Temperature-Sensitive Points Optimization of Spindle on Vertical Machining Center with Improved Fuzzy C-Means Clustering

Hu Shi *, Qiangqiang Qu, Yao Xiao, Qingxin Liu and Tao Tao

School of Mechanical Engineering, Xi’an Jiaotong University, Xi’an 710049, China
* Correspondence: tigershi@xjtu.edu.cn

Abstract: The heat generated by motors and bearings of machine tools has a significant impact on machining accuracy. Error modeling and compensation has proven to be effective ways to reduce thermal errors and improve accuracy. An improved fuzzy c-means (FCM) clustering algorithm is proposed to determine the optimized temperature sensitive points for thermal error modeling of a spindle on the vertical machining center. The sensors are deployed to measure the temperature of different positions of machine tools, and the improved FCM algorithm is used to classify the measured data. Combined with the F-test statistics of multiple linear regression, the optimal temperature points of each group are selected. The improved FCM clustering algorithm significantly reduces the multicollinearity problem among temperature measuring points and avoids them falling into local optimization. The modeling method was verified through experiments on two types of vertical machining centers. The results show that the accuracy of the spindle in Y and Z directions of the machine tools was increased by more than 75%, and the model has good robustness, demonstrating application prospects in the selection of temperature measuring points of the spindle system of vertical machining centers.

Keywords: improved FCM algorithm; optimization of thermal points; multiple linear regression; thermal error modeling of spindle; vertical machine center

1. Introduction

Thermal error has proven to be the most important factor affecting the precision stability of machine tools [1]. Machine tools are a complicated thermal system including internal external heat sources. There are many factors affecting the temperature field of machine tools, such as machining conditions, the use of coolants, and the surrounding environment, and the thermal error shows strong nonlinearity. Previous investigations suggest that the thermal error accounts for 40–70% of the total errors in the machining process [2]. Thermal error compensation can improve the accuracy of machine tools effectively [3]. The accuracy of thermal error modeling will directly affect the effectiveness of compensation in field applications, while the selection of temperature measurement points is the basis of thermal error modeling.

In order to obtain data for error modeling, a large number of temperature sensors are usually arranged on the machine tool, but the temperature variables need to be screened when modeling. If there are too many temperature variables, the modeling accuracy will be seriously decreased due to the collinearity. On the other hand, the problem of poor robustness will be generated during compensation if the key thermal points of the machine tool are not selected as the model variable. Scholars from all over the world have conducted extensive research on the selection of thermal points. Many approaches have been proposed for the selection and optimization of temperature sensor deployment in the past decades [4]. Li et al. [5] used grey system theory to optimize the temperature measuring points of machine tools; Miao et al. [6] obtained the key thermal points by
using the fuzzy clustering method and grey correlation analysis; HAN et al. [7] applied the FCM algorithm to select temperature-sensitive points. Wei et al. [8] used the fuzzy clustering method and correlation analysis to select temperature measurement points of the gantry machine feed system, and used the feature extraction method to obtain the independent variables of the thermal error prediction model, which effectively eliminated the collinearity between the temperature measurement points and determined the optimal temperature measurement points. Ramesh et al. [9] proposed a hybrid support vector machine (SVM) and Bayesian network (BN) model. The BN model was used first to classify the temperature measurement points, and the SVM model was used to predict the errors. Li et al. [10] used fuzzy clustering and grey relational analysis to screen temperature-sensitive points. Then they used the beetle antennae search algorithm (BAS) to optimize the weights and thresholds of the back propagation neural network (BP).

The above methods greatly simplify the selection of temperature measurement points and achieve good modeling results. However, grey correlation only reduces the correlation between temperature variables, and cannot overcome the multicollinearity of temperature variables. Due to the random initialization of the membership matrix, the traditional FCM cluster analysis will fall into the local optimal situation when selecting thermal points. Many researchers have improved the traditional FCM algorithm for the screening of temperature-sensitive points [11]. Hu et al. [12] constructed the hybrid AFCM algorithm by combining the objective functions of FCM with a SA-PSO algorithm, and FCM with a weight-assignment technique, and declared that hybrid algorithm has the significant superiority to overcome the main deficiency caused by the initial centers and outliers. However, the algorithm introduced many adaptive exponents to control both the convexity of the objective function and the ambiguity of the cluster and must be optimized beforehand. Zhao et al. [13] standardized Euclidean distance to improve the distribution of two temperature points in each component. However, the improvement of Euclidean distance cannot change the initial clustering center correspondingly. Liu et al. [14] suggested that the thermal error model can be updated adaptively by supplementing new data, which can effectively improve the prediction accuracy of thermal error at the temperature-point selection level. However, the conflict between the collinearity and the correlation among temperature sensitive points was ignored.

In this paper, an improved FCM clustering algorithm is used to initialize the membership matrix and classify the temperature measurement points. On this basis, the progressive regression forward screening is carried out to obtain the temperature points that have great influence on the thermal error of the machine tool spindle, and the multiple linear regression model is established. In addition to avoiding the multicollinearity between temperature points, the significance of the regression model is improved. It not only simplifies the modeling process, but also improves the accuracy and robustness of the thermal error prediction model.

2. Improved FCM Algorithm

2.1. Traditional FCM Clustering Algorithm

Different from c-means clustering, each sample point in FCM indicates the closeness between the sample point and each cluster through membership degree, which is the most widely used fuzzy separation clustering method [15]. The sum of the membership degree of each sample relative to each cluster center is 1. The result of clustering is to make the similarity within the class as large as possible and the difference between the classes as obvious as possible. Following the principle, the given data

\[ X = (x_1, x_2, \ldots, x_n) \]

is divided into \( c \) categories and the obtained clustering center minimizes the objective function. The function [16] is defined as:

\[
J_m(U, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m d_{ik}^2 = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m \| x_k - v_i \|_A^2
\]

(1)

where \( c \) is the number of clusters, \( v_i \) is the \( i \)th cluster center, \( u_{ik} \) is the membership degree of the \( k \)th data in the \( i \)th cluster. \( m \) is usually taken as 2 in most cases, which defines the
fuzziness of the resulting clusters, \( d_{ik} \) is the Euclidean distance from data \( x_k \) to cluster center \( v_i \).

In order to obtain the minimum value of the above objective function under constraints, the Lagrange multiplier method is used:

\[
F = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m d_{ik}^2 + \lambda \left( \sum_{i=1}^{c} u_{ik} - 1 \right)
\]

The result of the solution is obtained as follows:

\[
v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m}
\]

\[
u_{ik} = \frac{1}{\sum_{i=1}^{c} (\frac{d_{ik}}{d_i^m})^{-1}}
\]

2.2. Initializing the Membership Matrix

The multi-collinearity existing among temperature measuring points can be solved by using the traditional FCM algorithm. However, the iteration result may fall into the local optimal condition due to the random initialization of the membership matrix. WU et al. [17] suggested that the clustering can be formulated as a nonlinear optimization problem with constraints. They proved that modifying the membership matrix before the FCM algorithm can improve the local optimal problem caused by random initialization.

In this research, the membership matrix is initialized first and then the iteration is conducted. Taking the temperature rise stage of machine tool as an example, after the machine tool reaches the heat balance, the data of each temperature point at this state is sorted from high to low, and divided into two categories on average. The membership matrix initialization is shown in Table 1.

**Table 1. Initialization of membership matrix.**

<table>
<thead>
<tr>
<th>Temperature Measurement Points</th>
<th>Membership Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_1 )</td>
<td>( \frac{1}{N} )</td>
</tr>
<tr>
<td>( T_2 )</td>
<td>( \frac{N-1}{N} )</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( T_n )</td>
<td>( \frac{1}{N} )</td>
</tr>
</tbody>
</table>

The membership matrix can be initialized as above when the temperature measurement points of machine tools are divided into two categories. Compared with random initialization, this approach has the advantage of avoiding the improper clustering center. Although the Euclidean distance between the temperature curves of some temperature measurement points is large, it is different from the significance of thermal error. Random selection will cause the loss of model accuracy. The initialization membership matrix ensures that the sum of the membership degrees of each temperature point belonging to various types is 1. On the other hand, because the temperature points are ordered according to temperature value, the Euclidean distance between each temperature point is more uniform than that under random initialization. The above initial membership matrix was substituted into Equation (3) to calculate the initial cluster center, and the obtained cluster center was substituted into Equation (4) to update the membership matrix and iterate repeatedly. The termination condition of the iteration is:

\[
\max \left\{ \left| u_{ik}^{(t+1)} - u_{ik}^{(t)} \right| \right\} < \varepsilon
\]

where \( u_{ik}^{(t+1)} \) represents the membership value of data \( x_k \) in class \( i \) after the iteration of \( t + 1 \), \( \varepsilon \) is the error which is usually taken as \( 10^{-5} \).
The processing steps of the improved FCM clustering algorithm are illustrated as show in Figure 1.

Figure 1. Flow chart of improved FCM clustering algorithm.

3. Key Temperature Points Selection

The temperature data collected in the experiment is used to describe the temperature distribution of the whole machine as much as possible, therefore a lot of temperature sensors are set to collect the data. However, excessive temperature data will cause collinearity problems in the modeling process [18]. In order to improve the accuracy of the model and reduce the amount of calculation, it is very important to optimize the temperature measurement points. This paper uses the above improved FCM algorithm to screen the temperature data, and the process is shown in Figure 2.

Figure 2. The process of FCM screening temperature points was improved.

3.1. F-Test of Multiple Linear Regression

F-test examine that the statistical values follow the F distribution under the original hypothesis, and it is a test for the significance difference of the overall regression equation. For a set of characteristic data \( X = \{x_1, x_2, \cdots, x_n\} \) and \( Y = \{y_1, y_2, \cdots, y_m\} \), the variance of these two sequences is expressed as:

\[
S_X^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{X})^2
\]

(6)
\[ S_Y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (y_i - \bar{Y})^2 \]  \hspace{1cm} (7)

Thus, the F-test statistics can be obtained:

\[ F = \frac{S_X^2}{S_Y^2} \]  \hspace{1cm} (8)

### 3.2. F-Test for Key Thermal Points

F-test statistics indicate the extent to which independent variables explain dependent variables in linear regression. In the thermal error modeling, a single linear regression equation can be established between each temperature measuring point and the thermal error. The unitary linear regression equation is described as:

\[ Y = \beta_0 + \beta_j X_j + \xi, j = 1, 2, \ldots, p \]  \hspace{1cm} (9)

\[ \hat{Y} = \beta_0 + \beta_j X_j, j = 1, 2, \ldots, p \]  \hspace{1cm} (10)

where \( Y \) represents error data and \( \hat{Y} \) represents the calculated thermal error data of the empirical regression equation. \( X_j \) represents \( j \)th temperature point. \( \beta_0 \) and \( \beta_j \) represent the regression coefficient of the linear regression equation of one variable. \( p \) is the number of temperature points in a class. \( \xi \) is the error of fitting.

\[ \bar{y} = \frac{1}{n} \sum_{i}^{n} y_i \]  \hspace{1cm} (11)

\[ F_j = \frac{\sum_{i=1}^{n} (\hat{y}_{ji} - \bar{y})^2 / k}{\sum_{i=1}^{n} (y_i - \hat{y}_{ji})^2 / (n - k - 1)} \]  \hspace{1cm} (12)

where \( \hat{y}_{ji} \) represents the \( i \)th fitting thermal error data of the \( j \)th temperature point, \( \bar{y} \) is the mean value of the thermal error data. \( k \) is the degree of freedom of the linear regression equation. \( F_j \) represents the F-test statistic value of the \( j \)th thermal point.

In summary, after obtaining the temperature data of the machine tool and the thermal error data of the spindle, the improved FCM cluster analysis was carried out first, the temperature measurement points were divided into several categories, and then the F-test statistics were calculated for the progressive regression forward screening of the data. Finally, the key temperature points of the machine tool spindles were selected in each category according to the F-test statistics.

### 4. Experimental Verification on Vertical Machining Centers

The spindle is one of the most precise parts of the vertical machining center, and its precision directly affects the machining accuracy of the vertical machining center. In the thermal characteristic test, it is usually necessary to record the error, temperature, and running state of the machine tool, which is the basis of thermal error modeling and compensation [19]. Theoretically, the feed drive system generates heat in the working process, leading to the contour error of the machined parts [20]. In order to verify the practicability of the modeling method, this paper carried out verification tests of two different machine tools operating in different environments.

The experiment was carried out on the vertical machining center of model \( \mu \)1000/460VF. In the experiment, pt100 temperature sensors were used to measure the temperature data of the machine tool and the atmosphere. The thermally induced error of the machine tool spindle was measured by DT3005 eddy current sensor. A total of 32 temperature sensors were arranged considering the basic structural characteristics of the machine tools and the factors that may affect the thermal error of the spindle. Since the vertical machining
center has a certain thermal symmetry structure design in the X direction, this paper only analyzes the thermal error modeling in the Y direction and the Z direction. The location of the thermal sensors is shown in Figure 3.

![Figure 3](image-url)

**Figure 3.** Location of thermal sensors on the machine tool. (a) a certain thermal symmetry structure design in the Y direction. (b) a certain thermal symmetry structure design in the Z direction.

In addition to the thermal points shown in the Figure 3, there are additional thermal points (30,31) for the coolant inlet and outlet, and thermal points (19,32) for ambient air in workshop. After the improved FCM cluster analysis and F-test step-to-step regression forward screening, 32 thermal points were divided into two categories, and each category was sorted according to F-test statistics. The results are shown in Table 2.

Table 2. The analysis result of optimization algorithm for the first machine tool.

<table>
<thead>
<tr>
<th>Class of Clustering</th>
<th>Thermal Points Sorted by Test Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>F_{15} &gt; F_{20} &gt; F_{9} &gt; F_{24} &gt; F_{22} &gt; F_{21} &gt; F_{23} &gt; F_{14} &gt; F_{10} &gt; F_{13} &gt; F_{12}</td>
</tr>
<tr>
<td>Group 2</td>
<td>F_{19} &gt; F_{27} &gt; F_{26} &gt; F_{12} &gt; F_{6} &gt; F_{8} &gt; F_{5} &gt; F_{7} &gt; F_{10} &gt; F_{4} &gt; F_{2} &gt; F_{11} &gt; F_{17} &gt; F_{3} &gt; F_{1} &gt; F_{16} &gt; F_{26} &gt; F_{29} &gt; F_{31} &gt; F_{30}</td>
</tr>
</tbody>
</table>

Taking T20 and T27 as the key thermal points, the triple linear regression equation was established with the remaining thermal points. By means of calculating the F-test statistics of the regression coefficient, the ranking order from the largest to the smallest is shown as follows:

\[
F_{11} > F_{24} > F_{18} > F_{16} > F_{25} > F_{17} > F_{32} > F_{13} > F_{20} > F_{23} > F_{9} > F_{22} > F_{1} > F_{29} > F_{28} > F_{21} > F_{3} > F_{4} > F_{12} > F_{10} > F_{2} > F_{6} > F_{5} > F_{30} > F_{31} > F_{8} > F_{14} > F_{27} > F_{7} > F_{26}
\]

Improved FCM clustering analysis and F-test analysis are conducted to determine the input variables of the model to be T11 (the upper headstock), T15 (motorized spindle) and T19 (workshop environment). The temperature curve of the temperature sensitive points is plotted in Figure 4. It should be mentioned that ambient temperature is mostly selected as the key temperature point for modeling in some of the literature. However, this experiment was conducted in a constant-temperature plant, and the collected data shows that the ambient temperature did not change much during the experiments. Therefore,
combined with the clustering results, the ambient temperature was not regarded as the key temperature.

\[\begin{align*}
\text{KY} &= -23.3788 + 7.001 \times T_{11} - 4.5815 \times T_{15} - 1.3206 \times T_{19} \\
\text{KZ} &= -62.2603 - 6.5943 \times T_{11} + 5.7578 \times T_{15} - 2.6547 \times T_{19}
\end{align*}\]

where KY and KZ represent the output of error model. These two values are taken as the predicted error to implement compensation. The compensation effect is verified as shown in Figure 5.

**Figure 4.** Temperature data of the selected key points for the first machine tool.

The vertical machining center is thermally symmetrical in the X direction, so the thermal error in the X direction is relatively small. This paper only analyzes the thermal error in the Y and Z directions of the spindle. As can be seen from Figure 4, T11 (the upper headstock), T15 (motorized spindle), and T19 (workshop environment) are selected as the key temperature points for modeling. The temperature rises rapidly in the initial stage of machine operation and quickly reaches thermal equilibrium. There was a significant difference in the temperature between the headstock and the spindle over time, which produced an uneven temperature field and thermal deformation. The experiment was carried out in the temperature-rising stage for 500 min, and the error of the Y direction and Z direction reached 4.45 µm and 15.38 µm, respectively. The multiple linear regression model was established between the temperature data of key temperature points and the axial thermal errors. Therefore, the final thermal error compensation model of the machine tool can be expressed as

\[\begin{align*}
\text{KY} &= -23.3788 + 7.001 \times T_{11} - 4.5815 \times T_{15} - 1.3206 \times T_{19} \\
\text{KZ} &= -62.2603 - 6.5943 \times T_{11} + 5.7578 \times T_{15} - 2.6547 \times T_{19}
\end{align*}\]

**Figure 5.** The fitting results and residual error of the first machine tool.
According to the analysis in Figure 5, after thermal error compensation, the accuracy in the Y direction and Z direction increased by 75% and 93%, respectively, which verifies the effectiveness of the modeling theory.

In order to verify the universality of the modeling method, experiments were carried out again on a larger vertical machining center. These two machine tools have a similar structural type, and the temperature sensor layout is similar. The experiments were carried out with the machine tool experiencing the rise and fall in temperature, lasting a total of 1000 min. After improved FCM cluster analysis and F-test step-to-step regression forward screening, 27 thermal points were divided into two categories, and each category was sorted according to F-test statistics. The results are shown in Table 3.

Table 3. The analysis result of optimization algorithm for the second machine tool.

<table>
<thead>
<tr>
<th>Class of Clustering</th>
<th>Thermal Points Sort by Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>F_{16} &gt; F_{3} &gt; F_{27} &gt; F_{19} &gt; F_{18} &gt; F_{20} &gt; F_{25} &gt; F_{17} &gt; F_{1} &gt; F_{2} &gt; F_{12}</td>
</tr>
<tr>
<td>Group 2</td>
<td>F_{4} &gt; F_{14} &gt; F_{15} &gt; F_{22} &gt; F_{13} &gt; F_{11} &gt; F_{21} &gt; F_{7} &gt; F_{26} &gt; F_{6} &gt; F_{24} &gt; F_{23} &gt; F_{5} &gt; F_{9} &gt; F_{10} &gt; F_{8}</td>
</tr>
</tbody>
</table>

Taking T4 and T16 as the key thermal points, the triple linear regression equation was established with the remaining thermal points. By means of calculating the F-test statistics of the regression coefficient, the ranking order from largest to smallest is as follows:

$$F_{19} > F_{27} > F_{15} > F_{20} > F_{25} > F_{17} > F_{2} > F_{5} > F_{9} > F_{12} > F_{14} > F_{13} > F_{26} > F_{22} > F_{18} > F_{7} > F_{3} > F_{11} > F_{23} > F_{24} > F_{1} > F_{10} > F_{6} > F_{21} > F_{8}$$

Improved FCM clustering analysis and F-test analysis were conducted to determine the input variables of the model to be T4 (column on the side of the electric cabinet), T16 (workshop environment), and T19 (front bearing seat). The temperature curve of the machine tool can be expressed as

$$KY = 178.7066 - 6.1153 \times T_{4} - 2.3156 \times T_{16} + 1.8812 \times T_{19}$$  
(15)

$$KZ = -668.3694 - 14.4577 \times T_{4} + 0.8808 \times T_{16} + 38.5027 \times T_{19}$$  
(16)

Figure 6. Temperature data of the selected key points for the second machine tool.
It can be seen that the modeling method has good generalization performance on the vertical machining center with the same structure type. The error prediction result is shown in Figure 7.

Figure 7. The fitting results and residual error for the second machine tool.

According to the analysis in Figure 7, the accuracy in the Y direction and Z direction increased by 85.3% and 92.2%, respectively, which verifies the effectiveness of the modeling and prediction. It can be seen from the above analysis that the method has achieved good error-prediction performance on two machine tools of a similar type.

5. Conclusions

This paper mainly proposed a thermal error modeling method based on improved FCM cluster analysis and F-test statistics, which greatly reduced the temperature measurement data. The following conclusions can be drawn:

(1) Using the initial membership matrix to improve FCM clustering can effectively screen out the temperature-sensitive points and avoid collinearity between temperature points. The key temperature variables can be screened for multiple linear regression using the F-test.

(2) Multiple linear regression was conducted between the selected thermal points data and thermal error data, and the obtained model has high accuracy and good robustness. It indicates that the temperature variables selected with the proposed method formulate the thermal characteristics exactly.

(3) The model can be a duplicate for the machine tool with a similar configuration and different operating conditions to predict the thermal error successfully, and the generalizing ability of the modeling method was verified.

Author Contributions: Conceptualization, H.S. and Q.Q.; methodology, Q.Q.; software, Q.L.; validation, Y.X.; formal analysis, Q.L. and T.T.; investigation, Y.X.; data curation, Q.Q.; writing—original draft preparation, Q.Q.; writing—review and editing, H.S.; supervision, H.S.; project administration, T.T.; funding acquisition, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Program of China, grant number 2018YFB1703202.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.
References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.