Industrial Time Series Data Forecasting of LSTM Neural Network Based on Attention Mechanism

Zhiwei Jiang*, Xiudong Peng, Yuanfeng Sun and Mengyun Zhu

School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou, Zhejiang Province, 310018, China

*Corresponding author’s e-mail: jzw2020@yeah.net

Abstract. At present, the maintenance mode of industrial equipment is still based on regular maintenance and after-the-fact maintenance, and with the development of industrial production intelligence, the production data of equipment has increased dramatically. In order to reasonably carry out maintenance activities in the use stage of CNC machine tools and other industrial equipment, this paper proposes an industrial time series data prediction method based on LSTM of attention mechanism. Firstly, based on the LSTM recurrent neural network, combined with the complex historical data of CNC machine tools, the overall characteristics of important time series are obtained. Secondly, the mechanism of attention analysis is introduced, and the algorithm of network structure design and prediction process is given. Finally, the accuracy of prediction is compared with experiments with standard BP neural network. The experimental results show that, compared with the standard BP neural network, the LSTM model based on the attention mechanism has higher prediction accuracy, which verifies the effectiveness of the method.

1. Introduction

The development of new technologies, such as the Industrial Internet of Things and Big Data, has greatly promoted the process of intelligent and automated industrial production. In the process of industrial production, it is very important to find reliable indicators that can represent the stability of production from a large amount of information collected by various sensors. As a kind of data with typical time series characteristics, fault time series data can effectively predict the reliability of the use stage of CNC machine tools and other equipment, avoid blindness to equipment maintenance, enhance the pertinence of equipment maintenance, and improve the reliability and continuity of equipment, at the same time, this is of great significance for formulating reliable production plans and carrying out reasonable maintenance activities.

At present, there are mainly two kinds of methods for predicting equipment reliability through predictive reasoning based on existing knowledge experience such as time series data. One is based on statistical model, such as ARMA model[1], ARIMA model[2] and so on. The other is the prediction method based on artificial intelligence, such as artificial neural network, support vector machine[3] and deep learning[4]. Compared with the statistical model method, the prediction method based on artificial intelligence can well deal with the nonlinear relationship between data and the overall logic of data and information.

With the rapid development of deep learning, some deep learning models are used in the research of CNC machine tool fault prediction. LSTM neural network is a special recurrent neural network with strong sequence information processing capabilities. Considering the gradual failure of the CNC
machine tool during operation and the strong time series characteristics of the collected data, this paper proposes an attention-based\(^5\) LSTM industrial time series data prediction method, including 4 layers (input layer, hidden layer, fully connected layer, output layer) detailed design of the network structure, as well as the realization algorithm of network training and network prediction. In this paper, the simulation experiment is carried out using the data set of centerless grinding machine and compared with the BP neural network model. The experimental results show that the LSTM neural network time series data prediction method with attention mechanism has high prediction accuracy and is suitable for industrial time series data prediction.

2. Research process

This section gives the LSTM neural network industrial time series data prediction model based on the attention mechanism in view of the time series characteristics of industrial time series data. See Figure 1 below. The model can be divided into a four-layer network structure and two functional modules, namely the input layer, hidden layer, fully connected layer, output layer, model training module and model prediction module. In the study, supervised learning is used to train the fault prediction model. The input layer preprocesses the original fault time series data to meet the input requirements of the model. The hidden layer uses several LSTM memories cells to build a recurrent neural network. Through repeated iterations, the weight parameters are adjusted to reduce the error until it converges. Attention mechanism is introduced to give different reference weights to the data at different points in time and add dropout algorithm to alleviate the over fitting problem in the process of multivariate data learning. The fully connected layer reduces the data dimension and improves the prediction accuracy. The output layer outputs prediction results. The model training module adopts the Adam optimization algorithm based on gradient. The model prediction module adopts iterative method to use the trained LSTM model for prediction.

![Figure 1. LSTM industrial time series data prediction model diagram based on attention mechanism.](image-url)
2.1. Network model training

The network model training mainly focuses on the hidden layer. Before inputting the data into the prediction model, the original time series data set is preprocessed at the input layer. The original time series data set $D = \{d_1, d_2, \cdots, d_n\}$ is divided into training set $D_{tr} = \{d_1, d_2, \cdots, d_m\}$ and test set $D_{te} = \{d_{m+1}, d_{m+2}, \cdots, d_n\}$, which satisfies the constraints $m < n$ and $m, n \in N$. Then use the Z-Score standardization method to standardize the element $d_t$ in the training set, speed up the model convergence speed, and facilitate later data processing. The average value of the processed data is 0, and the standard deviation is 1. Equation (1) is the equation of the standardized method, where $x$ is the observation value of individual data, and $x'$ is the standardized individual data.

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

$$d_t' = \frac{(d_t - \frac{1}{n} \sum_{t=1}^{n} d_t)}{\left(\frac{1}{n} \sum_{t=1}^{n} (d_t - \frac{1}{n} \sum_{t=1}^{n} d_t)^2\right)^{1/2}}, 1 \leq t \leq m, t \in N \quad (2)$$

$d_t'$ is the sample data of the standardized training set, and $D_{tr}' = \{d_1', d_2', \cdots, d_m'\}$ represents the standardized training set. After the standardization of training set is completed, the $D_{tr}'$ of standardized training set is segmented by the method of data segmentation, and the length of segmentation window is set to $L$, then the model input after segmentation can be expressed as:

$$X = \{X_1, X_2, \cdots, X_L\} \quad (3)$$

$$X_p = \{d_p', d_{p+1}', \cdots, d_{m-L+p-1}'\}, 1 \leq p \leq L, L \in N \quad (4)$$

Then the theoretical output corresponding to the model input can be expressed as:

$$Y = \{Y_1, Y_2, \cdots, Y_L\} \quad (5)$$

$$Y_p = \{d_{p+1}', d_{p+2}', \cdots, d_{m-L+p}'\} \quad (6)$$

Integrate the multivariate data in the standardized training set into $[\text{samples, time_steps, features}]$ three-dimensional data, and input into the single hidden layer structure of the model. The hidden layer contains $L$ LSTM memory cells connected in series, where $H_L$ represents the output of the previous LSTM memory cell, and $C_L$ represents the state information. In order to reduce the problem of gradient disappearance and speed up the training, the ReLU function is used as the threshold activation function. Adding the Dropout algorithm to the hidden layer and randomly dropping the cell and its connection during the training process can alleviate the problem of over-fitting. Introduce the attention mechanism into the hidden layer and using the softmax activation function to calculate the attention probability distribution values $\alpha_i, \alpha_2, \cdots, \alpha_L$ assigned to each input, so that its output weight sum is 1, and generate an attention weight matrix $\alpha$ and context vector $V$. The calculation formulas are as follows:

$$q_i = \tanh(W'h_i + b_z), i \in N \quad (7)$$

$$\alpha_i = \frac{\exp(\text{score}(q_i, h_i))}{\sum_{k=1}^{L} \exp(\text{score}(q_k, h_k))}, \sum_{i=1}^{L} \alpha_i = 1 \quad (8)$$

$$V = \sum_{t=1}^{L} \alpha_t h_t, t \leq L \in N \quad (9)$$
In the above formulas, $W_s$ represents the weight matrix, $b_s$ represents the offset, and the score function is a similarity measurement function. The generated context vector $V$ is input into the softmax function for normalization, so that the sum is 1. Then, information is extracted through a fully connected layer.

After training in the model, the output of training set $D_{tr}$ can be expressed as follows:

$$P = \{P_1, P_2, \cdots, P_L\}$$

$$P_p = F(X_p, C_{p-1}, H_{p-1})$$

The error test between the model output $P$ and the theoretical output $Y$ is controlled by the loss function. The minimum loss function is taken as the optimization goal. Given the random seed number, learning rate, and training steps initialized by the network, the model continuously updates and corrects the network weights through the BPTT forward calculation method and the Adam stochastic gradient optimization algorithm, so as to get a more optimized hidden layer. The loss function uses the mean square error, the calculation formulas are as follows:

$$S_i = (p_i - y_i)^2$$

$$\text{loss} = \sum_{i=1}^{L} \left( \frac{S_i}{(m-L)} \right)$$

In the formulas, $p_i$ is the true value, $y_i$ is the predicted value, $L$ is the window length, and $m - L$ represents the number of elements in each sample sequence.

### 2.2. Network model prediction

Network model prediction is to use an iterative method to make predictions in the model trained in the previous section. First, input the last sequence of training set $D_{tr}$ into the trained attention-based LSTM network model, and obtain the predicted value sequence $P_{p}$ and the predicted value $P_{m+1}$ at time $m + 1$.

$$Y_d = \{d'_{m-L+1}, d'_{m-L+2}, \cdots, d'_{m}\}$$

$$P_p = \{p_{m-L+2}, p_{m-L+3}, \cdots, p_{m+1}\}$$

Then combine $P_{m+1}$ and the last $L-1$ data of $Y_d$ into a new row of data $Y_{d+1}$,

$$Y_{d+1} = \{d'_{m-L+2}, d'_{m-L+3}, \cdots, p_{m+1}\}$$

Input $Y_{d+1}$ into the trained model, you can get the predicted value $P_{m+2}$ at time $m + 2$, and so on, you can get the predicted value sequence $P_0 = \{P_{m+1}, P_{m+2}, \cdots, P_n\}$.

De-normalize the predicted value sequence $P^*$ and restore the predicted value to the time series data format consistent with the input, and finally obtain the predicted sequence corresponding to the test set sequence as:

$$P_{te} = \{p^*_{m+1}, p^*_{m+2}, \cdots, p^*_n\}$$

$$p^*_y = d_t \left( \frac{1}{n} \sum_{t=1}^{n} (d_t - \frac{1}{n} \sum_{t=1}^{n} d_t)^2 \right)^{1/2} + \frac{1}{n} \sum_{t=1}^{n} d_t$$

$$m + 1 \leq y \leq n, y \in N$$
The training fitting sequence $P_{tr}$ can be obtained by processing the model input sequence $X$ of the training set $D_{tr}$ in the same way. By calculating the deviation between $D_{tr}$ and $P_{tr}$, $D_{te}$ and $P_{te}$, the fitting accuracy and prediction accuracy of the model can be given quantitatively.

3. Simulation experiment

In this section, combined with the operation data of CNC machine tools, the prediction model proposed in Section 2 is used for simulation experiments.

3.1. Experiment preparation

This section mainly introduces the data set used in the simulation experiment and the configuration of the experimental environment.

3.1.1. Data set.

The data required for the experiment is provided by a sensor data set of a CNC machine tool in a mechanical processing plant. The data set contains data of 8 different types of sensors such as equipment power and spindle speed in the same time period. The data set has a time span of 6 days and a collection interval of 1 minute. After filtering the data set, there are 8 groups of valid data, 6000 in total. The 8 groups of data are divided into 2 groups of label data and 6 groups of continuous data. Figure 2 shows the visual features of 3 groups of data. Considering that the three data of equipment power, voltage and current play a key role in the normal operation of CNC machine tools, we decided to predict these three data. In order to make the input data multivariable, 8 groups of data are cleaned and integrated into three-dimensional data: [samples, time_steps, features], which are input into the hidden layer with time steps as the unique index.

![Figure 2. Visual features of time series data.](image)

3.1.2. Experimental environment.

The software and hardware environment used in this simulation experiment is Inter (R) CPU G630 2.70GHz, 4GB RAM, Windows10 (64-bit), Python 3.5, Keras 2.1.2, the backend is TensorFlow, and the integrated development environment is Visual Studio Code 1.27. The experiment uses the BP neural network model and the LSTM model as the comparative experimental model and uses MAPE and RMSE as the measurement standard to verify the prediction effect.
3.2. Experimental results

After comparative experiments, the results are shown in Table 1:

| Algorithm                          | RMSE  | MAPE  |
|------------------------------------|-------|-------|
| BP neural network                  | 8.223 | 3.553 |
| LSTM neural network                | 3.536 | 0.914 |
| LSTM-Attention neural network      | 1.814 | 0.556 |

Figure 3 shows the comparison results of the predicted and actual values of the equipment power of the two models (BP neural network does not participate in the comparison because of poor fitting). Blue is the predicted value, and red is the actual value. It can be seen from the figure that the prediction result of the time series data of the LSTM model based on the attention mechanism has a better fit with the actual data and the prediction is more accurate.

![Figure 3. Comparing results.](image)

4. Summary

Relying on the background of industrial big data and industrial intelligence, this paper deeply analyzes the overall logical sequence of industrial time series data, mines the complex temporal correlation between information, and proposes an LSTM neural network industrial time series data prediction method based on attention mechanism, which effectively improves the prediction accuracy of industrial time series data. The results show that the prediction effect of the proposed method is superior to the BP neural network model and the LSTM neural network model. The future direction of the paper considers to improve the method, improve its parallel computing capabilities, and shorten training time.

Acknowledgment

This paper was supported by Zhejiang provincial Key R&D Program of China (Grant No.2018C01084) and Zhejiang Gongshang University Graduate Research Innovation Fund.

References

[1] CHEN, Y.Z., FANG, M.Y., GUO, W.Z., GUO, K. (2015) Topic Popularity Prediction of Microblog Based on Wavelet Transformation and ARIMA[J]. Pattern Recognition and Artificial Intelligence, 28(7):586-594.

[2] LI, X.L., TAN, C.D., TAN, P.F., ZHANG, S.L. (2017) Application of ARIMA Model in Time Series Analysis to Predict the Efficiency of Pumping Well Production System[J]. Peak Data Science, 006(001):41-44.

[3] WANG, Y.M., MU, H.X., XU, G.N. (2019) Prediction of Output Power of Photovoltaic Cells Based on the Improved Particle Swarm Optimization Support Vector Machine[J]. Electrical Automation, 41(3):63-65.
[4] TAN, J., WANG, S.C. (2015) Research on prediction model for traffic congestion based on deep learning[J]. Application Research of Computers, 32(10):2951-2954.

[5] LIANG, Y., LIU, H.Z. (2010) Study of Image Retrieval Based on Vision Attention Mechanism[J]. Journal of Beijing Union University (Natural Sciences), 024(001):30-35.