AI for Real-Time Bus Travel Time Prediction in Traffic Congestion Management

Yuming Ou

Abstract This chapter presents a methodology of bus travel time prediction, which is driven by the state-of-the-art machine learning technologies and involves real-time bus GPS location data collection and processing. Public transport plays a vital role in the development of our societies, providing the mobility to access to jobs, education, housing, services and recreation. Due to the rapid global urbanization trend, public transport suffers from the increasing traffic congestion and delay. The proposed methodology can predict bus travel time in real time to help mitigate the impact of traffic congestion by providing timely information of bus arrival time and delay. A case study of prediction of bus travel time in an area of Sydney has been carried out to evaluate our approach. The results show that our approach can effectively predict bus travel time and consistently outperforms the benchmark methods in a variety of scenarios. This research work demonstrates the power of AI technologies to promote productivity in traffic congestion management.

1 Introduction

1.1 Urbanization and Traffic Congestion

According to a report [37] by United Nations, in 1990 there were 2.3 billion people - 43% of the world's population - living in urban area. In 2018, the urban population has increased to 4.2 billion, which was 55% of the world's population. This urbanization trend is expected to continue. In 2050, the global urban population is projected to 6.7 billion. In other words, in mid-century, about 68% of the world's population will be living in urban area. The rapid urbanization brings opportunities

Y. Ou
Data Science Institute, University of Technology Sydney, Australia
e-mail: `yuming.ou@uts.edu.au`

as well as challenges to us. If managing it well by improving productivity and allowing innovation, we can benefit from the urbanization and enjoy the sustainable growth as more than 80% of global GDP is generated in cities. However, the urbanization also imposes the challenges to meet the accelerated demands such as affordable housing, more jobs and efficient transport systems.

Transport is vital to urban development. Transport systems provide essential mobility for citizens to access to jobs, education, housing, services and recreational facilities. Transport systems also move goods in the cities and significantly contribute to the economic growth. Urban growth and transport are strongly related to each other. Transport has a big influence on urban development. Efficient transport systems can attract more people and boost the urbanization. On the other hand, population growth can cause an increase in travel demand and thereby an increase in the need for transport infrastructure.

With the rapid urbanization trend, one critical problem in transport systems is traffic congestion. Traffic congestion has significantly negative impact on economy. It imposes additional costs to the communities and businesses by longer and less predictable travel times, reduces economic opportunities and lowers quality of life. It is shown that traffic congestion has been increasing over the world in the past decades. Effectively and efficiently managing traffic congestion is a pressing need for many cities.

1.2 Significance of Bus Travel Time Prediction

In the modern cities, public transport systems play the key role of moving people, increasing business productivity and improving air quality. It is the most popular transport means for commuters who regularly travel to work in the rush hour. Public transport can help riders avoid the stress that results from the daily driving in highly congested areas. Conveying more people in much less space than individual cars, public transport also helps to lower traffic congestion and reduce harmful air emission. Public transport provides an economical and environmentally friendly way of travel in cities.

However, public transport is also impacted by the congestion and suffering from traffic delay. In order to enhance the satisfaction of transit users and attract more people to use public transport, it is significant to improve public transport services, for example, by reducing delays and timely updating passengers with useful information when delays happen.

Timely and accurate bus travel time prediction is important to the public transport operations. It helps the transit operators to plan effective and robust schedules resulting in less congestion and delay. Early knowing the delay can enable transit operators to promptly respond and take action to the unexpected events. For the transit users, this type of information is also of importance. By keeping the passengers well-informed, the impact of delay and the consequent anxieties are largely relieved. The travellers can optimize their travel plans, mitigate traffic delay, and avoid traffic congestion as much as possible based on the up-to-date information. Therefore, the overall quality of transit services can be improved by providing such information to the transit users.

1.3 Research Problem

The research problem of this study is to predict bus travel time in real time. Bus travel time is the time for a bus to travel from one place to another place, which usually means bus stops. Technically, travel time, arrival time and delay have the same meaning in the context of public transport as any of them can be easily inferred by others. In this chapter, we use these three terms and they are interchangeable if not specified.

Bus travel time prediction has always been an active research topic over the past decades due to its importance to our real-life applications. With the advance of technologies, the methods for bus travel time prediction are progressing. Nowadays, Automatic Vehicle Location (AVL) systems have been widely adopted by many transit agencies, which make use of the Global Positioning System (GPS) automatically determining and transmitting the geographic location of a vehicle in a real-time fashion. This technology advance provides the transit agencies with an effective way to track their transit vehicles. Thanks for the AVL systems, a wealth of real-time information about the movements of vehicles is available and can be used for travel time prediction.

Artificial Intelligence (AI), particularly Machine Learning technologies can provide solutions to this problem. Utilising Machine Learning technologies, we can build prediction models on the historical vehicle movement data collected by the AVL systems, and then make predictions by feeding latest data into the models.

In this chapter, we propose an AI-based approach to address the research problem. The proposed approach is an end-to-end solution including real-time data retrieving and parsing, GPS data map matching, and travel time prediction. Our approach can be used in the systems that provide real-time bus arrival time and delay information.

1.4 Research Challenges

Accurately predicting bus travel time in real time is a very challenging task. Firstly, as the nature of transport systems, there are so many stochastic variables that can affect the travel time. For instance, travel speed fluctuates over time due to the ever change of traffic conditions. A broken vehicle or a major car incident can block a road and cause the congestion on the upstream road segments. Traffic signals can impact the traffic flow and cause intersection delays if they are not well configured. It is expected to take a longer travel time if the weather is bad.

Secondly, the dwell time depending on travel demand also affect the travel time. The unexpected surge of travel demand caused by events such as concerts and sports can largely increase the dwell time and result in delays. The stochastic passenger arrival at the the bus stops makes the prediction more difficult.

Thirdly, it requires a real-time prediction for providing timely bus arrival information. This means that the real-time data needs to be retrieved, processed and fed into the prediction model for quickly responding to the new situations within a short period of time window.

1.5 Organization of the Chapter

The rest of this chapter is organized as follows. The related work will be introduced in Section 2, in which five categories of methods for travel time prediction will be presented. Section 3 will propose the overall framework of bus travel time prediction. The data used in this study and the method to collect the data will be introduced in Section 4. Section 5 will present the problem formulation with a number of definitions, the approach to correct GPS location data, and our proposed method for travel time prediction. The case study that we have applied our approach to predict bus travel time in an area of Sydney will be introduced in Section 6. The discussion about the implication of our proposed approach to the humanity will be given in Section 7. Finally, Section 8 will conclude this chapter.

2 Related Work

Over the past decades, many pieces of research have been conducted to address the problem of bus travel time prediction due to its significance. A variety of approaches have been proposed, which can be categorized into the following five types: 1) historical average methods, 2) time series methods, 3) regression methods, 4) Kalman filter methods, and 5) machine learning methods. The five categories of approaches are introduced in the following subsections..

2.1 Historical Average Methods

As pointed out in [53], traffic conditions normally follow consistent daily and weekly patterns, which indicates that a reasonable forecast of future traffic conditions at a particular time of day and day of week can be given by the historical average of conditions at the same time of day and day of week. Based on this finding, historical average methods assume that the future traffic condition is consistent with previous journeys in the same time period and then predict the future travel time by observed historical travel time.

The basic idea of this type of methods is to find the previous journeys under similar traffic conditions and then use historical average travel time of previous journeys to predict the future travel time. The variation of this type of methods is the way to choose similar journeys. A naive approach is to simply use the journeys at the same location and in the same time period, which is usually used as a baseline method for benchmarking other methods. K-Nearest Neighbour (KNN) [8, 41, 44] is a popular approach to choose similar journeys, which select the K-nearest neighbours of previous journeys. However, determining the optimal size of nearest neighbours is very tricky. The size of nearest neighbours largely influences the prediction performance [8]. Apart from that, K-NN is computationally intensive if a large-scale number of historical journeys is present.

In general, the historical average methods are reliable only when the traffic patterns in the area of interest is relatively stable, such as the rural areas.

2.2 Time Series Methods

Time series methods [9, 43] assume that there is a pattern or a mixture of patterns in the historical time series data and the patterns will remain the same in the future time period. Based on the assumption, time series methods try to model the historical time series data by mathematical functions and used the mathematical functions to forecast the future.

In [9], a non-linear time series model is used to predict the travel time on a highway section in Orlando, Florida. In this study, two models including single-variable model and multiple-variable model have been built and compared with each other. Interestingly, the results showed that the single-variable model based on speed time series data outperformed the multiple-variable model based on speed, occupancy and volume time series data.

The accuracy of this type of methods highly depends on the fitness of the mathematical functions to model the historical data and the similarity between historical and real-time traffic patterns. Both the variation of historical data and changes of real-time from historical traffic patterns can largely impact the accuracy of the prediction results.

2.3 Regression Methods

Regression methods assume that the bus arrival time is an output of a function of different variables such as traffic circumstances, number of passengers, number of bus stops, and climate situations. Therefore, this type of methods uses a mathematical function to describe the relationship between the dependent variable - travel time - with a set of independent variables.

There were many studies [5, 24, 27, 36, 40] that used regression models for bus travel time prediction. The major difference is the independent variables used for building the regression models. One of advantages of regression methods is that the importance of each independent variable to the dependent variable can be known by the built regression models. For example, in a study [36] a set of multiple linear regression models has been developed using independent variables including distance, number of stops, dwell time, boarding and alighting passengers and weather to predict bus arrival time. According to the results, weather is less important than other inputs in the models.

The major limit of regression methods is that variables in transport systems are likely to be inter-correlated rather than completely independent [6].

2.4 Kalman Filter Methods

Originated from the state-space representations in modern control theory, Kalman filter is a recursive procedure that estimates the future states of dependent variables. It is introduced to travel time prediction because of its advantage in continuously updating the state variable using new observations [6].

Many studies based on Kalman filter algorithm have been reported for travel time prediction [7, 26, 34, 47, 49]. For instance, Chu et al. [7] developed a method for travel time estimation based on Kalman filtering. The proposed method can dynamically estimate noise statistics of the system by adapting to the new observations. The Kalman filtering based algorithm was evaluated under recurrent and non-recurrent traffic congestion conditions. The results showed that the proposed method outperformed the benchmark method for both situations. In Yang's study [26], a discrete-time Kalman filter was used to predict arterial travel time in the scenarios of special events such as graduation ceremony.

2.5 Machine Learning Methods

Machine Learning [31] is a branch of artificial intelligence which is based on the idea that systems can learn from data and make decisions. It focuses on studying computer algorithms that build models based on historical data and improve the models automatically through experience. Typically building a machine model consists of four phases including 1) preparing training data set, 2) choosing a candidate algorithm, 3) training a model by the selected algorithm on the training data set, and 4) using and improving the model.

Support Vector Machine (SVM) is one of popular Machine Learning algorithms that are reported in the literature of bus travel time prediction. SVM uses kernel functions to find a hyperplane or set of hyperplanes that can be used for classification, regression or outliers detection. Yu et al. [2, 56] used SVM to predict bus arrival time by considering the segment-level travel time and four traffic conditions including peak time and sunny day, off-peak time and sunny day, peak time and rainy day, and off-peak time and rainy day. The model used three inputs consisting of segment, travel time of current segment, and the latest travel time of next segment, to output the predicted travel time.

Artifical Neural Network (ANN) [1] is also a popular Machine Learning algorithm for bus travel time prediction. ANN is inspired by biological neural networks, in which there are multiple layers of processing units called artificial neurons. Each neurons has an activation function and is connected with other neurons. The connection between two neurons means the output of a neuron as the input of another neuron. Each connection is assigned a weight which represents its importance. Through learning process, the initial weights are adjusted to capture the relationship between inputs and outputs usually by backpropagation algorithm [28]. ANN-based methods have gained popularity in predicting bus travel time because of their ability to solve complex non-linear relationships [5, 6, 25, 38] . For example, Ramakrishna et al. [38] developed a Multiple Layer Perceptron (MLP) for predicting bus travel time using vehicle speed data and passenger data, which achieved better performance over the linear regression approach. In the study of Jeong et al. [25], the ANN model outperformed both historical data model and regression model in predicting bus travel time using actual vehicle location data in Houston, Texas.

One advantage of Machine Learning methods is that they can deal with large volume of data sets. Another advantage is that they can discover the complex relationships between predictors, such as non-linear relationships. The ability to tolerate noisy data is also an advantage of Machine Learning methods.

3 Proposed System Framework

The proposed system framework of bus travel time prediction is illustrated in Fig. 1, in which each parallelogram represents a function component, each rectangle is the output of each function component, solid line stands for the process of model training and dashed line stands for the prediction process. There are four major function components which are described as follows.

- *Real Time GTFS Data Collection*: the component to collect real-time GTFS data which will be introduced in Section 4 of this chapter.
- *GPS Data Correction*: the component to correct GPS data points and match them to road segments, which we will introduce in Subsection 5.2 of this chapter.
- *Model Training*: the component to train a prediction model using historical GPS data, which will be introduced in Subsection 5.3 of this chapter.
- *Model Prediction*: the component to predict bus travel time using the trained prediction model, which will be introduced in Subsection 5.3 of this chapter.

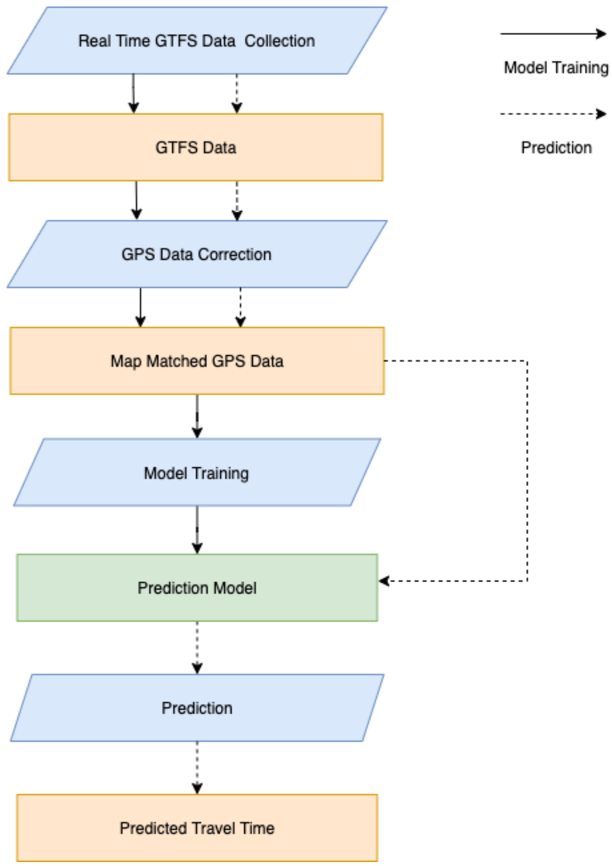

Fig. 1 *Framework of Travel Time Prediction*: Each parallelogram represents a function component, each rectangle is the output of each function component, solid line stands for the process of model training and dashed line stands for the prediction process. Best viewed in color.

4 Real Time Data Collection

This section introduces the data used for this research work and the workflow to collect the real-time data through RESTful data APIs.

4.1 GTFS Data

The General Transit Feed Specification (GTFS) [16] defines a common data format to allow public transit agencies to publish their transit data so that the data can be consumed by various applications. Generally, GTFS is divided into two streams inlcuding GTFS static and GTFS real-time. The former contains public transportation schedules and associated geographic information while the latter contains the real-time vehicle positions and all trip updates.

GTFS has been used as an industry standard for majority of transit agencies to publish their transit data around the world [17]. As GTFS data contains both scheduled and real-time information about transit operations, it has been actively used for many research problems such as transit accessibility [11, 10, 14, 19, 35], transit network analysis [20, 51], performance evaluation [4, 50], delay prediction [45, 46, 55], and transit trip inference [32, 57].

4.2 GTFS Data Collection

The data used for this study is the GTFS data published by the local transport agency: Transport for NSW [33]. We collect the following three data sets.

- *Real-time bus position data*: the real-time buses' movements with longitudes, latitudes, and associated time stamps. The real-time bus positions are captured by the GPS devices mounted on the buses. There are always errors associated with the GPS data. We need to correct the GPS data by a map matching algorithm which we introduce in Subsection 5.2 of this chapter.
- *Bus timetable data*: containing the scheduled bus trips and scheduled arrival times at bus stops.
- *Bus network data*: containing the geolocations of all bus stops and the physical geometry of the bus routes.

As the bus position data is published in real time, we need to develop a data collection service for continuously collecting the data through RESTful data APIs. Fig. 2 illustrates the workflow of data collection. To collect the real-time bus position data, the data collection service sends a data pulling request to the data APIs every 5 seconds. After receiving the data returned from the data API the service then parses the data and checks whether it is exactly the same to the previous data points. If so then it discards the data, otherwise it appends the data to stored data files. The purpose of removing the duplicate records is to save space as well as to reduce the computation cost in the following step of data processing. In the entire Sydney metropolitan region there are around 24,000 bus stops and more than 25,000 bus trips are being scheduled during a 24-hour day, which leads to more than 3GB of real-time bus position data being collected every day. Apart from the bus position data, the data collection service also collects timetable and network data daily in a similar fashion, in order to have up-to-date timetable and network data.

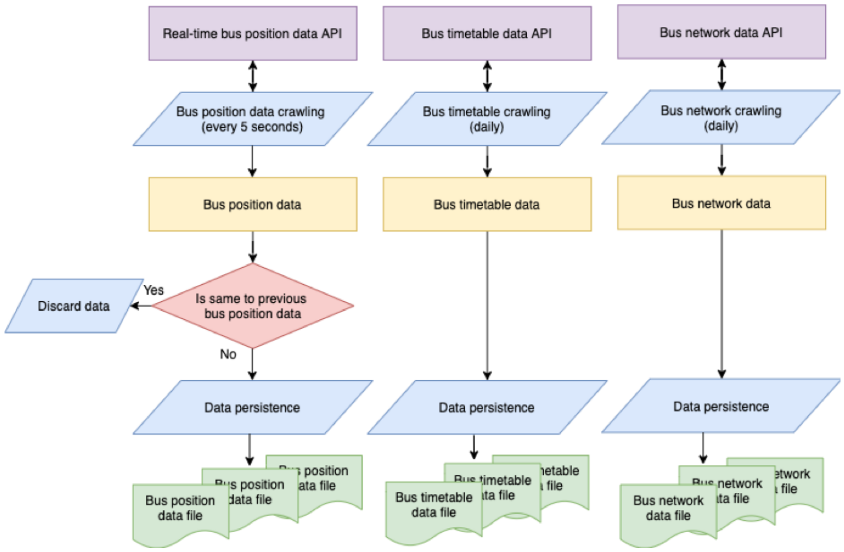

Fig. 2 *Workflow of data collection service*: The data collection service sends a data pulling request to the real-time bus position data API every 5 seconds. After receiving the data returned from the data API the service then parses the data and checks whether it is exactly the same to the previous data points. If so then it discards the data, otherwise it appends the data to stored data files. Apart from the bus position data, the data collection service also collects timetable and network data daily in a similar fashion.

5 Methodology

5.1 Problem Formulation

In this subsection, we first give the definitions of $Road\ Segment$, $Route$, $Bus\ Stop$ and $Trip$. On top of the definitions, we then propose the equation to calculate the travel time between two bus stops. Finally, we formulize the research problem of travel time prediction.

Definition 1. (Road Segment): A road segment seg is a portion of the road between two consecutive bus stops, which is represented by a tuple consisting of segment ID id and its length l.

$$seg = (id, l) \tag{1}$$

Definition 2. (Route): A route r is a vector of road segments from the origin bus stop to the destination bus stop,

$$r = [seg_1, seq_2, ..., seg_i, ..., seg_n] \tag{2}$$

in which seg_i is the i^{th} road segment of the route r and n is the total number of road segments of the route r.

Definition 3. (Bus Stop): A bus stop $stop$ is the end point of a road segment and is also the starting point of the successive road segment. There is a mapping function f for returning a road segment for a given bus stop:

$$f : stop_j \mapsto seg_i \tag{3}$$

in which $stop_j$ is the starting point of seg_i. Obviously, $stop_j$ is also the end point of seg_{i-1}.

Definition 4. (Trip): A trip $trip$ contains the information about the segments that the bus travels and their corresponding travel times. It is a vector of tuples consisting of road segment and corresponding travel time and time stamp,

$$trip = [(seg_1, tt_1, ts_1), (seg_2, tt_2, ts_2), ..., (seg_i, tt_i, ts_i), ..., (seg_n, tt_n, ts_n)] \tag{4}$$

in which tt_i is the travel time on road segment seg_i and ts_i is the time stamp that $trip$ starts to travel on seg_i

Based on the above definitions, we have the following theorem for calculating travel time between two bus stops.

Theorem 1. *(Travel Time between Two Bus Stops) The travel time between two bus stops for a given trip is the sum of corresponding travel times of road segments that the trip travels between the two bus stops:*

$$tts^{jk} = \sum_{i=g(f(stop_j))}^{g(f(stop_k))-1} tt_i \tag{5}$$

in which tts^{jk} is the travel time from $stop_j$ to $stop_k$, and g is a function returning the sequence of a road segment:

$$g : seg_i \mapsto i \tag{6}$$

Definition 5. (Travel Time Prediction): The research problem of this study is to build a model Θ that predicts the road segment travel times so that the travel time between two bus stops for a given trip can be calculated by above Eq. 5:

$$\Theta : (trip^1, ..., trip^{m-1}, seg_i, stop_j, t_j, stop_k) \mapsto \hat{tt}_i^m \tag{7}$$

in which $\hat{tt}_i^m$ is the predicted travel time for m^{th} trip on road segment seg_i, $trip^1$, ..., $trip^{m-1}$ are the previous trrips that have passed seg_i, $stop_j$ is the last bus stop that the m^{th} trip has passed, t_j is the arrival time at $stop_j$, and the $stop_k$ is the next bus stop that the m^{th} trip will arrive at.

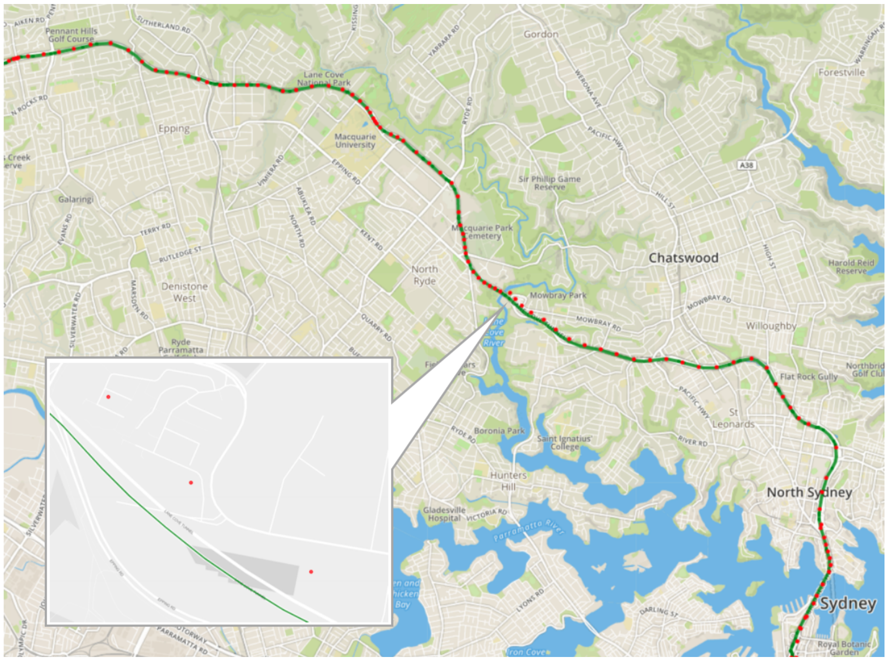

Fig. 3 *An example of GPS errors*: The red dots are the GPS data points sent from the GPS device on a bus while the green line is the actual bus trajectory along the main road. It can be observed that many GPS locations are falling further away from the green line (road centre line) instead of exactly being on it. Best viewed in color.

5.2 GPS Data Correction

Due to the well-known issue of GPS accuracy [54], the GPS data is always associated with an error which is a deviation from what the real position of the bus vehicle is. The errors are variable depending on the circumstances, the road network geometry layout and continuity of data transmission in real-time. Many other sources could contribute to GPS errors, such as clock error, signal jamming, weather and building blocking. An example of GPS errors is shown is Fig. 3 in which the red dots are the GPS data points sent from the GPS device on a bus while the green line is the actual bus trajectory along the main road. It can be observed that many GPS locations are falling further away from the green line (road centre line) instead of exactly being on it. Consequently, before using the bus GPS data to train the prediction model, we need to correct the GPS data through map matching algorithms by matching every GPS coordinate transmitted by the bus to a correct location on the road centreline.

There are various methods that have been used in the literature for map matching [3, 21, 30, 52]. One native way is the point-to-curve method, which projects GPS points to their closest edges. This method is simplistic and lacks robustness especially when the road network has a complicated structure such as in the CBD

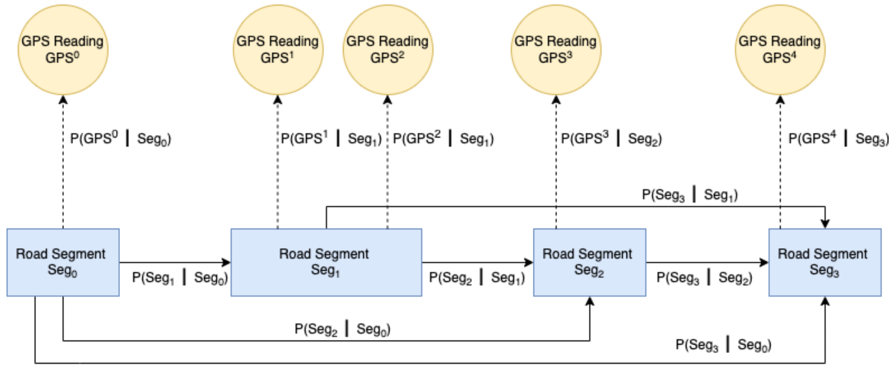

Fig. 4 *Hidden Markov Model for Map Matching*: The blue rectangles are the hidden states of the road segments on which the bus is while the yellow circles are the observations of the GPS readings. Best viewed in color.

areas. An improved method is the curve-to-curve method which considers the closeness and similarity between the curve formed by GPS points and the candidate path. However, it still has the same problems under the circumstances of large GPS errors and complicated overlayed networks. Other approaches include using the geometry and topology of the road network [42], Kalman filters [29], and Fuzzy rules [39].

To achieve a high accuracy of GPS data correction in real-time, our map matching method is based on a Hidden Markov Model (HMM) [12, 48]. HMMs usually model a system by considering their unobserved states and their observations. In the system one hidden state can change to any other hidden state by following a state transition probability. Instead of the hidden states, one can observe the values generated from the hidden states with emission probabilities. In this work, we model the road segments on which the bus is as the hidden states and the GPS readings as the observations as shown in Fig. 4. Under this setting, the emission probability is defined in the following Eq. 8,

$$P(GPS^t|seg_i^t) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{gd(GPS^t,GPS_i^t)^2}{2\sigma^2}} \tag{8}$$

in which GPS^t is the bus GPS reading at time t, seg_i^t is the road segment i that the bus is on at time t, GPS_i^t is the projection of GPS^t on seg_i^t, gd is the great circle distance between two geolocations, and σ is the stand deviation of the GPS device error.

Furthermore, the transition probability is defined in the following Eq. 9,

$$P(seg_j^{t+1}|seg_i^t) = \frac{gd(GPS^t, GPS^{t+1})}{rd(GPS_i^t, GPS_j^{t+1})} \tag{9}$$

in which rd is the distance between two geolocations along the road segment path.

Given a sequence of GPS readings as the observations, we can utilize the Viterbi algorithm [13] to find out the most likely sequence of road segments as the hidden states.

5.3 LSTM-Based Travel Time Prediction

Our approach to predict the arrival time at next bus stop is composed of two steps including 1) predicting the travel time for each segment that is between current location to the next bus stop, 2) and then summing up the travel times for all above segments. One advantage of our approach is that predicting the segment-based travel time can capture the characteristics of each segment at a finer level of granularity than directly predicting the travel time from current location to the next bus stop as a whole. Another advantage of our approach over the method of simply predicting travel time between two bus stops is that it can be used for real-time prediction. Our approach can keep updating the prediction when a bus is travelling by updating the bus' location.

In order to predict the travel time on a road segment, we build a model based on Long Short-Term Memory (LSTM) network [15, 18, 23]. LSTM networks are a type of Recurrent Neural Network (RNN) which are well-suitable for time series data. LSTM networks are improved for dealing with the issue of vanishing gradient [22] that the traditional RNNs usually suffer from. When the gradient values become extremely small during the training of RNNs, the weights are prevented from changing their values and the neural networks stop further learning. LSTM networks overcome the vanishing gradient problem by using a mechanism called gates to control the information flow into and out the memory of the network.

Fig. 5 shows the LSTM unit that is used in our approach. It consists of a cell which is the memory of the network and three gates including forget gate, input gate and output gate.

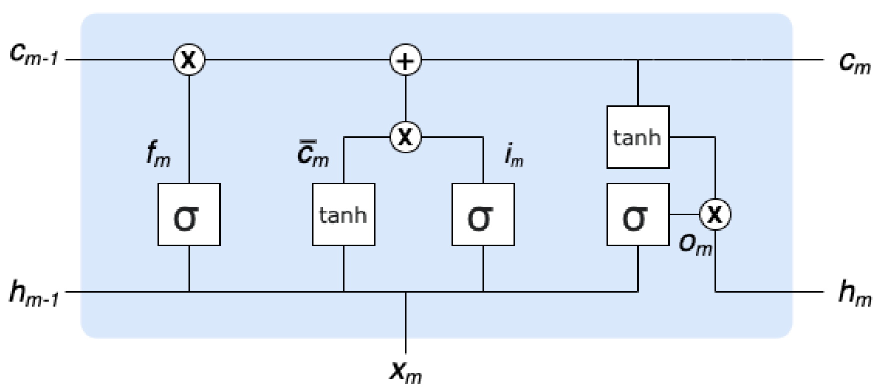

Fig. 5 *Long Short-Term Memory Unit*: The structure of a LSTM unit

The sequence data set Ψ used for training the LSTM network for a road segment is defined as follows:

$$\Psi = \{..., (X_m, tt_{m+1}), ...\} \tag{10}$$

$$X_m = (x_{m-n},\ x_{m-n+1}, ...,\ x_m) \tag{11}$$

$$x_m = [tt_m, s_m, stop_j, t_j] \tag{12}$$

in which tt_m is the travel time of the m^{th} trip $trip^m$ on the road segment, s_m is the seconds from midnight derived from the time stamp that $trip^m$ starts to travel on the road segment, $stop_j$ is the last bus stop $trip^m$ has passed, t_j is the arrival time at bus stop $stop_j$, and n is the length of a sequence.

After building the sequence data set Ψ, we train the LSTM network by the following equations.

$$f_m = \sigma(W_{xf}x_m + W_{hf}h_{m-1} + b_f) \tag{13}$$

$$i_m = \sigma(W_{xi}x_m + W_{hi}h_{m-1} + b_i) \tag{14}$$

$$\bar{c}_m = \phi(W_{xc}x_m + W_{hc}h_{m-1} + b_c) \tag{15}$$

$$c_m = \bar{c}_m \odot i_m + f_m \odot c_{m-1} \tag{16}$$

$$o_m = \sigma(W_{xo}x_m + W_{ho}h_{m-1} + b_o) \tag{17}$$

$$h_m = o_m \odot \phi(c_m) \tag{18}$$

in which f_m is the forget gate, i_m is the input gate, $\bar{c}_m$ is the cell input, c_m is the cell state, o_m is the output gate, h_m is the output, σ is the sigmoid activation function, ϕ is the tanh activation function, W_* is the weight matrices, and b_* is the bias vectors.

6 Case Study

6.1 Case Study Setting

Our proposed methodology has been applied to an area of Sydney to predict bus travel time. The area for our case study is shown in Fig. 6. We focused on the road segments that are highlighted in blue. There are multiple bus routes which are operating on these road segments. In total, there are sixteen bus stops on the road segments as represented by the purple dots. Fourteen bus stops including the stops from the first to the fourteenth are on a main road and the remaining two bus stops are on a motorway. One major reason why we choose these road segments is because part of them is on a main road and part of them is on a motorway. We can test our method performance for both types of road.

We collected the GTFS real-time data using the method introduced in Subsection 4.1 of this chapter. The data covers six months of real-time bus GPS location data in the study area. There are more than 2.1 million GPS data points generated from

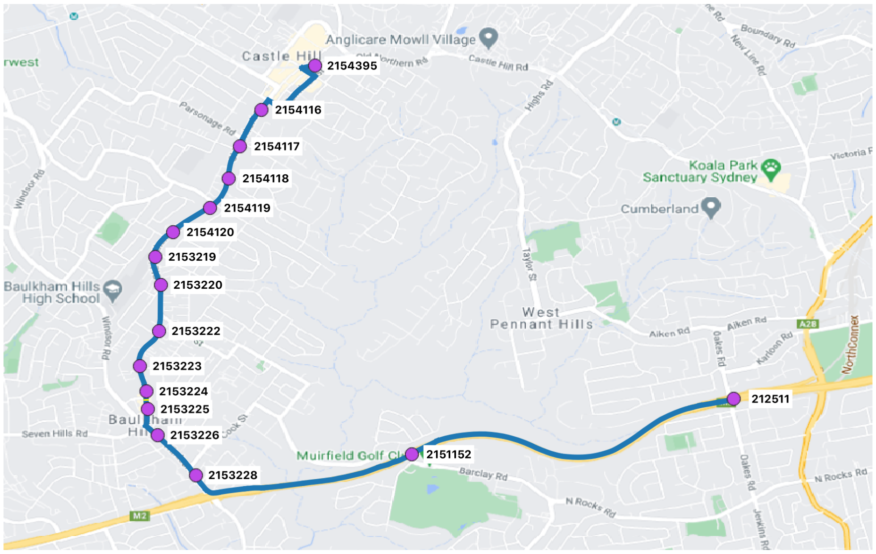

Fig. 6 *Area for the case study*: Blue line represents the road segments while the purple dots stand for the bus stops. The numbers beside the bus stops are the bus stop IDs. Best viewed in color.

37,622 bus trips from May 2019 to October 2019. We used the method presented in Subsection 5.2 of this chapter to correct the GPS data by map-matching them to the corresponding locations on the road segments.

The six-month data was split into a training data set for training the model and a test data set for evaluating the model's performance. They cover four months and two months of time period respectively. We evaluated our approach against other three methods including Moving Average, Linear Regression and Support Vector Machine.

6.2 Experimental Results

Three metrics were used for evaluating the model performance, including Mean Absolute Error (MAE), Symmetric Mean Absolute Percentage Error (SMAPE), and Root Mean Squared Error (RMSE). They are defined by the following equations respectively.

$$MAE = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N-1} |\hat{tts}_{mn} - tts_{mn}|}{M(N-1)} \tag{19}$$

$$SMAPE = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N-1} \frac{|\hat{tts}_{mn} - tts_{mn}| \times 2}{|\hat{tts}_{mn}| + |tts_{mn}|}}{M(N-1)} \times 100\% \tag{20}$$

$$RMSE = \sqrt{\frac{\sum_{m=1}^{M}\sum_{n=1}^{N-1}(tt\hat{s}_{mn} - tts_{mn})^2}{M(N-1)}} \tag{21}$$

in which M is the total number of bus trips and N is the total number of GPS data points for a bus trip, $tt\hat{s}_{mn}$ is the prediction of travel time in minutes for the n^{th} GPS data point of m^{th} bus trip, and tts_{mn} is the corresponding actual travel time in minutes.

We used the training data set to train a model following the approached proposed in Section 5 of this chapter, and then used the test data set to evaluate the model. We compared our approach with three benchmark methods including Moving Average (MA), Liner Regression (LR), Support Vector Regression (SVR), using the above evaluation metrics. Table 1 provides the evaluation results, which shows that our approach outperforms the other methods for all evaluation metrics.

Table 1 Comparison of prediction errors for four models

Methods	MAE (min)	SMAPE (%)	RMSE (min)
MA	0.72	19.42	0.94
LR	0.55	18.32	0.86
SVR	0.54	18.77	0.84
Our approach	0.50	17.37	0.72

Table 2 Comparison of prediction errors for four models (weekday vs weekend)

	Methods	MAE (min)	SMAPE (%)	RMSE (min)
Weekday	MA	0.83	19.19	0.97
	LR	0.59	17.91	0.92
	SVR	0.55	17.97	0.89
	Our approach	0.51	17.31	0.74
Weekend	MA	0.59	19.55	0.81
	LR	0.51	18.54	0.72
	SVR	0.48	18.82	0.73
	Our approach	0.46	17.65	0.64

To further investigate the performance of our approach, we compared it with the three benchmark methods in different scenarios. The first scenario is that we divided the test data set into two data sets for weekday and weekends respectively and used them to evaluate the methods. The second scenario is to split the time period from 6AM to 22PM into three parts including morning peak hours (from 6:30AM to 10AM), afternoon peak hours (from 3PM to 7PM) and non-peak hours

Table 3 Comparison of prediction errors for four models (weekday peak hours vs weekday non-peak hours)

	Methods	MAE (min)	SMAPE (%)	RMSE (min)
Morning Peak Hours	MA	0.77	19.39	1.02
	LR	0.59	18.27	0.91
	SVR	0.57	18.16	0.87
	Our approach	0.53	16.31	0.78
Non-peak Hours	MA	0.69	18.95	0.88
	LR	0.51	17.75	0.79
	SVR	0.52	17.79	0.77
	Our approach	0.46	16.64	0.65
Afternoon Peak Hours	MA	0.57	19.72	0.97
	LR	0.56	19.47	0.93
	SVR	0.55	19.34	0.84
	Our approach	0.51	19.16	0.73

Table 4 Comparison of prediction errors for four models (motorway vs non-motorway)

	Methods	MAE (min)	SMAPE (%)	RMSE (min)
Motorway	MA	1.01	18.41	1.20
	LR	0.92	17.87	1.14
	SVR	0.93	18.01	1.12
	Our approach	0.89	13.00	1.13
Non-motorway	MA	0.67	20.13	0.89
	LR	0.51	19.19	0.81
	SVR	0.47	19.23	0.77
	Our approach	0.40	18.57	0.56

(the remaining). The third is to evaluate the methods in the scenario that the bus stops are on a motorway. The evaluation results for the above three scenarios are given in Table 2, 3, and 4 respectively, which show that our approach consistently beat other methods.

7 Discussion

Artificial Intelligence is regarded as one of the most revolutionary developments in human history. Nowadays we are witnessing its transformative power. There are so many AI-based cutting-edge solutions solving the most critical challenges faced by the society.

The research work presented in this book chapter is one of the examples - AI technologies are used in solving the challenging problem of bus travel time prediction. The proposed AI-based solution can process large amount of vehicle movement data and predict bus travel time in real time, which helps manage the critical issue of traffic congestion. It demonstrates that AI technologies can largely improve the efficiency of our workplace and empower high-performance organisations, governments and communities.

8 Conclusion

In this chapter, we study the research problem of bus travel prediction which is significant to our societies as it helps to improve our daily lives. In order to address this research problem, we propose an approach to predict bus travel time using real-time bus GPS location data. The proposed method involves real-time data collection and processing and adopts the state-of-the-art machine learning technologies.

To verify our approach, a case study was carried out to predict the bus travel time in an area of Sydney. In the case study, three benchmark methods are used to compare with our approach. The evaluation results based on three evaluation metrics show that our proposed approach consistently outperforms the three benchmark methods in a variety of scenarios. In the future, we will further improve our approach by applying Graph Neural Networks.

The proposed method in this chapter can support traffic congestion management by providing the information of real-time bus arrival time and delay. This information can help not only transport operators proactively manage traffic congestion and take actions for mitigating the impact of delay, but also commuters better schedule their travel plans accordingly. The research work shows the power of AI technologies to promote productivity in traffic congestion management.

References

1. Agatonovic-Kustrin, S., Beresford, R.: Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis **22**(5), 717–727 (2000)
2. Bin, Y., Zhongzhen, Y., Baozhen, Y.: Bus arrival time prediction using support vector machines. Journal of Intelligent Transportation Systems **10**(4), 151–158 (2006)
3. Brakatsoulas, S., Pfoser, D., Salas, R., Wenk, C.: On map-matching vehicle tracking data. In: Proceedings of the 31st international conference on Very large data bases, pp. 853–864 (2005)
4. Braz, T., Maciel, M., Mestre, D.G., Andrade, N., Pires, C.E., Queiroz, A.R., Santos, V.B.: Estimating inefficiency in bus trip choices from a user perspective with schedule, positioning, and ticketing data. IEEE Transactions on Intelligent Transportation Systems **19**(11), 3630–3641 (2018)
5. Chen, M., Liu, X., Xia, J., Chien, S.I.: A dynamic bus-arrival time prediction model based on apc data. Computer-Aided Civil and Infrastructure Engineering **19**(5), 364–376 (2004)

6. Chien, S.I.J., Ding, Y., Wei, C.: Dynamic bus arrival time prediction with artificial neural networks. Journal of transportation engineering **128**(5), 429–438 (2002)
7. Chu, L., Oh, J.S., Recker, W.: Adaptive kalman filter based freeway travel time estimation (2005)
8. Coffey, C., Pozdnoukhov, A., Calabrese, F.: Time of arrival predictability horizons for public bus routes (2011). DOI 10.1145/2068984.2068985
9. D'Angelo, M.P., Al-Deek, H.M., Wang, M.C.: Travel-time prediction for freeway corridors. Transportation Research Record **1676**(1), 184–191 (1999)
10. Farber, S., Fu, L.: Dynamic public transit accessibility using travel time cubes: Comparing the effects of infrastructure (dis) investments over time. Computers, Environment and Urban Systems **62**, 30–40 (2017)
11. Farber, S., Morang, M.Z., Widener, M.J.: Temporal variability in transit-based accessibility to supermarkets. Applied Geography **53**, 149–159 (2014)
12. Fine, S., Singer, Y., Tishby, N.: The hierarchical hidden markov model: Analysis and applications. Machine learning **32**(1), 41–62 (1998)
13. Forney, G.D.: The viterbi algorithm. Proceedings of the IEEE **61**(3), 268–278 (1973)
14. Fransen, K., Neutens, T., Farber, S., De Maeyer, P., Deruyter, G., Witlox, F.: Identifying public transport gaps using time-dependent accessibility levels. Journal of Transport Geography **48**, 176–187 (2015)
15. Gers, F.A., Schmidhuber, J., Cummins, F.: Learning to forget: Continual prediction with lstm (1999)
16. Google: Google transit apis. URL `https://developers.google.com/transit`
17. Google: Public feeds wiki. URL `https://code.google.com/archive/p/googletransitdatafeed/wikis/PublicFeeds.wiki`
18. Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J.: Lstm: A search space odyssey. IEEE transactions on neural networks and learning systems **28**(10), 2222–2232 (2016)
19. Guthrie, A., Fan, Y., Das, K.V.: Accessibility scenario analysis of a hypothetical future transit network: social equity implications of a general transit feed specification–based sketch planning tool. Transportation research record **2671**(1), 1–9 (2017)
20. Hadas, Y.: Assessing public transport systems connectivity based on google transit data. Journal of Transport Geography **33**, 105–116 (2013)
21. Hashemi, M., Karimi, H.A.: A critical review of real-time map-matching algorithms: Current issues and future directions. Computers, Environment and Urban Systems **48**, 153–165 (2014)
22. Hochreiter, S.: The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **6**(02), 107–116 (1998)
23. Huang, Z., Xu, W., Yu, K.: Bidirectional lstm-crf models for sequence tagging. arXiv preprint arXiv:1508.01991 (2015)
24. Jeong, R., Rilett, R.: Bus arrival time prediction using artificial neural network model. In: Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749), pp. 988–993 (2004). DOI 10.1109/ITSC.2004.1399041
25. Jeong, R., Rilett, R.: Bus arrival time prediction using artificial neural network model. In: Proceedings. The 7th international IEEE conference on intelligent transportation systems (IEEE Cat. No. 04TH8749), pp. 988–993. IEEE (2004)
26. Jiann-Shiou Yang: Travel time prediction using the gps test vehicle and kalman filtering techniques. In: Proceedings of the 2005, American Control Conference, 2005., pp. 2128–2133 vol. 3 (2005). DOI 10.1109/ACC.2005.1470285
27. Kwon, J., Coifman, B., Bickel, P.: Day-to-day travel-time trends and travel-time prediction from loop-detector data. Transportation Research Record **1717**(1), 120–129 (2000)
28. Leung, H., Haykin, S.: The complex backpropagation algorithm. IEEE Transactions on signal processing **39**(9), 2101–2104 (1991)
29. Li, L., Quddus, M., Zhao, L.: High accuracy tightly-coupled integrity monitoring algorithm for map-matching. Transportation Research Part C: Emerging Technologies **36**, 13–26 (2013)

30. Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., Huang, Y.: Map-matching for low-sampling-rate gps trajectories. In: Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems, pp. 352–361 (2009)
31. Mitchell, T.M., et al.: Machine learning (1997)
32. Nassir, N., Khani, A., Lee, S.G., Noh, H., Hickman, M.: Transit stop-level origin–destination estimation through use of transit schedule and automated data collection system. Transportation research record **2263**(1), 140–150 (2011)
33. Transport for NSW, A.: Transport for nsw's open data portal. URL `https://opendata.transport.nsw.gov.au/`
34. Okutani, I., Stephanedes, Y.J.: Dynamic prediction of traffic volume through kalman filtering theory. Transportation Research Part B: Methodological **18**(1), 1–11 (1984)
35. Owen, A., Levinson, D.M.: Modeling the commute mode share of transit using continuous accessibility to jobs. Transportation Research Part A: Policy and Practice **74**, 110–122 (2015)
36. Patnaik, J., Chien, S., Bladikas, A.: Estimation of bus arrival times using apc data. Journal of public transportation **7**(1), 1 (2004)
37. Publications, U.N.: World Urbanization Prospects: The 2018 Revision. UN (2019). URL `https://books.google.com.au/books?id=Kp9AygEACAAJ`
38. Ramakrishna, Y., Ramakrishna, P., Lakshmanan, V., Sivanandan, R.: Use of gps probe data and passenger data for prediction of bus transit travel time. In: Transportation Land Use, Planning, and Air Quality, pp. 124–133 (2008)
39. Ren, M., Karimi, H.A.: Movement pattern recognition assisted map matching for pedestrian/wheelchair navigation. The journal of navigation **65**(4), 617–633 (2012)
40. Shalaby, A., Farhan, A.: Bus travel time prediction model for dynamic operations control and passenger information systems. Transportation Research Board **2** (2003)
41. Sinn, M., Yoon, J.W., Calabrese, F., Bouillet, E.: Predicting arrival times of buses using real-time gps measurements. In: 2012 15th International IEEE Conference on Intelligent Transportation Systems, pp. 1227–1232. IEEE (2012)
42. Srinivasan, D., Cheu, R.L., Tan, C.W.: Development of an improved erp system using gps and ai techniques. In: Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems, vol. 1, pp. 554–559. IEEE (2003)
43. Stephanedes, Y.J., Kwon, E., Michalopoulos, P.: On-line diversion prediction for dynamic control and vehicle guidance in freeway corridors. 1287 (1990)
44. Sun, D., Luo, H., Fu, L., Liu, W., Liao, X., Zhao, M.: Predicting bus arrival time on the basis of global positioning system data. Transportation Research Record **2034**(1), 62–72 (2007)
45. Sun, F., Dubey, A., Samal, C., Baroud, H., Kulkarni, C.: Short-term transit decision support system using multi-task deep neural networks. In: 2018 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 155–162. IEEE (2018)
46. Sun, F., Pan, Y., White, J., Dubey, A.: Real-time and predictive analytics for smart public transportation decision support system. In: 2016 IEEE International Conference on Smart Computing (SMARTCOMP), pp. 1–8. IEEE (2016)
47. Vanajakshi, L., Subramanian, S.C., Sivanandan, R.: Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses. IET intelligent transport systems **3**(1), 1–9 (2009)
48. Varga, A., Moore, R.: Hidden markov model decomposition of speech and noise. In: International Conference on Acoustics, Speech, and Signal Processing, pp. 845–848. IEEE (1990)
49. Wall, Z.R.: An algorithm for predicting the arrival time of mass transit vehicles using automatic vehicle location data. Ph.D. thesis, University of Washington (1998)
50. Wessel, N., Allen, J., Farber, S.: Constructing a routable retrospective transit timetable from a real-time vehicle location feed and gtfs. Journal of Transport Geography **62**, 92–97 (2017)
51. Wessel, N., Widener, M.J.: Discovering the space–time dimensions of schedule padding and delay from gtfs and real-time transit data. Journal of Geographical Systems **19**(1), 93–107 (2017)

52. White, C.E., Bernstein, D., Kornhauser, A.L.: Some map matching algorithms for personal navigation assistants. Transportation research part c: emerging technologies **8**(1-6), 91–108 (2000)
53. Williams, B., Hoel, L.: Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. Journal of Transportation Engineering **129**, 664–672 (2003). DOI 10.1061/(ASCE)0733-947X(2003)129:6(664)
54. Williams, S.D., Bock, Y., Fang, P., Jamason, P., Nikolaidis, R.M., Prawirodirdjo, L., Miller, M., Johnson, D.J.: Error analysis of continuous gps position time series. Journal of Geophysical Research: Solid Earth **109**(B3) (2004)
55. Wu, J., Zhou, L., Cai, C., Dong, F., Shen, J., Sun, G.: Towards a general prediction system for the primary delay in urban railways. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 3482–3487. IEEE (2019)
56. Yu, B., Yang, Z.Z., Wang, J.: Bus travel-time prediction based on bus speed. In: Proceedings of the Institution of Civil Engineers-Transport, vol. 163, pp. 3–7. Thomas Telford Ltd (2010)
57. Zahabi, S.A.H., Ajzachi, A., Patterson, Z.: Transit trip itinerary inference with gtfs and smartphone data. Transportation Research Record **2652**(1), 59–69 (2017)