processes

Article

Electric Load Prediction of Electric Vehicle Charging Stations Based on Moving Average–Gated Recurrent Unit

Wei Huang [1,*], Chuanhong Ru [1], Jian Qin [1], Yong Lin [1], Qingxi Cai [1] and Bing Song [2]

1 State Grid Zhejiang Electric Power Company, Ltd., Hangzhou 310063, China
2 Key Laboratory of Advanced Control and Optimization for Chemical Processes, East China University of Science and Technology, Shanghai 200237, China
* Correspondence: tiankongyiran90@163.com

Abstract: The load prediction of electric vehicle charging stations is the basis of their static safety, which directly affects the safety of operation, the rationality of planning, and the economy of supply. However, various factors lead to drastic changes in short-term power consumption, which makes the data more complicated and difficult to predict. In this paper, the moving average–gated recurrent unit method is proposed to predict the electric load of electric vehicle charging stations. A prediction model is established based on the historical data of electric load of electric vehicle charging stations to realize the accurate prediction of future electric loads. Firstly, considering the problems of noise in the historical data of electric vehicle charging stations, the moving average method is used for smoothing. Secondly, the smoothed data are modeled by the gated recurrent unit, and the future prediction results are obtained. Finally, the validity and practicability of the proposed method are proved by the research and testing of the actual electric vehicle charging station power load dataset. Compared with the classic LSTM prediction model, the proposed MA-GRU method can achieve more accurate prediction performance.

Keywords: load prediction; electric vehicle charging station; data driven; moving average; gated recurrent unit

Academic Editor: Takuya Oda

Received: 28 September 2024
Revised: 22 October 2024
Accepted: 28 October 2024
Published: 28 February 2025

Citation: Huang, W.; Ru, C.; Qin, J.; Lin, Y.; Cai, Q.; Song, B. Electric Load Prediction of Electric Vehicle Charging Stations Based on Moving Average–Gated Recurrent Unit. *Processes* **2025**, *13*, 706. https://doi.org/10.3390/pr13030706

1. Introduction

Due to global climate change and the requirements of environmental policies in various countries, the popularity of electric vehicles is gradually increasing. The widespread use of electric vehicle charging stations would cause fluctuations in the power grid during charging. Therefore, accurately predicting the load of electric vehicle charging stations is crucial for the planning, operation, and management of the power system. It can effectively balance the load of the power grid, improve reliability, and maintain safe and stable operation of electric vehicle charging stations.

Compared to traditional power system load forecasting, electric vehicle charging station load prediction is more challenging due to its high randomness and complexity. The prediction methods include model-based methods and data-driven methods. Model-based methods require the establishment of accurate mathematical models, such as Monte Carlo simulations and the Markov travel chain. Due to the high complexity of predicting the load of electric vehicle charging stations, it is difficult to establish an accurate mathematical model. Therefore, data-driven methods are widely used. Data-driven prediction methods can be further divided into machine learning methods and deep learning methods, which can be found in Figure 1. The prediction methods based on typical machine learning mainly include support vector regression (SVR), principal component regression (PCR), and partial least squares (PLS) and so on.

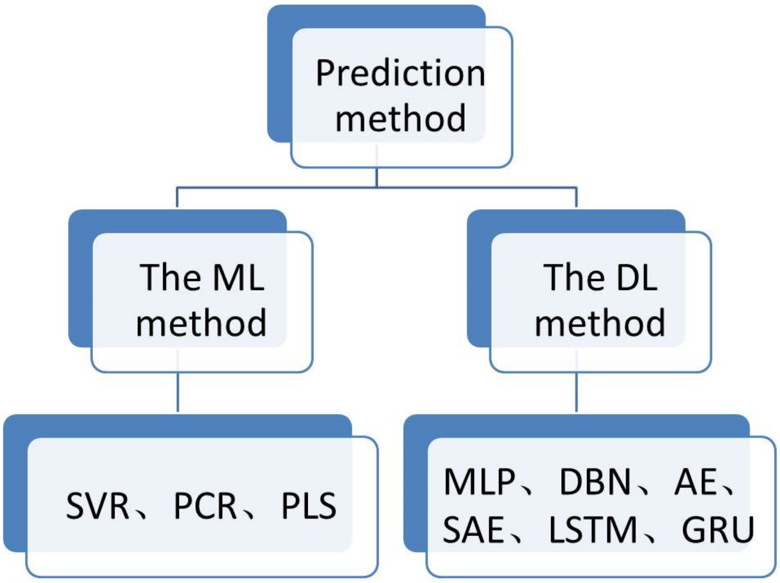

Figure 1. Classification chart of prediction methods.

Machine learning methods [1–3] are one of the data-driven approaches that focus on using data to mimic the way humans learn through algorithms. Machine learning is an important part of the growing field of data science by using statistical methods to train algorithms to complete prediction tasks. In machine learning methods, PLS [4,5] is a method commonly used to construct prediction models by projecting process variables and mass variables into a new space to find linear regression relationships to complete the construction of prediction models. Wang et al. [6] conducted a comprehensive evaluation of different variable selection methods based on PLS predictive models and proposed a new metric method to evaluate the performance of different variable selection methods. SVR [7] is also a common prediction method, and its principle is similar to that of support vector machines (SVMs). SVMs aim to find a hyper-plane in N-dimensional space that can categorize the number of data points explicitly, which is mainly used for classification tasks, while SVR aims to find the best hyper-plane that can fit the most data points, which is more suitable for predictive regression tasks. Moreover, most of the above prediction methods of machine learning are linear models. When the system presents serious nonlinearity, prediction accuracy is greatly affected.

In recent years, deep learning (DL) methods [8–10] have been further developed for prediction tasks. Faced with the high dimension, strong correlation, and high redundancy of raw data, DL methods have superior feature extraction ability to process these data, so it has great potential in prediction modeling. Among them, a neural network [11] consists of multiple hidden layers that can be activated using functions such as RELU [12], SELU [13], tanh [14], or softmax [15] to cope with the strong nonlinearity of the variable. In addition, neural networks show a strong ability to deal with large-scale, high, and nonlinear regression problems, so they are widely used in the construction of prediction models.

With the further development of neural-network-related research, more and more new network structures have been developed and applied to prediction tasks. A multi-layer perceptron (MLP) [16] is a forward-structured artificial neural network consisting of an input layer, multiple hidden layers, and an output layer, each of which is fully connected and activated by a function, and it is the most basic type of neural network. Deep belief nets (DBNs) [17,18] are composed of a multi-layer restricted Boltzmann machine (RBM) and a layer of back propagation (BP) neural networks. Their essence is a kind of specially constructed neural network. In DBN, the input data are first passed through multiple RBM layers, and in each layer, the output of the RBM of the upper layer is used as the input of the RBM of the next layer, and the output information of the RBM of the last layer is trained as the input data of other networks. Multi-layer RBMS can better capture the joint probability distribution of observation data and labels, which can make better predictions. Wang et al. [19] proposed a prediction modeling method based on DBN and an extreme learning machine, which used DBN to extract the features of process variables and input the extracted features into the learning machine for training to obtain a prediction model. Zhu et al. [20] developed a prediction model of a polymer melt index using DBN. Compared with traditional neural networks, DBN training includes both supervised and unsupervised training stages. In order to mine valuable information from the data of the polymerization process, in the unsupervised training phase, DBN is trained using process data without variable labels to improve the estimated performance. An auto encoder (AE) [21] is an unsupervised learning model that uses the input data themself as the label to obtain the reconstructed output through the fully connected layer and uses the difference between the reconstructed output and the input for training. The smaller the difference, the better the features extracted from the hidden layer can represent the information contained in the process variables. A stacked auto encoder [22,23] (SAE) is a deep network form of AE, consisting of multiple AEs stacked together. SAE can be seen as a change to replace the RBM layer of DBN with the AE layer. The features extracted from the previous layer of AE are used as input and reconstructed output for the next layer of AE, while the features extracted from the last layer can be trained as input data for other networks. A multi-layer AE can make the features extracted by the network contain deeper input information, so SAE is often used to extract the features of variables. Liu et al. [24] proposed an SAE that could preserve both non-local and local structures, considering that the extraction of features by minimizing the global reconstruction error would lead to the loss of intrinsic geometric structure information in the original data. Yuan et al. [25] proposed a novel variable weighted SAE. Through correlation analysis with quality variables, important variables were identified from variables in each AE input layer, and different weights were assigned according to their correlation degrees, thus designing an improved AE model. Then, the improved AE model is stacked to form a deep network to complete the construction of the entire prediction model.

Considering the temporal relationship between variables, long short-term memory (LTSM) neural networks [26,27] are also widely used in the construction of prediction models. LSTM is a special recurrent neural network [28,29] that sets the weight of the information of the past moment and the current moment through the forgetting gate, the input gate, and the output gate, so as to find the temporal relationship between the output and the input. Zhou et al. [30] combined the dynamic time features of the difference variables with the nonlinear features of the sequence data by introducing the dynamic information of the input, established a new network unit, and applied it to the LSTM, so as to build a prediction model. Considering that LSTM never considers the correlation between process variables and quality variables, Yuan et al. [31] proposed a prediction modeling method based on spatiotemporal attention LSTM. By considering the interaction of spatiotemporal mass, the predictive performance of the prediction model is improved. A gated recurrent unit is also a typical recurrent neural network, which is used to solve the problem of long-term memory and gradient disappearance in back-propagation like LSTM, and it is widely used in time-series data prediction. The above prediction network method based on deep learning has good prediction performance for nonlinear systems. However, the input of these networks often lacks the preprocessing of the original data.

In order to predict the electric load of electric vehicle charging stations, a moving window–gated recurrent unit method is proposed. Based on the historical data of electric vehicle charging stations, a big data-driven electric vehicle charging station electric load forecasting model is established to predict the electric vehicle charging station electric load, so as to ensure the safe, high-quality, and economical operation of the electric vehicle charging station. First of all, given that the collected historical data of electric loads of electric vehicle charging stations always contains some noise points, the moving average is used to smooth it, so that the smoothed dataset can effectively eliminate noise. Secondly, the prediction model is established based on a gated recurrent unit, which can effectively consider the time-series relationship between data. Finally, the proposed method is applied to electric vehicle charging stations and compared with the traditional methods to show the effectiveness and superiority of this method.

The main contributions of our work in this manuscript include the following:

(1) The moving window–gated recurrent unit method is proposed to predict the electric load of electric vehicle charging stations.
(2) A big data-driven electric vehicle charging station electric load forecasting model is established.
(3) The moving average is used to smooth the collected historical data of electric loads of electric vehicle charging stations.

2. The Proposed Moving Average–Gated Recurrent Unit Method

2.1. The Smoothing Method Moving Average

The nature of noise is a random error or deviation in a measured variable, including the wrong value and deviation from the expected isolated point. When noise is involved in the process of prediction analysis, it not only increases the amount of data but also increases the amount of calculation and the calculation error. The moving average smooths the data sequence by calculating the average value of the data within the specified window. Its core idea is to calculate the average value of all data points contained in a fixed size window. Then, this window slides forward in the data series and repeatedly calculates the average value of each position. In this way, the data can be smoothed, and the influence of random fluctuation can be reduced.

The time-series data are $X = [x_1, x_2, \cdots, x_n]$. First of all, they are normalized in the following ways:

$$x_t = \frac{\max(X) - x_t}{\max(X) - \min(X)} \tag{1}$$

where $\max(X)$ is the maximum value in X, and $\min(X)$ is the maximum value in X.

Let the window size be w, and then the calculation formula of the moving average X_t at time point t is the following:

$$X_t = \frac{1}{w}\sum_{i=0}^{w-1} x_{t-i} \tag{2}$$

where X_t is the moving average at time t, and x_{t-i} is the data at the time point $t-i$.

2.2. The Electric Load Prediction Model Construction of Electric Vehicle Charging Stations

The gated recurrent unit (GRU) is a recursive neural network which aims to solve the problems of gradient disappearance and gradient explosion in traditional RNNs, especially when dealing with long sequence data. Compared with the LSTM method, the GRU has fewer training parameters, is easier to train, and can greatly improve the training efficiency. So, GRU can achieve better training results.

The structure of GRU is shown in Figure 1, which consists of an update gate and a reset gate. The reset gate determines the number of old states retained, and the update gate determines the number of old states contained in the new state. These gating mechanisms help the model decide when to update or reset the hidden state.

In Figure 2 below, R_t is the reset gate, Z_t is the update gate, W_z, U_z, W_t, U_t are weight coefficient matrixes, and b_z, b_r are offset vectors. $\widetilde{H}_t$ represents the candidate hidden state at time t, H_t represents the hidden state at time t, and X_t is the network input at time t. H_{t-1} is the hidden states at time $t-1$, and H_t is jointly decided by H_{t-1} and $\widetilde{H}_t$. Z_t determines how much information in H_{t-1} comes from H_{t-1} and how much information comes from $\widetilde{H}_t$. A larger value of Z_t means that more new information will be brought to the next state, and a larger value of R_t means that more information from the previous state can be ignored.

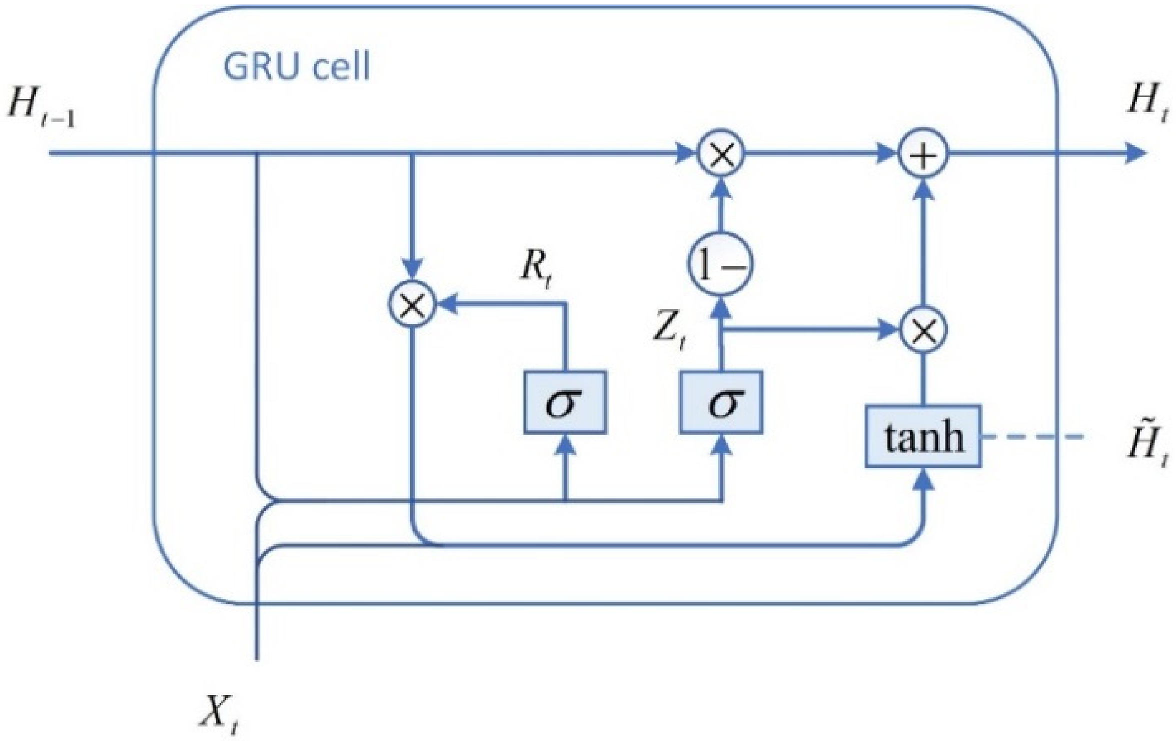

Figure 2. The structure of GRU.

The update gate determines how the hidden state of the current time step should be updated from the hidden state of the previous time step. It controls the information flow between the previous state and the current state. The calculation formula for the update gate is as follows:

$$Z_t = \sigma(W_z H_{t-1} + U_z X_t + b_z) \tag{3}$$

where Z_t is the output of the update gate, σ is the sigmoid function, W_z is the weight matrix, H_{t-1} is the hidden state of the previous time step, and X_t is the input of the current time step.

The reset gate determines the degree to which the hidden state of the previous time step should be forgotten in the current time step. It allows the model to selectively discard past information, thus enabling the model to capture short-term dependencies. The calculation formula of the reset gate is as follows:

$$R_t = \sigma(W_t H_{t-1} + U_t X_t + b_r) \tag{4}$$

where R_t is the output of the reset gate, and W_t is the weight matrix.

The hidden state is updated by combining the outputs of the update gate and the reset gate. The reset gate controls the influence of the previous hidden state, and the update gate determines the final value of the hidden state. The update formula for hidden status is as follows:

$$\widetilde{H}_t = \tanh(W_h[R_t \times H_{t-1}] + U_h X_t + b_h) \tag{5}$$

$$H_t = (1 - Z_t) \times H_{t-1} + Z_t \times \widetilde{H}_t \tag{6}$$

where $\widetilde{H}_t$ is the candidate hidden state, and H_t is the final hidden state.

The advantages of GRU are listed as follows:

(1) Less parameters: compared with LSTM, GRU has fewer parameters, which makes it easier to train and performs better in some cases.
(2) Capturing long-term and short-term dependencies: GRU can effectively capture long-term and short-term dependencies through a gating mechanism, especially when processing long-sequence data, which is superior to traditional RNN.
(3) High computational efficiency: GRU is more efficient than LSTM because of its simpler structure, and it has advantages when dealing with large-scale data.

When X_t is input into GRU, H_t can be obtained, which is the time-series feature. Then, according to this feature H_t, the future value of the input sequence X_t can be predicted. The flow chart of the proposed method is given in Figure 3.

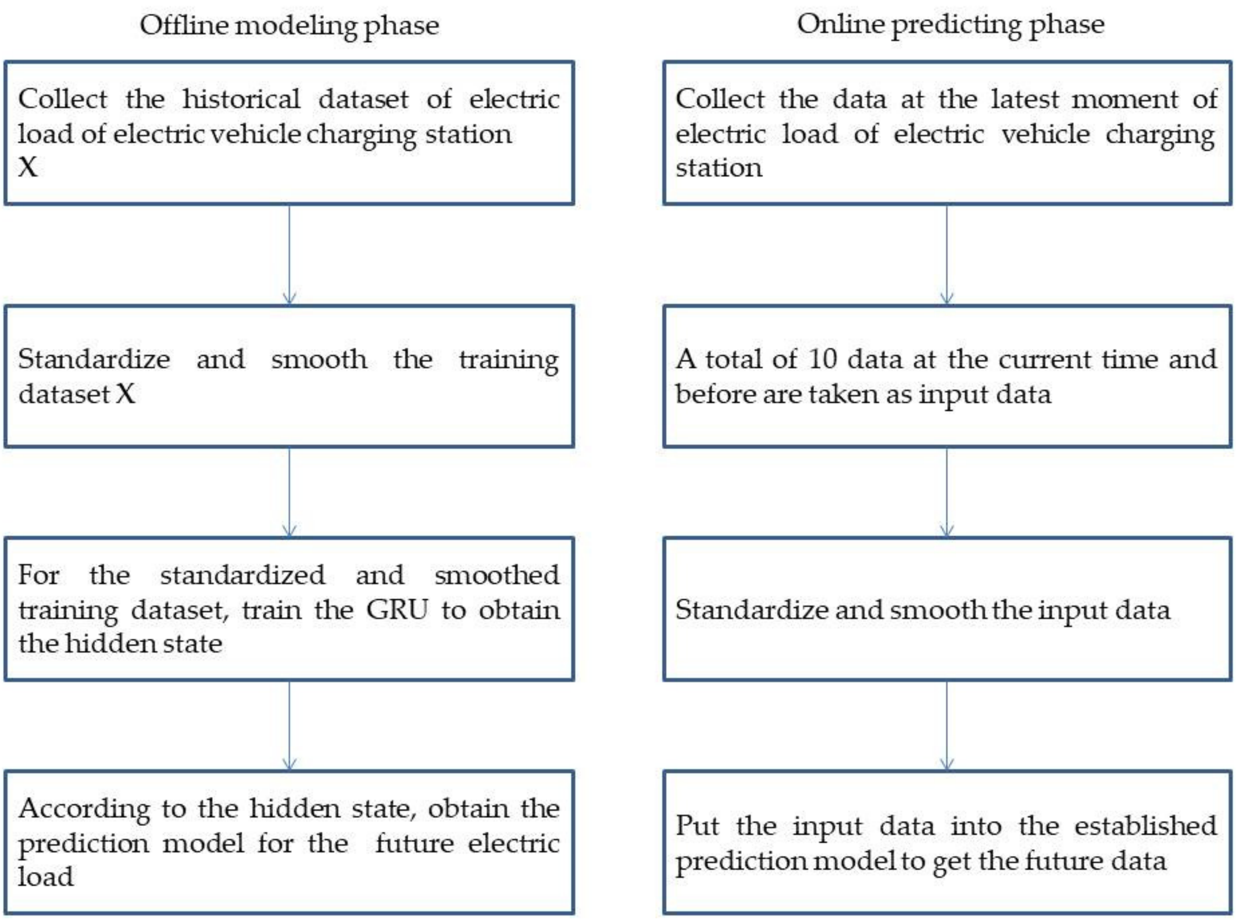

Figure 3. Classification chart of prediction methods.

For planners of electric vehicle charging infrastructure, the proposed MA-GRU model can help to effectively balance the load of the power grid and improve the reliability of the electric vehicle charging station. The complexity of the proposed MA-GRU method is mainly reflected in the offline modeling stage, that is, the training stage of the network. For practical applications, online forecasting only needs to input the corresponding data into the constructed model. In addition, in order to enhance the accuracy of the prediction model, in practical applications, the latest data can be incrementally updated for the offline-built model.

3. Case Study

The proposed MA-GRU method was tested and verified on the UK EV energy task force dataset (https://github.com/CoogyEoin/EV-charger-power-forecast-model/blob/master/ev_data.csv, accessed on 20 August 2024), which is a practical dataset in the real world. The dataset is released by the UK EV energy task force, which contains the charging behavior data of electric vehicles, especially the charging habits of users in different regions of the UK. The purpose of this dataset is to analyze the charging mode of electric vehicle users in Britain, study the charging demand in different regions, and evaluate the infrastructure layout and power grid load. The dataset consists of the following attributes:

Charging session ID: uniquely identifies each charging session.

Date and time: the specific date and time of charging.

Charging station ID: indicates the charging station where the charging session is located.

Charging time: the duration of each charging session in minutes.

Charging power: the recorded power changes during charging.

Charging capacity: the total energy of each charge in kWh.

Geographical location: the geographical coordinates of the charging station.

The charging power is selected as the variable to be predicted, and a total of 200 samples are collected as the training dataset for prediction model establishment. The input of the proposed method is the data of the 10 moments before the current moment, and the output is the data of the next moment after the current moment. In order to verify the effectiveness of the proposed MA-GRU method, it is compared with the LSTM method, which is widely used in electric vehicle charging station electric load prediction. In addition, in order to prove the effectiveness of the moving average method for data smoothing, the proposed MA-GRU method is compared with the method directly using GRU, and the MA-LSTM method is compared with the method directly using LSTM.

There are 50 data to be predicted in the testing dataset. This paper evaluates the model performance by calculating the root-mean-square error (RMSE) of the model on the testing dataset. RMSE reflects the deviation between the predicted output and the actual output of the testing dataset. The smaller the RMSE value, the more accurate the predicted value will be. Therefore, RMSE can reflect the accuracy of the prediction model. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(X_t - \widetilde{X}_t\right)^2} \quad (7)$$

where N represents the number of samples in the testing dataset, and X_t and $\widetilde{X}_t$ represent the actual value and predicted value of the charging power variable, respectively.

Figure 4 gives prediction results of the LSTM method. Figure 5 lists prediction results of the GRU method. Figure 6 shows the prediction results of the MA-LSTM method. Figure 7 shows the prediction results of the MA-GRU method. In these four figures, 50 samples in the testing dataset are predicted, respectively, according to the changes of 200 samples in the training dataset. In Figures 4–7, the horizontal axis denotes the sample point, and the vertical axis is the charging power. Figure 8 gives the bar plot of RMSE for four methods, and Table 1 lists RMSE of four methods.

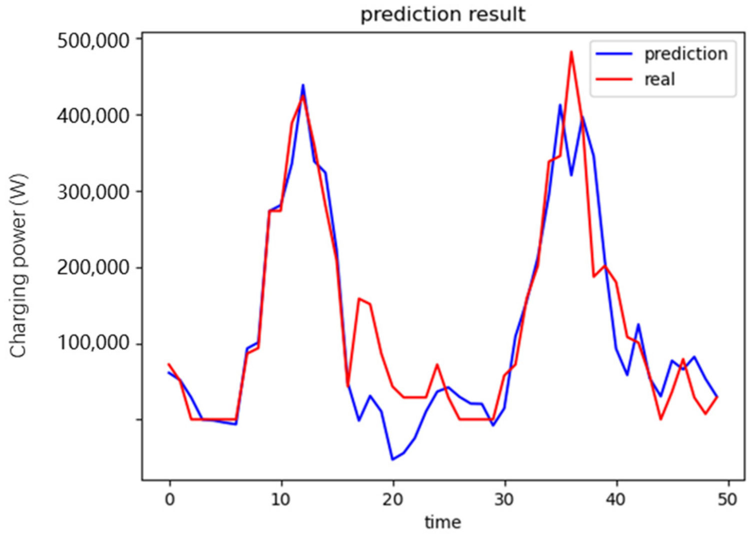

Figure 4. Prediction results of the LSTM method.

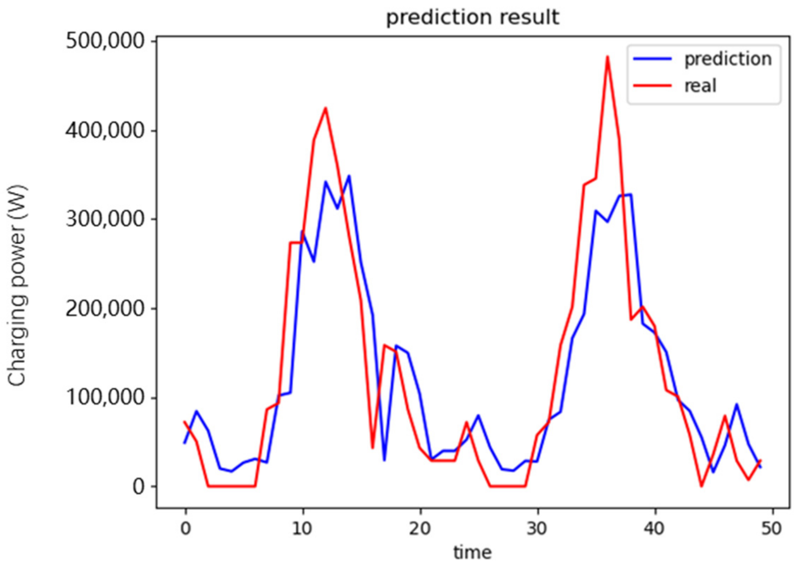

Figure 5. Prediction results of the GRU method.

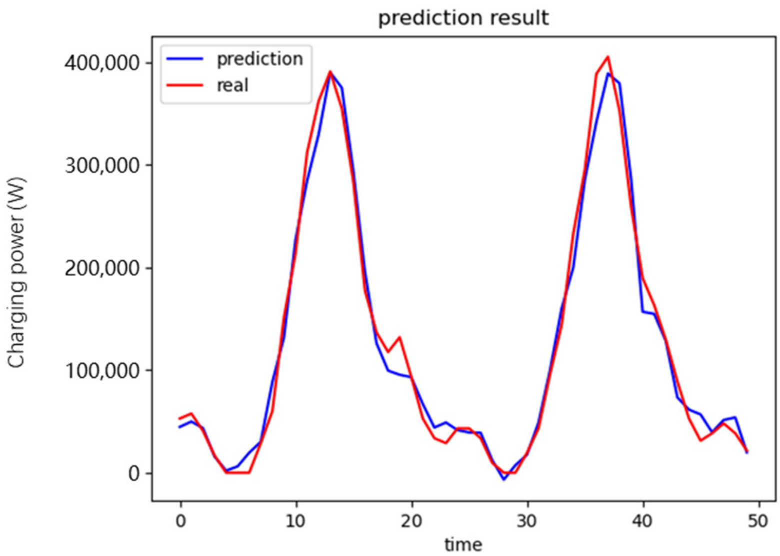

Figure 6. Prediction results of the MA-LSTM method.

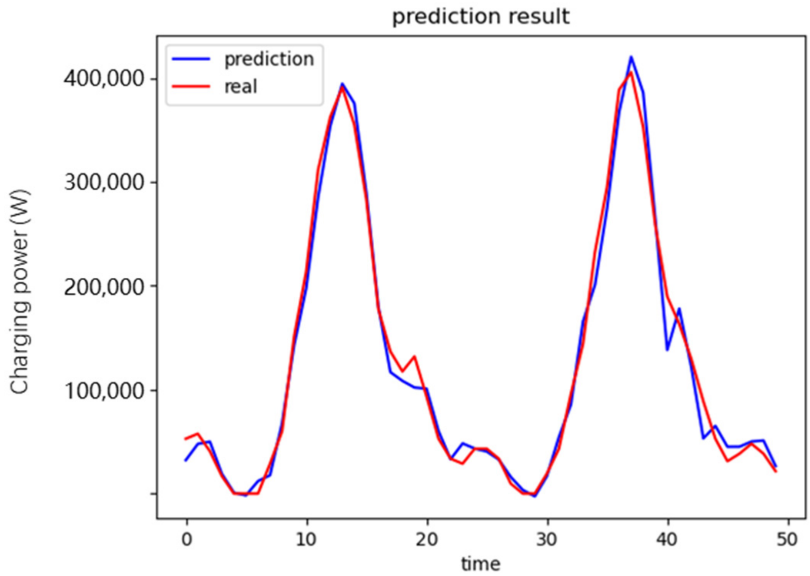

Figure 7. Prediction results of the MA-GRU method.

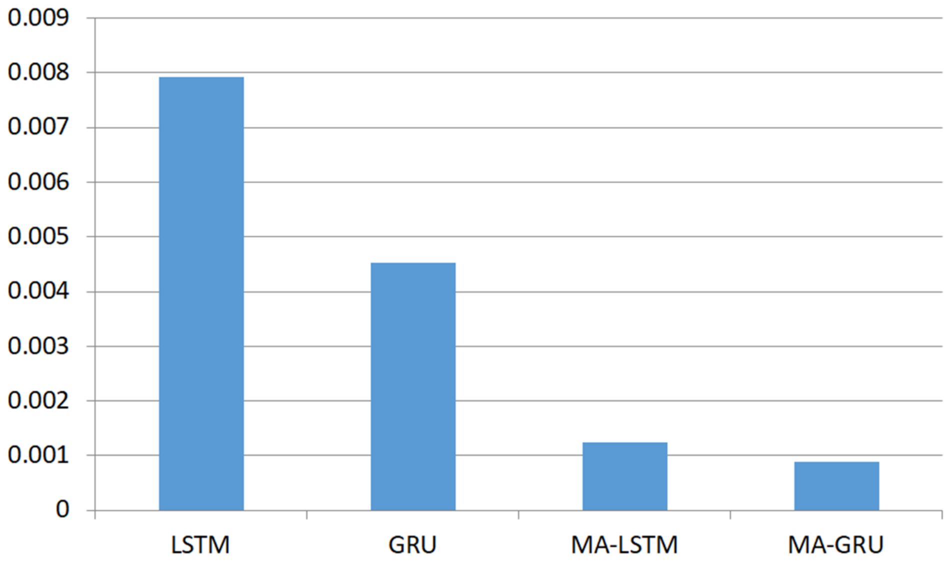

Figure 8. The bar plot of RMSE.

Table 1. RMSE of four methods.

	LSTM	GRU	MA-LSTM	MA-GRU
RMSE	0.00792	0.00453	0.00125	0.00089

According to Figures 4–7, the four methods of LSTM, GRU, MA-LSTM, and MA-GRU can all predict charging power. In Figure 4, the RMSE of the LSTM method is 0.00792. Compared with the LSTM method, the GRU method in Figure 5 has made some progress, and its RMSE is 0.00453. The smaller RMSE obtained by the GRU method shows the advantages compared with LSTM. As shown in Figure 6, the RMSE of the MA-LSTM method is 0.00125. Compared with the results of the LSTM method in Figure 4, the smaller RSEM obtained by the MA-LSTM method can illustrate the effectiveness of the data-smoothing MA method. Among all the four methods, the proposed MA-GRU method can obtain the smallest RMSE, and its value is 0.00089, which shows the effectiveness of the MA method and the superiority of GRU.

In summary, LSTM is a classical network prediction method considering the time series of data. By comparing with it, the effectiveness of the proposed MA-GRU method in this paper can be illustrated. In addition, the proposed MA-GRU method is compared with the GRU-only method. A better RMSE can illustrate the advantages of the MA strategy.

4. Conclusions

In this paper, a method of predicting the electric load of electric vehicle charging stations with a moving average–gated recurrent unit is proposed, and the prediction model is established through the historical data of electric vehicle charging stations, thus realizing the accurate prediction of future electric loads. This method adopts the moving average strategy to smooth the historical data of electric vehicle charging stations to eliminate the unsmooth phenomenon caused by noise, etc. Then, the proposed method establishes a prediction model based on the gated recurrent unit and the smoothed data and obtains the future prediction results of an electric vehicle charging station's electric load. Finally, the effectiveness and superiority of the proposed method are proved by the research and testing of the power load data of the UK EV energy task force dataset and comparison with the classical prediction model. Based on the results, the prediction accuracy of the proposed model is obviously higher than that of the contrast model.

In future work, in order to establish a more accurate prediction model of electric vehicle charging stations under larger datasets or different charging behaviors, the integration of GRU and a more effective data-smoothing technique can be studied. Moreover, given that reinforcement learning has a self-learning mechanism by giving rewards, hybrid models like the combination of MA-GRU with reinforcement learning can be established to update the model with different charging behaviors. In addition, the parallel MA-GRU model can be established for large-sized datasets.

Author Contributions: Methodology, W.H., C.R. and B.S.; Validation, J.Q., Y.L. and Q.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Science and Technology Project of State Grid Zhejiang Electric Power Co., Ltd. (B311TZ230002).

Data Availability Statement: https://github.com/CoogyEoin/EV-charger-power-forecast-model/blob/master/ev_data.csv (accessed on 20 August 2024).

Conflicts of Interest: Authors Wei Huang, Chuanhong Ru, Jian Qin, Yong Lin and Qingxi Cai were employed by the State Grid Zhejiang Electric Power Company, Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declare that this study received funding from Science and Technology Project of State Grid Zhejiang Electric Power Co., Ltd. (B311TZ230002). The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Brunton, S.L.; Noack, B.R.; Koumoutsakos, P. Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.* **2020**, *52*, 477–508. [CrossRef]
2. Surucu, O.; Gadsden, S.A.; Yawney, J. Condition Monitoring using Machine Learning: A Review of Theory, Applications, and Recent Advances. *Expert Syst. Appl.* **2023**, *221*, 119738. [CrossRef]
3. Singh, N.K.; Yadav, M.; Singh, V.; Padhiyar, H.; Kumar, V.; Bhatia, S.K.; Show, P.-L. Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. *Bioresour. Technol.* **2023**, *369*, 128486. [CrossRef] [PubMed]
4. Facco, P.; Doplicher, F.; Bezzo, F.; Barolo, M. Moving average PLS soft sensor for online product quality estimation in an industrial batch polymerization process. *J. Process Control.* **2009**, *19*, 520–529. [CrossRef]

5. Hasnen, S.H.; Shahid, M.; Zabiri, H.; Taqvi, S.A.A. Semi-supervised adaptive PLS soft-sensor with PCA-based drift correction method for online valuation of NOx emission in industrial water-tube boiler. *Process Saf. Environ. Prot.* **2023**, *172*, 787–801. [CrossRef]
6. Wang, Z.X.; He, Q.P.; Wang, J. Comparison of variable selection methods for PLS-based soft sensor modeling. *J. Process Control* **2015**, *26*, 56–72. [CrossRef]
7. Desai, K.; Badhe, Y.; Tambe, S.S.; Kulkarni, B.D. Soft-sensor development for fed-batch bioreactors using support vector regression. *Biochem. Eng. J.* **2006**, *27*, 225–239. [CrossRef]
8. Hua, L.; Zhang, C.; Sun, W.; Li, Y.; Xiong, J.; Nazir, M.S. An evolutionary deep learning soft sensor model based on random forest feature selection technique for penicillin fermentation process. *ISA Trans.* **2023**, *136*, 139–151. [CrossRef] [PubMed]
9. Wang, S.; Tian, W.; Li, C.; Cui, Z.; Liu, B. Mechanism-based deep learning for tray efficiency soft-sensing in distillation process. *Reliab. Eng. Syst. Saf.* **2023**, *231*, 109012. [CrossRef]
10. Jia, M.; Xu, D.; Yang, T.; Liu, Y.; Yao, Y. Graph convolutional network soft sensor for process quality prediction. *J. Process Control* **2023**, *123*, 12–25. [CrossRef]
11. Kurniawan, E.D.; Effendy, N.; Arif, A.; Dwiantoro, K.; Muddin, N. Soft sensor for the prediction of oxygen content in boiler flue gas using neural networks and extreme gradient boosting. *Neural Comput. Appl.* **2023**, *35*, 345–352. [CrossRef]
12. Ansari, S.; Naghdy, F.; Du, H.; Pahnwar, Y.N. Driver mental fatigue detection based on head posture using new modified reLU-BiLSTM deep neural network. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 10957–10969. [CrossRef]
13. Abdin, A. Applying self-normalizing neural networks to tackle data-driven soft sensing problems in manufacturing lines. In Proceedings of the IEEE International Conference on Big Data, Orlando, FL, USA, 15–18 December 2021.
14. Li, Z.; Tian, L.; Yan, X. A multi-objective optimization based deep feature multi-subspace partitioning method for process monitoring. *Expert Syst. Appl.* **2023**, *225*, 120097. [CrossRef]
15. Chikkam, S.; Singh, S. Condition monitoring and fault diagnosis of induction motor using DWT and ANN. *Arab. J. Sci. Eng.* **2023**, *48*, 6237–6252. [CrossRef]
16. Fan, Y.; Tao, B.; Zheng, Y.; Jang, S.-S. A data-driven soft sensor based on multilayer perceptron neural network with a double LASSO approach. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 3972–3979. [CrossRef]
17. Hu, C.-H.; Pei, H.; Si, X.-S.; Du, D.-B.; Pang, Z.-N.; Wang, X. A prognostic model based on DBN and diffusion process for degrading bearing. *IEEE Trans. Ind. Electron.* **2019**, *67*, 8767–8777. [CrossRef]
18. Lu, Z.; Liu, H.; Chen, F.; Li, H.; Xue, X. BOF steelmaking endpoint carbon content and temperature soft sensor based on supervised dual-branch DBN. *Meas. Sci. Technol.* **2023**, *35*, 035119. [CrossRef]
19. Wang, X.; Hu, W.; Li, K.; Song, L.; Song, L. Modeling of soft sensor based on DBN-ELM and its application in measurement of nutrient solution composition for soilless culture. In Proceedings of the 2018 IEEE International Conference of Safety Produce Informatization, Chongqing, China, 10–12 November 2018.
20. Zhu, C.H.; Zhang, J. Developing soft sensors for polymer melt index in an industrial polymerization process using deep belief networks. *Int. J. Autom. Comput.* **2020**, *17*, 44–54. [CrossRef]
21. Tian, Y.; Zhu, Q.X.; He, Y.L. Novel Deep layers-extended Autoencoder with Correlation and Its Industrial Soft Sensing. *IEEE Trans. Ind. Inform.* **2023**, *20*, 596–604. [CrossRef]
22. Zabalza, J.; Ren, J.; Zheng, J.; Zhao, H.; Qing, C.; Yang, Z.; Du, P.; Marshall, S. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. *Neurocomputing* **2016**, *185*, 1–10. [CrossRef]
23. Wang, W.; Wang, H.; Chen, L.; Hao, K. A novel soft sensor method based on stacked fusion autoencoder with feature enhancement for industrial application. *Measurement* **2023**, *221*, 113491. [CrossRef]
24. Liu, C.; Wang, Y.; Wang, K.; Yuan, X. Deep learning with nonlocal and local structure preserving stacked autoencoder for soft sensor in industrial processes. *Eng. Appl. Artif. Intell.* **2021**, *104*, 104341. [CrossRef]
25. Yuan, X.; Huang, B.; Wang, Y.; Yang, C.; Gui, W. Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE. *IEEE Trans. Ind. Inform.* **2018**, *14*, 3235–3243. [CrossRef]
26. Yuan, X.; Li, L.; Wang, Y. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. *IEEE Trans. Ind. Inform.* **2019**, *16*, 3168–3176. [CrossRef]
27. Ma, L.; Zhao, Y.; Wang, B.; Shen, F. A Multi-Step Sequence-to-Sequence Model with Attention LSTM Neural Networks for Industrial Soft Sensor Application. *IEEE Sens. J.* **2023**, *23*, 10801–10813. [CrossRef]
28. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [CrossRef]
29. Duchanoy, C.A.; Moreno-Armendáriz, M.A.; Urbina, L.; Cruz-Villar, C.A.; Calvo, H.; Rubio, J.d.J. A novel recurrent neural network soft sensor via a differential evolution training algorithm for the tire contact patch. *Neurocomputing* **2017**, *235*, 71–82. [CrossRef]

30. Zhou, J.; Wang, X.; Yang, C.; Xiong, W. A novel soft sensor modeling approach based on difference-LSTM for complex industrial process. *IEEE Trans. Ind. Inform.* **2021**, *18*, 2955–2964. [CrossRef]
31. Yuan, X.; Li, L.; Shardt, Y.A.W.; Wang, Y.; Yang, C. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Trans. Ind. Electron.* **2020**, *68*, 4404–4414. [CrossRef]