WILEY | Hindawi

Research Article

Charging-Related State Prediction for Electric Vehicles Using the Deep Learning Model

De Zhao [ID],[1,2,3] **Hua Wang** [ID],[4] **and Zhiyuan Liu** [ID][1,2,3]

[1]*Jiangsu Key Laboratory of Urban ITS, Southeast University, Nanjing, China*
[2]*Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Southeast University, Nanjing, China*
[3]*School of Transportation, Southeast University, Nanjing, China*
[4]*Department of Civil and Environmental Engineering, National University of Singapore, Block E1A, #07-03, 1 Engineering Drive 2, 117576, Singapore*

Correspondence should be addressed to De Zhao; zhaode@seu.edu.cn

Electric vehicles (EVs) are becoming the potential contender for the conventional gasoline vehicles in view of the environment-friendly and energy-efficient characteristics. The prediction of EV charging-related states (defined in this study as home charge, outside charge, home stop, outside stop, low-battery travel, and high-battery travel) could help to identify the future charging demand (power consumption) of EV individuals. Specifically, it could guide the operation and management of charging facilities and also provide tailored charger availability information based on users' real-time locations. This study aims to predict charging-related states of individual EVs using a deep learning approach. We first propose a tangible approach to convert EV trajectory data into state sequences and then develop a bidirectional gated recurrent unit model with attention mechanism (Bi-GRU-Attention) to forecast EV states. A sensitivity analysis is conducted to tune and/or calibrate parameters in the model based on plug-in hybrid EV trajectories dataset collected in Shanghai, China. Experiment results show that (i) the proposed method could achieve an average accuracy of 77.15% with a 1-hour prediction length and it outperforms the baseline models for all tested prediction lengths; (ii) it is also revealed that the prediction accuracy varies dramatically with different states and time periods. Among all states, the proposed model has a higher prediction accuracy on "home stop" (89.0%). As for time periods, the EV states around 08:00 am and 04:00 pm are hard to predict, and a comparatively low prediction accuracy (close to 60%) is obtained; and (iii) the stability and robustness analysis implies that the proposed model is stable and insensitive to SOC noise or season.

1. Introduction

Electric vehicles (EVs) have been recently advocated by policy-makers in view of their environment-friendly characteristics [1]. Both financial and nonfinancial incentives are used worldwide to promote the market penetration of EVs [2–5]. In particular, the development of battery technology is accelerating the adoption of EVs. It is estimated that 25 million units of EVs will be sold by the year 2025 around the world [6]. In other words, EVs are expected to be a strong competitor to the traditional gasoline vehicles (GVs) in the market, especially when combined with vehicular communications [7].

One of the obstacles for EV promotion is how to properly deploy and manage charging facilities [8, 9]. Unlike the refueling for GVs, pure EV and plug-in hybrid EV users have to replenish electricity energy more frequently due to a limited driving range with a long duration [10]. They prefer to charge their EVs without many detours from their customary routes [11, 12]. Hence, travel habits and charging pattern of EV users are worth much more consideration in the operation of charging facilities [13, 14]. Herein, the travel habit and charging pattern means that how the traveling and charging states of an EV (such as home charge, outside charge, home stop, outside stop, travel with low battery, and travel with high battery) change over time and location dynamically.

It is imperative to seek a viable approach for understanding and predicting the EV states at the individual level because it can substantially provide valuable information for both charging facility operators and users. For operators, they are concerned about possible overloading of electricity grid incurred by a large number of EVs' charging practices simultaneously [15, 16]. The prediction of EV states for a realistic scenario could let the operators anticipate different types of charging demand (home charging or public charging) in the next few hours and make some necessary control measures in advance (such as peak shaving to avoid highly variable load [16]). For users, we could also estimate users' next charging time and charging type based on their history of charging behavior and suggest future available charging facilities from the users' perspective. Since these states of EV are all highly related to the charging behavior, we refer to them as "charging-related states of EV" hereafter in this article. It is worth noting that the "charging-related states" here are generalized, covering both charging and discharging process.

To investigate the charging-related states of EV, the first research line is assessing the EV users through field surveys. For example, Davies and Kurani [17] made the first attempt to examine the charging state of EV based on data of charging time and charging frequencies of 40 households. They found that daily charging times varied widely across different households. Afterward, Franke and Krems [18] analyzed EV charging preference, and their findings also showed that there are significant individual differences between EV users. In this sense, individual EV charging state cannot be described in an aggregate way.

The second research path focuses on the mathematical model to formulate the charging state. Dong and Lin [19] built a Poisson-gamma distribution model to characterize travel distances between two sequential charges and used GPS-based survey data to validate the model. A few regression models were proposed then to find out the latent predictors of EV charging state, such as the linear regression model [20], logit model [11, 21, 22], and machine learning models [23]. The latent predictors include but are not limited to the state of charge (SOC), the availability of charging station, detour distance for the charging, time of day, charging cost, the travel plan of next day, etc. More models and specific factors have been explicitly summarized and discussed by Hardman, Jenn [24]. Except for charging state estimation, researchers also build mathematical models to directly control the battery charging process. Liu et al. applied a constrained generalized predictive control strategy to the battery charging process. A coupled thermoelectric model was introduced to estimate the battery behaviors [25]. Subsequently, they derived a distributed average tracking approach to achieve the optimal charging control of EV battery [26]. This approach significantly reduced the computational burden for the charger controller.

The rise of deep learning approach makes it possible to predict time-series sequences accurately. The sequence prediction is widely applied in biochemistry-molecular biology and natural language processing (NLP) [27]. A bunch of models has been developed in the above research areas, for example, n-gram, hidden Markov chain (HMM), recurrent neural networks (RNN), etc [28]. In this decade, the RNN becomes the most prevailing tool for NLP applications in industry due to its excellent performance in terms of both prediction accuracy and computational efficiency [29]. Unlike the n-gram or HMM which can only predict the current state based on limited numbers of history states, the RNN could theoretically work with unbounded history states [28]. The RNN provides a powerful solution for processing time-series data as well as sequence data. It has also been used for traffic analyses, for example, traffic flow and travel speed prediction [30], driving behavior analysis [31], and GPS data-based vehicle classification [32]. These applications are all numerical prediction or data-driven clustering.

Recently, it has spawned a vast body of literature that uses the deep learning approach to estimate/predict EV state, including state-of-charge (SOC) [33], state-of-health (SOH) [34], state-of-available-power (SoAP) [35], etc. For example, Liu et al. have proposed a transferred RNN-based framework to achieve the battery calendar health prognostics [36]. In this task, the RNN model proved to outperform the other two typical feedforward neural networks—BP NN and RBF NN. The research team also combined the long short-term memory model (LSTM, a special type of RNN) with Gaussian process regression model to predict the future capacities and remaining useful life of EV battery [37]. It turns out that the combination model can achieve accurate results. As for SOC prediction, deep learning models also showed good performance in many cases, such as LSTM [38], BP NN [39], or RNNs-CNNs [40]. Among them, the RNN model is most prevailing since it can work with very long history states.

From the above literature review, it can be seen that previous studies have paved a solid way to investigate EV charging state by finding out the latent predictors or method framework. When analyzing EV charging state, however, these studies have not paid much attention on the other important states such as driving, stopping, etc., and overlooked their inherent correlations. In fact, these states are also closely related to the charging state in that a sequence of EV states can reflect the complete travel habit and charging pattern of an individual EV. For example, a "low-battery traveling" state followed by an "outside charging" state could provide more information for succeeding state inference and prediction than a sole state "outside charging." It indicates that the EV user probably realized a low battery and then charged his EV in a public charging station. By in-depth learning a time-series sequence of EV states, we can decode its charging pattern and predict future states. However, the accurate prediction of the state sequence is always a challenging task since conventional time-series models cannot be used in the nonnumerical sequence.

Considering the aforementioned research gap, a tangible data-driven framework through devising RNN approach (bidirectional gated recurrent unit model with attention mechanism, Bi-GRU-Attention) is proposed to predict EV states. We predict the charging-related states instead of directly forecasting the charging demand because the other

states, such as low-battery travel or home stop, also imply the latent charging behavior. The charging-related states as a whole are more informative than merely the charging behavior. Besides, the time allocated to each specific state can unveil the time-use pattern of EV, especially the discriminative time intervals for "outside charging" and "home charging" states which could indicate different types of loads on electric grid.

The main contributions of this study are as follows: Firstly, this article makes the first attempt to tackle the charging-related states of EV as a consecutive sequence rather than separate states and proposes an effective method to covert EV trajectory data to EV state sequences for model training. Secondly, we enrich the charging-related states with traveling and location information and define six types of EV charging-related states. Then, we provide a tangible RNN-based approach (Bi-GRU-Attention) to predict EV charging-related states. Thirdly, plug-in hybrid EV trajectory data from Shanghai of China are used to validate the proposed model and tune the related parameters. The result indicates that the proposed approach outperforms the traditional sequence prediction models in terms of prediction accuracy.

The rest of the study is organized as follows: Section 2 elaborates our research methodology, including the EV state definition and the RNN approach for EV state sequence prediction. Section 3 illustrates the process of converting EV trajectory data to state sequence data and validates the proposed model using real-world data. Section 4 concludes the article and gives a few suggestions for the future study.

2. Research Methodology

2.1. The Set of Charging-Related States. According to the intrinsic battery status, there are three states of an EV, namely charging, traveling (discharging), and stopping (self-discharging, a slow charge leakage phenomenon of batteries when not in use). These three types of state information could be identified by monitoring the current in the battery. However, location and SOC information is not revealed. To better investigate charging-related states, we further divide them into six subtypes according to the location and SOC of EV: (1) home charge (HC); (2) outside charge (OC); (3) home stop (HS); (4) outside stop (OS); (5) low-battery travel (LT, SOC $< S_0$); and (6) high-battery travel (HT, SOC $\geq S_0$). S_0 is the charging threshold, lower than which the EV user would feel an urgent need to charge the EV. The parameter S_0 is expected to vary with the risk preference of EV users and their EV conditions (e.g., battery degradation level). However, the exact value of S_0 for each user is difficult to estimate since S_0 is not very definitive. Besides, in this study, we only use S_0 to differentiate two states regarding traveling, that is, LT and HT. Hence, we assume the same value of S_0 for all users in this study. The specific setting of S_0 is given in Section 3.1.

We differentiate the six types of EV charging-related states because these states respectively indicate different urgency of charging and convenience of home charging. For example, LT means that the EV is still running on the road

when SOC is below the charging threshold S_0, which implies an urgency of charging later. Accordingly, HT means that the EV is working properly with the SOC higher than S_0 (less urgent need for charging), whereas "home" or "outside" can measure the availability of home charging or the desire for public charging facility.

2.2. Sequence Prediction with RNN. The RNN belongs to a powerful family member of artificial neural network models. Unlike the traditional neural network, the RNNs can store the previous state information to process sequence data. However, the simple RNNs are not capable of training long sequence data because the gradient may vanish quickly for parameters across the sequence [30], while EV state prediction is just the case with a long sequence. The sequence length of EV state prediction may even excess 100 if we use whole-day history data as input. To solve the long sequence issue (vanishing gradient problem), two variants of simple RNN were proposed successively: long short-term memory (LSTM) [41] and gated recurrent unit (GRU) [42].

The LSTM has added a "forget" mechanism to the basic RNN, which enables error to backpropagate through the long sequences. It makes the long sequence prediction possible. The LSTM has very wide applications in the transportation field [29, 30, 32, 43] since it is proposed in early 1997 [41]. The GRU, introduced late in 2014 by Cho, Merri¨enboer [42], is another popular variant of RNN like LSTM. Nevertheless, the GRU contains fewer parameters than the LSTM when training the same data. It is demonstrated to be more efficient than the LSTM with empirically similar performances in most tasks [29]. Therefore, in this study, we choose the GRU as the core part of our model. The specific structure of GRU for sequence prediction is illustrated in Figure 1.

Similar to the basic RNN or LSTM, based on the input $\mathbf{x_t}$ and the previous hidden state $\mathbf{h_{t-1}}$, the GRU can produce the current hidden state $\mathbf{h_t}$ as well as the output $\hat{y}_t$. The main difference is that the GRU has added two gates: reset gate $\mathbf{r}$ and update gate $\mathbf{z}$. On the one hand, the update gate is used to determine how much of the previous state information should be passed to the future. The update gate for time t can be calculated by

$$z = \sigma\left(U_z \cdot x_t + W_z \cdot h_{t-1}\right). \tag{1}$$

On the other hand, the reset gate determines how much of the previous state information should be forgotten/dropped. The reset gate for time t can be calculated as:

$$r = \sigma\left(U_r \cdot x_t + W_r \cdot h_{t-1}\right), \tag{2}$$

where U_z, W_z, U_r, and W_r in (1) and (2) are the weight matrices, and σ is the sigmoid activation function given by

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \tag{3}$$

Then, reset gate $\mathbf{r}$ and update gate $\mathbf{z}$ work together to affect the output according to the following mechanisms:
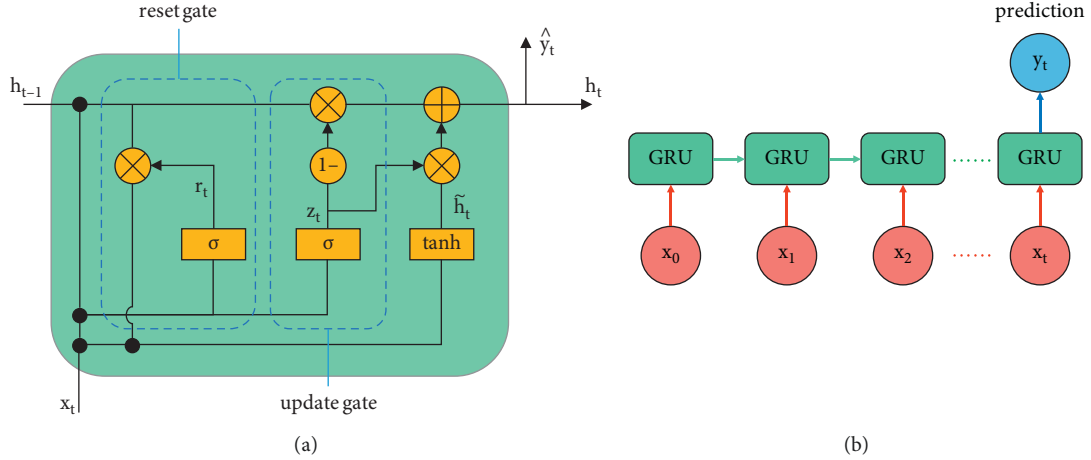
FIGURE 1: The structure of GRU for sequence prediction. (a) The structure of GRU. (b) Sequence prediction using GRU.

$$\widetilde{h}_t = \tanh\left(U_z \cdot x_t + W_s \cdot \left(h_{t-1} \odot r\right)\right),$$
$$h_t = (1 - z) \odot \widetilde{h}_t + z \odot h_{t-1}, \tag{4}$$

where $\odot$ is the Hadamard Product, $\widetilde{h}_t$ is the middle memory content, and tanh is the nonlinear activation function calculated by:

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}. \tag{5}$$

The final estimated $\widehat{y}_t$ is a probability distribution over the predefined classes (EV state), which can be calculated by the softmax function (multinomial logit):

$$\widehat{y}_t = \text{softmax}\left(W_y \cdot h_t\right),$$
$$\text{softmax}\left(x_j\right) = \frac{\exp\left(x_j\right)}{\Sigma_i \exp\left(x_i\right)}. \tag{6}$$

2.3. Bidirectional GRU with the Attention Model. Attention mechanism proved to be effective in enhancing performance of RNN with long input sequence [44]. Hence, we also added the attention architecture in this study. In models with attention mechanism, the GRU network is usually replaced by a bidirectional GRU network, which can read input sequence in both the forward and backward directions [45]. The specific architecture of bidirectional GRU with attention (Bi-GRU-Attention) mechanism used in the study is depicted in Figure 2. Three sources of data are encoded and fused as the model input, including historical SOC data, distance to home data, and the sequential EV state data, because all the three sources of data contain potential information about the future state of EV. Particularly, distance to home data could be derived based on the real-time location of EV and the EV driver's home location. The final output of the Bi-GRU-Attention model would be the most likely estimation for the future EV state (the focus of our study).

The Bi-GRU-Attention models are built based on Keras, a neural network API on top of TensorFlow. The categorical cross-entropy function is used as the loss function, which is the objective function when training the deep learning model. The cross-entropy loss function measures the performance of classification models with discrete output. It can be calculated by:

$$L_{\text{cross-entropy}}\left(\widehat{y}, y\right) = -\sum_i y_i \log\left(\widehat{y}_i\right). \tag{7}$$

There are four parameters in the model: epochs, optimizer, batch size, and dropout rate. Epochs determine how many rounds of training will be run. In this study, we choose the early stopping mechanism to determine the epochs rather than fixing a specific value. Specifically, when the prediction results do not have any improvements for 5 times, the training of the model will terminate [46]. The early stopping mechanism can save the total training time.

The optimizer is the algorithm used to choose a proper learning rate and avoid getting trapped in local minima for the deep learning model. We use the Adaptive Moment Estimation (Adam) algorithm as the optimizer, which adaptively computes the learning rates for each parameter. Adam is proved to be quite robust to the input parameters [32].

During each epoch, the entire dataset is divided into a number of batches for training because the entire dataset is usually massive. The batch size is the number of samples in each batch, which should be specified beforehand. To improve the convergence speed and avoid the overfitting of the model, we also add the dropout layer to the model, because the dropout layer, as a regularization approach, could thin the network by ignoring some units at random. The proper value of batch size is picked up based on the sensitivity analysis, which is extensively discussed in the next section.

3. Data Sources and Experimental Results

3.1. Electric Vehicle Trajectory Datasets. The Bi-GRU-Attention model for state prediction in this study is fit for both pure EV and plug-in hybrid EV. The datasets used comprise
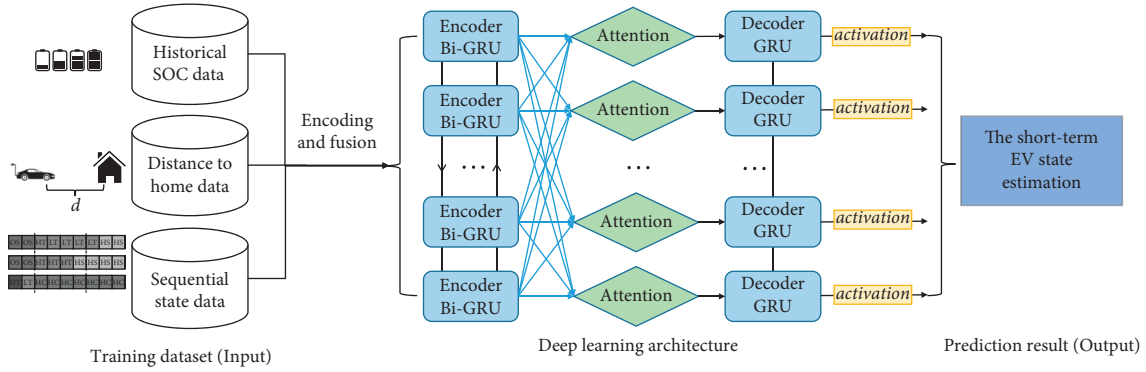
FIGURE 2: Bidirectional GRU with attention mechanism.

the trajectories of 50 personal plug-in hybrid EVs (non-commercial) in Shanghai, China from 26 May 2015 to 26 May 2016. These 50 EVs are randomly selected and their 1-year datasets cover all the districts of Shanghai. In each dataset, a series of information fields are recorded with an updating frequency approximately 30 secs, including the real-time location (longitude and latitude), velocity, direction, SOC, the status of EV (traveling, charging, or idling stop), etc. These pieces of information are very common and could also be found in EV trajectories from other cities, such as in Aichi prefecture [47] and other prefectures in Japan [8]. The data recording system embedded in the EV works all the time, except that the vehicle has completely stalled. Thus, we could complete the datasets by adding the status of "stopping" after the vehicle has completely stalled. It is estimated that there are totally 5,084 charging piles in Shanghai by the end of 2016 [48].

Based on the above datasets, we could find out the low-battery threshold S_0, that is, when the drivers usually feel urgent to charge their EVs. We plot the histogram of SOC when the drivers start to charge to estimate the low-battery threshold S_0, as shown in Figure 3. It can be found that as the start-to-charge SOC goes down, the probability density increases dramatically, especially when the SOC is below 25%, which indicates that the potential urgency for EV charging is increasingly strong with SOC less than 25%. Therefore, we set the low-battery threshold S_0 as 25% in this experimental study.

3.2. Home Location Identification and Dataset Converting. Because the EV trajectory datasets do not have the exact information about home location of an EV user, we in this section identify the home location by their trajectory patterns. The home location of each EV user is estimated by the following rules shown in Figure 4 (the trajectory of one EV user). Firstly, we extract the data rows of one EV with the EV status "charging" or "stopping" between 2:00 a.m. and 4:00 a.m, because during this time interval, the majority of EV users probably charge or park their EVs at home. Secondly, based on the spatial location of each data point in the extracted subset, we filter out the data point with the maximal number of proximity points in its walkable radius. We set this data point as the home location of this EV user.
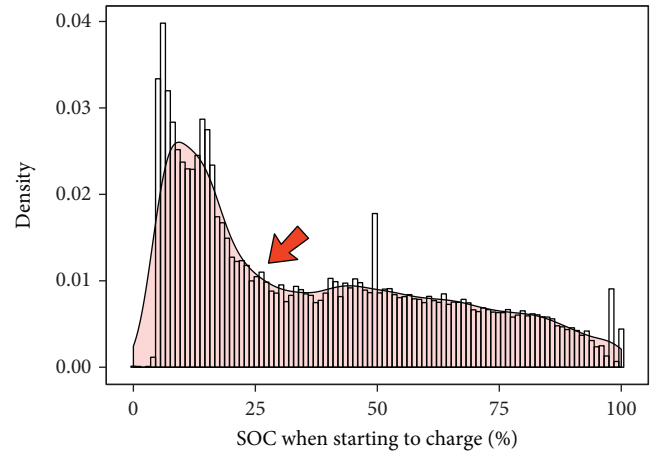

FIGURE 3: Histogram of SOC when starting to charge.

The walkable radius is set to be 500 m [49]. Based on the estimated home location, we can calculate the Euclidean distance to home of each trajectory data point. All the data points situated in the walkable 500 m radius to home are regarded as at home in this study since many drivers do not have a dedicated parking spot around their homes.

Next, we convert the EV trajectory dataset into EV state sequences for model training and prediction. By filtering the Euclidean distance to home, status, and SOC, we can easily define a state for each data row. For example, the data row with Euclidean distance to home <500 m (walkable radius) and status "charging" is set as "HC," while the data row with status "traveling" and SOC<25% is set as "LT." Then, we cut the dataset into pieces with an equal time window. The latest state in each piece is set as the state of this time window and all the states are stitched together into the state sequence. The excerpt of derived state sequences with a time window of 60 min is illustrated in Figure 5. We choose the 24-hour data before the predicting point as the input sequence across the entire study. Therefore, the length of the state sequence depends on the time window we use. For example, if we set the time window as 30 min, the length of input state sequence would be 24 h/30 min = 48. Therefore, the time window is another parameter that should be calibrated, besides the dropout rate and batch size mentioned in Section 2.

The identified
home location

500 m radius

The original EV trajectory

The charging or stopping data points
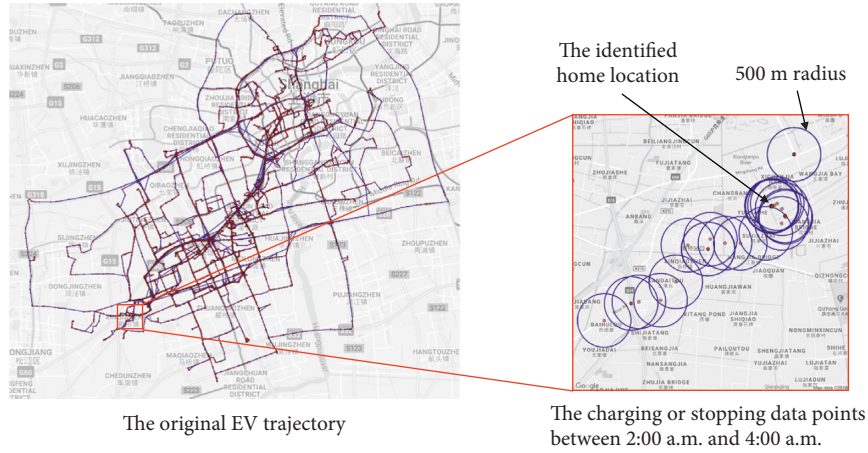between 2:00 a.m. and 4:00 a.m.

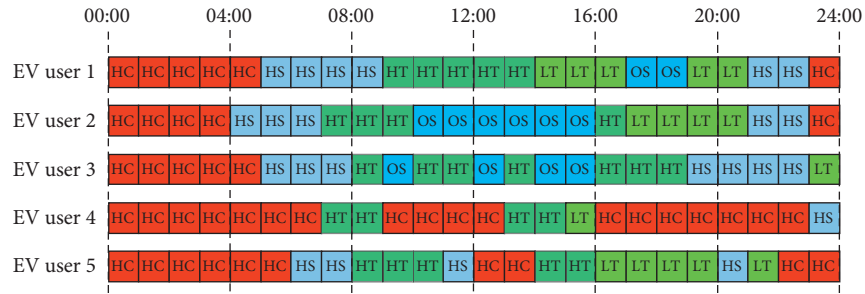FIGURE 4: Home location identification based on EV trajectory data.

FIGURE 5: Randomly selected state sequences of five EVs with a time window of 60 min.

3.3. Performance Metrics. We cut the EV sequence dataset into two parts. The first 90% of the dataset is used as training data and the remaining 10% is used as the referenced data for validation. The prediction accuracy is selected as the performance metrics in training the model. The prediction accuracy is calculated as follows:

$$\text{acc} = \frac{N_{\text{true}}}{N_{\text{total}}}, \tag{8}$$

where N_{true} is the number of samples that are correctly predicted (true positive) and N_{total} is the total number of samples.

3.4. Predictive Performance. A preliminary analysis of the EV states at the individual level reveals that there exists a large variation of individual charging pattern. Figure 6 shows the EV state distribution of four different users at different times of the day. From the figure, we can find that the majority of users have a high probability of night charging at home, which is in line with the previous literature [50]. While, the other states display a strong individual pattern, such as outside charging (OC) and low-battery travel (LT). EV user 2 shown in Figure 6(b) has a high proportion of outside charging. By contrast, EV user 1 shown in Figure 6(a) charges the vehicle nearly all at home. EV users 1 and 4 have a very sharp increase in the probability of home charging after 22:00. That is because electricity

tariffs will go down by 50% from 22:00 to 6:00 in Shanghai and these two users are probably very sensitive to the charging cost. This phenomenon agrees with the conclusion from the previous study by Sun, Yamamoto [21].

We also calculate the entropy of states for each EV user to measure the uncertainty or unpredictability of each user's state. Higher entropy means weaker regularity of the state distribution, which is more difficult to predict [51, 52]. The entropy of the 50 EV users ranges from 1.83 to 2.39, which is a very high variation. Therefore, the prediction results of different users could vary a lot. For example, the state prediction for EV user 3 (entropy 2.39) would be more difficult than that of EV user 4 (entropy 1.83).

Next, we make the parameter sensitivity analysis in the training process and discuss the performance of our proposed model. The dropout rate is not sensitive to the prediction accuracy based on our preliminary analysis and thus we set it as 0.3. To make the simple GRU and proposed Bi-GRU-Attention model more suitable for real applications in this case, two parameters in the model were tuned thoroughly, namely batch size and time window. For batch size, we only consider cases with batch size 64 or above. Because when the batch size is 32 or smaller, the prediction accuracy fluctuates heavily and does not even converge finally. The prediction length is set as 60 min during the calibration, which means that we predict the state of EV an hour ahead using the history state sequence. The parameter sensitivity analysis of GRU and Bi-GRU-Attention is shown in
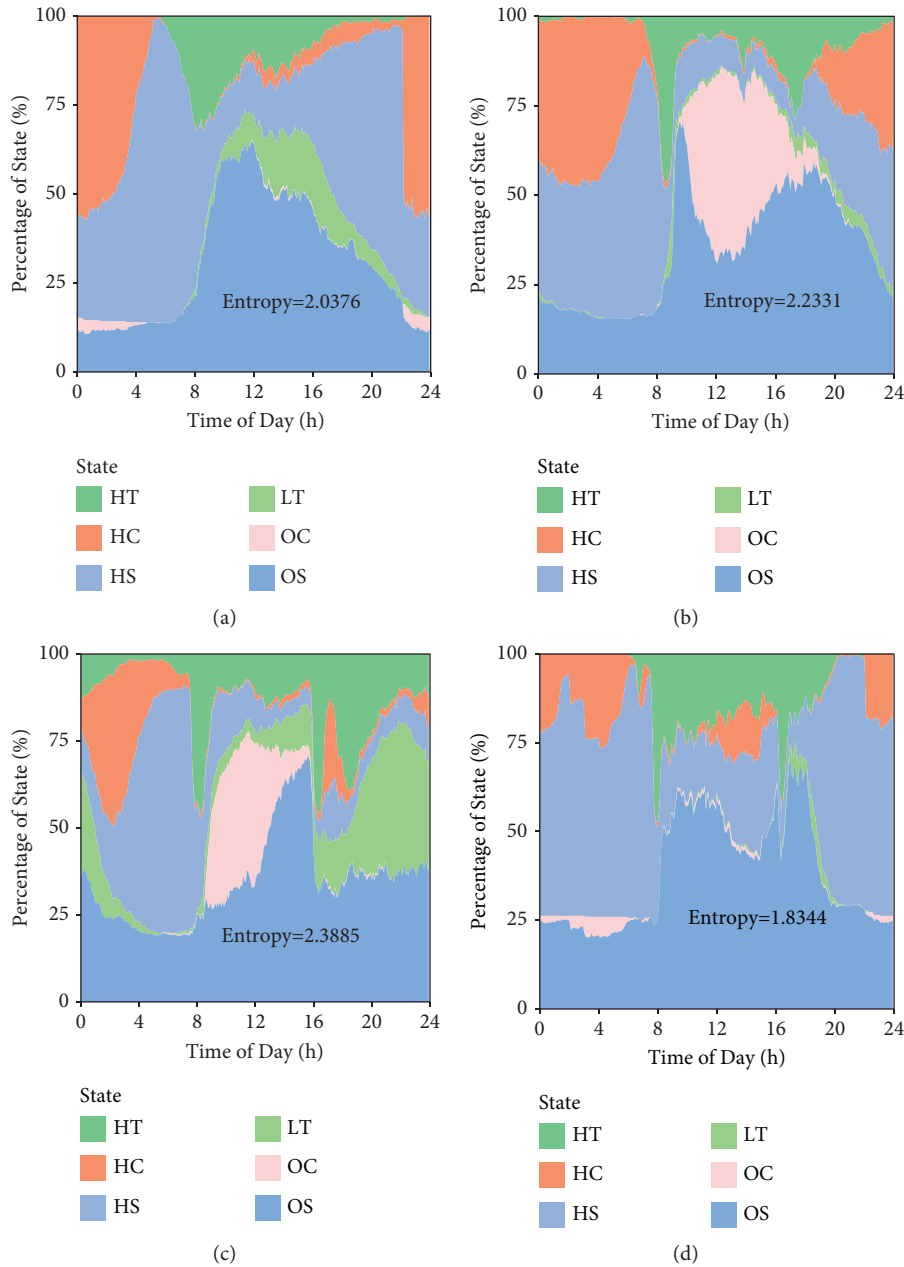
Figure 6: EV state distribution at different times of day. (a) EV user 1. (b) EV user 2. (c) EV user 3. (d) EV user 4.

Figure 7, and there are totally 20 combinations of parameters for each model. From the figures, we can find that both the time window and batch size show a significant relationship with the mean prediction accuracy of the EV fleet. It is interesting to find that, as the time window increases, the prediction accuracy increases a little (at 10 min) and then drops slowly. Regarding batch size, the prediction accuracy gradually decreases while batch size increase with the time window bigger than 10 min. This phenomenon is due to that bigger batch size requires only fewer iterations and consequently loses some accuracy. In general, both GRU and Bi-GRU-Attention achieve the highest prediction accuracy (76.75% and 77.15%, respectively) with batch size 64 and time window 10 min as marked in Figure 7. The following

GRU and Bi-GRU-Attention models are trained under this group of parameter settings unless otherwise stated. It is worth noting that the two parameters need recalibrating when using this model in other datasets/cities.

Based on the calibrated parameters, we compared the Bi-GRU-Attention model with two baseline models: GRU and n-gram. The performances of the three models (two baseline models) under various prediction lengths are shown in Figure 8. The prediction performances for all three models deteriorate slowly over the prediction length. It can be seen from Figure 8 that the Bi-GRU-Attention model outperforms GRU and n-gram in terms of prediction accuracy for all considered scenarios, especially when the prediction length is long (i.e., 180 min). When the prediction length is
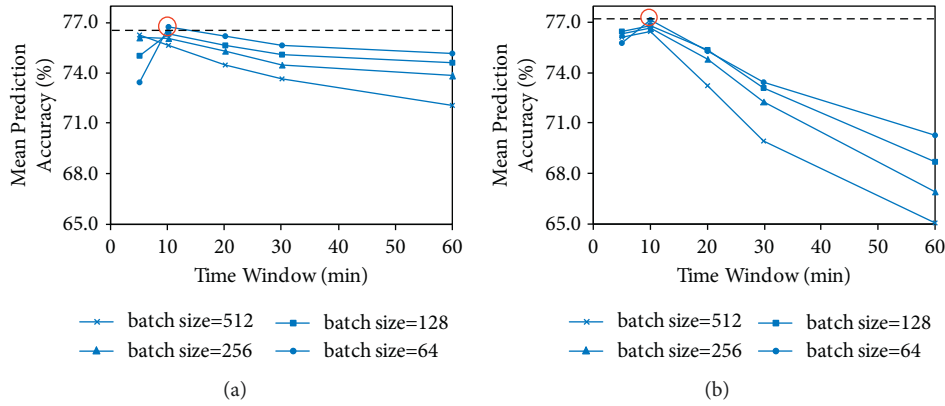
(a)

(b)

FIGURE 7: Parameter tuning and sensitivity analysis. (a) GRU. (b) Bi-GRU-Attention.
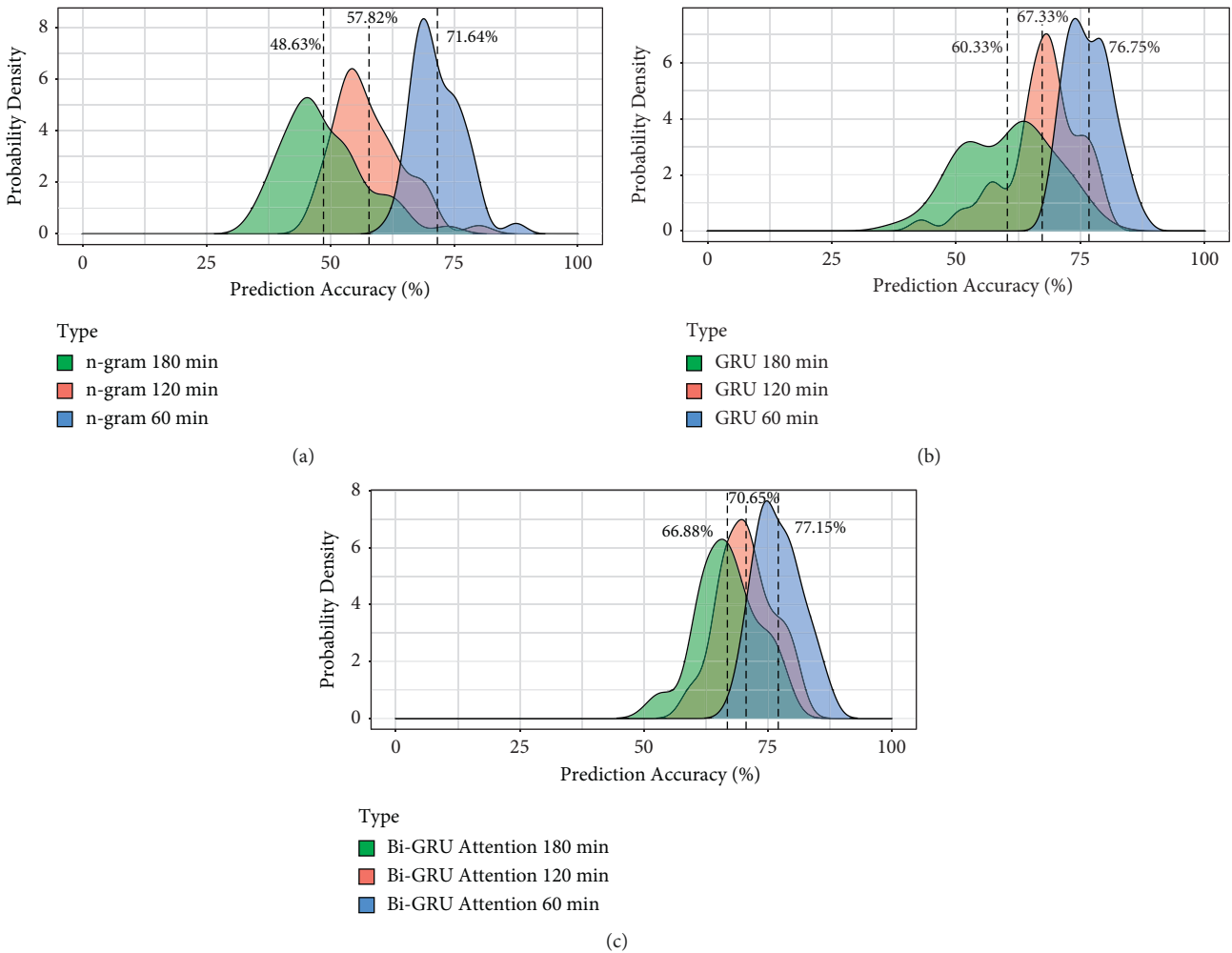

(a)

(b)

(c)

FIGURE 8: The performance of Bi-GRU-Attention compared with baseline models. (a) *n*-gram prediction accuracy. (b) GRU prediction accuracy. (c) Bi-GRU-Attention prediction accuracy.

180 min, the average prediction accuracy of the Bi-GRU-Attention model could still keep at 66.9% but that of GRU and n-gram drop down to 60.3% and 48.6%, respectively. Furthermore, the prediction accuracies of different EVs vary greatly, ranging from 67.7% to 87.5% when the prediction

length is 60 min, which is consistent with the findings of entropy variation as in Figure 6.

The prediction accuracies on different states are shown in Figure 9(a). The prediction accuracy varies greatly with different states, ranging from 38.4% to 89.0%. It can be
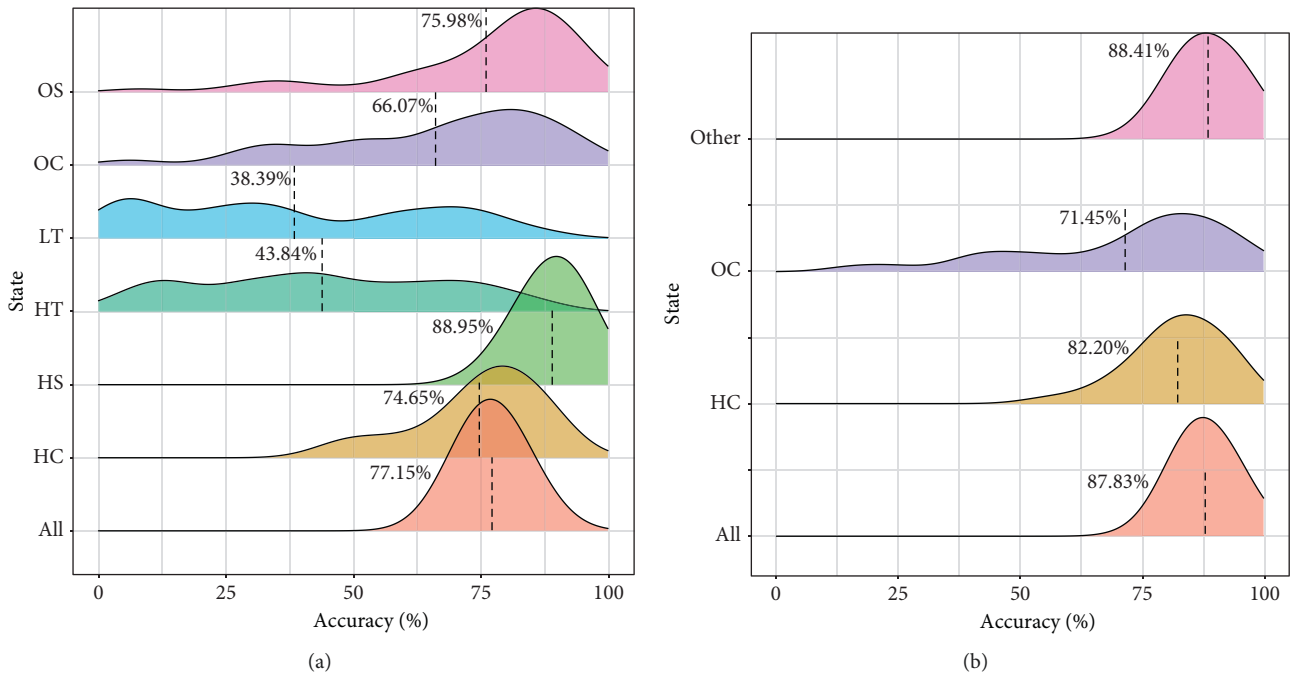
FIGURE 9: The prediction accuracy on different states. (a) Six types of state prediction. (b) Three types of state prediction.

observed that compared with the other states, the prediction accuracy of "HS" is much higher based on the proposed model. It indicates that the proposed model has a stronger prediction ability for state "HS" or the state "HS" is more predictable. By contrast, the state "LT" has a low prediction accuracy, which is probably because the frequency of "LT" state is very low as shown in Figure 6 and there is not much history for the model to learn about. Furthermore, both "LT" and "HT" have a wide prediction accuracy range for different users, indicating that the regularity of EV traveling varies a lot from individual to individual.

For cases where the operators only care about two EV charge states (home charge or outside charge), the Bi-GRU-Attention model could just predict 3 states ("OC," "HC," and others) to improve the prediction accuracy, as shown in Figure 9(b). The prediction accuracy for "OC" increases greatly from 66.07% to 71.45%, and the prediction accuracy for "HC" increases from 74.65% to 82.20%.

We also compared the prediction accuracy at different times of day to find out the more predictable time intervals as Figure 10 illustrated. Obviously, the prediction accuracies at 08:00 and 16:00 are relatively low, especially on weekdays (even lower than 60%). This is probably because at the outset of the day (around 08:00 am) and during the late afternoon (around 16:00) there are not many clues about the day's or the night's trip plans. On the contrary, the accuracy after midnight (02:00 am-05:00 am) is pretty high, even over 90% at 03:00 am. This result is as expected since most people would just charge or park their EVs at home during this time interval, which makes it easy to predict. Besides, the afternoon predictions differ a lot between weekdays and weekends, and the EV states on weekend afternoons are more predictable.
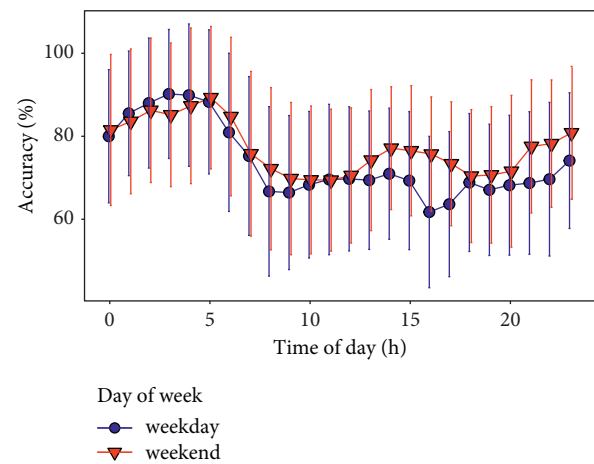

FIGURE 10: The prediction accuracy at different times of the day.

3.5. Stability and Robustness. Measurement noise is common in the real-world EV battery state monitoring process. To validate the performance of the proposed Bi-GRU-Attention model in the presence of measurement noise, Gaussian white noises with different variances are added in the measured SOC value. According to the state-of-art SOC estimation technology, the mean absolute error (MAE) range is 0.004–0.024 [53]. Hence, we set the maximum Gaussian noise variance as 0.06 (much bigger than the upper bound of estimation error). Figure 11 shows the mean prediction accuracy under different Gaussian noise variances of SOC. The embedded subplots visualize the Gaussian noises in one sample EV. We can find that as the Gaussian noise variance increases, the prediction accuracy decreases slightly. When the Gaussian noise variance reaches 0.06, the
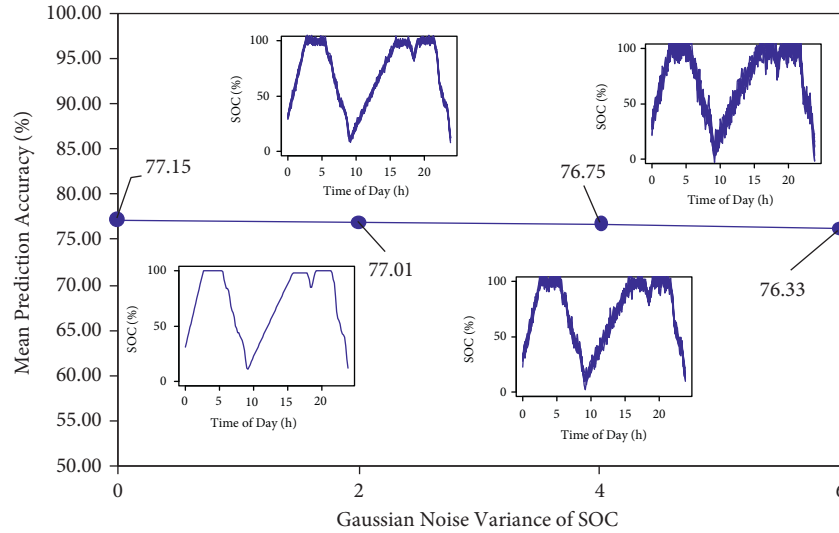
FIGURE 11: The prediction accuracy under different Gaussian noise variances of SOC.
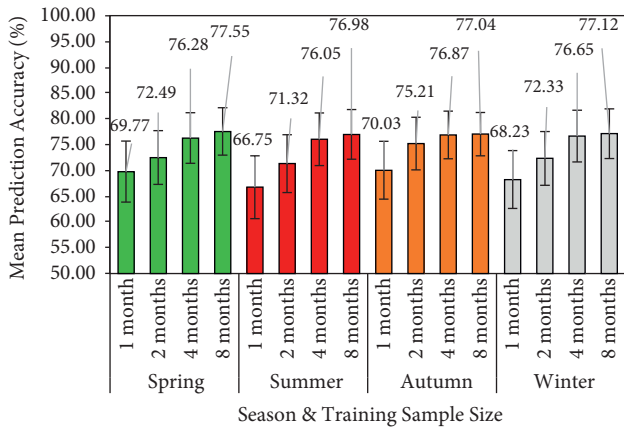

FIGURE 12: Model performance for different seasons and sample size.

mean prediction accuracy is 76.33% (0.82% reduction compared with no noise). Although the prediction accuracy decreased 0.82%, it still holds a good performance, indicating that the proposed Bi-GRU-Attention model is stable and insensitive to SOC noise. This is probably because we only predict the discrete EV states instead of the exact value.

To verify robustness of the trained Bi-GRU Attention model, four weeks' data (April 11–17, July 11–17, October 11–17, and January 11–17) are taken as the testing data to represent four seasons, and the other data are used to training the model. Furthermore, to find out the proper sample size (data requirements) when applied to other cities, we also select different sample sizes for training. The results from 16 scenarios (4 seasons ∗ 4 sample sizes) are compared in Figure 12. The prediction accuracy values for four seasons with 8 months' training data are 77.55%, 76.98%, 77.04%, and 77.12%, respectively, indicating that the proposed model has stable prediction performance in all seasons. In each season, the prediction accuracy increases significantly with the sample size. When the sample size is bigger than 4

months, the performance improvement for the model gradually slows down. It means that when using Bi-GRU Attention model in other cities, the sample size is suggested over 4 months.

4. Conclusions

In this study, we proposed the Bi-GRU-Attention model to predict the charging-related state of EV, which could improve the operation and management of charging facilities and guide nearby charging services for EV users. The charging-related states of EV were classified into six types, including "home charge," "outside charge," "home stop," "outside stop," "low-battery travel," and "high-battery travel." We also presented a solution to convert EV trajectory data into the above sequential state data. The proposed data converting solution and Bi-GRU-Attention model for charging-related state prediction have been validated by plug-in hybrid EV trajectory data collected in Shanghai, China. The results showed that the prediction accuracy of the proposed approach is sensitive to the length of time interval (time window) and batch size (parameter in Bi-GRU-Attention model training). We have obtained the highest prediction accuracy (77.15%) with the time window of 10 min and batch size 64 by using the proposed method. It outperforms the traditional sequence prediction model, n-gram, and simple GRU under all tested prediction lengths.

Numerous analyses also show that the prediction accuracy varies greatly in different states. Among all charging-related states, the prediction accuracy of "HS" derived by the proposed model ranks in the top (89.0%), while the prediction accuracy of "LT" is very low (38.4%). Furthermore, the prediction accuracy also varies widely over time of day. In detail, we obtain the lowest prediction accuracy around 08:00 am and 04:00 pm, and the highest one for midnight (02:00 am - 05:00 am). The stability and robustness analysis implies that the proposed Bi-GRU-Attention model is stable and insensitive to SOC noise or season. In summary, this

study contributes substantially to the prediction of short-term EV states. The proposed deep learning approach for EV state prediction opens an interesting direction for future research on transportation sequential data analysis.

There are some limitations to this study. Firstly, we use only 1-year trajectory data of 50 plug-in hybrid EVs to validate our approach. A larger number of samples from other cities and pure EVs can be used to test the robustness of the proposed deep learning approach in the future. Secondly, EV charging behavior is divided into "home charge" and "outside charge" in this study. More granular divisions (i.e., workplace charge and public charge) are not considered in this study due to data limitation. Hence, this study could be extended and improved by considering more specific charging states in the sequence or training with larger datasets to enhance the predictive power of the Bi-GRU-Attention model.

Data Availability

The data used to support the findings of this study were supplied under license and so cannot be made freely available. The data can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] O. Veneri, *Technologies and Applications for Smart Charging of Electric and Plug-In Hybrid Vehicles*, Springer, Berlin, Germany, 2016.

[2] S. Bakker, K. Maat, and B. van Wee, "Stakeholders interests, expectations, and strategies regarding the development and implementation of electric vehicles: the case of The Netherlands," *Transportation Research Part A: Policy and Practice*, vol. 66, pp. 52–64, 2014.

[3] R. Rao, H. Cai, and M. Xu, "Modeling electric taxis' charging behavior using real-world data," *International Journal of Sustainable Transportation*, vol. 12, no. 6, pp. 452–460, 2018.

[4] S. Hardman, "Understanding the impact of reoccurring and non-financial incentives on plug-in electric vehicle adoption - a review," *Transportation Research Part A: Policy and Practice*, vol. 119, pp. 1–14, 2019.

[5] B. Yu, Y. Ma, M. Xue et al., "Environmental benefits from ridesharing: a case of Beijing," *Applied Energy*, vol. 191, pp. 141–152, 2017.

[6] G.A.&T.T. Global, *Electric Vehicle Market Outlook 2018*, Frost & Sullivan, San Antonio,TX, USA, 2018.

[7] X. Qu, Y. Yu, M. Zhou, C.-T. Lin, and X. Wang, "Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach," *Applied Energy*, vol. 257, Article ID 114030, 2020.

[8] M. Xu, Q. Meng, K. Liu, and T. Yamamoto, "Joint charging mode and location choice model for battery electric vehicle users," *Transportation Research Part B: Methodological*, vol. 103, pp. 68–86, 2017.

[9] N. Shahraki, H. Cai, M. Turkay, and M. Xu, "Optimal locations of electric public charging stations using real world vehicle travel patterns," *Transportation Research Part D: Transport and Environment*, vol. 41, pp. 165–176, 2015.

[10] R. Philipsen, T. Brell, W. Brost, T. Eickels, and M. Ziefle, "Running on empty - u," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 59, pp. 475–492, 2018.

[11] X.-H. Sun, T. Yamamoto, and T. Morikawa, "Fast-charging station choice behavior among battery electric vehicle users," *Transportation Research Part D: Transport and Environment*, vol. 46, pp. 26–39, 2016.

[12] P. Huang, Z. Ma, L. Xiao, and Y. Sun, "Geographic Information System-assisted optimal design of renewable powered electric vehicle charging stations in high-density cities," *Applied Energy*, vol. 255, Article ID 113855, 2019.

[13] X. Hao, Z. Lin, H. Wang, S. Ou, and M. Ouyang, "Range cost-effectiveness of plug-in electric vehicle for heterogeneous consumers: an expanded total ownership cost approach," *Applied Energy*, vol. 275, Article ID 115394, 2020.

[14] L. Luo, W. Gu, Z. Wu, and S. Zhou, "Joint planning of distributed generation and electric vehicle charging stations considering real-time charging navigation," *Applied Energy*, vol. 242, pp. 1274–1284, 2019.

[15] M. E. Biresselioglu, M. Demirbag Kaplan, and B. K. Yilmaz, "Electric mobility in Europe: a comprehensive review of motivators and barriers in decision making processes," *Transportation Research Part A: Policy and Practice*, vol. 109, pp. 1–13, 2018.

[16] C. Zhang, J. B. Greenblatt, P. MacDougall, S. Saxena, and A. Jayam Prabhakar, "Quantifying the Benefits of Electric Vehicles on the Future Electricity Grid in the Midwestern United States," *Applied Energy*, vol. 270, 2020.

[17] J. Davies and K. S. Kurani, "Recharging behavior of households' plug-in hybrid electric vehicles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2191, no. 1, pp. 75–83, 2010.

[18] T. Franke and J. F. Krems, "Understanding charging behaviour of electric vehicle users," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 21, pp. 75–89, 2013.

[19] J. Dong and Z. Lin, "Stochastic modeling of battery electric vehicle driver behavior," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2454, no. 1, pp. 61–67, 2014.

[20] N. Daina, J. W. Polak, and A. Sivakumar, "Patent and latent predictors of electric vehicle charging behavior," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2502, no. 1, pp. 116–123, 2015.

[21] X. H. Sun, T. Yamamoto, and T. Morikawa, "Charge timing choice behavior of battery electric vehicle users," *Transportation Research Part D: Transport and Environment*, vol. 37, pp. 97–107, 2015.

[22] Y. Yang, E. Yao, Z. Yang, and R. Zhang, "Modeling the charging and route choice behavior of BEV drivers," *Transportation Research Part C: Emerging Technologies*, vol. 65, pp. 190–204, 2016.

[23] Y. W. Chung, B. Khaki, T. Li, C. Chu, and R. Gadh, "Ensemble machine learning-based algorithm for electric vehicle user behavior prediction," *Applied Energy*, vol. 254, Article ID 113732, 2019.

[24] S. Hardman, A. Jenn, G. Tal et al., "A review of consumer preferences of and interactions with electric vehicle charging infrastructure," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 508–523, 2018.

[25] K. Liu, K. Li, and C. Zhang, "Constrained generalized predictive control of battery charging process based on a coupled thermoelectric model," *Journal of Power Sources*, vol. 347, pp. 145–158, 2017.

[26] Q. Ouyang, Z. Wang, K. Liu, G. Xu, and Y. Li, "Optimal charging control for lithium-ion battery packs: a distributed average tracking approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3430–3438, 2020.

[27] X. Y. Zhang, F. Yin, Y. M. Zhang, C. L. Liu, and Y. Bengio, "Drawing and recognizing Chinese characters with recurrent neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 849–862, 2018.

[28] S. Liu, N. Yang, M. Li, and M. Zhou, "A recursive recurrent neural network for statistical machine translation," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 1491–1500, Baltimore, MD, USA, June 2014.

[29] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based Natural Language processing [review article]," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.

[30] X. L. Ma, Z. M. Tao, Y. H. Wang, H. Y. Yu, and Y. P. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[31] X. Wang, R. Jiang, L. Li, Y. L. Lin, X. H. Zheng, and F. Y. Wang, "Capturing car-following behaviors by deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 910–920, 2018.

[32] M. Simoncini, L. Taccari, F. Sambo, L. Bravi, S. Salti, and A. Lori, "Vehicle classification from low-frequency GPS data with recurrent neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 91, pp. 176–191, 2018.

[33] Y. Tian, R. Lai, X. Li, L. Xiang, and J. Tian, "A combined method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network and an adaptive cubature Kalman filter," *Applied Energy*, vol. 265, Article ID 114789, 2020.

[34] K. Liu, Z. Wei, C. Zhang, Y. Shang, R. Teodorescu, and Q.-L. Han, "Towards long lifetime battery: AI-based manufacturing and management," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 7, pp. 1139–1165, 2022.

[35] A. Farmann and D. U. Sauer, "A comprehensive review of on-board State-of-Available-Power prediction techniques for lithium-ion batteries in electric vehicles," *Journal of Power Sources*, vol. 329, pp. 123–137, 2016.

[36] K. Liu, Q. Peng, H. Sun, M. Fei, H. Ma, and T. Hu, "A transferred recurrent neural network for battery calendar health prognostics of energy-transportation systems," *IEEE Transactions on Industrial Informatics*, p. 1, 2022.

[37] K. Liu, Y. Shang, Q. Ouyang, and W. D. Widanage, "A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3170–3180, 2021.

[38] J. Hong, Z. Wang, W. Chen, L.-Y. Wang, and C. Qu, "Online joint-prediction of multi-forward-step battery SOC using LSTM neural networks and multiple linear regression for real-

[39] W. Houlian and Z. Gongbo, "State of charge prediction of supercapacitors via combination of Kalman filtering and backpropagation neural network," *IET Electric Power Applications*, vol. 12, no. 4, pp. 588–594, 2018.

[40] F. Zhao, Y. Li, X. Wang, L. Bai, and T. Liu, "Lithium-Ion batteries state of charge prediction of electric vehicles using RNNs-CNNs neural networks," *IEEE Access*, vol. 8, Article ID 98168, 2020.

[41] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[42] K. Cho, B. V. Meerienboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, Doha, Qatar, October 2014.

[43] M. Wollmer, C. Blaschke, T. Schindl et al., "Online driver distraction detection using long short-term memory," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 574–582, 2011.

[44] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," 2014, https://arxiv.org/abs/1409.0473?context=cs.NE.

[45] S. Hao, D.-H. Lee, and D. Zhao, "Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system," *Transportation Research Part C: Emerging Technologies*, vol. 107, pp. 287–300, 2019.

[46] X. L. Huang, J. Sun, and J. Sun, "A car-following model considering asymmetric driving behavior based on long short-term memory neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 346–362, 2018.

[47] K. Liu, T. Yamamoto, and T. Morikawa, "Impact of road gradient on energy consumption of electric vehicles," *Transportation Research Part D: Transport and Environment*, vol. 54, pp. 74–81, 2017.

[48] Research-in-China, "China EV Charging Station and Charging Pile Market Report," Research-in-China, Beijing, China, JAF007, 2017.

[49] D. Potoglou, "Vehicle-type choice and neighbourhood characteristics: an empirical study of Hamilton, Canada," *Transportation Research Part D: Transport and Environment*, vol. 13, no. 3, pp. 177–186, 2008.

[50] M. Baresch and S. Moser, "Allocation of e-car charging: assessing the utilization of charging infrastructures by location," *Transportation Research Part A: Policy and Practice*, vol. 124, pp. 388–395, 2019.

[51] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Individual mobility prediction using transit smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 19–34, 2018.

[52] C. Song, Z. Qu, N. Blumm, and A. L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[53] P. Xu, B. Liu, X. Hu, T. Ouyang, and N. Chen, "State-of-Charge estimation for lithium-ion batteries based on fuzzy information granulation and asymmetric Gaussian membership function," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 7, pp. 6635–6644, 2022.