Thermal error modeling based on BiLSTM deep learning for CNC machine tool

Pu-Ling Liu¹ · Zheng-Chun Du¹ · Hui-Min Li¹ · Ming Deng¹ · Xiao-Bing Feng¹ · Jian-Guo Yang¹

Received: 1 September 2020 / Revised: 21 October 2020 / Accepted: 26 December 2020 / Published online: 21 February 2021
© The Author(s) 2021

Abstract The machining accuracy of computer numerical control machine tools has always been a focus of the manufacturing industry. Among all errors, thermal error affects the machining accuracy considerably. Because of the significant impact of Industry 4.0 on machine tools, existing thermal error modeling methods have encountered unprecedented challenges in terms of model complexity and capability of dealing with a large number of time series data. A thermal error modeling method is proposed based on bidirectional long short-term memory (BiLSTM) deep learning, which has good learning ability and a strong capability to handle a large group of dynamic data. A four-layer model framework that includes BiLSTM, a feedforward neural network, and the max pooling is constructed. An elaborately designed algorithm is proposed for better and faster model training. The window length of the input sequence is selected based on the phase space reconstruction of the time series. The model prediction accuracy and model robustness were verified experimentally by three validation tests in which thermal errors predicted by the proposed model were compensated for real workpiece cutting. The average depth variation of the workpiece was reduced from approximately 50 μm to less than 2 μm after compensation. The reduction in maximum depth variation was more than 85%. The proposed model was proved to be feasible and effective for improving machining accuracy significantly.

Keywords Thermal error · Error modeling · Bidirectional long short-term memory (BiLSTM) · Phase space reconstruction · Computer numerical control (CNC) machine tool

1 Introduction

With the continuous improvement of production automation, manufacturing industries have increasingly higher requirements for the machining accuracy of computer numerical control (CNC) machine tools. An increase in temperature results in thermal deformation, and thermal deformation leads to a relative displacement at the cutting point and, thus, influences the accuracy of the workpiece being produced [1]. Thermally induced errors have become one of the largest machine tool error sources in precision and ultraprecision machining. Up to 75% of the overall geometrical errors of machined workpieces can be induced by the effects of temperature [2]. Thermal error plays a significant role in limiting the accuracy of the produced workpiece.

To eliminate the negative influence of thermal error on the machining accuracy, a variety of thermal error modeling approaches have been proposed by many researchers worldwide. For most modeling methods, the relationships between the temperatures and thermal errors are studied based on either the numerical simulation results or experimental data [3], namely, the principal-based model (PBM) and empirical-based model (EBM). The PBM builds the relationship between thermal errors and heat generated by a system of nonlinear differential equations, which frequently employs the principles of friction-induced heat, heat conduction, heat convection, heat radiation, and thermal expansion in the model construction. Liu et al. [4]

✉ Zheng-Chun Du
zcd@sjtu.edu.cn

¹ School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

established a thermal resistance network model of the motorized spindle system of a CNC machining tool. The influences of the radial and axial thermal conduction resistance and thermal convection resistance between the cooling system and the components of the spindle system on the temperature rise are considered in the model. Li et al. [5] studied the modal characteristics under both a steady state and a static state of the spindle by using the finite-element method. The thermal characteristics, vibration modes, and natural frequencies were also analyzed. Liu et al. [6] analyzed the thermal behavior of the machine-workpiece system and established a thermal error prediction model based on heat transfer theory. The comprehensive thermal error of the machine-workpiece system was reduced to approximately 20%.

The EBM is a kind of black-box methodology that assumes thermal errors can be considered as a function of some critical thermal discrete temperature points on the machine. Regression analysis and artificial neural networks (ANNs) are the most commonly used EBM methods. Ye et al. [7] proposed a thermal error regression model for determining the thermal deformation coefficient of the moving shaft of a gantry milling machine. Grama et al. [8] developed a thermal error compensation model in a linear regression framework, where the thermal key point was selected using principal component analysis and k -means clustering. Li et al. [9] established a thermal error prediction model based on an improved backpropagation (BP) neural network. The temperature measurement points were clustered by a self-organization mapping neural network. Improved particle swarm optimization is used to optimize the parameters of the BP neural network. The prediction accuracy of the improved model for the spindle thermal error is 93.1%.

The prediction accuracy of the PBM and EBM methods has been demonstrated to be satisfactory in many applications. However, the existing modeling methods have encountered unprecedented challenges because of the significant impact of Industry 4.0 on machine tools. On the one hand, the accuracy of the PBM model mainly depends on the accuracy of the heat source model and heat transfer model. It is difficult to obtain an accurate thermal error model because it depends highly on a number of factors, such as machine working cycles, the use of coolants, and environmental conditions [10]. The PBM models become highly complicated if they consider all the essential influential effects within a machine tool, and they finally reach their limit owing to the increasing complexity and high demand from production [11]. On the other hand, the thermal error data of CNC machine tools are a group of sequential, large, fast, and continuous nonlinear data sequences, which are constantly generated with the operation of the machine tools. The existing EBM models are

incapable of processing such a large group of data with high efficiency and good robustness. Therefore, a thermal error model with a relatively simple framework, as well as the ability to handle big data with learning abilities, is desired.

Deep learning has rapidly advanced and has achieved state-of-the-art performance in various fields. Deep learning can learn complex features by combining simple features learned from data. It takes advantage of large datasets and computationally efficient training algorithms to outperform other approaches in various machine learning tasks [12]. The state of the bidirectional long short-term memory (BiLSTM) deep learning network changes with time. It can essentially be regarded as a dynamic system. The thermal error of the CNC machine tool changes with the change in the machine temperature field. The thermal error data have time sequence and continuity characteristics, and the current and historical states are interrelated. Therefore, the thermal error accords with the deep learning network in terms of the dynamic time series. The use of BiLSTM deep learning to extract the temporal and spatial characteristics of dynamic time series data is conducive to the accurate prediction of thermal error.

In this article, a thermal error modeling method based on BiLSTM deep learning is proposed. The model consists of one BiLSTM layer, two feedforward neural network (FFN) layers, and one max-pooling layer. A training algorithm is designed to realize the model training. The window length of the input sequence is determined by the proposed input sequence construction method based on the phase space reconstruction of the time series. In addition, verification tests were conducted theoretically and practically to prove the accuracy and robustness of the proposed BiLSTM-based thermal error model. The basic principle of BiLSTM is introduced in Sect. 2. The process of establishing the thermal error model, which includes the model framework, training algorithm, and input sequence construction method, is given in Sect. 3. In Sect. 4, the verification tests and application results of the proposed BiLSTM-based thermal error model are described. Finally, some conclusions are presented in Sect. 5.

2 Basic principle of BiLSTM

The concept of long short-term memory (LSTM) was proposed by Hochreiter and Schmidhuber [13]. The basic LSTM unit consists of three gates and two conveyor belts that protect the state of each neuron. It controls the transfer path of information through the gating mechanism. Four neural network layers interact with each other in a special way in one LSTM unit, as shown in Fig. 1.

The state calculations for each step are presented as [14]

Fig. 1 Illustration of LSTM unit structure

$$F_t = \sigma(W_f \cdot [H_{t-1}, X_t] + b_f), \quad (1)$$

$$I_t = \sigma(W_i \cdot [H_{t-1}, X_t] + b_i), \quad (2)$$

$$O_t = \sigma(W_o \cdot [H_{t-1}, X_t] + b_o), \quad (3)$$

$$C'_t = \tanh(W_c \cdot [H_{t-1}, X_t] + b_c), \quad (4)$$

$$C_t = F_t \odot C_{t-1} + I_t \odot C'_t, \quad (5)$$

$$H_t = O_t \odot \tanh(C_t), \quad (6)$$

where F_t represents the forget gate at time t , to decide the information needs to be thrown away from the cell state. In addition, I_t represents the input gate, to decide the new information needs to be stored in the cell state, and O_t represents the output gate, to decide the information needs to be output. These three gates use the sigmoid function to filter information. A value of zero means “let nothing through”, while a value of one means “let everything through”. Moreover, C'_t is the candidate cell state; C_t is the cell state; H_t is the hidden layer state; X_t is the model inputs; W_f , W_i , W_o and W_c are the weight matrices; b_f , b_i , b_o and b_c are the biases; and “ $\cdot$ ” and “ $\odot$ ” denote the matrix multiplication and element multiplication, respectively.

The above operating mechanism of LSTM is explicitly designed to avoid the long-term dependency problem, which is common for recurrent neural networks (RNNs). LSTM can remove or add information to the cell state, which is carefully regulated by different gates. Therefore, learning from information over a long period of time is the default behavior of LSTM.

The BiLSTM consists of two LSTM layers with opposite directions, as shown in Fig. 2. The hidden layer state $\overrightarrow{H_t}$ encodes the information features in the forward direction, while the hidden layer state $\overleftarrow{H_t}$ encodes the information features in the backward direction. BiLSTM can learn information more precisely by utilizing the forward and backward orders of the information sequence. Thus, it is especially suitable for dealing with time series data.

Fig. 2 Illustration of BiLSTM model structure

3 Thermal error modeling based on BiLSTM

3.1 Model framework

The proposed model framework is illustrated in Fig. 3. The input layer contains the model input sequence at and before time t . The window length n of the input sequence is discussed in Sect. 3.3.2. The output layer gives the thermal error value predicted by the model at time $t + 1$, which is marked as Y_{t+1} , and h is the hidden size of the model. The hidden layer consists of four layers: one BiLSTM layer, two FFN layers, and one max-pooling layer.

- (i) The BiLSTM layer is the core part of the model. The main function of the BiLSTM layer is feature extraction from the input sequence. Thorough a series of operations, training, and learning using Eqs. (1)–(6), the weight parameters W_f , W_i , W_o and W_c and bias parameters b_f , b_i , b_o and b_c are obtained to finalize the structure of the model.
- (ii) The FFN layers are used to realize dimensionality reduction by linear transformation. The first FFN layer reduces the hidden layer state of the BiLSTM from $2h$ to h , and the second FFN layer reduces the prediction vector from h to 1, i.e., a predicted single thermal error value.
- (iii) The max-pooling layer is used after the first FFN layer to transform an h -dimension hidden state of the BiLSTM into an h -dimension hidden space vector, namely the prediction vector in Fig. 3, by its characteristic of translation invariance.

3.2 Training algorithm

The model training is carried out according to an elaborately designed algorithm based on the above framework, which is shown in Fig. 4. There are four main steps.

Fig. 3 Illustration of model framework

Fig. 4 Flow chart of the model training algorithm

- (i) Hyperparameters must be set before model training. These parameters are used to define high-level concepts of the model, such as complexity or learning ability. Details are discussed in Sect. 3.3.3.
- (ii) Data preprocessing helps the model converge more efficiently and accurately. Different variables often have different dimensions and units, which negatively influence the results of data analysis. The method of deviation standardization is used to eliminate the dimensional influence between variables, which makes the data processing more convenient and faster. Deviation standardization is a linear transformation of the

original variable x , so that the transformed result x^* is mapped between $[0,1]$. The conversion function is given as

$$x^* = \frac{x - S_{\min}}{S_{\max} - S_{\min}}, \quad (7)$$

where $S_{\min}$ is the minimum value of the sample data, and $S_{\max}$ is the maximum value of the sample data.

- (iii) Weight initialization has a significant influence on the final training results of the model. If the initial value of the weight is too large, it causes a gradient explosion and makes the model unable to converge. If the initial value of the weight is too small, the gradient disappears, and the model converges slowly or converges to the local minimum value. The method of Kaiming initialization is adopted because the activation function sigmoid in the BiLSTM model is not symmetric about zero, which improves the convergence of the model. Kaiming initialization is a Gaussian distribution with a mean value of zero and variance of $\sqrt{\frac{2}{F_{\text{in}}}}$, shown as

$$w \sim G\left[0, \sqrt{\frac{2}{F_{\text{in}}}}\right], \quad (8)$$

where F_{in} is the number of input neurons, and w is the weight.

- (iv) Model parameter training determines the weight parameters and bias parameters of the model. The determination of model parameters means the finalization of the model. First, a loss function that evaluates the inconsistency between the predicted value and the real value of the model must be defined. The model established is a thermal error prediction model, which belongs to the category of regression problems (as opposed to

classification problems). Therefore, mean squared error is selected as the loss function, shown as

$$J = \frac{\sum_{i=1}^k (y_i - y'_i)^2}{k}, \quad (9)$$

where J is the loss function, k the sample size, y_i the measured value, and y'_i the predicted value.

Second, the parameter gradient is calculated by the loss function, and the gradient is optimized by appropriate methods to minimize the loss function and make the model converge. In the gradient calculation, the L^2 regularization method is used to limit the scale of the parameters to prevent overfitting, that is, to add a regularization term after the loss function, shown as

$$J' = J + \frac{\lambda}{2k} \|\mathbf{w}\|_2^2, \quad (10)$$

where J' is the loss function after regularization, and λ is the regularization parameter.

Next, a gradient clipping process is added to prevent the gradient explosion and ensure the convergence of the training process. By setting the clipping threshold, i.e., the maximum gradient norm, the gradient exceeding the threshold is regulated, shown as

$$\mathbf{t}^* = \begin{cases} \mathbf{t}^* \times \frac{\theta}{\|\mathbf{t}^*\|_2}, & \text{if } \|\mathbf{t}^*\|_2 \geq \theta, \\ \mathbf{t}^*, & \text{otherwise,} \end{cases} \quad (11)$$

where $\mathbf{t}^*$ is the tensor, θ the clipping threshold, and $\|\mathbf{t}^*\|_2$ the L^2 norm of $\mathbf{t}^*$.

Thereafter, the Adam optimization algorithm is used to adapt the learning rate of each parameter to achieve a better and faster convergence of the training process. The iteration formula of the Adam algorithm is shown as

$$\hat{\mathbf{v}}_t \leftarrow \frac{\mathbf{v}_t}{1 - \beta_1^t}, \quad (12)$$

$$\hat{\mathbf{s}}_t \leftarrow \frac{\mathbf{s}_t}{1 - \beta_2^t}, \quad (13)$$

$$\mathbf{g}'_t \leftarrow \frac{\eta \hat{\mathbf{v}}_t}{\sqrt{\hat{\mathbf{s}}_t + \varepsilon}}, \quad (14)$$

where $\mathbf{v}_t$ is the momentum variable (first moment), $\mathbf{s}_t$ the exponential weighted moving average variable by the square of elements (second moment), $\hat{\mathbf{v}}_t$ and $\hat{\mathbf{s}}_t$ the revised $\mathbf{v}_t$ and $\mathbf{s}_t$, respectively, after deviation correction, β_1 a constant of the first-order decay rate, β_2 a constant of the second-order decay rate, η the learning rate, ε the stability constant, and $\mathbf{g}'_t$ the calculation update after Adam optimization.

Finally, the parameter weight $\mathbf{w}$ is iterated according to

$$\mathbf{w}_t \leftarrow \mathbf{w}_{t-1} - \mathbf{g}'_t. \quad (15)$$

One-time training of the model parameters is completed. Then, a validation dataset is put in the model to calculate the loss function. If the loss function of the validation dataset cannot meet the requirements after all iterations are completed, hyperparameters must be optimized until a satisfactory loss function value of the validation dataset is obtained.

3.3 Thermal error model establishment

3.3.1 Data collection

Thermal error exists in all three axes of the machine coordinate system. In this article, only the z -direction is taken as an example to illustrate the modeling process. The studied machine tool is a horizontal machining center. The simplified machine tool structure is shown in Fig. 5. The axial direction of the spindle is in the z -direction. The spindle moves up in the x - and y -directions. The workpiece is fixed on the workbench and moves up in the z -direction. The modeling object is the comprehensive axial spindle thermal error at the tool center point (TCP), which includes the thermal error of the machine tool and the thermal error of the tool.

A computer-controlled testing setup was built to measure the error data automatically at the TCP and temperature data at various locations in the machine tool. The sampling interval was 2 s. An Omron eddy current sensor and a PT101 thermal sensor were used. Figure 6 shows the placement of the eddy current sensor during the measurement. The mandrel is used before the blade making instead of the real tool because the tip of the three-blade milling cutter is not a flat plane that cannot be measured by an eddy current sensor.

The error measured at the TCP, marked as E_{TCP} , includes the thermal error of the machine tool E_{TM} , thermal error of the tool E_{TT} , static errors of the machine tool E_{SM} , and static errors of the tool E_{ST} . The comprehensive thermal error E_{TMT} at the TCP can be obtained using

$$E_{TCP} = E_{TM} + E_{SM} + E_{TT} + E_{ST}, \quad (16)$$

Fig. 5 Schematic diagram of machine tool structure

Fig. 6 Placement of eddy current sensor for thermal error measurement

$$E_{TMT} = E_{TM} + E_{TT}, \quad (17)$$

$$E_{TMT} = E_{TCP} - E_{SM} - E_{ST}, \quad (18)$$

where E_{SM} is determined experimentally, and E_{ST} is calculated according to linear-elastic physical principles [15].

Figure 7 shows the measurement result of E_{TMT} . The blue line is the rotation speed of the spindle during the measurement. The temperature data of each position are shown in Fig. 8. Table 1 gives the details of the temperature sensor positions, which are shown in Fig. 5. Except for the temperature sensor of room temperature, all the temperature sensors were sensors built into the machine tool. The measurement was repeated under different room temperatures for another three times to obtain enough datasets for model training and model validation. Two datasets were used for model training; one dataset was used for model validation; and another one was used for model testing.

3.3.2 Input sequence construction

In the input layer, as shown in Fig. 3, $\{X_{t-n+1}, X_{t-n+2}, \dots, X_t\}$ is the input sequence of the model, where n is the window length. Moreover, $n - 1$ is the number of input sequences, and it is also the number of steps to form a deep learning network in the time dimension. Steps that are too short lead to information loss, while steps that are too long

Fig. 7 Measuring data of thermal error E_{TMT}

Fig. 8 Temperature data measuring results

Table 1 Position of temperature sensor

Parameter	Definition
T_s	Motor of spindle
T_x	Motor of x-axis
T_y	Motor of y-axis
T_z	Motor of z-axis
T_{B1}	Front bearing of spindle
T_{B2}	Front bearing of spindle
T_{B3}	Back bearing of spindle
T_{room}	Ambient temperature

lead to information redundancy and even invalid information, which affect the model accuracy.

A window length selection method based on the phase space reconstruction of a time series is proposed [16]. According to the theorem of Takens, there exists a best embedding dimension of the reconstructed phase space that has the same geometric characteristics as the original space. This best embedding dimension m is the optimal number of input sequences, i.e., $m = n - 1$. When two nonadjacent points in a high-dimensional phase space are projected into a low-dimensional phase space, they may become two adjacent points, that is, false adjacent points. With the increase in the dimension of the reconstructed phase space, the false adjacent points are eliminated gradually.

The time series of input parameter x in X_t can be expressed as $\{x_i | i = 1, 2, \dots, N\}$, where N is the length of the time series. Then, the reconstructed phase space of x with embedding dimension m and delay time τ can be expressed as

$$\mathbf{x}' = \{x'_1, x'_2, \dots, x'_i, \dots, x'_M\}, \quad (19)$$

$$\mathbf{x}'_i = \{x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}\}, \quad (20)$$

where $1 \leq i \leq M$ and $M = N - (m - 1)\tau$.

For each $\mathbf{x}'_i$, there is an adjacent point $\mathbf{x}'_j$ based on the Euclidean norm $R_i(m)$, shown as

$$R_i(m) = \|\mathbf{x}'_i(m) - \mathbf{x}'_j(m)\|_2, \quad (21)$$

where $1 \leq j \leq M$. When the dimension changes from m to $m+1$, the distance becomes

$$R_i(m+1) = \|\mathbf{x}'_i(m+1) - \mathbf{x}'_j(m+1)\|_2. \quad (22)$$

The relation between $R_i(m)$ and $R_i(m+1)$ is shown as

$$(R_i(m+1))^2 = (R_i(m))^2 + \|\mathbf{x}'_{i+m\tau} - \mathbf{x}'_{j+m\tau}\|_2^2. \quad (23)$$

$\delta_i(m)$ represents the degree of distance change and is described as

$$\delta_i(m) = \frac{\|\mathbf{x}'_{i+m\tau} - \mathbf{x}'_{j+m\tau}\|_2}{R_i(m)}. \quad (24)$$

If $\delta_i(m)$ exceeds a certain threshold δ_τ , then $\mathbf{x}'_j(m)$ is a false adjacent point of $\mathbf{x}'_i(m)$. The corresponding embedding dimension m is the best embedding dimension of the reconstructed phase space, δ_τ is set as 30%, and the calculation of $\delta_i(m)$ starts from $m=2$. The calculation results of the best embedding dimension are shown in Table 2. The maximum embedding dimension of all input parameters is 4; therefore, the window length of the input sequence n is set to 5.

3.3.3 Hyperparameter setting

Setting the appropriate hyperparameters is helpful to improve the performance and effectiveness of model learning. Based on the training algorithm proposed in Sect. 3.2, the following hyperparameters, which are shown in Table 3, must be defined before training. The window length is determined by the proposed method based on phase space reconstruction. Hidden size, batch size, epoch,

Table 3 Hyper-parameters setting

Hyper-parameter	Setting value
Hidden size	20
Window length	5
Input length	9
Batch size	128
Epoch	1 000
Shuffle	True
Learning rate	0.001
Regulation norm	0.01
Maximum gradient norm	10
Adam first order decay rate	0.9
Adam second order decay rate	0.999
Adam stability constant	10^{-8}

Table 4 Optimal loss function of validation dataset and test dataset

Loss function validation dataset	Loss function test dataset	Loss function requirement
0.072	0.024	< 1

Fig. 9 Comparison between the model predicted thermal error and the measured thermal error

Table 2 Embedding dimension for each input parameter

Input parameter		Embedding dimension
x_1	T_s	3
x_2	T_x	2
x_3	T_y	2
x_4	T_z	2
x_5	T_{B1}	3
x_6	T_{B2}	3
x_7	T_{B3}	4
x_8	T_{room}	2
x_9	Rotation speed	4

Table 5 Evaluation result of each model

Model	RMSE	MAE/ μm	MAPE	Max absolute residual error/ μm	Prediction accuracy/%
FFN	0.605	0.508	1.29	1.74	90.8
RNN	0.423	0.331	0.91	1.56	91.3
BiRNN	0.378	0.313	0.85	1.63	91.8
BiLSTM	0.154	0.111	0.34	0.98	95.7

Fig. 10 Thermal error prediction results of the other three models

Fig. 11 A zoom-in observation of the comparison between four models

Table 6 Operation procedures of robustness verification test

Sequence number	Rotation speed of spindle /(r·min ⁻¹)	Running time/s
1	12 000	3 600
2	0	1 200
3	6 000	600

Fig. 12 Thermal error measurement data and model prediction data

shuffle, learning rate, regulation norm, and maximum gradient norm are determined by a simple searching method. Other hyperparameters, including those not listed in the table, are set according to experience value or default value. The best group of hyperparameters that has the lowest loss function of the validation dataset is also listed in Table 3.

Fig. 13 Temperature measurement data of robustness verification test

Fig. 14 Real picture of a machined workpiece

Table 7 Dimension of each hole

Hole number	Diameter/mm	Depth/mm
1	30	8
2	30	5
3	30	10
4	50	8
5	50	10
6	50	5
7	80	5
8	80	8
9	80	10

3.3.4 Model training results

The model parameters are determined according to the optimal loss function value of the validation dataset, and the model was further verified by the test dataset. Table 4 shows the optimal loss function of the validation dataset and test dataset. Based on the model that is established with the above hyperparameters, the loss function of the validation dataset meets the requirement, and the loss function level of the test dataset shows the good prediction accuracy.

Fig. 15 Picture of test site and CNC machine tool

Table 8 Descriptions on the cutting process

Step	Spindle speed/ (r·min ⁻¹)	Time/ min	Feed rate/ (mm·min ⁻¹)
Facing milling	2 000	4	600
Rough milling	5 000	8	3 000
Semi-finish milling	2 000	5	250
Finish milling	2 000	10	50

of the model. A comparison between the model-predicted thermal error and the measured thermal error of the test dataset is shown in Fig. 9. The residual error is less than 1 μm , and the prediction accuracy is more than 95.7%, which is a fairly good performance.

4 Verification of BiLSTM-based thermal error model

4.1 Prediction accuracy comparison with models based on FFN/RNN/BiRNN

The proposed BiLSTM model was compared with the traditional neural networks FFN and RNN, as well as a bidirectional RNN (BiRNN) deep learning network. The comparison test was performed with the same dataset as in Section 3.3. The input sequence and hidden size of the network were kept the same. Three evaluation indices that are commonly used in the field of machine learning in regression problems, are shown as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N |y_i - y'_i|^2}{N}}, \quad (25)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - y'_i|, \quad (26)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - y'_i}{y_i} \right| \times 100\%, \quad (27)$$

where y_i is the measured value, y'_i the predicted value, and N the sample size. Equations (25)–(27) were used to evaluate the model performance.

The calculation results of the evaluation indices, as well as the maximum absolute residual error and prediction accuracy, are listed in Table 5. The thermal error prediction results of the other three models are shown in Fig. 10. For a clearer presentation, a zoom-in observation of the comparison of these four models is shown in Fig. 11. The comparison test result shows that the proposed BiLSTM model outperforms the other three models in all three evaluation indices, and it has the best performance in terms of both the residual error and prediction accuracy.

Fig. 16 Schematic diagram of the thermal error compensation system

Fig. 17 Error compensation scheme

4.2 Robustness verification

The training, validation, and test datasets used previously were obtained under the same running program of the machine tool. To further confirm the prediction accuracy and robustness of the proposed BiLSTM-based model and avoid unnecessary cost loss before the workpiece cutting test, the thermal errors under different operation procedures were measured and predicted by the established model. The operation procedures are listed in Table 6. The thermal errors are shown in Fig. 12, and the temperature measurement data are shown in Fig. 13. The maximum absolute residual error was $1.29\ \mu\text{m}$, and the prediction accuracy was 94.9%. The prediction accuracy of the proposed model was maintained at a relatively high level under different running programs, verifying the good robustness of the established model.

4.3 Thermal error compensation on workpiece cutting process

The effectiveness of the proposed thermal error model based on BiLSTM deep learning has been proved theoretically in previous verifications. In this section, the validation of the model through practical verification by an actual workpiece cutting test is discussed. An artificially designed workpiece made of aluminum alloy was used to check the effectiveness of the proposed model in real applications. Figure 14 shows a real picture of a machined

workpiece, which has nine holes with different diameters and depths. The dimensions of each hole are shown in Table 7. Only the accuracy of depth was considered in this study, because the direction of the hole depth was the z -direction.

The workpiece cutting test was performed in a horizontal machining center (HMC-C100P), as shown in Fig. 15. Detailed descriptions of the cutting process are listed in Table 8. Workpieces were machined with and without a thermal error compensation system. The sample size of each group needed to be at least 30 pieces to achieve statistical significance. The changes in the workpiece at the depth of the hole before and after compensation were compared to verify the validity of the proposed BiLSTM-based thermal error model.

4.3.1 Error compensation system

A schematic diagram of the thermal error compensation system is illustrated in Fig. 16. This compensation system was developed based on a fast Ethernet data interaction technique and the external machine original coordinate shift (EMZPS) function of the machine tool [17]. The error prediction value calculated by the error prediction model was stored in the R address of the programmable machine controller (PMC) through data interaction. When the data between the machine tool and PMC were scanned and interacted, the error prediction value in the R address was automatically obtained and combined with the

Fig. 18 Comparison on the measurement results of the depth of hole

Table 9 Main differences in workpiece accuracy before and after compensation

Group	Variation range/ μm			Variation average/ μm			Accuracy improvement/%		
	Depth = 5 mm	Depth = 8 mm	Depth = 10 mm	Depth = 5 mm	Depth = 8 mm	Depth = 10 mm	Depth = 5 mm	Depth = 8 mm	Depth = 10 mm
	Depth = 5 mm	Depth = 5 mm	Depth = 5 mm	Depth = 5 mm	Depth = 5 mm	Depth = 5 mm	Depth = 5 mm	Depth = 5 mm	Depth = 5 mm
Before compensation	8.0–103.0	–1.0–94.0	24.0–90.0	45.3	46.1	54.4	86.7	85.6	85.5
After compensation	–13.6–13.7	–9.5–13.5	–9.5–13.0	1.8	1.2	1.9			

interpolation algorithm in the next stage. The compensation value in each scanning cycle was iterated to the interpolation step to realize the modification of the interpolation vector and complete the error compensation.

The compensation system consisted of a hardware platform and a software platform. The compensation scheme is shown in Fig. 17. The hardware platform was implemented on an embedded computer, which included an error compensation unit and a data acquisition unit. The software platform was developed with LabVIEW, which is associated with the embedded MATLAB procedure, which includes an error modeling module and an error compensation module.

4.3.2 Thermal error compensation result

The depth of the hole of all machined workpieces was measured using a coordinate measuring machine (CONURA-G2) in a standard environment. A comparison of the measurement results is shown in Fig. 18. The vertical coordinate Δh is the relative deviation of the measured hole depth value to the center value. The horizontal coordinate is the serial number of each hole according to the cutting order.

The main differences in workpiece accuracy before and after thermal error compensation are listed in Table 9. The average depth variation is reduced from approximately 50 μm to less than 2 μm after compensation. The maximum depth variation range is significantly reduced to $\pm 15 \mu\text{m}$. The workpiece accuracy in the depth of the hole is improved by more than 85%. These results acknowledge the excellent prediction accuracy of the proposed BiLSTM-based thermal error model.

5 Conclusions

A thermal error modeling method was proposed based on BiLSTM deep learning for a horizontal turning center. The model is composed of one BiLSTM layer, two FFN layers, and one max-pooling layer. With a specially designed gate mechanism and bidirectional structure in the BiLSTM layer, the model has a superior learning ability from both short-term and long-term information to extract the temporal and spatial characteristics of dynamic time series temperature data and thermal error data.

A training algorithm was elaborately designed for model training to obtain an optimal loss function of the validation data, which is crucial to the thermal error modeling. The training algorithm adopts the gradient clipping method and Adam optimization algorithm to prevent gradient explosion and ensure the convergence of the training process.

A window length selection method for input sequences was proposed based on the phase space reconstruction of a time series. The best embedding dimension of the reconstructed phase space is used to determine the window length of the input sequence because it has the same geometric characteristics as the original space, which benefits the efficiency of model learning.

Three validation tests were performed to verify the accuracy and robustness of the proposed BiLSTM-based thermal error model. (i) Compared with the traditional modeling method based on FFN/RNN/BiRNN, the proposed model outperformed the other three models in all three evaluation indices, as well as in residual error and prediction accuracy. (ii) The prediction accuracy of the proposed model was verified with a different running program of the machine tool. The test results show that the proposed model has good robustness under different machining conditions. (iii) The effectiveness of the proposed thermal error model was validated by a real workpiece cutting test. The average depth variation of the workpiece was reduced from approximately 50 μm to less than 2 μm after compensation. The reduction in maximum depth variation was more than 85%, which is a significant improvement in workpiece accuracy. The proposed thermal error model based on BiLSTM deep learning has good prediction accuracy and satisfactory robustness, so it can significantly improve the workpiece accuracy.

Acknowledgements The research was sponsored by the National Natural Science Foundation of Major Special Instruments (Grant No. 51527806) and the National Natural Science Foundation Projects of the People's Republic of China (Grant No. 51975372).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ramesh R, Mannan MA, Poo AN (2000) Error compensation in machine tools—a review Part II: thermal errors. *Int J Mach Tools Manuf* 40(9):1257–1284
2. Mayr J, Jedrzejewski J, Uhlmann E et al (2012) Thermal issues in machine tools. *CIRP Annals-Manuf Technol* 61(2):771–791
3. Li Y, Zhao W, Lan S et al (2015) A review on spindle thermal error compensation in machine tools. *Int J Mach Tools Manuf* 95:20–38
4. Liu Y, Ma YX, Meng QY et al (2018) Improved thermal resistance network model of motorized spindle system considering temperature variation of cooling system. *Adv Manuf* 6:384–400
5. Li SS, Shen Y, He Q (2016) Study of the thermal influence on the dynamic characteristics of the motorized spindle system. *Adv Manuf* 4(4):355–362
6. Liu K, Liu H, Li T et al (2019) Intelligentization of machine tools: comprehensive thermal error compensation of machine-workpiece system. *Int J Adv Manuf Technol* 102(9/12):3865–3877
7. Ye WH, Guo YX, Zhou HF et al (2020) Thermal error regression modeling of the real-time deformation coefficient of the moving shaft of a gantry milling machine. *Adv Manuf* 8:119–132
8. Grama SN, Mathur A, Aralaguppi R et al (2017) Optimization of high speed machine tool spindle to minimize thermal distortion. *Procedia CIRP* 58:457–462
9. Li B, Tian X, Zhang M (2019) Thermal error modeling of machine tool spindle based on the improved algorithm optimized BP neural network. *Int J Adv Manuf Technol* 105(9):1497–1505
10. Ni J (1997) CNC machine accuracy enhancement through real-time error compensation. *J Manuf Sci Eng* 119(4B):717–725
11. Wuest T, Irgens C, Thoben KD (2014) An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *J Intell Manuf* 25(5):1167–1180
12. Xu J, Guo L, Jiang J et al (2019) A deep learning methodology for automatic extraction and discovery of technical intelligence. *Technological Forecasting & Social Change* 146:339–351
13. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Computation* 9(8):1735–1780
14. Wei J, Liao J, Yang Z et al (2020) BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. *Neurocomputing* 383(28):165–173
15. Putz M, Regel J, Wenzel A et al (2019) Thermal errors in milling: comparison of displacements of the machine tool, tool and workpiece. *Procedia CIRP* 82:389–394
16. Kim HS, Eykholt R, Salas JD (1999) Nonlinear dynamics, delay times, and embedding windows. *Physica D: Nonlinear Phenomena* 127(1/2):48–60
17. Li Z, Yang J, Fan K et al (2015) Integrated geometric and thermal error modeling and compensation for vertical machining centers. *Int J Adv Manuf Technol* 76(5/8):1139–1150

Pu-Ling Liu is a PhD candidate at the School of Mechanical Engineering, Shanghai Jiao Tong University, China. Her research interests include machine tool error measurement, error modeling, and error compensation, with a focus on thermal error modeling and compensation.

Zheng-Chun Du is an Associate Professor at the School of Mechanical Engineering, Shanghai Jiao Tong University, China. He received his PhD in Mechanical Engineering from Southeast University. His research interests include error measurement, error modeling and error compensation of machine tools, precision measurement, and processing.

Xiao-Bing Feng is an Assistant Professor at the School of Mechanical Engineering, Shanghai Jiao Tong University, China. He received his PhD in precision manufacturing from the National University of Singapore. He finished his post-doctoral research in manufacturing metrology at the University of Nottingham, UK. His research interests include dimensional, surface, and machine tool metrology, with a focus on in situ metrology.

Hui-Min Li is a PhD candidate at the School of Mechanical Engineering, Shanghai Jiao Tong University, China. Her research interests include machine tool error measurement, error modeling, and error compensation, with a focus on volumetric error modeling and compensation.

Jian-Guo Yang is a professor at the School of Mechanical Engineering, Shanghai Jiao Tong University, China. He received his PhD degree in mechanical manufacturing from Shanghai Jiao Tong University. His research interests include precision machining and testing. He has received a number of awards, including the author of National Excellent Doctoral Dissertation of China and the National Technology Invention Award. He is a vice president of

Chinese Colleges and Universities Society of Machine Tools and has published more than 300 papers.

Ming Deng is a PhD candidate at the School of Mechanical Engineering, Shanghai Jiao Tong University, China. His research interests include machine tool error measurement, error modeling, and error compensation, with a focus on error measurement of five axis machine tools.