IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 19, 2019, accepted January 5, 2020, date of publication January 9, 2020, date of current version January 21, 2020.
Digital Object Identifier 10.1109/ACCESS.2020.2965094

Bus Arrival Time Prediction Based on LSTM and Spatial-Temporal Feature Vector

HONGJIE LIU, HONGZHE XU, YU YAN, ZAISHANG CAI, TIANXU SUN, AND WEN LI
School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
Corresponding author: Hongjie Liu (hj_popel@stu.xjtu.edu.cn)

This work was supported in part by the China Scholarship Council under Grant 201906280274.

ABSTRACT Bus arrival prediction has important implications for public travel, urban dispatch, and mitigation of traffic congestion. The factors affecting urban traffic conditions are complex and changeable. As the predicted distance increases, the difficulty of traffic prediction becomes more difficult. Forecast based on historical data responds quite slowly for changes under the short-term conditions, and vehicle prediction based on real-time speed is not sufficient to predict under long-term conditions. Therefore, an arrival prediction method based on temporal vector and another arrival prediction method based on spatial vector is proposed to solve the problems of remote dependence of bus arrival and road incidents, respectively. In this paper, combining the advantages of the two prediction models, this paper proposes a long short-term memory (LSTM) and Artificial neural networks (ANN) comprehensive prediction model based on spatial-temporal features vectors. The long-distance arrival-to-station prediction is realized from the dimension of time feature, and the short-distance arrival-to-station prediction is realized from the dimension of spatial feature, thereby realizing the bus-to-station prediction. Besides, experiments were conducted and tested based on the entity dataset, and the result shows that the proposed method has high accuracy among bus arrival prediction problems.

INDEX TERMS Artificial neural networks, bus arrival prediction, LSTM, spatial-temporal feature vector.

I. INTRODUCTION

As it is difficult for road supply capacity to meet the rapid growth of traffic demand, traffic congestion has already become a serious problem in many regions. The public transport bus system, as an economical, efficient, low-carbon and environmentally-friendly vehicle, can greatly alleviate traffic pressure and effectively alleviate environmental pressures in terms of energy consumption, which is recognized as the best option to solve urban traffic problems. In the public transportation field, how to improve the service of the public transportation system, save the travelers' time and improve the waiting experience for urban residents are urgent problems to be solved.

With the development of the Global Positioning System (GPS), by installing positioning sensors on buses, managers can obtain their spatial location in real-time, and calculate vehicle arrival time based on information such as the speed of the buses and distance between stations. Therefore, citizens can arrange their travels more reasonably according to the estimated arrival time of the vehicle, which greatly saves their waiting time. However, due to many factors such as GPS signals, road traffic, passenger flow, and incidents situations, the problem of bus arrival prediction becomes extremely complicated. The single-input prediction method cannot balance the accuracy between long-distance prediction and short-term prediction, because the prediction method only based on historical data and real-time input data cannot effectively respond to the sudden situation and predict long-distance arrival time, respectively [1].

In order to effectively cope with the influences of the incidents in the bus arrival prediction problem, this paper proposes a forecasting analysis method based on space feature vector. In this method, the bus driving path is divided into several segments according to the intersection of the road and calculates the current speed of all buses on each road [2]. Finally, the weighted average is calculated as the instantaneous speed of the bus at the current section, and the required time for the bus to pass the current section is calculated which will be treated as the eigenvalue of the spatial vector during the stage of prediction.

However, arrival prediction based on spatial eigenvectors could not effectively solve the long path forecast problem. The longer the travel path of the bus is, the more obvious the

The associate editor coordinating the review of this manuscript and approving it for publication was Junxiu Liu.

fluctuation of the road condition accumulated by time will be. In order to solve the above problems, this paper proposes a station arrival prediction method based on spatial feature vector, which can solve the problem of remote dependence of bus by analyzing and calculating bus historical data. In the previous arrival analysis based on historical data, the analysts divided the historical data roughly and only divided the day into rush hours and daily hours. But according to the actual observation of the authors of this paper, in addition to the obvious rush hours, the state of the bus is not static, there is obvious fluctuation as time passes by. Therefore, we obtain the optimal value of time cycle division based on the statistical results, divide the day into 188-time slices by using time-slice method and calculate the actual situation of the bus within the time cycle, which was taken as the eigenvalue based on the time eigenvector.

Time-based feature vector analysis can effectively solve the problem of long-range dependencies, but it cannot deal with road incidents. Spatial eigenvector analysis can effectively deal with the impact of highway incidents on the arrival prediction, but it cannot solve the problem of bus remote dependence. Therefore, this paper proposes an arrival prediction method based on time and space eigenvectors. We combine long-term prediction with short-term prediction to build a hybrid neural network based on LSTM and ANN to accurately predict the arrival time of vehicles. Based on the long-term prediction of historical data, a long-term rolling time window is established to complete data update and forgetting. Based on the real-time analysis of spatial feature vectors, the instantaneous short distance arrival prediction is completed. Finally, after a series of trainings, the model obtained good experimental results in test sets.

In this paper, a hybrid model of LSTM and ANN based on spatial-temporal feature vector is proposed. Long-distance arrival-to-station prediction is performed by the temporal feature vector [3]. The time-slicing method is used to calculate the time required to arrive at adjacent stations for all buses in the current forecast time slice. The prediction method can be used to effectively avoid prediction accuracy reduction, which caused by changes in road conditions on long-distance bus lines as the vehicles travel for a long time. The short-distance arrival-to-station prediction is carried out by using the spatial feature dimension. The instantaneous average speed of a single road vehicle is calculated by the space pavement segmentation. The instantaneous time is used to calculate the time between stations in real-time, and the prediction accuracy problem under complex and variable road conditions is solved. Finally, the LSTM-A model based on an integrated spatial-temporal vector is used for bus-to-station prediction. After a lot of training and learning, the model has achieved great results on the test set.

II. RELATED WORK

There are many representative methods for bus arrival time prediction. We present an overview of the last 3 years in this area. An overview of methods for bus arrival time prediction over the last 3 years is presented. Wu *et al.* [4], MatiurRahman *et al.* [5] presented reviews about several common methods of location prediction based on trajectory data. Technically, these methods can be divided into five categories: Support Vector Machines (SVM) [6]–[11] based, Kalman Filter (KF) [12], [13], [14] based, Global Positioning System (GPS) [15], [16] based, Particle Filtering (PF) [17], [18] based, and Neural Network [19]–[31] based.

A. SUPPORT VECTOR MACHINE

SVM firstly maps the input data into higher dimensional space with a specifically designed kernel such that the relationship between modified input data and the target variable is linear. Yang, M., *et al.* presents a prediction model of bus arrival time based on Support Vector Machine with a genetic algorithm (GA-SVM) [6]. Peng, Z., *et al.* proposed a forecasting method based on principal component analysis-genetic algorithm-support vector machine (PCA-GA-SVM) to improve the precision of bus arrival time prediction [7]. B. Z. Yao *et al.* proposed a support vector machine model (single - step prediction model) composed of spatial and temporal parameters [8]. Bai, Cong *et al.* proposed a dynamic bus travel time prediction model for multi-bus routes, which based on support vector machines (SVMs) and Kalman filtering-based algorithm [9]. Moridpour, Sara, *et al.* proposed a Least Squares SVM (LS-SVM) method that expedites the training process by simplifying the quadratic programming problem using a linear regression technique [10]. However, the non-linearity of SVM and SVR comes from the kernel trick which is not scalable for a large-scale problem [11].

B. KALMAN FILTERING

KF has been widely applied to this task. Abidin *et al.* considered the effect of utilizing information acquired from social networks in the Kalman Filter model [12]. Li *et al.* considered KF combined with other methods, and proposed a three-stage mixed model which includes K-means, real-time adjusted Kalman filter, Markov historical transfer model [13]. KF -based method needs a probe to estimate the dynamic term [14], and it's hard to build reliable dynamics of buses on our data set, which combined space and time information.

C. GLOBAL POSITIONING SYSTEM

GPS signal positioning is the most direct method. Based on the bus riders' smartphone Wi-Fi information, Liu *et al.* presented a model to track and predict the arrival time of a city bus [15]. Automatic Vehicle Location (AVL) and smartphone location can also predict bus arrival time. Farooq *et al.* presented a prediction system relying on real-time AVL. Those methods could not make good use of historical information, and it would ignore space features [16].

D. PARTICLE FILTERING

PF technique has been widely applied to deal with historical GPS information and could predict bus arrival time.

Dhivyabharathi *et al.* proposed a method to predict stream travel time using a particle filtering approach which considers the predicted stream travel time as the sum of the median of historical travel times, random variations in travel time over time, and a model evolution error [17]. In order to fix the heterogeneous traffic conditions that exist in India, Dhivyabharathi et al. developed a model based on particle filtering technique which is better than the existing method with MAPE values around 17% with the accuracy of +/- 2 minutes, wherein inputs are obtained using K-NN ((k-nearest neighbors) algorithm (The core of KNN is that a sample belongs to most categories of k samples adjacent to it) [18]. However, the particle filtering algorithm used in these two papers is only suitable for a nonlinear stochastic system with state-space model, but the time property of bus arrival prediction is not considered.

E. NEURAL NETWORK

Neural networks become more and more popular because of their non-linearly modeling ability. C. H. Chen proposed an arrival time prediction method (ATPM) based on RNNs to predict the stop-to-stop travel time for motor carriers [19]. J. B. Pang, *et al.* proposed to exploit the long-range dependencies among the multiple time steps for bus arrival prediction via a recurrent neural network (RNN) [20]. Zhang *et al.* proposed a model based on MapReduce combining clustering with the neural network [21]. H. F. Yang, Tharam S. Dillon, and Y. P. Chen. proposed a novel model, stacked autoencoder Levenberg-Marquardt model, which is a type of deep architecture of neural network approach aiming to improve forecasting accuracy [22]. Polson, G. Nicholas, and O. S. Vadim. develop a deep learning model to predict traffic flows [23]. Y. K. Wu, *et al.* proposed a DNN based traffic flow prediction model (DNN-BTF) to improve the prediction accuracy [24]. Heghedus *et al.* tried five neural network models and identified the best performing deep learning model [25]. A method based on a deep neural network (DNN) model was proposed by Treethidtaphat *et al.* [26]. Kunpeng Zhang *et al.* proposed an end-to-end multi-task learning temporal convolutional neural network (MTL-TCNN) to predict the short-term passenger demand in a multi-zone level [27]. In the same year, Zhang et al. also proposes a deep learning based multitask learning (MTL) model using Bayesian optimization to tune parameters of MTL [28] to predict short-term traffic speed. Sari, F. Riri. proposed a new approach based on a Bayesian mixture model for the prediction [29]. L. Zheng, *et al.* proposed a feature selection-based approach to identify reasonable spatial-temporal traffic patterns related to the target link, in order to improve the online-prediction performance [30].In order to resolve the problem of empirical approaches cannot sufficiently capture diverse travel time distributions, K. P. Zhang, *et al.* proposed a deep learning based Trip Information Maximizing Generative Adversarial Network (T-InfoGAN) [31]. However, in their works, timely GPS data is not combined with the neural network model.

The above several models can solve bus-to-station prediction problem to some degree, but the influence factors that these models considered have one-sidedness. SVM relies too much on kernel tricks to achieve prediction of different scale. KF needs a probe to monitor the state dynamically, which ignores the influence of the past. GPS over-emphasizes the current state of the bus, and the prediction accuracy becomes bad as the predicted distance increases. The PF only considers the time of the bus to the station and ignores the spatial effect of the bus. The input used in the NN network is too one-sided and doesn't consider the comprehensive effect of time and space characteristics.

Therefore, this paper proposes an LSTM-A algorithm. This model combined LSTM and ANN, by adjusting the temporal vectors and spatial vectors to achieve the bus arrival time prediction.

III. METHODOLGY

This section introduces the data set (see Section III-A) used for the case study, as well as the methods for the data preprocessing (see Section III-B), temporal-based feature vector analysis (see Section III-C), spatial-based feature vector analysis (see Section III-D), and arrival prediction based on LSTM and ANN (see Section III-E).

TABLE 1. Data description.

	FIELD	DESCRIPTION	FORMAT	UNIT
1	LINE_UID	LINE NUMBER	STRING	
2	LINE_TYPE	DIRECTION	STRING	
3	BUS_UID	BUS NUMBER	STRING	
4	STATION_SEQ	STATION SEQUENCE	STRING	
5	C_DATA_TIME	ACTUAL ARRIVAL TIME	STRING	
6	REA_TIME_STAMP	ACTUAL ARRIVAL TIMESTAMP	INTEGER	SECOND
7	POSITION_UP_TIME	DATA COLLECTION TIME	STRING	
8	POSITION_SPEED	SPEED	FLOAT	KILOMETER /HOUR
9	POSITION_LATITUDE	LATITUDE	FLOAT	
10	POSITION_LONGITUDE	LONGITUDE	FLOAT	
11	…	…	…	…

A. DATASET DESCRIPTION

The research data used in this paper come from the Xingtai bus company. The data is from September 1st to September 30th, 2018, containing a total of 5,804,504 pieces. The meaning of the field of the dataset is shown in Table 1 and the data sample is shown in Table 2.

B. DATA PREPROCESSING

This section presents the method to preprocess the data before the experiment. Data collected from the transit system

TABLE 2. Data sample.

1	LINE_UID	0120
2	LINE_TYPE	2
3	BUS_UID	BDBB442B9C424F65A08B
4	STATION_SEQ	4
5	C_DATA_TIME	2018-09-28 18:52:10:00
6	REA_TIMESTAMP	1538131930
7	POSITION_UP_TIME	2018-09-28 18:51:44
8	POSITION_SPEED	16.00
9	POSITION_LATITUDE	37.029716584244376
10	POSITION_LONGITUDE	114.5346751069406
11	…	…

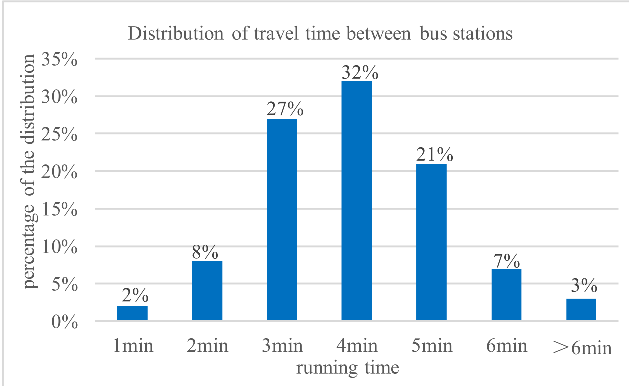


FIGURE 1. Distribution of travel time between bus stations.

exists the situations of lost, duplicated, biased and dirty. Therefore, the data need to be pre-processed before used. This paper uses the following two methods to preprocess the data.

During actual driving, the vehicle may encounter sudden accidents such as road maintenance, vehicle breakdown, car refueling, and temporary parking, etc. Therefore, in the experimental statistics, there are some abnormal points in which running time is greater than the normal driving time. According to the statistical results mentioned in the first section of this chapter, the upper limit of the time window is set to 300 seconds. In addition, it can be seen from the experimental results that during the actual time segmentation process, there are some vehicles spanning two -time slices when driving between the two stations. This running time between two stations of this kind of vehicle travels is far shorter than normal running time. Therefore, according to the experimental result, the lower limit of the time window is set to 25 seconds, thereby filtering the abnormal point to reduce noise interference. The processing result is shown in Fig. 4. The figure shows the sequence of time lengths for the line No.0120 running from 6:00 A.M. to 23:00 P.M. in one day. In this section, the slice of the temporal-based feature vector is obtained through experiments. The result is the time value from station 3 to station 4 in different time slices from 288-time slices.

C. TEMPORAL-BASED FEATURE VECTOR ANALYSIS

The historical data of bus arrivals has a very important research significance for future bus arrival prediction. Researchers have proposed a variety of prediction methods based on historical data, which can be roughly divided into the simple average method, moving average method and exponential smoothing method. When the above methods are used to process time slices, the time slice granularity is relatively simple and rough. A more common way is to divide the whole day time period into the peak hours and other time periods, perform corresponding data processing for different time periods, and combining GPS information of the current bus to predict arrival time. This historical prediction method solves the bus arrival prediction problem to some extent. However, the time-slice arrival prediction of large particles will inevitably lead to the relative roughness of the prediction results. Compared with the fluctuation of the running time of the bus throughout the day, the prediction accuracy is slightly worse. Therefore, this paper proposes a temporal dimension division method based on the actual travel time slice between bus stations.

i *Temporal Slice:* For most cities in China, the distance between the two stations is roughly between 1 and 2 km, and the average speed of city buses is 20 km/h. Therefore, the bus would take about 3 to 6 minutes between two stations. Based on this assumption, the author of this article conducted statistics about the distribution of the time between two stations based on the data of bus line No.0120 of Xingtai in September 2018. The statistical results are shown in Fig. 1. The X-axis represents the bus travel time between two stations, and the Y-axis represents the probability that the travel time between two adjacent stations falls in a certain time slice. It can be seen from the above statistical results that 90% of the buses run less than five minutes between adjacent stops. Therefore, the author divided the 24 hours of the day into 288-time slices at a time interval of 5 minutes. In each time slice, the running time between two stations will be calculated from the data get within 30 days. Considering the different conditions of working days, weekends and holidays, the experimental data were divided into two groups according to working days and off days. The example data mentioned in this paper are the working day experimental data.

ii *Driving State Division:* Usually, the bus has two states when running which is approach the station and stop at the station. To calculate the time period between the two stations, it should be count from the time when the bus leaves the last station to the time when it leaves the next station, as shown in Fig. 2. The state of the bus in the stations will be marked as 'stop' and the state of the bus during running will be marked as 'approach'. For example, the total running time from station 3 to station 4 is t_0, and the running time from station 4 to station 5 is t_1.

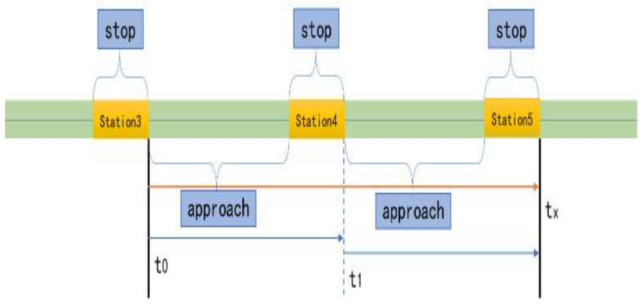


FIGURE 2. Driving state division.

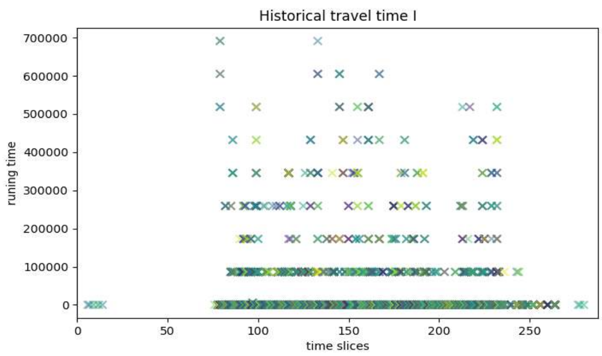


FIGURE 3. Historical travel time I.

iii *Historical Running Time:* The bus driving records between station 3 and station 4 of the bus line No.0120 were chosen for statistical analysis of their performance on different time slices. The result is shown in Fig. 3, The X-axis means 288-time slices and the Y-axis means the running time between the adjacent stations.

During actual driving, the vehicle may encounter sudden accidents such as road maintenance, vehicle breakdown, car refueling, and temporary parking, etc. Therefore, in the experimental statistics, there are some abnormal points in which running time is greater than the normal driving time. According to the statistical results mentioned in the first section of this chapter, the upper limit of the time window is set to 300 seconds. In addition, it can be seen from the experimental results that during the actual time segmentation process, there are some vehicles spanning two -time slices when driving between the two stations. This running time between two stations of this kind of vehicle travels is far shorter than normal running time. Therefore, according to the experimental result, the lower limit of the time window is set to 25 seconds, thereby filtering the abnormal point to reduce noise interference. The processing result is shown in Fig. 4. The figure shows the sequence of time lengths for the line No.0120 running from 6:00 A.M. to 23:00 P.M. in one day. In this section, the slice of the temporal-based feature vector is obtained through experiments. The result is the time value from station 3 to station 4 in different time slices from 288-time slices.

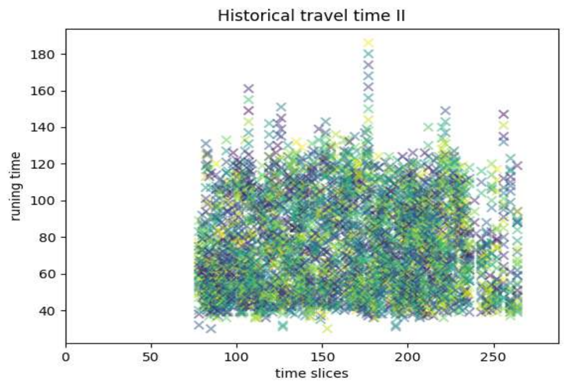


FIGURE 4. Historical travel time II.

In this section, the bus-to-station prediction is divided into small time slice predictions by slicing, and the vehicle arrival time length predictions in the time slice is obtained by comparing the time at the same time slice of different days.

The time-based feature vector analysis is completed to provide data support for the LSTM-ANN comprehensive prediction in the following paragraphs. These variables are combined into a vector to describe the temporal information of the stop p_{ij}:

$$p_{ij} = \left[p_i, p_j, s_k, [t_1, t_2 \ldots t_n]\right]^T \tag{1}$$

p_i, p_j: The station order is *i, j*.

s_k: Time slice k.

$[t_1, t_2 \ldots t_n]$: The time taken from station i to station j on day n when the time slice is k.

p_{ij} : The feature vector from site p_i to site p_j in s_k.

Equation (1) is the mathematical representation of feature vectors based on temporal's slice.

D. SPATIAL-BASED FEATURE VECTOR ANALYSIS

The current location of the bus directly determines the distance to the next target station and subsequent stations. If the road incidents are ignored, the bus arrival time can be simply obtained by dividing the distance by the current speed. Bus arrival time prediction models used this method in the early period, but this method is too idealized and the prediction results were different from the actual ones. Later, researchers used the difference equation to establish an autoregressive moving average time sequence model, and finally realized bus arrival time prediction through residual analysis and data fitting. However, the white noise of the residual sequence in this model affects the result seriously, and the complexity and variability of city traffic are not considered, so the prediction is not accurate. Therefore, this paper proposes a spatial dimension division method based on the road space slice model to predict the time at the local road.

i *Spatial Slice:* According to the principle of the reservoir, as the accumulation rate of water in a pool depends on the difference between the inflow and outflow, the excessive inflow velocity and the low outflow

FIGURE 5. The result of bus route segment.

velocity are the main cause of pool siltation. Similarly, whether a section of road is congested depends on the speed of the upstream vehicle and the speed of the downstream vehicle. In addition to incidents, vehicle traffic speeds are only relevant to driving vehicles on the road segment. Therefore, this paper proposes a spatial segmentation method based on the traffic convergence of traffic signals. For the bus line with line number 0120 in Xingtai in September 2018, the line is divided by the road network traffic light, and the segmentation result is shown in Fig. 5. In this figure, the 22-kilometer-long bus line passes through 17 traffic light intersections. Considering the start and endpoints of the line, the bus line can be divided into 18 slices and the spatial feature vector is initialized through the above steps.

ii *Space Slice Travel Speed Calculation:* From the discussion in the previous section, we can know that the speed of the vehicle on the road is only related to the vehicle on the current road. On the one hand, the vehicle will reduce the speed when the road is crowded. On the other hand, the instantaneous speed of the single-vehicle will change in a different position. For example, the first half of the road is congested due to road construction, vehicle scratching, parking and road narrowing, whereas the latter half is smooth because the car misses the congestion area. In this situation, if only calculate the time until the end of the road according to the instantaneous speed, there will be a large deviation from the actual arrival time. Thus, in this section, a method for calculating the instantaneous speed of a vehicle based on space slice is proposed. The basic formula is as (2):

$$v_i = \frac{1}{n}\sum\nolimits_{j=0}^{n} v_j \tag{2}$$

In the above formula, v_i is the average speed of the vehicle whose number is i, n is the number of buses on the current road, v_j is the instantaneous speed of vehicle numbered j.

iii *Spatial Slice Arrival Time Prediction:* If bus numbered i is in the current space slice we could estimate the distance to the endpoint of this slice by GPS coordinates, and divide this distance by the current vehicle speed v_i to predict the arrival time. If bus numbered i is not in the current space slice, we should predict the time of the current spatial slice though divided the length of the slice by the average vehicle speed v_i. Taking the line No.0120 as the measured line and extracting all the bus information (position, speed, etc.) on the path at a certain moment (assuming bus i is at the originating station at this time), the calculation result is as shown in the Fig. 6.

FIGURE 6. Time cost of different line segments.

The numbers in the figure represent the arrival prediction data of the spatial slice at the current time (the value accurate to one decimal place), and the total length of the route is 63.2 minutes, and the current calculation result is saved as the spatial feature vector of the time.

In this section, the bus-to-station prediction is divided into small spatial slice predictions by slicing, and the prediction time in the slice is obtained by comprehensively considering the driving speed of all the public transportation vehicles in the spatial slice. The space-based feature vector analysis is completed by calculating the time when the bus arrives at the endpoint in different time slices, and this process provides data for the LSTM model to predict comprehensive information later. These variables are assembled into a vector to describe the spatial information of the segment of road s_{ij}:

$$\mathrm{s}_{ij} = [\mathrm{i}, \mathrm{j}, \mathrm{l}_{ij}, [\mathrm{v}_1, \mathrm{v}_2 \ldots \mathrm{v}_n]]^T \tag{3}$$

In this formula, s_{ij} is the road whose starting point is i and ending point is j, l_{ij} is the length of s_{ij} and $[\mathrm{v}_1, \mathrm{v}_2 \ldots \mathrm{v}_n]$ are speeds of all buses on the road.

E. ARRIVAL PREDICTION BASED ON LSTM AND ANN

The temporal-based feature vector analysis can predict the arrival time of the current vehicle with a high probability by mining historical data, but it cannot effectively respond to sudden situations. The spatial-based feature vector analysis can accurately predict the short-distance bus arrival time by analyzing the real-time situation of the road vehicles at the current time, but the prediction accuracy of the long-distance vehicles' arrival will decrease with the error accumulation. Therefore, in this section, a hybrid neural network (LSTM-A)

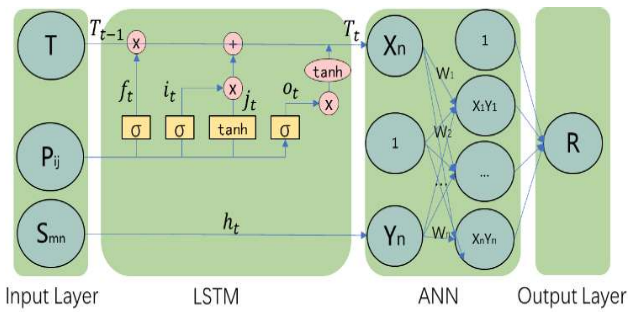


FIGURE 7. The structure of hybrid neural network.

based on LSTM and ANN is proposed, which combines the feature vectors based on time-space.

The structure is shown in Fig. 7. In the input layer, T is the input parameter of the current moment, P_{ij} is the temporal feature vector and S_{mn} is the spatial feature vector. Output layer output the current forecast result.

i *The Calculation of the Temporal Feature Vector Based LSTM*: The current time and temporal feature vector are used as the input parameters, and the time feature value based on the historical prediction information is dynamically updated by the LSTM. If the number of days exceeds the maximum time threshold, the value in the initial time feature vector can be forgotten by the forget gate, and the added time feature vector is updated through the input gate. Finally, the LSTM calculation output is confirmed by the output gate.

Forget Gate:

$$f_t = \sigma(W_f \cdot [\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_n] + b_f) \tag{4}$$

Input Gate:

$$i_t = \sigma(W_i \cdot \mathrm{d}_{n+1} + b_i) \tag{5}$$

$$j_t = tanh(W_j \cdot \mathrm{d}_{n+1} + b_j) \tag{6}$$

Hidden Gate:

$$T_t = (f_t * T_{t-1} + i_t * j_t) * o_t \tag{7}$$

Output Gate:

$$o_t = \sigma(W_f \cdot [\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_{n+1}] + b_o) \tag{8}$$

W_* is the weight matric, and b_* is the bias.
T is the old cell state.
P_{ij} is the input temporal feature vector.
S_{mn} is the input spatial feature vector.
R is the predict result of the computation.
σ is the sigmoid function.

For example, regarding 28 days as the maximum threshold of the time feature vector, the current time moves from September 28th to September 29th, and the time feature of September 1st will exceed the maximum time window. The slice data of September 1st is forgotten through the Forget Gate and then the time feature of September 29 is added to the Input Gate, finally, the result output is controlled by the matrix 0 1 of the Output Gate.

The pseudo-code as Algorithm 1 and Algorithm 2:

Algorithm 1 ALSTM

```
Input: Time window W, R (Temporal t, Spatial s)
Output: Result 1: if t in w then
    for input = 1 : t do
        t1 = LSTMCELL (ct, ht, input)
    end for
end if
s1 = s / 600;
p (t2, s2) = PCA (t1, s1);
if R then
    for (x, y) in p (t2, s2) do
    network = Train (x, y)
    end for
end if
Result = Predict (network, row x, row y);
Return Result;
```

Algorithm 2 LSTM Cell

```
Input: Cell Status prevct, Hide Status prevht, input
Output: Cell Staus ct, Hide Staus ht
combine = prevht +input:
ft = forgetLayer(combine);
candidate = candidatelayer(combine);
ct = prevct * ft + candidate * it
ot = outputLayer(combine);
ht = ot * tanh(ct);
Return ct, ht;
```

ii *Spatial-Temporal Feature Vector Homogenization:* The temporal feature vector obtained by time slice is segment via the station interval while the spatial feature vector obtained by the spatial slice is segmented via the joint point of the road, so the measurement units of the two kinds of feature vectors are not uniform. Therefore, when performing predictive analysis for a specific line, it is necessary to perform the same operation on the two feature vectors. In this paper, the Principal Component Analysis (PCA) is used to analyze and do a projection to the data, so that the eigenvalue normalization is realized by projecting the spatial dimension into the time dimension.

iii *ANN Calculation Based on Temporal-Spatial Feature Vector:* Temporal-based feature vector, the arrival prediction problem is decomposed into point-to-point vector depends only on the length of the historical vector, whereas it is independent of the previous point of the vector starting point coordinates. Therefore, the arrival prediction based on the time vector is more accurate when dealing with the subsequent prediction station with a long distance from the bus starting point. In the face of incidents, however, the response is relatively slow. The spatial-based feature vector turned the answer to the prediction problem into the ratio between the length of the current road and the average driving speed of the current road. This ratio is more accurate in predicting the arrival time of the short distance to the station [30], [32]. Over time, however, the predicted value deviates greatly from the real value after the traffic conditions of different sections change.

This section combines the advantages of the two analytical methods, and proposes an ANN calculation method based on temporal-spatial feature vector, The Linear Regression model as (9):

$$A_j(\bar{x}, \bar{w}) = \sum_{i=0}^{n} x_i w_{ji} \tag{9}$$

Sigmoid function:

$$O_j(\bar{x}, \bar{w}) = \frac{1}{1 + e^{-A_i(\bar{x}, \bar{w})}} \tag{10}$$

Formula (9) is the formula for the Linear Regression model which is used to train and predict the time-spatial feature vector.

Formula (10) is the Sigmoid activation function which converts the result from linear to non-linear.

IV. EXPERIMENT AND RESULT

The experiment uses the data of bus line (line_uid = 0120), and select the data from September 1st to September 20th, 2018 as a training set (excluding 6 days' holiday, we got 14 days' data, the time window is set to 14), the data of September 21st, 2018 will be treated as a test set, the test results are in Fig. 8, Fig. 9, Fig. 10, Fig. 11.

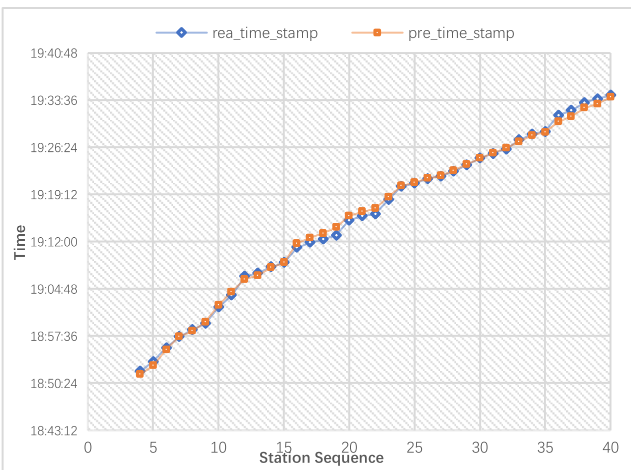


FIGURE 8. Actual and predict time comparison.

Our method has been implemented by TensorFlow. To verify the feasibility and suitability of the calculation method presented in this paper, the data of line_uid 0114 of Xingtai

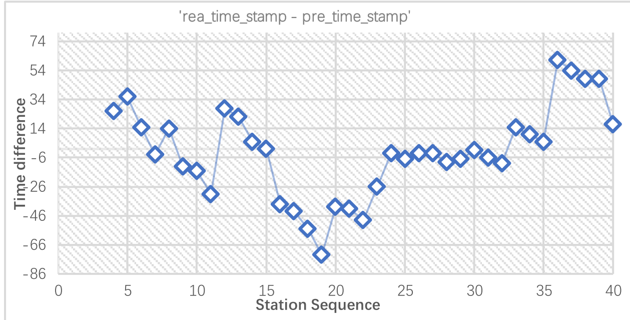


FIGURE 9. Difference value of rea_time_stamp and pre_time_stamp.

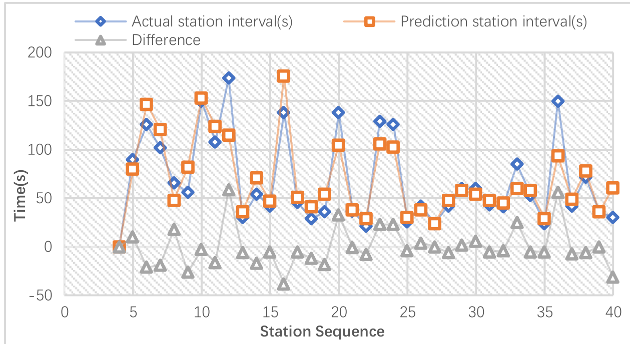


FIGURE 10. The comparison between actual and predicted value of station interval.

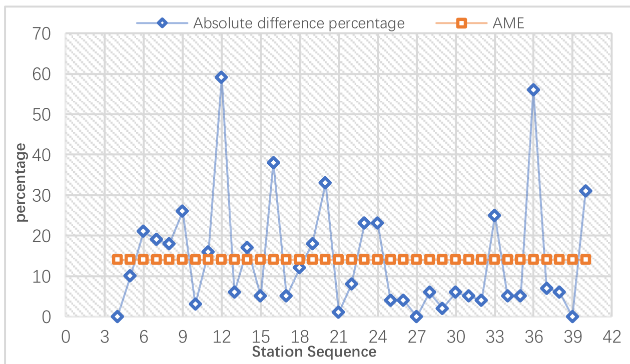


FIGURE 11. The result of ADP and AME.

in China is selected to do the analysis for the buses' running time. After 1000 times sampling, the data is compared and analyzed according to the up and down of the line.

In Fig.8-11, rea_time_stamp represents the timestamp for bus arrives at each bus station and pre_time_stamp is the predicted timestamp for bus arrives at each station.

Figure 8 shows the actual travel time and predicted travel time of a single bus on route 0114, the X axis shows the number of the bus station and the Y axis shows the moment that the bus at the station.

Figure 9 shows the actual travel time and predicted travel time difference of a single bus on route 0114, the X-axis shows the number of the bus station and the Y-axis shows the

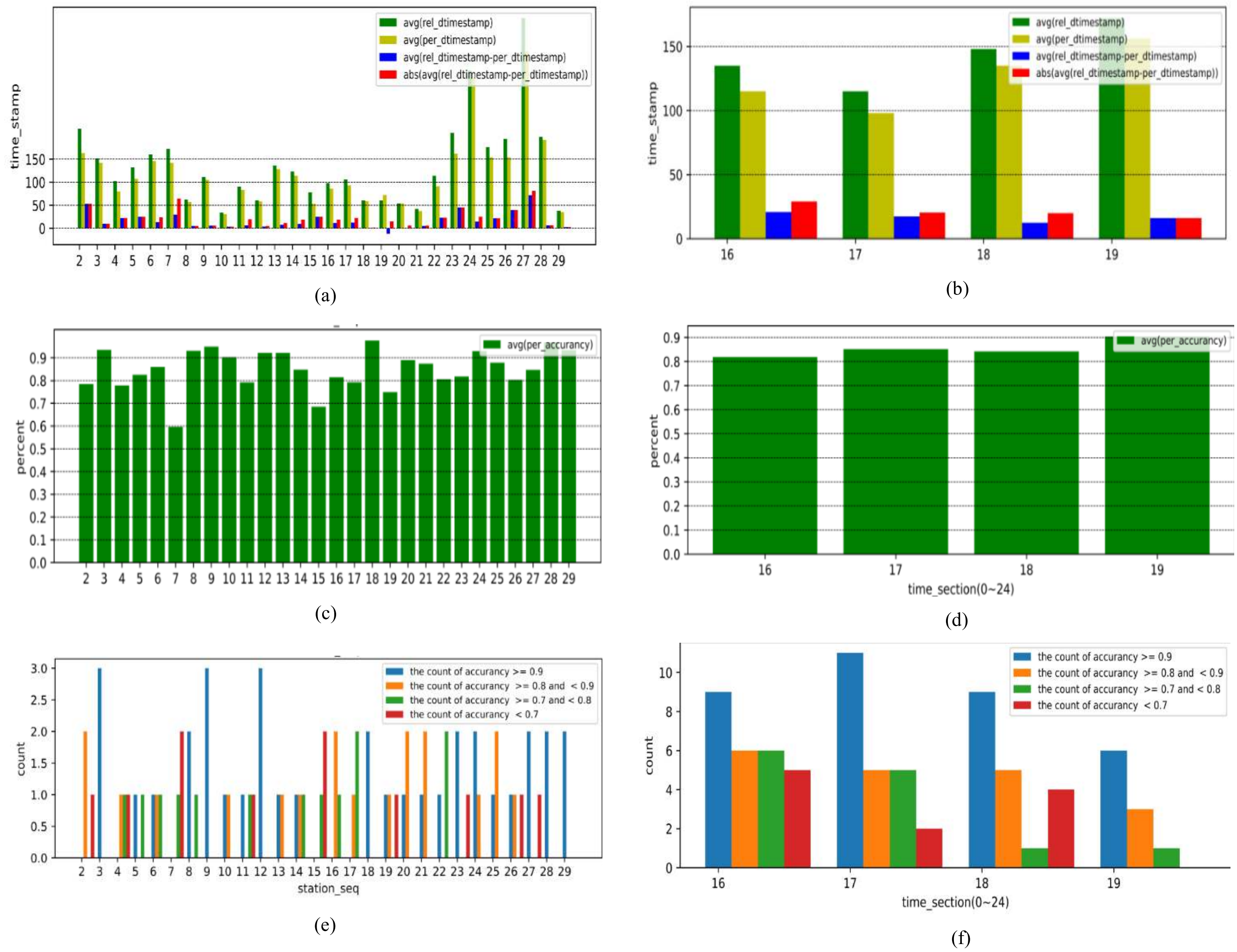


FIGURE 12. **Prediction results.**

difference between the predicted arrival time and the actual arrival time.

Figure 10 shows the comparison between the actual and predicted time of the interval of a bus stop on route 0114, the X-axis shows the number for the bus station and the Y-axis shows the time period between adjacent station.

Figure 11 shows the Absolute difference percentage and AME (Average Magnitude of Error) of the actual and predicted time consumption at a bus stop on route 0114, the X-axis shows the number for the bus station and the Y-axis is the average value of the error.

The experiment mentioned above did analysis and prediction of the chosen dataset. The detailed analysis was done from the aspect of the real-time of arrival, the difference between the real and predicted bus arrival time, the time period between adjacent bus stations, and the mean value of the error. From the result of the experiment, it can be noticed that the proposed method gets a great result and the prediction result has small deviation from the actual result.

This paper demonstrates the performance of the proposed algorithm from the perspectives of average actual running time, average prediction time, average difference, absolute difference average, time period prediction accuracy, and time period distribution frequency and tests these on the data that we got from the bus route No.0114. The specific parameters are explained as follows:

Fig. 12(a) is the predicted comparison of each station on the 0114 line which includes:

avg(rel_dtimestamp): Average actual travel time

avg(per_dtimestamp): Average predicted time

avg (rel_dtimestamp – per_dtimestamp): Average difference

avg (abs (rel_dtimestamp – per_dtimestamp)):

Absolute difference average (for each difference the value is taken as an absolute value and then averaged).

Fig.12(b) is time sampling, the comparison of the operational data calculation results at the time of 16, 17, 18 and 19 on the day is extracted. Fig.12(c) is the Mean Absolute Error for each station, the absolute value of the deviation between the predicted value of a single site and the arithmetic mean of the overall prediction. Fig.12(d) is the sampling period prediction accuracy. Fig.12(e) is a measurement accuracy frequency distribution for each station. Taking the second station as an example, the current forecast is distributed between 0.8 and 0.9. Fig. 12(f) is measurement accuracy frequency distribution.

In this section, the line_uid 0114 with the longest travel distance is selected as a case study to analyze our method. The length of the line is 25 Km, and the driving time is about 1 hour and 26 minutes. It passes through the congested road sections such as the city center and hospitals and also passes through complex environmental sections such as the suburb. It is very suitable for this comprehensive prediction model of long path prediction and a short period with incidents situation prediction. In order to test the performance of the experimental model, this experiment demonstrated the prediction results of the mixed model in a complex environment from the aspects of prediction result, MAE and the prediction distribution.

A. PREDICTION RESULT

The first line of Fig. 12(a) is a complete prediction for bus No. 0114, including the predicted results of all stations from the starting point of the route to the endpoint and the actual running time of the day. According to the difference analysis, except for some special stations, the prediction time error of each station is less than 1 minute, and the performance is good. From the figure, we notice that the station (26, 27) has a large error, and it is the spatial-temporal feature vector that leads to this problem because the number of buses between stations 26 and 27 is sparse. When the bus is running, due to we get statistic data of up-going bus and down-going bus separately, it is reasonable that there are few or no buses at some stations in a certain period of the time slice, and the relatively abandoned road will also have an impact on the experimental results because there are fewer cars running on it. Generally speaking, among the total 29 stations' prediction, most of the deviation of arrival time prediction is less than 30 seconds, and the prediction results are reliable.

B. MEAN ABSOLUTE ERROR

In the second line of Fig. 12(b), the deviation of prediction accuracy was calculated for all stations along the route. The deviation value of more than 93.1% in the whole prediction process was greater than 70%, and 79% of the deviation value is greater than 80%, If the spatial and temporal eigenvectors with insufficient individual data are removed and the sampling periods with good data characteristics are extracted for prediction analysis, the overall prediction accuracy is as high as 80%. Different time and space dimensions are selected to enhance the correlation between feature values, and repeated experiments can bring further improvement to the model.

C. PREDICTION DISTRIBUTION

The third line of Fig. 12(c) selects the predicted hit rate of the same site for different spatial-temporal feature vectors multiple times. For example, in the above figure, the three prediction results at the same time of the day are selected, and site 2 is taken as an example. Among the three prediction results, the accuracy of the two predictions is between 0.8 and 0.9, and the accuracy of one prediction is less than 0.7. In the measurement data of the sampling period, the value of the predicted value distribution higher than 0.8 is much larger than the value lower than 0.8. Therefore, from the perspective of the accuracy distribution, the overall prediction effect is good.

In this section, the author makes a comprehensive analysis of the overall prediction performance from the prediction results, MAE and prediction distribution. Experimental results show that under the good data conditions, the prediction results will not jitter significantly with the change of path length, and the overall prediction performance is stable. To some extent, the model solves the problem of long and short path predictions and promotes the further development of the research on bus arrival prediction.

D. CONTRAST EXPERIMENT

In this section, a series of comparative experiments are carried out, which proved that the performance of the LSTM-A algorithm is better than the traditional prediction method. The eigenvalues used in this paper mainly include historical arrival data, road segmentation data, and real-time vehicle speed data. The above eigenvalues are taken as the input of the model for contrast experiments. The results are shown in Fig. 13. Average (AVG), AMM, and ElasticnetCV denote the prediction results of the arithmetic mean, the Sklearn linear regression prediction, and the Sklearn elastic network regression prediction respectively. Backpropagation (BP) neural network is a network using error square as the target function and using the gradient descent method to calculate the minimum value of the target function. Alstm is time recurrent neural networks, which is proposed to solve the genera long-term dependence problems of the general RNN [33]. All RNN has a form of a chain of the repeated neural network module. Moreover, six types of prediction models are described as follows.

i *AVG:* It is the most intuitive prediction method to use arithmetic mean value to predict bus arrival time. Ma*et al.* [34] and Zhang *et al.* [35] uses this method in their experiment. The time period between two stations is treated as input, then predict the time series for the target. In this paper, 27 bus stations are included, so the input variable is 27, and the output is 1.

ii *AMM:* Using linear regression to predict the passing time for the target between stations. Yu *et al.* [36], Fadaei *et al.* [38]. utilize this method to finish their experiment. The input is the historical time for the bus passing the target station. The number of the input and output equals to bus numbers minus one. So, there are 27 inputs and 1 output.

iii *ElasticnetCV:* This method using an elastic network model. Zeng *et al.* [39] and Lánská, Lánská [40]

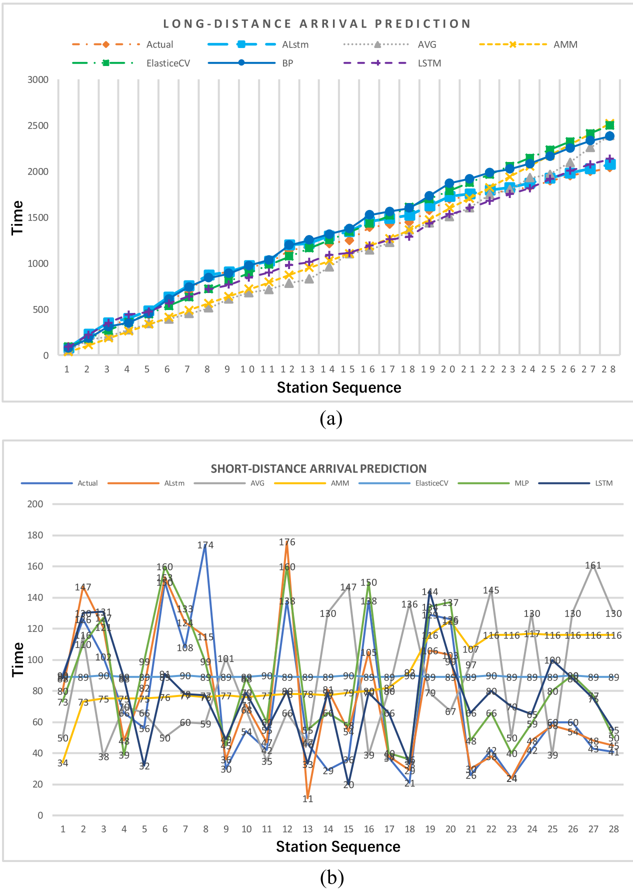


FIGURE 13. The results of comparative experiment.

mention this model in their article. The number of input and output are the same with linear regression which is 27 and 1 respectively.

iv *BP:* This method using backpropagation to estimate the arrival time. Wang *et al.* [41] and Pan *et al.* [42] use this model to predict bus arrival prediction. The number of the input parameters is 27.

v *LSTM:* Using LSTM to finish bus arrival prediction. Petersen *et al.* [43] and Zhao *et al.* [44] mentions this in their article. The influence of prediction time on prediction results is controlled by the time window.

vi *ALSTM:* In this paper, the hybrid model is used for bus arrival prediction. The input contains 27-times feature vector and a 1space feature vector.

The bus GPS data from Sep 1st to Sep 20th are used as the experiment data. A detailed structure is described in Table 3. The experiment parameters are as follows.

Without considering other external influences, the prediction results of the above methods are shown in Fig.13. From Fig. 13 (a), it is obvious that the prediction curve of LSTM-A algorithm in long-distance arrival prediction is closer to the actual arrival time of the bus, so its result is more accurate. From Fig. 13(b) we can know that the prediction curve of the LSTM-A algorithm in short-distance arrival prediction is closer to the actual arrival time of the bus, so its result is more accurate. All in all, although some forecast results are not good, from the overall forecast value it can be noticed that the forecast results are close to the real value.

TABLE 3. Structural description.

Hyperparameters	AMM/ ElasticnetCV	BP	LSTM	ALSTM
Hidden units	80	60	80	60
Learning rate	0.005	0.005	0.001	0.001
Loss weights	1	1	1	1
Batch size	0.25	0.25	0.25	0.25
Dropout rate	0.1	0.1	0.1	0.1
Epoch	10	10	10	10

TABLE 4. Results of short-distance arrival prediction.

Algorithm	MAE	RMSE	MAPE
AMM	52.74	57.67	1.15
AVG	63.70	73.67	1.34
ElasticeCV	45.96	48.64	1.01
BP	32.00	37.45	0.57
LSTM	22.70	26.45	0.41
LSTM-A	**17.11**	**22.82**	**0.27**

TABLE 5. Result of long-distance arrival prediction.

Algorithm	MAE	RMSE	MAPE
AMM	199.11	223.08	0.22
AVG	189.59	215.95	0.21
ElasticeCV	141.63	191.69	0.11
BP	130.25	163.88	0.09
LSTM	98.07	111.04	0.086
LSTM-A	**39.56**	**45.63**	**0.04**

TABLE 4 is the experimental result of short-distance arrival prediction, and the distance between any two stations will be used in prediction. It can be seen from the experimental results that LSTM-A has better performance. TABLE 5 is the experimental result of long-distance arrival prediction, and the distance from the start point to the endpoints will be used in prediction. It can be seen from the experimental result that LSTM-A has better performance.

In this section, the bus arrival is predicted based on short-distance and long-distance respectively, and the predicted results are evaluated from MAE, RMSE, MAPE and other aspects. From the experimental results, it can be seen that the LSTM-A algorithm proposed in this paper is better.

V. CONCLUSION

This paper presents a bus arrival prediction method and has the following contributions. First, we divided the 24 hours into 188 time periods to generate a time feature vector.

After this, the LSTM time window model was established based on the time period and this model can be used to solve the window moving problem when handling time prediction problems.

Second, this paper divided the long path into several spatial segments and used the co-average speed of the current segment as the instantons speed. The predicted time of each segment will be used as the spatial feature vector and it will be sent to the ALSTM.

Third, a novel hybrid neural network ALSTM was proposed to solve the bus station arrival prediction problem. The spatial-temporal feature vector was established by some former work and they were sent to the network to finish the prediction task and achieve great results.

The main contribution of this paper is as followed, based on LSTM, ANN and the spatial-temporal feature vector, this paper achieved the goal of solving the bus arrival prediction problem, and avoided the remote dependency and error accumulation of public transport vehicles. Besides, this paper divides the bus station prediction problem into road section predict subproblems. The concept of real-time calculation is introduced for each related sub-problem, so as to avoid the prediction error caused by complex road conditions. Last but not the least, as shown in the experimental results, the proposed algorithm outperforms the single neural network model in the accuracy and travel time prediction.

In the future, we aim to develop this approach by adding more feature vector, such as the duration and frequency of traffic signals, climate characteristics, temporary stops of public transportation vehicles and the proportion of time spent on boarding and disembarking, etc. [45]. A high-dimensional vector analysis model is established by introducing more feature vectors to further improve the accuracy and reliability of bus-to-station prediction.

REFERENCES

[1] H. Xu and J. Ying, "Bus arrival time prediction with real-time and historic data," *Cluster Comput.*, vol. 20, no. 4, pp. 3099–3106, Dec. 2017, doi: 10.1007/s10586-017-1006-1.

[2] A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, and M. Weidlich, "Traveling time prediction in scheduled transportation with journey segments," *Inf. Syst.*, vol. 64, pp. 266–280, Mar. 2017, doi: 10.1016/j.is.2015.12.001.

[3] B. A. Kumar, L. Vanajakshi, and S. C. Subramanian, "Bus travel time prediction using a time-space discretization approach," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 308–332, Jun. 2017, doi: 10.1016/j.trc.2017.04.002.

[4] R. Wu, G. Luo, J. Shao, L. Tian, and C. Peng, "Location prediction on trajectory data: A review," *Big Data Mining Anal.*, vol. 1, no. 2, pp. 108–127, 2018, doi: 10.26599/BDMA.2018.9020010.

[5] M. M. Rahman, S. Wirasinghe, and L. Kattan, "Analysis of bus travel time distributions for varying horizons and real-time applications," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 453–466, Jan. 2018, doi: 10.1016/j.trc.2017.11.023.

[6] M. Yang, C. Chen, L. Wang, X. Yan, and L. Zhou, "Bus arrival time prediction using support vector machine with genetic algorithm," *Neural Netw. World*, vol. 26, no. 3, pp. 205–217, Jul. 2016, doi: 10.14311/nnw.2016.26.011.

[7] Z. Peng, Y. Jiang, X. Yang, Z. Zhao, L. Zhang, and Y. Wang, "Bus arrival time prediction based on PCA-GA-SVM," *Neural Netw. World*, vol. 28, no. 1, pp. 87–104, May 2018, doi: 10.14311/nnw.2018.28.005.

[8] B. Yao, C. Chen, Q. Cao, L. Jin, M. Zhang, H. Zhu, and B. Yu, "Short-term traffic speed prediction for an urban corridor," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 2, pp. 154–169, Feb. 2017, doi: 10.1111/mice.12221.

[9] C. Bai, Z.-R. Peng, Q.-C. Lu, and J. Sun, "Dynamic bus travel time prediction models on road with multiple bus routes," *Comput. Intell.and Neuroscience*, vol. 2015, pp. 1–9, 2015, doi: 10.1155/2015/432389.

[10] S. Moridpour, T. Anwar, M. T. Sadat, and E. Mazloumi, "A genetic algorithm-based support vector machine for bus travel time prediction," in *Proc. Int. Conf. Transp. Inf. Saf. (ICTIS)*, Jun. 2015, doi: 10.1109/ictis.2015.7232119.

[11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000, doi: 10.1017/CBO9780511801389.

[12] A. F. Abidin and M. Kolberg, "Towards improved vehicle arrival time prediction in public transportation: Integrating SUMO and Kalman filter models," in *Proc. 17th UKSim-AMSS Int. Conf. Model. Simul. (UKSim)*, Mar. 2015, doi: 10.1109/uksim.2015.63.

[13] J. Li, J. Gao, Y. Yang, and H. Wei, "Bus arrival time prediction based on mixed model," *China Commun.*, vol. 14, no. 5, pp. 38–47, May 2017, doi: 10.1109/cc.2017.7942193.

[14] Z. He, L. Zheng, P. Chen, and W. Guan, "Mapping to cells: A simple method to extract traffic dynamics from probe vehicle data," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 3, pp. 252–267, Mar. 2017, doi: 10.1111/mice.12251.

[15] W. Liu, J. Liu, H. Jiang, B. Xu, H. Lin, G. Jiang, and J. Xing, "WiLocator: WiFi-sensing based real-time bus tracking and arrival time prediction in urban environments," in *Proc. IEEE 36th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jun. 2016, doi: 10.1109/icdcs.2016.31.

[16] M. U. Farooq, A. Shakoor, and A. B. Siddique, "GPS based public transport arrival time prediction," in *Proc. Int. Conf. Frontiers Inf. Technol.(FIT)*, Dec. 2017, doi: 10.1109/fit.2017.00021.

[17] B. Dhivyabharathi, E. S. Hima, and L. Vanajakshi, "Stream travel time prediction using particle filtering approach," *Transp. Lett.*, vol. 10, no. 2, pp. 75–82, Mar. 2018, doi: 10.1080/19427867.2016.1192016.

[18] B. Dhivyabharathi, B. A. Kumar, and L. Vanajakshi, "Real time bus arrival time prediction system under Indian traffic condition," in *Proc. IEEE Int. Conf. Intell. Transp. Eng. (ICITE)*, Aug. 2016, doi: 10.1109/icite.2016.7581300.

[19] C.-H. Chen, "An arrival time prediction method for bus system," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 4231–4232, Oct. 2018, doi: 10.1109/jiot.2018.2863555.

[20] J. Pang, J. Huang, Y. Du, H. Yu, Q. Huang, and B. Yin, "Learning to predict bus arrival time from heterogeneous measurements via recurrent neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 9, pp. 3283–3293, Sep. 2019, doi: 10.1109/tits.2018.2873747.

[21] J. Zhang, J. Gu, L. Guan, and S. Zhang, "Method of predicting bus arrival time based on MapReduce combining clustering with neural network," in *Proc. IEEE 2nd Int. Conf. Big Data Anal. (ICBDA)*, Mar. 2017, doi: 10.1109/icbda.2017.8078828.

[22] H.-F. Yang, T. S. Dillon, and Y.-P.-P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2371–2381, Oct. 2017, doi: 10.1109/tnnls.2016.2574840.

[23] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017, doi: 10.1016/j.trc.2017.02.024.

[24] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 166–180, May 2018, doi: 10.1016/j.trc.2018.03.001.

[25] C. Heghedus, "PhD forum: Forecasting public transit using neural network models," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, May 2017, doi: 10.1109/smartcomp.2017.7947031.

[26] W. Treethidtaphat, W. Pattara-Atikom, and S. Khaimook, "Bus arrival time prediction at any distance of bus route using deep neural network model," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, doi: 10.1109/itsc.2017.8317891.

[27] K. Zhang, Z. Liu, and L. Zheng, "Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning," *IEEE Trans. Intell. Transp. Syst.*, to be published, doi: 10.1109/tits.2019.2909571.

[28] K. Zhang, L. Zheng, Z. Liu, and N. Jia, "A deep learning based multitask model for network-wide traffic speed predication," *Neurocomputing*, to be published, doi: 10.1016/j.neucom.2018.10.097.

[29] M. Misbahuddin and R. Fitri Sari, "Bayesian mixture model for prediction of bus arrival time," *Int. J. Technol.*, vol. 6, no. 6, p. 961, Dec. 2015, doi: 10.14716/ijtech.v6i6.1350.

[30] L. Zheng, C. Zhu, N. Zhu, T. He, N. Dong, and H. Huang, "Feature selection-based approach for urban short-term travel speed prediction," *IET Intell. Transp. Syst.*, vol. 12, no. 6, pp. 474–484, Aug. 2018, doi: 10.1049/iet-its.2017.0059.

[31] K. Zhang, N. Jia, L. Zheng, and Z. Liu, "A novel generative adversarial network for estimation of trip travel time distribution with trajectory data," *Transp. Res. C, Emerg. Technol.*, vol. 108, pp. 223–244, Nov. 2019, doi: 10.1016/j.trc.2019.09.019.

[32] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015, doi: 10.1016/j.trc.2015.03.014.

[33] J. Xu, R. Rahmatizadeh, L. Boloni, and D. Turgut, "Real-time prediction of taxi demand using recurrent neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2572–2581, Aug. 2018, doi: 10.1109/tits.2017.2755684.

[34] J. Ma, J. Chan, G. Ristanoski, S. Rajasegarar, and C. Leckie, "Bus travel time prediction with real-time traffic information," *Transp. Res. C, Emerg. Technol.*, vol. 105, pp. 536–549, Aug. 2019, doi: 10.1016/j.trc.2019.06.008.

[35] J. Zhang, D. Shen, L. Tu, F. Zhang, C. Xu, Y. Wang, C. Tian, X. Li, B. Huang, and Z. Li, "A real-time passenger flow estimation and prediction method for urban bus transit systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3168–3178, Nov. 2017, doi: 10.1109/tits.2017.2686877.

[36] B. Yu, H. Wang, W. Shan, and B. Yao, "Prediction of bus travel time using random forests based on near neighbors," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 4, pp. 333–350, Apr. 2018, doi: 10.1111/mice.12315.

[37] M. Fadaei, O. Cats, and A. Bhaskar, "A hybrid scheme for real-time prediction of bus trajectories," *J. Adv. Transp.*, vol. 50, no. 8, pp. 2130–2149, Dec. 2016, doi: 10.1002/atr.1450.

[38] M. Sinn, J. W. Yoon, F. Calabrese, and E. Bouillet, "Predicting arrival times of buses using real-time GPS measurements," in *Proc. 15th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2012, doi: 10.1109/itsc.2012.6338767.

[39] W. Zeng, M. Tomio, and M. Takayuki, "Exploring trip fuel consumption by machine learning from GPS and CAN bus data," *J. Eastern Asia Soc. Transp. Stud.* vol. 11, pp. 906–921, Dec. 2015, doi: 10.11175/easts.11.906.

[40] M. Lánskà, "Modelling of price demand elasticity for entry to bus terminals," *Trans. Transp. Sci.*, vol. 3, no. 2, pp. 67–70, Jun. 2010, doi: 10.2478/v10158-010-0008-y.

[41] L. Wang, Z. Zuo, and J. Fu, "Bus arrival time prediction using RBF neural networks adjusted by online data," *Procedia-Social Behav. Sci.*, vol. 138, pp. 67–75, Jul. 2014, doi: 10.1016/j.sbspro.2014.07.182.

[42] J. Pan, X. Dai, X. Xu, and Y. Li, "A self-learning algorithm for predicting bus arrival time based on historical data model," in *Proc. IEEE 2nd Int. Conf. Cloud Comput. Intell. Syst.*, vol. 3, Oct. 2012.

[43] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Syst. Appl.*, vol. 120, pp. 426–435, Apr. 2019, doi: 10.1016/j.eswa.2018.11.028.

[44] Z. Huang, Q. Li, F. Li, and J. Xia, "A novel bus-dispatching model based on passenger flow and arrival time prediction," *IEEE Access*, vol. 7, pp. 106453–106465, 2019, doi: 10.1109/access.2019.2932801.

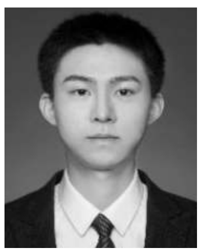
[45] B. A. Kumar, R. Jairam, S. S. Arkatkar, and L. Vanajakshi, "Real time bus travel time prediction using k-NN classifier," *Transp. Lett.*, vol. 11, no. 7, pp. 362–372, Jul. 2019, doi: 10.1080/19427867.2017.1366120.

HONGJIE LIU was born in Weinan, China, in 1988. He is currently pursuing the Ph.D. degree in computer science with Xi'an Jiaotong University. His main research direction is travel time optimization of urban residents based on high dimensional curved surface projection.

HONGZHE XU received the Ph.D. degree in mechanical engineering, in 2004. From 1999 to 2012, she was a Vice Professor with the Research School of Computing. Since 2012, she has been a Professor with the Research School of Computing, Xi'an Jiaotong University, Xi'an. She is the author of ten books, more than 50 articles, and more than five inventions. Her research interests include intelligent platform in cloud environment, object-oriented big data analysis, application of data mining, and medical big data analysis.

YU YAN was born in Xianyang, China, in 1994. She received the B.S. degree in the Internet of Things engineering from Northwest University, Xi'an, in 2016, and the M.S. degree in computing from Xi'an Jiaotong University, Xi'an, in 2019. Her main research directions are natural language processing and knowledge mapping in medical imaging.

ZAISHANG CAI is currently pursuing the master's degree with the School of Electronics and Information Science, Xi'an Jiaotong University. His research interests include trajectory clustering, trajectory feature point extraction, and transportation data mining.



TIANXU SUN received the bachelor's degree in computer science and technology from Harbin Engineering University, in 2018. He is currently pursuing the master's degree in computer technology with Xi'an Jiaotong University. He did some research about medical image diagnosis based on semantic segmentation. His research interest includes digital image processing using deep learning method.

WEN LI received the master's degree in computer science, in 1994. From 2000 to 2020, she was a Lecturer with the Research School of Computing, Xi'an Jiaotong University, Xi'an. She is coauthor of four books, five articles, and two inventions. Her research interests include intelligent platform in cloud computing, object-oriented big data analysis, and medical big data analysis.