 2 transfer points, where the first transfer point is selected in the same way as in $G_2$ (the selected transfer point is denoted as $tp_1$). Then the second transfer point is selected as the transfer point that is mostly used by the paths containing $tp_1$ in group $G_3$. All the alternative bus services will be recorded at the same time. Since the historical arrival times of all alternative bus services at the origin and the transfer stops are known, it is easy to recover the journey record by assuming that the passenger can always board the first bus that arrives. This can also be extended 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 to board the first bus at each transfer point. During the journey record generation, we can allow $\rho\%$ passengers who need to transfer to use the second bus trajectory. With the journey record, the ground truth of total journey time can be easily calculated.

*(4) Weather data:* Weather condition influences the bus travel speed by affecting the bus stopping time at bus stops as well as the moving speed of vehicles. Hourly-grained weather data are collected during the same time period, i.e. from May 06 to July 07, 2017[4]. There are 14 types of weather conditions, including thundershowers, strong thunderstorms, rain showers, light rain, sunny, etc.

## *B. Baseline Methods for Bus Travel Time Prediction*

Since there are no existing approaches for the same problem considered in this paper, we compare our method with the following baseline methods, which are briefly described below.

1) *Weighted-Segment Sum Method* (WSSM): WSSM partitions the entire journey route into several segments based on transfer points. For each segment, the segment journey time is predicted as the weighted sum of the travel time of all alternative bus services. The weights are calculated based on the frequency (the total number of bus trips in each day) of each bus service. For example, there are three alternative bus services for the first segment, i.e. $b_1$, $b_2$ and $b_3$, whose service frequencies are $n_1$, $n_2$ and $n_3$, respectively. Then the weight for service $b_i$ $(1 \leq i \leq 3)$ is calculated as $\frac{n_i}{\sum_{j=1}^{3} n_j}$.

2) *Weighted Complete-Path Method* (WCPM): This method first identifies all feasible complete journey paths (where the utilized bus service for each route segment is deterministic) and estimates the total travel time for each deterministic journey path. Then the final journey time is calculated as the weighted average of the journey time of all feasible journey paths, where the weights are calculated based on the frequencies of the involved bus services. Specifically, assume there are 3 alternative bus services for the first segment ($b_1$, $b_2$, $b_3$) and 4 alternative bus services for the second segment ($b_4$, $b_5$, $b_6$, $b_7$), then the weight for the path $< b_i, b_j >$ is calculated as $\frac{n_i \cdot n_j}{\sum_{1 \leq p \leq 3, 4 \leq q \leq 7} n_p \cdot n_q}$.

3) *Earliest Bus Service Method* (EBSM): The EBSM partitions the journey route into several segments based on the transfer point. For each segment, the bus service with shortest waiting time is used to estimate the travel time for the segment, then the total travel time of the entire journey is calculated as the sum of the travel time over all segments.

[3]https://www.mytransport.sg/content/mytransport/home/dataMall.html

[4]https://www.timeanddate.com/weather/singapore/singapore

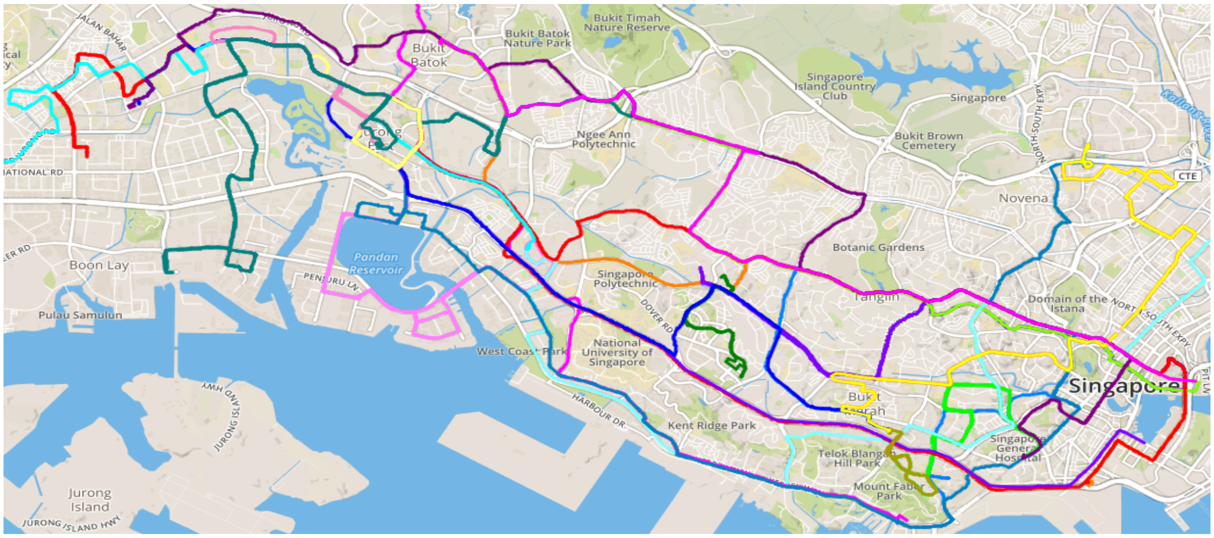



Fig. 3. The spatial distribution of 30 bus lines used in the experiments.

### C. Evaluation Metrics

The performance measures used are the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE),

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

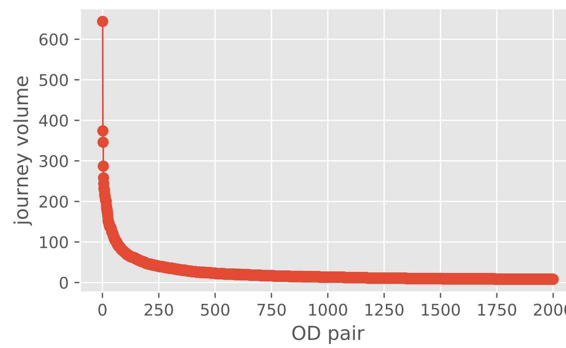
$$RMSE = \sqrt{\frac{1}{F}\sum_{i=1}^{F}(y_i - \hat{y}_i)^2}$$

where $F$ is the size of the testing set, $y_i \in R^F$ is the predicted value and $\hat{y}_i \in R^F$ is the actual value observed.

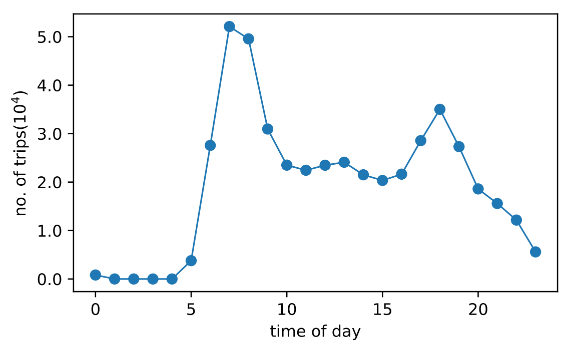
### D. Results Comparison

*1) Prediction Accuracy:* In the experiment, the data collected from May 06, 2017 to Jun. 30, 2017 are used for training the prediction model, while the data collected for the last week (from July 01 to July 07, 2017) are used for testing. During the training process, 25% of the training set is used for validation. For more clear illustration of the obtained results, we classify all the journey routes into three group: 1) in the first group (group 1), each of the journey paths uses a single bus service without transfering; 2) in the second group (group 2), each of the journey paths contains 2 bus service with one transfering; 3) in the third group (group 3), each of the journey paths contains 3 bus service with two transfering. As mentioned before, the journey records are generated based on real-world historical bus trajectories of 30 bus lines in Singapore. For each group, we randomly select 1000 journey routes (including information on the origin, destination, the set of transfer points, the set of alternative bus services) and generate 50 journey start times in the period from Jul. 01 to Jul. 07, 2017. The generated journey start times follow the distribution, as shown in Fig. 4(b)

Fig. 5 presents the performance comparison of our proposed method and the baseline methods, in terms of MAE and RMSE. It can be observed that the proposed method produces the best performance on MAE as well as RMSE on all of the three groups of journey routes. For example, the average improvements compared with the baseline methods on MAE over the group $G_1$ are 10.3%, 46.7%, 49.6%, respectively. The improvements are more significant on relatively shorter jour-

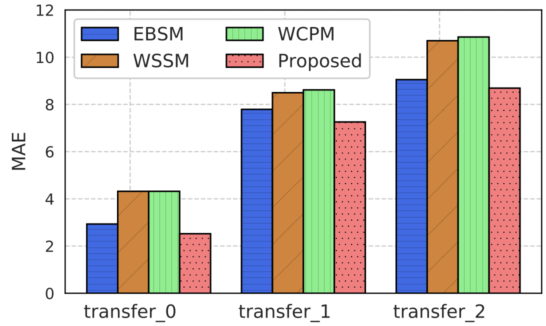



(a) Distribution of journey volumes on the top 2000 OD-pairs.

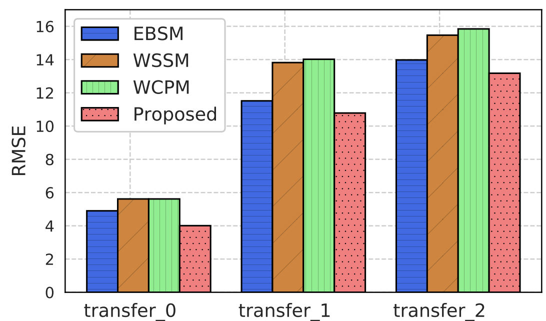



(b) Distribution of journey start times of historical journey records of the top 2000 OD-pairs.

Fig. 4. Statistical Information of journey records on the top 2000 OD-pairs.




(a) Comparison of MAE




(b) Comparison of RMSE

Fig. 5. Comparison of the results, the paths of group $G_1$ need not to transfer, each of the paths in group $G_2$ contains 2 segments with 1 transfer, and each paths in group $G_3$ contains 3 segments with 2 transfer points.

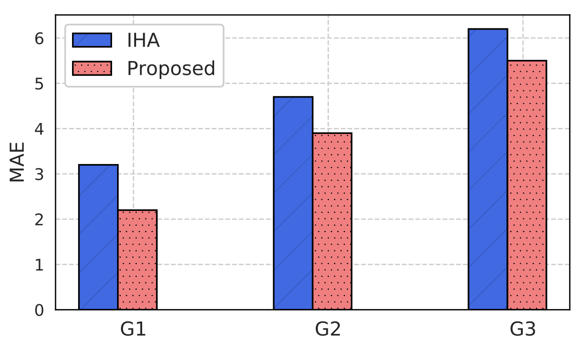



Fig. 6. Comparison of the waiting time obtained by the IHA method and the proposed method, the routes in $G_1$ contain waiting time at origin stop, routes in $G_2$ contain waiting time at origin stop and one transfer stop, routes in $G_3$ contain waiting time at origin stop and two transfer stops.

neys than those of longer ones. This is because the proposed method partitions a given journey into multiple components (riding time and waiting time components), and the journeys with transfers typically have more components than journeys without transfers thus leading to larger accumulated errors. Another reason is that, longer journeys contain more transfer points and route segments, and more bus services are involved which increases the prediction complexity. Similar results are observed on RMSE, as shown in Fig. 5(b).

Fig. 6 compares the accuracy of waiting time prediction at the origin stop in terms of MAE, in comparison to the IHA method proposed in [7]. It can be observed that the proposed method for waiting time prediction achieves better performance than IHA methods for all three scenarios. For the first group, i.e. predict the waiting time at the origin stop (online scenario), the proposed method achieves better results than IHA because the previous arrival times of the involved bus services are available and are used in the proposed method. For the group $G_2$ and $G_3$, the improvements decrease slightly because the waiting time at the transfer points are more challenging, as it can only be predicted in an offline scenario, which leads to relatively higher errors.

*Runtime of Prediction Model*: Despite the longer training time, the bus journey prediction is very efficient. The average predicttion time of the four algorithms are: 0.042s(WSSM), 0.320s(WCPM), 0.035s(EBSM), 0.051s (proposed), respectively.

## VI. Conclusions

In this paper, we investigated the problem of predicting bus journey time, considering that the journey route (segment) can be served by multiple alternative bus services. We proposed a framework that partitions the entire journey route into several segments, such that the travel time regarding each segment is predicted using a segment module and the total journey time is obtained by aggregating the travel time of all segments. For each segment, there are multiple alternative bus services, and an attention network is developed to take into account the different importance of different bus services. We also developed effective models to predict the bus riding time as well as the waiting time for buses at the transfer points. By conducting extensive experiments on large scale real-world bus travel data, we showed that our proposed method can predict the travel time for any given journeys and obtains better results than the baselines.

## 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. 1522, 2015.

[2] Zhou, Y., Jiang, G., Jiang, G., “Discover the road sequences of bus lines using bus stop information and historical bus locations,” *Int. J. Distrib. Sens. Netw.*, vol. 15, no. 2, pp. 1522, 2019, Art. no. 1550147719830552.

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

[4] 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.

[5] 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.

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

[7] He, P., Jiang, G., Lam, S. K., Tang, D., “Travel-Time Prediction of Bus Journey With Multiple Bus Trips,” *IEEE Trans. Intell. Transp. Syst.*, 2018, inpress.

[8] 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 *ACM SIGSPATIAL GIS 2015*, pp. 279-288, California, US, Nov., 2012.

[9] 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.

[10] 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, 2015.

[11] 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 *94th Annu. Meeting Transport. Res. Board*, pp. 1-18, paper no. 15-5553, Washington D.C., Jan. 2015.

[12] 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.

[13] 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.

[14] 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.

[15] Lin, C.J., "Projected gradient methods for nonnegative matrix factorization," *Neural Comput.*, vol. 19, no. 10, pp. 2756-2779, 2007.

[16] Nie, F., Huang, H., Cai, X., Ding, C. H., "Efficient and robust feature selection via joint $l_{2,1}$-norms minimization," in *NIPS 2010*, pp. 1813-1821, Hyatt Regency, Vancouver Canada, Dec. 2010.

[17] X. Yang, C. Deng, X. Liu, F. Nie, "New $l_{2,1}$-Norm Relaxation of Multi-Way Graph Cut for Clustering," in *AAAI 2018*, pp. 4374-4381, New Orleans, Louisiana, USA, Feb. 2018.

[18] W. Xu, X. Liu, Y. Gong, "Document clustering based on nonnegative matrix factorization," in *ACM SIGIR 2003*, pp. 267-273, New York, NY, USA, Aug. 2003.

[19] Deng, D., Shahabi, C., Demiryurek, U., Zhu, L., "Situation Aware Multi-Task Learning for Traffic Prediction," in *IEEE ICDM 2017*, pp. 81-90, Nov. 2017.

[20] He, K., Zhang, X., Ren, S., Sun, J., "Deep residual learning for image recognition," in *IEEE CVPR 2016*, pp. 770-778, Las Vegas, NV, USA, Jun. 2016.

[21] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929-1958, 2014.

[22] B. Pan, U. Demiryurek, C. Shahabi, "Utilizing real-world transportation data for accurate traffic prediction," in *IEEE ICDM 2012*, pp. 595-604, Nov. 2012.

[23] Vanajakshi, L., Rilett, L.R., "Support vector machine technique for the short term prediction of travel time," in *IEEE IV 2007*, pp. 600-605, Istanbul, Turkey, Jun. 2007.

[24] 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.

[25] Bin, Y., Zhongzhen, Y., Baozhen, Y., "Bus arrival time prediction using support vector machines," *J. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 151-158, 2006.

[26] Duan, Y., Lv, Y., Wang, F.Y., "Travel time prediction with LSTM neural network," in *IEEE ITSC 2016*, pp. 1053-1058, Rio de Janeiro, Brazil, Nov. 2016.

[27] Sun, Y., Jiang, G., Lam, S. K., Chen, S., He, P., "Bus Travel Speed Prediction using Attention Network of Heterogeneous Correlation Features," in *SIAM SDM 2019*, pp. 73-81, Hyatt Regency Calgary, Calgary, Alberta, Canada, May 2019.

[28] Zaki, M., Ashour, I., Zorkany, M., Hesham, B., "Online bus arrival time prediction using hybrid neural network and Kalman filter techniques," *Int. J. Mod. Eng. Res.*, vol. 3, no. 4, pp. 2035-2041, 2013.

[29] Durn-Hormazbal, 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. 2639, 2016.

[30] D. Wang, J. Zhang, W. Cao, J. Li, Y. Zheng, "When Will You Arrive? Estimating Travel Time Based on Deep Neural Networks," in *AAAI 2018*, pp. 2500-2507, New Orleans, Louisiana, USA, Feb. 2018.

[31] Gal, A., Mandelbaum, A., Schnitzler, F., Senderovich, A., Weidlich, M., "Traveling time prediction in scheduled transportation with journey segments," *Inf. Syst.*, vol. 64, pp. 266280, 2017.

[32] Jenelius, E., Koutsopoulos, H. N., 'Travel time estimation for urban road networks using low frequency probe vehicle data," *Transp. Res. Part B: Method.*, vol. 53, pp. 64-81, 2013.

[33] 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.

[34] Xu, J., Deng, D., Demiryurek, U., Shahabi, C., Van Der Schaar, M., "Mining the situation: Spatiotemporal traffic prediction with big data," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 4, pp. 702-715, 2015.