Tool Wear Prediction Based on Edge Data Processing and Deep Learning Model

Dong Yan¹, Xinyu Ding²*, Shuaiyu Pan³, Haoyang Huang⁴

¹School of Automation, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
²School of Automation, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
³School of Automation, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
⁴School of Automation, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

*s180301066@stu.cqupt.edu.cn

Abstract. In order to improve the accuracy of tool wear prediction and enhance the real-time application in industrial sites, a tool wear prediction method based on edge data processing and CNN-BiGRU neural network is proposed. This method first implements data preprocessing on edge nodes, effectively reducing the amount of data transmission to avoid network link congestion. After that, the CNN-BiGRU neural network was deployed in the cloud for model training. Experimental results show that the tool wear prediction method based on edge data processing and CNN-BiGRU neural network has good real-time performance and high prediction accuracy in simulated industrial field applications.

1. Introduction

In turning and milling operations, the tool has long-term contact with the workpiece to cause wear. The tool must be replaced in time to avoid severe wear that affects the processing quality and processing efficiency. Therefore, it is necessary to monitor the tool wear. Stopping the machine to detect tool wear is time-consuming and laborious, which is not practical in practice. With the emergence of the latest technologies such as artificial intelligence (AI), cloud computing and edge computing, cyber-physical systems (CPS), and big data, innovations using technology convergence are accelerating the development of smart manufacturing [1]. Technologies such as machine learning and neural networks have been widely used in tool wear monitoring and prediction.

So far, a variety of machine learning models have been applied to tool wear monitoring and prediction. Liu et al. combined genetic algorithm and SVM to establish a tool wear prediction model, and got a lower error than the BP neural network method [3]. Xu et al. used wavelet packets to extract tool wear characteristics under different conditions, and used SVM classifier to identify tool wear [4]. Geramifard et al. used Hidden Markov Model for tool wear monitoring [5]. In addition to the above-mentioned traditional machine learning methods, deep learning has also been used for tool wear prediction in recent years. Deep learning algorithms are a promising research approach to automatically extract features under highly abstract conditions [5]. For various time series signals such as force and
vibration monitored during workpiece processing, the long short-term memory network (LSTM) can memorize the time series information in the sequence data, and better preserve the characteristic information. Long and short-term memory networks are used and have better prediction effects than models such as SVM and HMM [6] [7]. Although LSTM has a good training effect on time series signals, the network structure is more complicated and training takes a long time. And when applied in an industrial field, a large amount of data collected by the sensor in a short time must be transmitted to the cloud for training. Big data will take up too much bandwidth and the transmission time will be too long, and the real-time prediction cannot be satisfied.

A tool wear prediction method based on edge data processing and CNN-BiGRU neural network is proposed. Aiming at the problem of long training time, GRU network is introduced. GRU (Gated Recurrent Unit) is a simplified version of LSTM, with fewer GRU parameters and faster convergence, so it actually takes a lot less time, which can greatly speed up our iterative process. At the same time, in order to ensure the accuracy of prediction, the CNN network is added to make feature extraction more comprehensive. In response to the problem of excessive data transmission, edge computing is introduced. As a supplement to cloud computing, edge computing provides a new solution for solving problems such as high latency, ensuring data security, and saving bandwidth costs. By collecting and preprocessing data at the edge nodes deployed on the edge side of the machine tool, the amount of data transmission is greatly reduced, which is conducive to real-time prediction of tool wear.

The rest of this article is arranged as follows. Chapter 2 introduces the model framework, Chapter 3 introduces experimental verification and result analysis, and finally concludes.

2. Tool wear prediction model
The cloud can provide almost unlimited computing power and storage capacity. Putting all computing tasks on the cloud has proven to be an effective method of data processing, because the computing power on the cloud exceeds the ability of edge things [8]. However, taking the tool wear data provided by PHM 2010 as an example, the total amount of data collected by the sensor installed on the CNC machine tool for each tool can reach 2.94GB. The massive data of all tools in the workshop transmitted to the cloud for analysis and calculation is a great challenge to the network bandwidth, and the real-time prediction cannot be satisfied due to the transmission delay. Therefore, edge computing is introduced to preprocess the data on the edge nodes close to the machine tool before uploading to the cloud. The combined network of CNN and BiGRU is constructed in the cloud to predict tool wear values. The specific process is shown in Figure 1.

![Figure 1. Tool wear prediction model framework](image)
2.1. Edge data preprocessing
At the edge node, the collected sensor data is down-sampled to reduce the amount of data. For a sample $X_i(L, N)$ with a time step of $L$ and a dimension of $N$, the down-sampling process is shown in Table 1 below.

### Table 1. Down-sampling algorithm

| Down-sampling process |
|------------------------|
| For $j \in (0, K)$ do |
| For $m \in \left( \frac{(j-1)(L-\Delta)}{K}, \frac{(j+1)(L-\Delta)}{K} + \Delta \right)$ do |
| $X_i^j(K, N) \leftarrow \frac{K}{L-(1-K)\Delta} \sum_i X_i^m$ |

$X_i^j$ is the new sample after being downsampled by $X_i$; $X_i^j(K, N)$ represents the $j$th time step of $X_i$; $X_i^m$ represents the $m$th time step of $X_i$; $X_i^j(K, N) \in \mathbb{R}^{1\times N}$; $K$ is the new time step; $\Delta$ represents the coincidence step before down-sampling at the $j$th time step and the $j-1$th time step.

2.2. Normalized
The multi-sensor data that characterizes the change in tool wear needs to be normalized to make the data of different dimensions comparable. Normalize the down-sampled data to the interval $[0,1]$, for the $n$th dimension data $X_i^j(K, n)$:

$$X_i^j(K, n) = \frac{X_i^j(K, n) - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$

$x_{\text{max}}$ and $x_{\text{min}}$ respectively represent the maximum and minimum values of the $n$th dimension data.

2.3. CNN-BiGRU model establishment
The hybrid network of CNN and BiGRU is built in the cloud. The normalized data will be fed into the CNN network first. Convolutional neural network (CNN) extracts different local features through multiple convolution kernel operations, and the pooling layer is used to reduce dimensionality. Set up two convolutional layers and two maximum pooling layers, the number of convolution kernels, convolution kernel size and maximum pooling size are $[100, 100, 10]$ and $[100, 5, 2]$ respectively; in CNN Build a layer of BiGRU network for feature extraction of time series sequence. BiGRU is a combination of two unidirectional GRUs. At any time $t$, the input will be fed into the two GRUs in opposite directions at the same time, and the output will be jointly determined by the two unidirectional GRUs. BiGRU can learn the characteristics of the forward state and the backward state. The number of neurons in the hidden layer of the BiGRU network is set to 100. Dropout is introduced between the two layers of CNN and between CNN and BiGRU to alleviate the network model overfitting, and the Dropout parameter is set to 0.2. Both CNN and BiGRU network activation functions are set to Relu. Using the Adam optimizer, the loss function uses the mean square error (MSE), the mean square error can be expressed as:

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^{n} (y_i^t - y_t)^2$$

$y_i^t$ and $y_t$ represent the predicted value and true value of the wear amount respectively.

3. Experiments and results
The tool wear data set provided by PHM 2010 [9] was used to verify the performance of the proposed tool wear prediction method. The data set includes six individual tool records c1~c6. C1, c4 and c6 all include 7-dimensional sensor data collected from 315 cuts and the corresponding three-dimensional tool wear value for each cut. The sensor data acquisition frequency is 50KHz, and data of more than 200,000 steps are collected each time. Use the data down-sampling method mentioned above to downsample c1, c4, and c6. The results are shown in Table 2.
Table 2. Data volume comparison

| Sample | Raw sensor data size (MB) | Processed data size (MB) |
|--------|--------------------------|--------------------------|
| c1     | 2959                     | 31                       |
| c4     | 3011                     | 31                       |
| c6     | 2949                     | 31                       |

After down-sampling, the reforming time step is 1000 and the dimension is 7, the data volume of each tool is reduced from about 3GB to 31MB, and the transmission data is reduced by 99%.

The cross-validation method is adopted. Two of the three tool data sets are selected as training data, and the other is used as test data. 630 training samples and 315 test samples are obtained. Calculate the average of the three-dimensional tool wear value corresponding to each sample as the label. The performance of tool wear prediction under different data sets is shown in Figure 2.

(a) C1

(b) C4
Figure 2. Tool wear prediction results under different data sets

It can be observed from Figure 2 that the prediction curve fits well with the real curve. The root mean square error (rmse) and mean absolute error (mae) are used to evaluate the performance of the model, and the prediction results under the CNN and GRU network models are used for comparison. The performance of each model is shown in Table 3 and Table 4.

Table 3. Rmse comparison

| Model     | C1     | C4     | C6     |
|-----------|--------|--------|--------|
| CNN       | 9.26   | 12.80  | 12.26  |
| GRU       | 7.89   | 13.40  | 9.89   |
| **CNN-BiGRU** | **7.49** | **10.04** | **7.63** |

Table 4. Mae comparison

| Model     | C1     | C4     | C6     |
|-----------|--------|--------|--------|
| CNN       | 7.08   | 8.64   | 10.45  |
| GRU       | 5.81   | 8.52   | 8.37   |
| **CNN-BiGRU** | **5.76** | **6.80** | **6.11** |

The comparative experiment shows that the CNN-BiGRU used has achieved the best prediction performance among the three models.

4. Conclusions

A tool wear prediction method based on edge data processing and CNN-BiGRU is proposed. The original data is down-sampled on the edge nodes to shorten the time step. Deploy the CNN-BiGRU network model in the cloud. Finally, the method was verified using public data sets. The experimental results show that the transmission data is reduced by 99%, and the edge data processing is beneficial to alleviate the bandwidth pressure and improve the real-time performance of industrial field applications. Under the model proposed in this paper, the wear prediction results of the three tools are better fitted to the actual measurement results, and the effect is better than the separate CNN and GRU models. In the future, we will explore the initial extraction of time-domain and frequency-domain features on the edge, and then transfer the features to the cloud network for training.

Acknowledgements

This work is grateful for the support of the Edge Computing Verification Platform and Discrete Industry Solution project (No.2018YFB700205).
References
[1] C. Y. Park, K. B. Laskey, S. Salim and J. Y. Lee, "Predictive situation awareness model for smart manufacturing," 2017 20th International Conference on Information Fusion (Fusion), Xi'an, 2017, pp. 1-8.
[2] W. Liu, L. Liu, H. Zhang and Z. Yang, "Research of Tool Wear Predictive Technique Based on Support Vector Machine," 2009 2nd International Congress on Image and Signal Processing, Tianjin, 2009, pp. 1-3.
[3] Xu Tao and Wang Tao, "Cutting tool wear identification based on wavelet package and SVM," 2010 8th World Congress on Intelligent Control and Automation, Jinan, 2010, pp. 5953-5957.
[4] O. Geramifard, J. Xu, J. Zhou and X. Li, "Multimodal Hidden Markov Model-Based Approach for Tool Wear Monitoring," in IEEE Transactions on Industrial Electronics, vol. 61, no. 6, pp. 2900-2911, June 2014.
[5] Najafabadi, M.M., Villanustre, F., Khoshgoftaar, T.M. et al. Deep learning applications and challenges in big data analytics. Journal of Big Data 2, 1 (2015).
[6] A. Al-Dulaimi, S. Zabihi, A. Asif and A. Mohammadi, "Hybrid Deep Neural Network Model for Remaining Useful Life Estimation," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 3872-3876.
[7] M. Wang, J. Zhou, J. Gao, Z. Li and E. Li, "Milling Tool Wear Prediction Method Based on Deep Learning Under Variable Working Conditions," in IEEE Access, vol. 8, pp. 140726-140735, 2020.
[8] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge Computing: Vision and Challenges," in IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637-646, Oct. 2016.
[9] X. Li, B.S.Lim,J.H.Zhou, S. Huang,S.J.Phua, K.C.Shaw, and M.J.Er,"Fuzzy neural network modelling for tool wear estimation in dry milling operation," in Proc.Annu.Conf. Prognostics Health Manage.Soc.(PHM),San Diego, CA,USA, Sep. 2009, pp.1-11.