Application of machine learning RNN network in signal Transmission Prediction

A. Project Introduction

For long time and long distance signal transmission, the temperature, humidity of the external environment and the visibility of the interference of equipment will affect the transmission efficiency and transmission accuracy of the transmission system, and affect the stability of signal transmission. Therefore, a real-time voltage prediction method based on gated recursive recurrent neural network is proposed in this paper, which can adapt to the long time continuous attention sequence model. Using more effective compression mechanism and model advantages, improve system transmission accuracy, self-feedback to enhance model stability, improve system performance.
Through consulting the data, due to the limitations of the original RNN model, we further extended RNN, using LSTM and GRU model variants for comparative tests, and introduced the third model Wavenet. We are going to input three dimensional/four dimensional array to get the training model, then output transmission voltage predicted value, compare the fitting degree of the predicted value and the original value, in the end evaluate the validity of these three application algorithms in data prediction.

B. Model Building

In this experiment, Faraday-Michelson quantum cryptographic transmission and distribution system is taken as the main research subject
In order to study the key distribution system real-time voltage prediction for experimental learning. After learning, it is found that the interference loop voltage conforms to the characteristics of the recurrent neural network, that is to say, the output value of the historical voltage has an effect on the phase voltage value at the later time point, which has a certain correlation and consistency.
But at the same time, the external environment of the interference ring equipment, such as temperature, humidity and other environmental factors, will also produce the voltage value
Impact. Therefore, the temperature, humidity and voltage output values of the interference ring equipment should be monitored in real time at the same time, and the data should be derived as the basic data set. Temperature-humidity-voltage 3d data were input into the model as influence parameters for learning and training, and finally the prediction and estimation level of temperature, humidity and voltage at historical time point to voltage at current time point was tested. Based on the above, we obtained three groups of 3D scanning data.
The dataset data is divided into two groups, one for training and one for testing. The training set is used to train the prediction model, and the test set is used to evaluate the final results. The procedure of interference ring voltage prediction experiment is described as follows:
1. Data preprocessing.

2. Divide the data set into training set and test set. The training set contains 70% of the data, while the test set contains the other 30%.
3. Construct a THREE-DIMENSIONAL array from the training set to input to the GRU network through spatial reconstruction.
4. Use MAE as the loss function and Adam algorithm as the optimization method to determine the GRU model parameters.
5. Determine the number of training iterations, i.e. the number of periods, according to the loss function.
6. Predict the test set.
7. Compare with other real-time voltage prediction models of different architectures and evaluate these models against selected performance indicators.
8. Visualization of prediction results after inverse normalization. Through the above 8 steps, the interference ring voltage prediction model based on GRU neural network is established.

Before performing the prediction of the time series, we first need to train the prediction model effectively. Enter the 3D parameter size matrix and set input_shape to (50, 3). There are 200 hidden neurons in the hidden layer of the GRU model. The activation function sigmoid function is located in the Dense The fully connected layer has 40 hidden neurons.

The optimizer is set to 'adam', i.e., the parameters of the GRU model are determined by using the Adam optimization algorithm, combined with the mean absolute error MAE as the loss function for continuous optimization. To avoid negative situations such as gradient explosion or gradient disappearance, the gradient threshold of the training model is set equal to 1. The number of training sessions is predefined as 200.

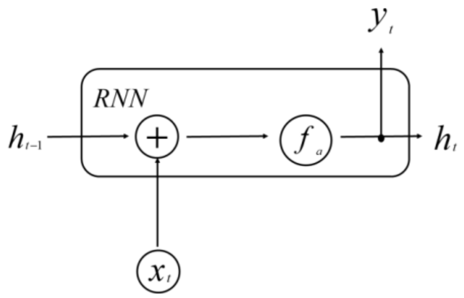
Finally, we set the step size, the number of training rounds, and the number of samples selected for one training session for the model. These are timestep=50, Epochs=50, and batch_size=64, respectively.

C. Algorithm Introduction

C.1: Recurrent neural network architecture

Recursive neural network and gating recursive unit related theory and technology. In 1982, Ropfield networks formally brought recurrent neural networks into the public domain. Recurrent neural network structure, namely RNN model, is the earliest deep learning neural network for processing time series data. By controlling the information model, only the information related to the optimization loss is saved, and combining with the current memory information, the output information is influenced to realize the effective prediction of the future value.

RNN neural network and hidden layer element structure are shown in figure:

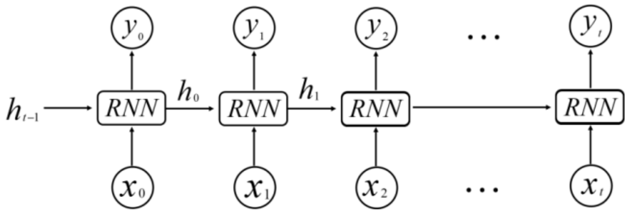


The model realization formula is as follows:

$$h_t = f_a(W_{xh} \bullet x_t + W_{hh} \bullet h_{t-1} + b_h) \quad (3.1)$$

$$y_t = W_{hy} \bullet h_t + b_y \quad (3.2)$$

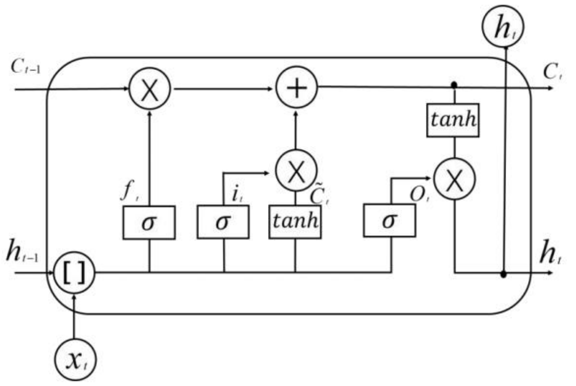
Although RNN is a neural network used to deal with time series problems, RNN has the disadvantage of gradient disappearance or gradient explosion on long time series due to its fully connected characteristics.
Full time series expansion diagram of RNN model:



In the process of long-distance transmission, the number of time units will increase with the increase of distance. Combined with the characteristics of sequential full connection processing of RNN model, the multiplication times of weight matrix factors increase exponentially.
Due to the nature of the S-type functions used by neural networks, the gradient of RNN may be lost by expanding RN into very deep feedforward neural networks as the time delay increases.
Inversely, when the input value is updated and the normalized value is greater than 1, the amplitude of the change will expand exponentially, resulting in gradient explosion. Moreover, when training RNN model, it is difficult to get the appropriate value of delay window length automatically.
Therefore, it is particularly important to further excavate the temporal correlation of signals and extract the semantic information effectively. Therefore, a more effective recurrent neural network structure LSTM is proposed to solve these problems.

C.2: Long short Term Memory Network (LSTM)
In order to solve the problems of gradient disappearance, gradient explosion and long-term dependence,Hochreiter and Schmidhuber introduced LSTM neural network structure in 1997. In the field of AI, the earliest and most widely used memory structure is the recurrent neural network (RNN), known as long and short-term memory (LSTM), which uses output gates on its state vectors to determine what information is stored or

forgotten from memory. It is suitable for processing and predicting important events with very long intervals and delays in time series.
The structural model of LSTM neural network is shown in Figure 3.3. Compared with RNN model, there are three more controllers, namely input gate, output gate and forgetting gate. Depending on the importance of information acquired during model training, these three gate structures allow LSTM neurons to update, maintain, or delete information contained in the cell state, thereby determining how much useful information to pass on to the next training.



$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.4)$$

$$\tilde{C}_t = \tanh(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.5)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3.6)$$

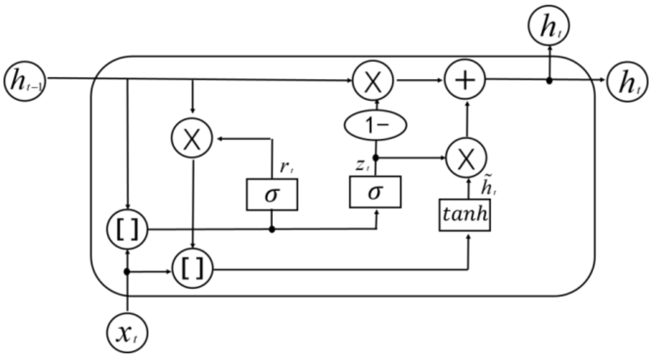
$$O_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_o) \quad (3.7)$$

$$h_t = O_t * \tanh(C_t) \quad (3.8)$$

Although THE prediction effect of LSTM is excellent, its calculation efficiency is low, and the number of learnable parameters in the model increases twice with the size of memory. An LSTM with a memory size of 64KB generates arguments with a size of 8GB.
With the deepening and evolution of research, people strive to further explore an adaptive model with fewer parameters and more simple and effective training process based on the LSTM structure. GRU model -- gated recursive unit is proposed. The overall structure of GRU is simpler and the network performance is closer.

C.3: Gated Recursive Unit (GRU)
LSTM model needs a lot of parameters and complex structure in operation. Therefore, it is easy to overfit. With continuous research and development, Cho proposed GRU model with simpler structure and similar network performance as a variant of LSTM in

2014 in order to overcome these shortcomings. The GRU architecture retains the characteristics of LSTM while having a simpler structure. The main change is to combine the forgetting gate and input gate of the LSTM model into an update gate.



$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{3.9}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{3.10}$$

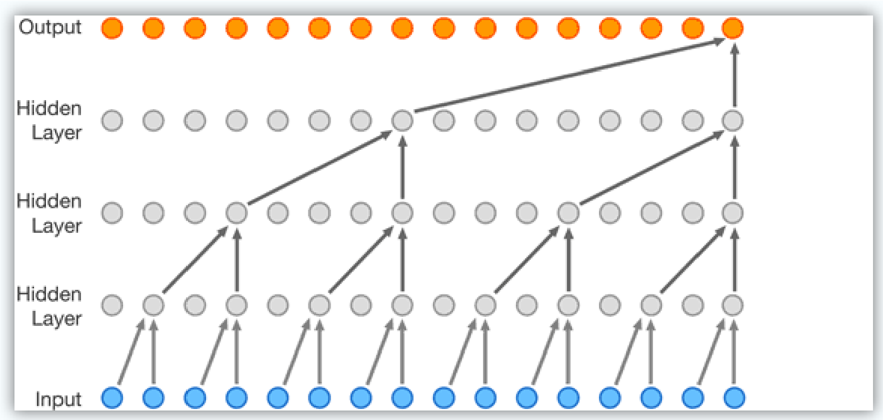
$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{3.11}$$

$$h_t = z_t * h_{t-1} + (1 - z_t) * \tilde{h}_t \tag{3.12}$$

As described above, the GRU model captures short-term dependencies and long-term dependencies by resetting the frequent activation of the gate
Associated with the activation of the update gate. Because the GRU architecture has only two types of control gates, it reduces the time wasted if other "gates" are present. Therefore, the calculation speed of GRU model is faster than that of LSTM model.

C.4: Wavanet

The core of the Wavenet model is to train the model and predict the given input sequence according to the previous data. Then, the predicted value is added to the input sequence, and then the original measured data and the newly added predicted value together form the new input parameters, and the same can be achieved. After each sample is predicted, it is fed back to the network to predict the next sample. Because models with causal convolution do not have cyclic connections, the training speed of Wavenet is theoretically faster than that of traditional RNN, especially when applied to very long sequences.



And, when Wavenet model is applied to text-to-speech conversion, it produces the most advanced performance, natural sound and high accuracy. Both sound signal and interference loop voltage are one-dimensional models. The realization process of text to speech is also based on the correlation between sound signals.
However, the later learning process reveals the shortcomings and limitations of Wavenet model. Specific analysis will be made in the following article.

D. Results and Discussion

Taking the output of data 1 as an example, the test set image and the training set fitting image of GRU using GRU, LSTM and Wavenet networks for interference loop voltage prediction are shown here respectively.

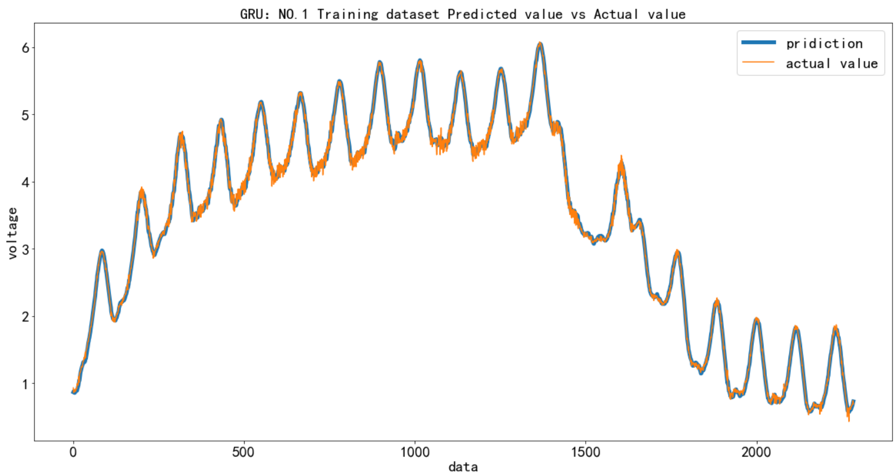


Figure D.1 Dataset 1 training set fit

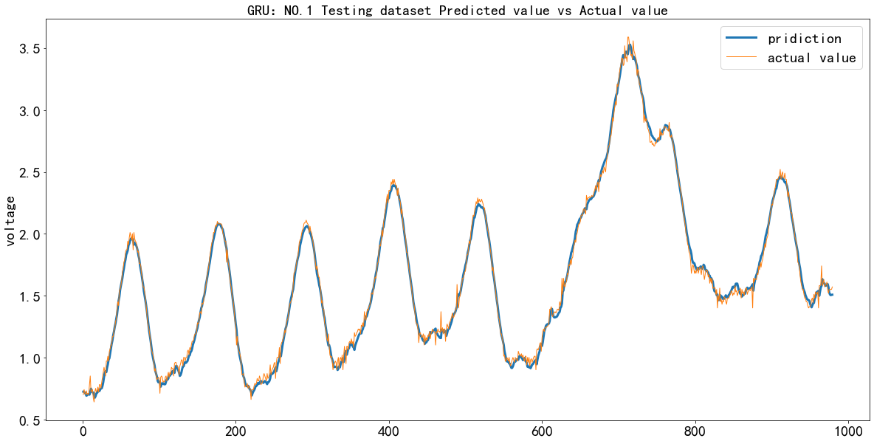


Figure D.2 GRU: Dataset 1 Test Set

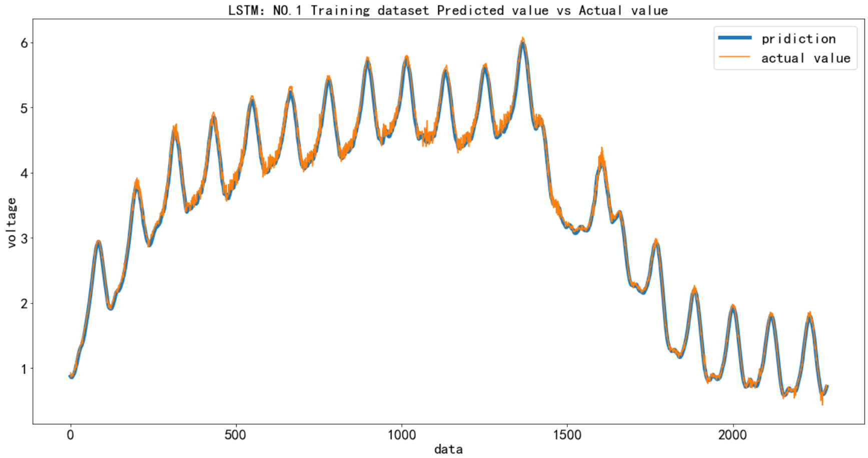


Figure D.3 LSTM: Dataset 1 Train Set

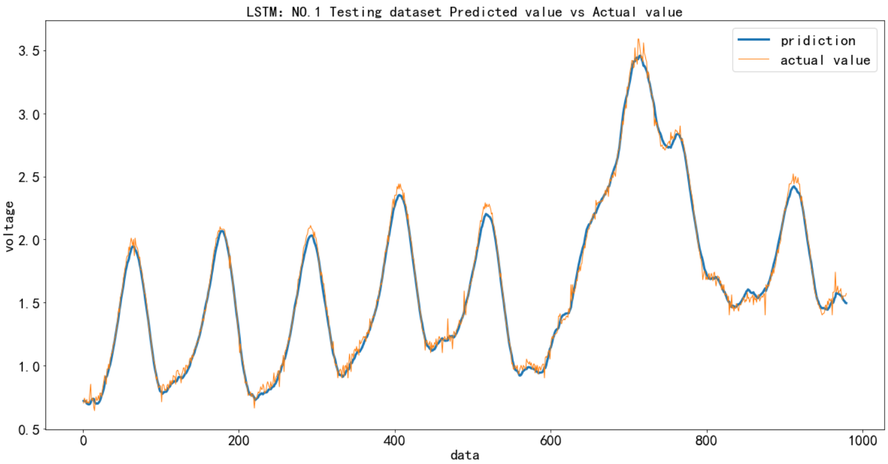


Figure D.4 LSTM: Dataset 1 Test Set

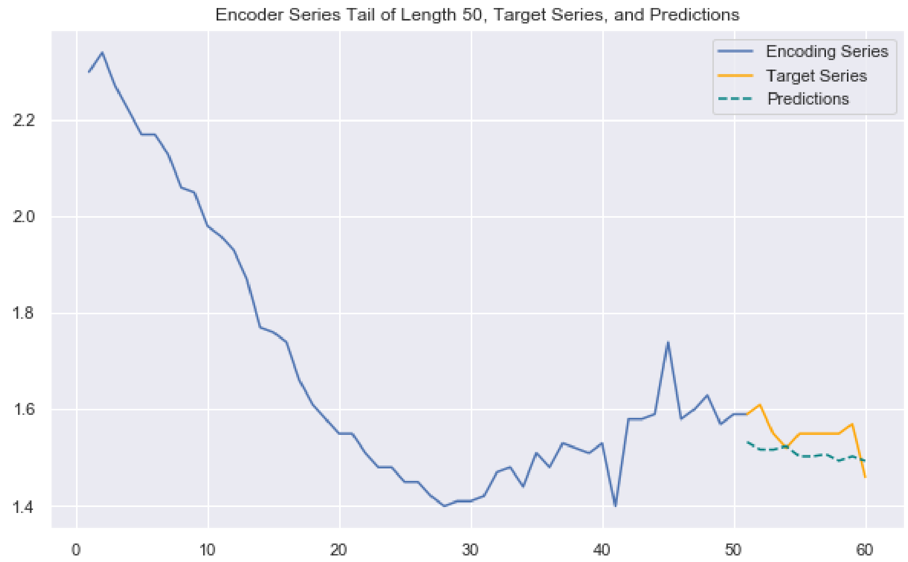


Figure D.5 Wavenet: Dataset1 Test Set

The coincidence effect of predicted value and actual value of GRU and LSTM network is very good. The blue predicted value trajectory curve and the yellow actual value curve are separated to some extent in the interval [780,850], and the rest of the curves basically coincide.

However, the prediction effect of Wavenet model is not good. If all the test sets are displayed, the prediction effect of the model can only reach the average value of the theoretical value, and there is no change basically. It is approximately a horizontal straight line on the function image and cannot show the amplitude change of the original theoretical value. Therefore, through later debugging, the ratio of the original data set was adjusted from 7:3 to 8:2, the step size was set to 100, and only the last 10 predicted values of the training set were displayed. As shown in the figure, it can be seen that the prediction function overlaps with the theoretical value in two places, which can basically show the change trend of the value. However, there is still a big difference with the test effect of GRU model in the figure above.

The prediction effect of GRU, LSTM, Wavenet and GRU training set fitting image in data 2 set

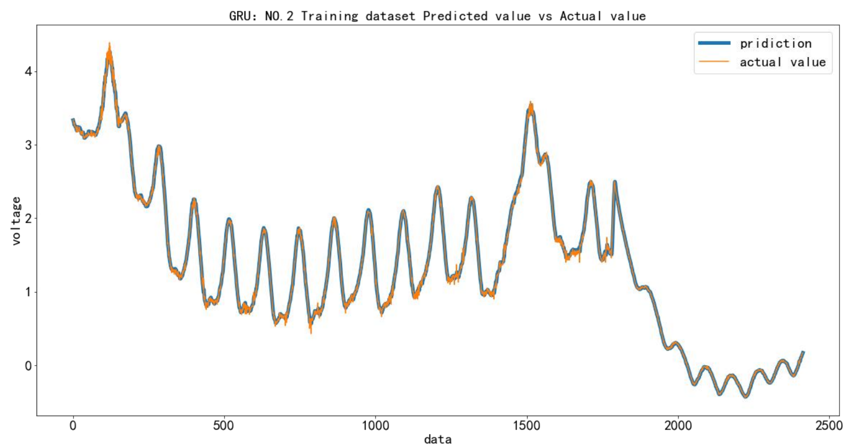


Figure D.6 GRU: Dataset 2 training set fit

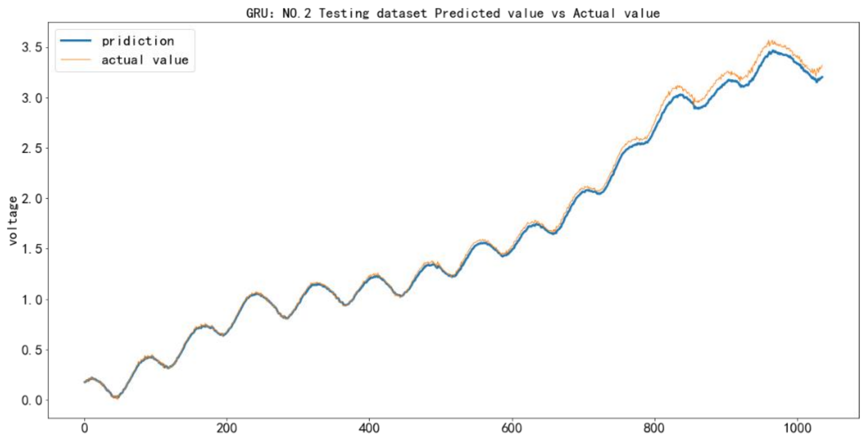


Figure D.7 GRU: Dataset 2 Test Set

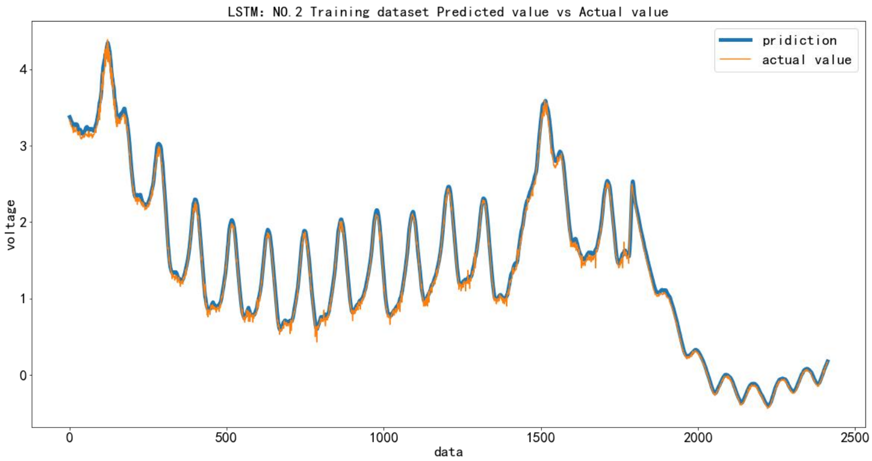


Figure D.8 LSTM: Dataset 2 test set

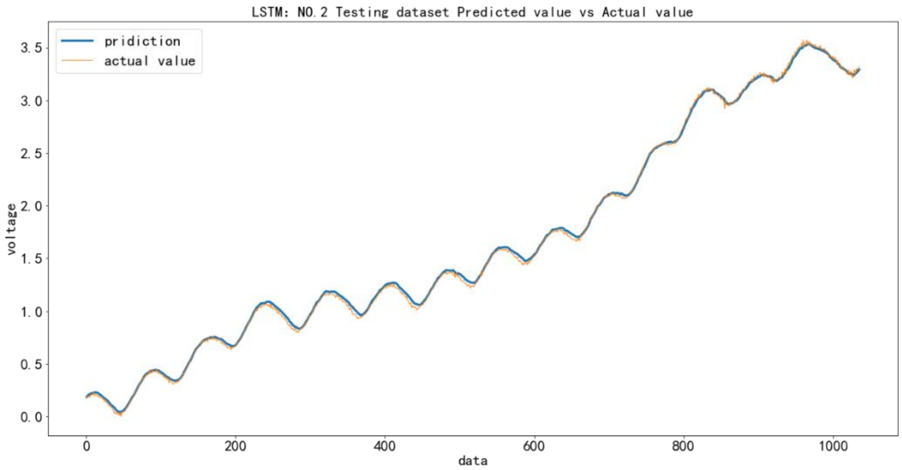


Figure D.9 LSTM: Dataset 2 test set

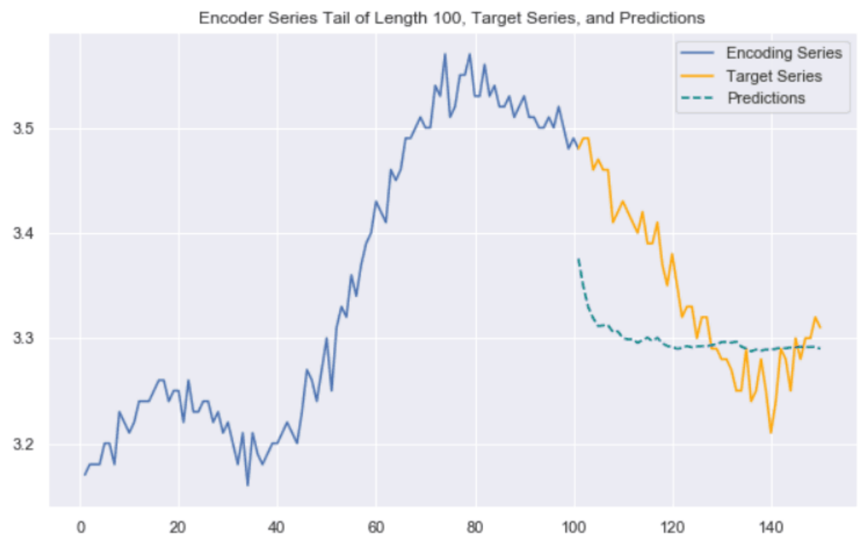


Figure D.10 Wavenet: Dataset 2 Test Set

It can be seen from the prediction curve of dataset 2 that LSTM and GRU models have a high coincidence rate and errors mainly exist
The yellow and blue curves show a small error in the data after test set 650. According to Figure 5.8Wavenet prediction curve, the difference between the green prediction curve and the yellow target curve is large, which can only predict the approximate direction of data development. In terms of numerical value, there are only six overlaps, and the effect is poor.

E. Model Evaluation

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(observed_t - predicted_t)^2 \qquad (5.1)$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|observed_t - predicted_t\right| \qquad (5.2)$$

$$Loss(k) = \frac{\sum_{i=L+1}^{m}(tem_i(k) - tem_i^{'}(k))^2 + (hum_i(k) - hum_i^{'}(k))^2 + (vol_i(k) - vol_i^{'}(k))^2}{m-L} \qquad (5.3)$$

First of all, we tested the model on a dataset composed of 3314 data points. After 50 rounds, the experiment showed that both LSTM model and GRU model had good fitting and prediction effect, and the two evaluation indexes obtained were also ideal. In addition, due to the limited computing capacity of CPU, it could not bear too much computation. Therefore, the number of cycles of both LSTM model and GRU model was set to 50 in this experiment, and the number of cycles was not increased to more than 200.

Comparison of data set 1 predictive performance:

	Training Set MSE	*Training Set MAE*	*Testing Set MSE*	*Testing Set MAE*	*Running Time*
LSTM	0.0039	0.047	0.003	0.041	4min24s
GRU	0.0013	0.0027	0.0026	0.0019	4min15s

It can be seen from Table 1 that, for the subset of 3,314 data points, the prediction accuracy of LSTM model is lower than that of fitted GRU model. LSTM model has better fitting ability for small data models such as the set with a data set length of 1000. When the data volume increases, the prediction accuracy and fitting speed of GRU show better results.
We performed the same experiment on dataset 2 consisting of 3500 data points. The MSE, MAE, and runtime for each model are listed in following table.

Comparison of data set 2 predictive performance:

	Training Set MSE	*Training Set MAE*	*Testing Set MSE*	*Testing Set MAE*	*Running Time*
LSTM	0.0015	0.0257	0.0021	0.0338	4min55s
GRU	0.0012	0.0278	0.0015	0.0278	4min28s

According to Table 5.2, the error of GRU model is still smaller than that of LSTM model. As the amount of data increases, the overall computation time increases. However, the time difference between the two models increases from 9s of data 1 to 13s of data 2. When the amount of data increases, GRU model can achieve faster fitting speed and better performance due to the reduction of gate function.
Below, the prediction of Wavenet model is evaluated, and the reasons for the inapplicability of the model are analyzed. MSE and MAE of the second group of data are shown in Figure. To be more obvious and intuitive, the MSE was magnified 100 times.

Wavenet:Dataset2 MSE

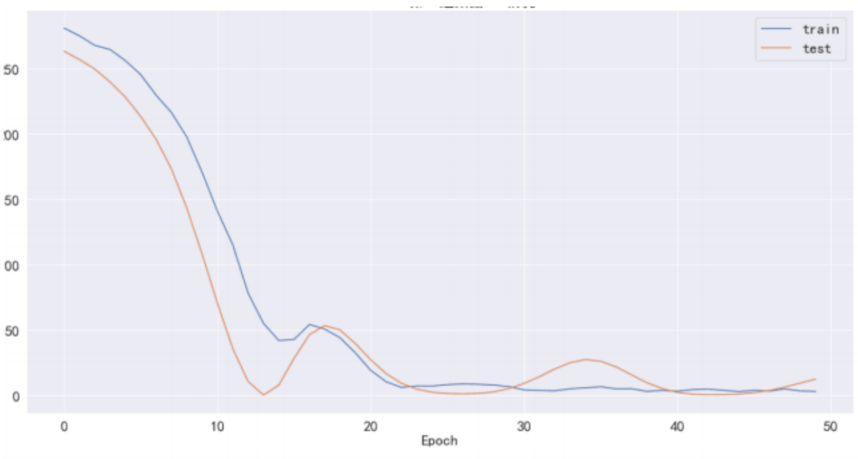


Figure D.11 Wavenet Dataset 2 MSE

MSE showed an obvious downward trend after 13 rounds of training, and then the curve was gradually flat, especially in the training set, where the optimal MSE reached 0.05. The MSE of the orange test set basically remained between 0.1 and 0.2.

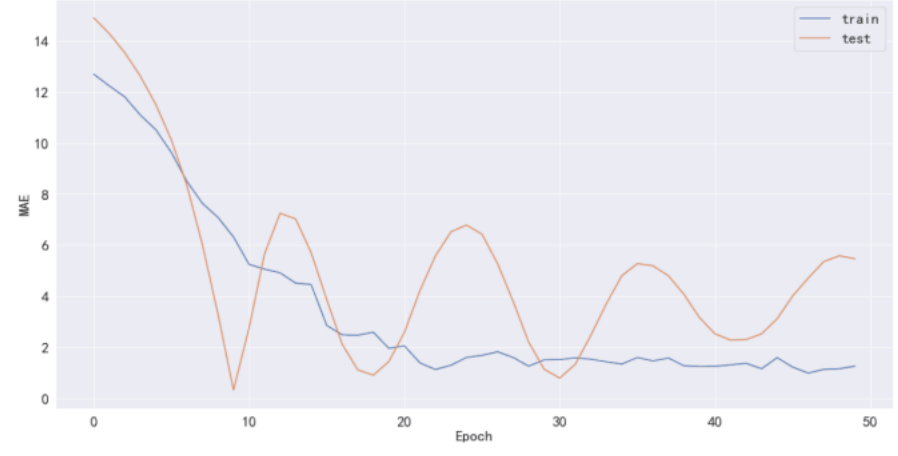


Figure D.12 Wavenet Dataset 2 MAE

In MAE, the performance of the training set is still better than that of the test set. After 20 rounds of the cycle, the curve of the training set gradually flattens out and basically remains at about 1.8. However, the mean absolute error of the test set is extremely unstable, with ups and downs and the highest value of 7.6. After the experiment, I reflect on and summarize the Wavenet model. The reason for Wavenet model's poor effect in this experiment is that the model is a very typical one-dimensional model with one-dimensional data as input and output. If the voltage prediction is solved, the input data only accepts the historical value of the voltage, and the predicted output value is transferred to the input layer for the subsequent calculation as the new prediction parameter.

Because the main purpose of this experiment is to find the temperature and humidity on the voltage of the influence, so should be temperature/humidity/voltage three-dimensional parameters input at the same time, but this and Wavenet itself structure characteristics is contradictory.

Secondly, Wavenet has a good expressive force in speech and text conversion because the sound signal itself has a good time correlation and time continuity, can be predicted through the history of speech signal simulation after the value. However, the voltage is greatly affected by the environment, and only the voltage itself can easily produce errors in predicting its future value.

Unlike dataset 1, 2, and 3, dataset 4 is a four-dimensional array with parameters such as temperature, humidity, interference visibility, and voltage, respectively. We set data set 4 with interference visibility as the control group, and compared the output prediction of interference loop voltage under the two conditions of interference visibility or not.

GRU: Data3: NO interference visibility effect

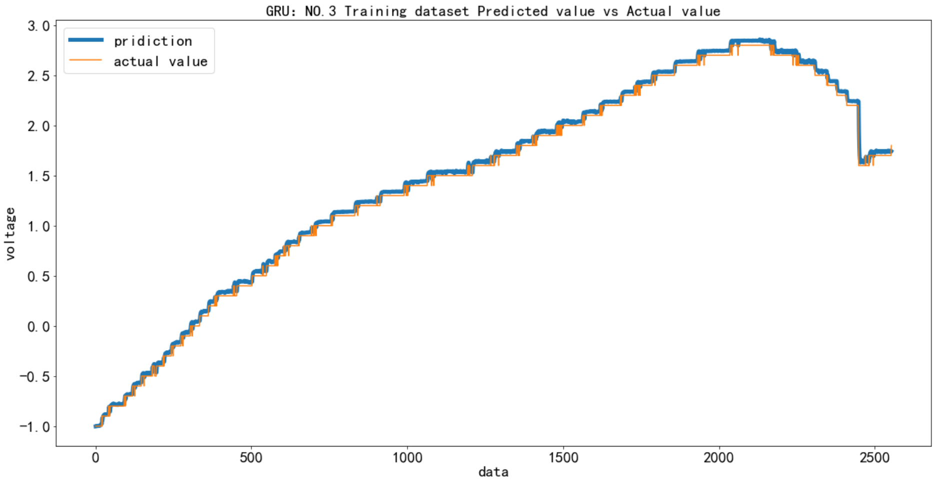


Figure D.13 GRU: Dataset3 training set

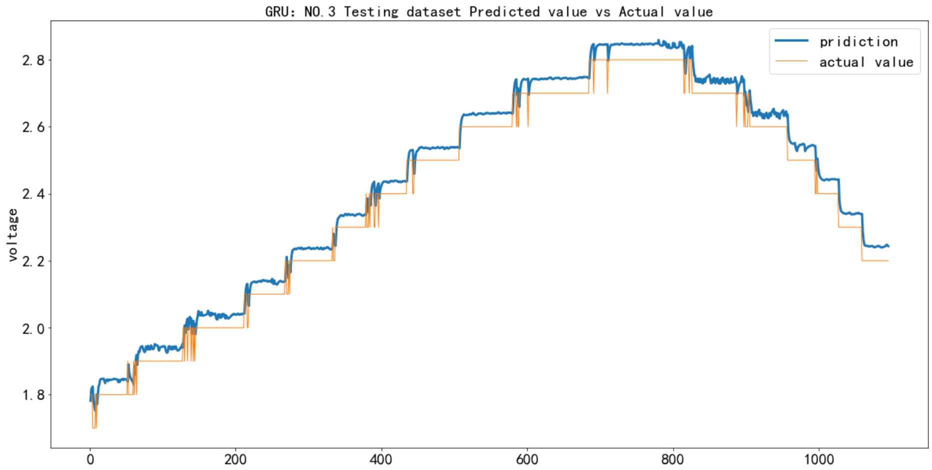


Figure D.14 GRU: Dataset3 testing set

LSTM: Data3: NO interference visibility effect

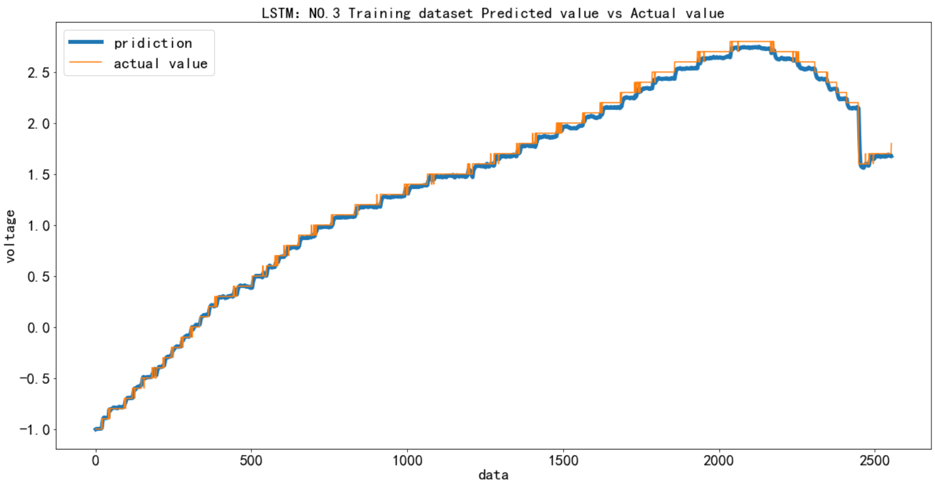


Figure D.15 LSTM: Dataset3 training set

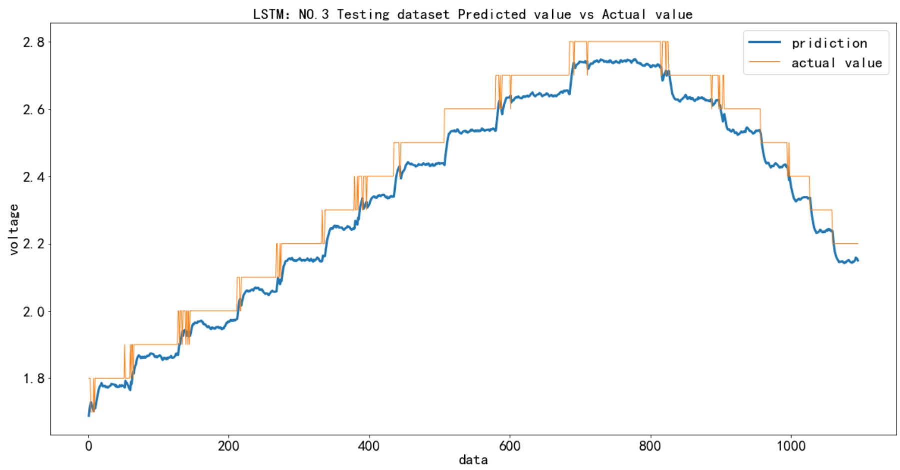


Figure D.16 LSTM: Dataset3 testing set

GRU: Data4: With interference visibility effect

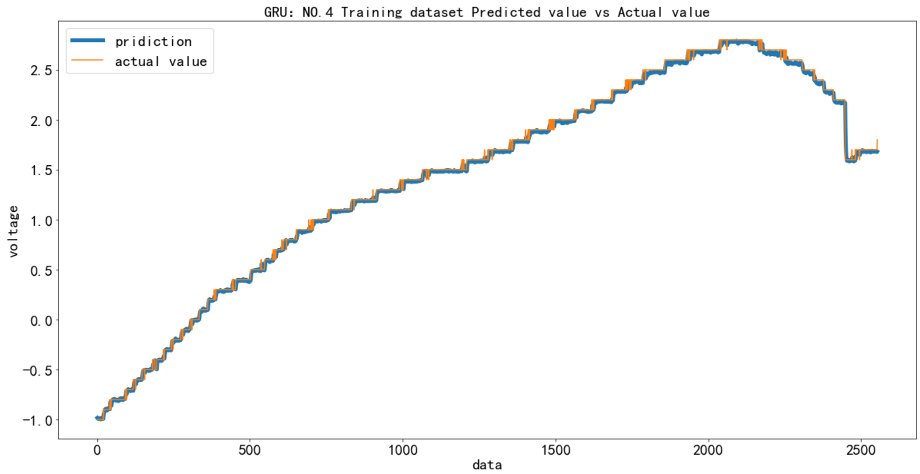


Figure D.17 GRU: Dataset4 training set

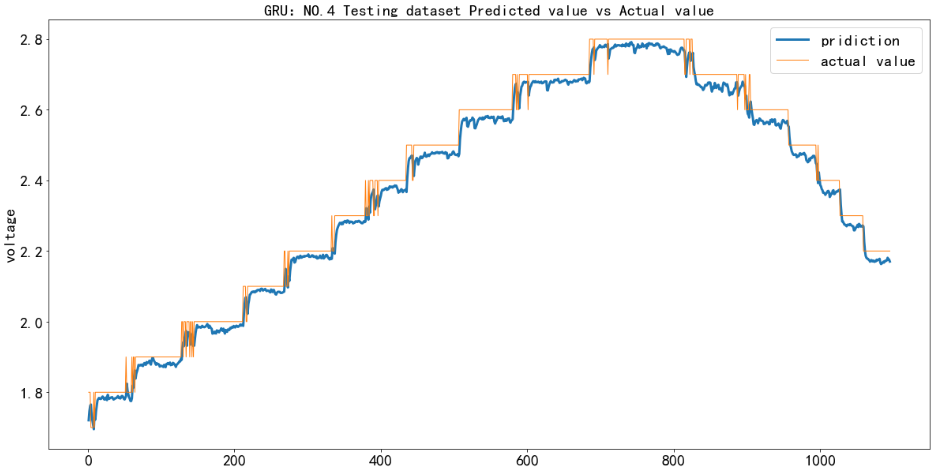


Figure D.18 GRU: Dataset4 testing set

	Training Set MSE	*Training Set MAE*	*Testing Set MSE*	*Testing Set MAE*	*Running Time*
LSTM	0.0014	0.0245	0.0014	0.0315	5min32s
GRU(3-D)	0.0013	0.00274	0.0012	0.0295	5min10s
GRU(4-D)	0.010	0.0206	0.0009	0.0245	5min28s

The comparison between GRU and LSTM models is consistent with the above analysis of dataset 1 and 2. When the input array of GRU model is 4-dimensional, it can be found that MSE and MAE of training set and test set show better prediction results. Among them, MSE and of training set and test set both decrease by 0.0003, and MAE of training set decreases by 0.0068, and MAE of test set decreases by 0.005, increasing by more than 20%.

F. Conclusion

Finally, we integrate the performance values in the above table, as shown below.

Table Comparison of the prediction performance of datasets 1, 2,3 and 4

Dataset		**Training set MES**	**Training set MAE**	**Test Set MSE**	**Test Set MAE**	**Operation time**
Dataset 1	**LSTM**	0.0039	0.047	0.003	0.041	4min24s
Size:3314	**GRU**	0.0013	0.0027	0.0026	0.0019	4min15s
Dataset 2	**LSTM**	0.0015	0.0257	0.0021	0.0338	4min55s
Size:3500	**GRU**	0.0012	0.0278	0.0015	0.0278	4min28s
Dataset 3	**LSTM**	0.0014	0.0245	0.0014	0.03153	5min32s
Size: 3700	**GRU(3-dimensions)**	0.0013	0.0274	0.0012	0.0295	5min10s
Dataset 4	**GRU(4-dimensions)**	0.001	0.0206	0.0009	0.0245	5min28s

We can draw the following conclusions:

1. As the data set increases, the operation time of the model also increases. As the data sets used this time are all larger than 3000 groups with large numerical quantities, GRU shows obvious similarity compared with LSTM

Speed advantage. And the larger the data set, the more obvious the advantage. In the three groups of experiments, the time difference was 9s, 17s and 22s respectively.

2. Another influence brought by more scanning points is the improvement of prediction accuracy. The prediction of GRU model in the case of 3-dimensional input parameters of three sets is respectively taken for longitudinal comparison. The MES of the test set decreased from 0.0030 in dataset 1 to 0.0014 in dataset 4, and the error decreased by 53.3%.

3. The 3d data group in dataset 4 was compared with the 4d control group. After adding one-dimensional parameters, the prediction accuracy of the model was improved, and the optimization ratios of training set MSE, training set MAE, test set MSE and test set MAE were 23.08%, 33%, 25% and 16.9%.

4. In this experiment, the mean square error (MSE) and mean absolute error (MSE) of the GRU prediction model remain at one thousandth and one hundredth respectively, proving that the GRU model is a robust cyclic neural network structure.

G. Optimization of GRU Model and Future Prospects

A layer of GRU network is used in this experiment, after which we can try to increase the complexity of the model and the wide range of applications. For small data sets, RNN and LSTM models are still preferred.

A two-tier network can be considered for the experiment. That is, it is divided into two phases. The first stage is the initialization training of the neural network. The input of environmental covariates from past time is used as training data to learn the GRU neural network parameters by batch processing. Finally, the output layer with three characteristic quantities called temperature, humidity, voltage, and interference visibility is calculated by hidden layer. And the trained network parameters are used as the initial parameter input of the second stage GRU network, when voltage is added as the fourth covariate. The second stage is then the real-time prediction of voltage. The parameters of the GRU neural network are updated adaptively using real-time voltage data.