

DOI: 10.13382/j.jemi.2017.12.019

基于核主成分分析的热误差模型自变量优化^{*}

魏 弦

(攀枝花学院 攀枝花 617000)

摘 要:为了改善主成分分析对温度场的非线性数据提取能力不足的问题,提出一种核主成分温测点优化法。引入非线性映射函数,将输入的温度数据向特征空间映射,选择高斯径向基为核函数,将特征空间的内积运算变换为输入空间的核函数运算,求出特征值和核特征向量,建立综合自变量。在一台数控加工中心上进行试验,将提出方法建立的预测模型和主成分分析获得的预测模型进行比较,均方根误差降低约36%,最大残差降低29%,结果表明,采用核主成分法建模,能更好提取温度数据特征,可以使试验机床进给系统的热误差预测能力显著提升。

关键词:数控机床;热误差;自变量优化;核主成分;误差补偿

中图分类号: TG502.15; TH161 **文献标识码:** A **国家标准学科分类代码:** 460.3599

Independence variable optimization of thermal error model based on KPCA

Wei Xian

(Panzhihua University, Panzhihua 617000, China)

Abstract:To address the issue that principle component analysis (PCA) shows a poor ability to acquire the characteristic of nonlinearity data, a kernel principle component analysis (KPCA) temperature point optimization method is proposed. Firstly, nonlinearity mapping function is introduced to map the input temperature data into the characteristic space, and a Gaussian radial basis is selected to be a kernel function. Secondly, inner product operation in characteristic space is transformed into kernel function operation in input space, eigenvalues and kernel eigenvectors are found. Finally, a comprehensive independent variable is formed. According to an experiment conducted on a CNC machine center, and compared with the PCA model, RMSE and Maximum residual error reduces by 36% and 29%, respectively. KPCA can preferably acquire the characteristic of temperature data, and the prediction ability of KPCA model has an obvious improvement.

Keywords: CNC machine tool; thermal error; independence variable optimization; KPCA; error compensation

0 引 言

在机床的精密加工过程中,热误差占整个机床制造误差的40%~70%。如何克服热误差对机床加工精度的影响是亟待解决的问题^[1]。

目前,降低热误差影响的最为经济的手段是误差补偿^[2]。误差补偿通过热关键点的温度数据创建误差预测模型,再利用预测值修改数控程序中的坐标偏置,实现热

误差的修正。国内外常见的建模方法为神经网络^[3-4]、支持向量机^[5-6]、多元线性回归^[7-8]、分段拟合^[9]及误差解耦综合建模等^[10-11]。然而,温度测点的优化和热误差模型的建立同样关键,因为温度测点的选择及优化是影响模型鲁棒性和预测精度的核心因素。国内外学者对测点优化也进行了广泛的研究,主要方法有模糊聚类^[12-13],但此法选出的热关键点容易出现多元共线性,从而影响预测模型的鲁棒性^[8];主成分分析法(principle component analysis, PCA)^[7, 14]的提出,有效地解决了多元共线性的

收稿日期:2017-05 Received Date: 2017-05
^{*} 基金项目:国家自然科学基金(51375382)、四川省科技厅科技支撑计划(2016GZ0205)、四川省教育厅重点项目(16ZA0415)、攀枝花学院博士基金(BKQJ2017007)资助项目

影响。主成分分析法应用于线性问题时效果很好,但机床温度场属于时变、非线性系统,主成分分析应用于非线性时有一定缺陷,如:导致原始信息的保留量大幅降低,运行效率低等问题^[15]。因此,主成分优化温度测点会造成热关键点信息丢失,从而影响模型的预测精度。为了更高效地保留温度信息,本文提出了核主成分分析(kernel principle component analysis, KPCA)测点优化法,在温度测点数目一定的条件下,只需选取适宜的核函数和核参数,便可以使 KPCA 比 PCA 降维后保留的信息更丰富,即 KPCA 的优化效果好。在一台数控加工中心上进行实验,结果表明,KPCA 模型比 PCA 模型能获得更高的预测精度。

1 KPCA

KPCA 是一种新的特征提取方法,它通过核函数将传统的 PCA 推广到非线性应用^[16]。相较于 PCA, KPCA 可以高效提取数据的非线性特征,所以其在模式识别、图像视觉处理和文本分类等众多领域取得良好应用效果。

1.1 基本原理

针对输入空间中 M 个温度样本 $\mathbf{x}_k (k = 1, 2, \dots, M)$, $\mathbf{x}_k \in R^N$, 令 $\sum_{k=1}^M \mathbf{x}_k = 0$, C 为协方差, 计算如下:

$$C = 1/C \sum_{j=1}^M \mathbf{x}_j \mathbf{x}_j^T \quad (1)$$

传统的主成分分析法,通过求解 $\lambda \mathbf{v} = C\mathbf{v}$ 方程, λ 是 C 的特征值,对应的特征向量为 $\mathbf{v}$,从而得到较大贡献率的特征值和其特征向量。

如将非线性映射函数 φ 引入,则可将 $x_1, x_2, \dots, x_M$ 这些输入空间样本点转换到特征空间中,表示为 $\varphi(x_1), \varphi(x_2), \dots, \varphi(x_M)$, 令:

$$\sum_{k=1}^M \varphi(\mathbf{x}_k) = 0 \quad (2)$$

可得到特征空间中的协方差矩阵:

$$\bar{C} = \frac{1}{M} \sum_{j=1}^M \varphi(\mathbf{x}_j) \varphi(\mathbf{x}_j)^T \quad (3)$$

求解方程为:

$$\lambda \mathbf{v} = \bar{C}\mathbf{v}, \mathbf{v} \in F \setminus \{0\} \quad (4)$$

因而得到:

$$\lambda(\varphi(\mathbf{x}_k) \cdot \mathbf{v}) = \varphi(\mathbf{x}_k) \cdot \bar{C}\mathbf{v} \quad (5)$$

线性表示特征向量 $\mathbf{v}$ 为:

$$\mathbf{v} = \sum_{j=1}^M \alpha_j \varphi(\mathbf{x}_j) \quad (6)$$

由式(3)、(5)、(6)可得:

$$\lambda \sum_{i=1}^M \alpha_i (\varphi(\mathbf{x}_k) \cdot \varphi(\mathbf{x}_i)) =$$

$$\frac{1}{M} \sum_{i=1}^M \alpha_i (\varphi(\mathbf{x}_k) \cdot \sum_{i=1}^M \varphi(\mathbf{x}_i)) (\varphi(\mathbf{x}_j) \cdot \varphi(\mathbf{x}_i))$$

$$k = 1, 2, \dots, M \quad (7)$$

把 $M \times M$ 的矩阵定义为 K :

$$K_{i,j} = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) \quad (8)$$

可以将式(7)转化为:

$$M\lambda\alpha = K\alpha \quad (9)$$

通过式(9)可求得映射空间上的特征向量和特征值。在 F 空间向量 $\mathbf{V}^k$ 上样本的投影为:

$$\mathbf{V}^k \cdot \varphi(\mathbf{x}) = \sum_{i=1}^M \alpha_i^k (\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x})) = \sum_{i=1}^M \alpha_i^k K(\mathbf{x}_i, \mathbf{x}) \quad (10)$$

当特征空间上的数据无法实现中心化条件,需修正矩阵,将式(9)中的 K 用式(11)中的 $\bar{K}$ 代替。

$$\bar{K}_{i,j} = K_{i,j} - \frac{1}{M} \sum_{m=1}^M l_{im} k_{mj} - \frac{1}{M} \sum_{n=1}^M l_{in} k_{nj} + \frac{1}{M^2} \sum_{m,n=1}^M l_{im} k_{mn} l_{nj} \quad (11)$$

式中:所有 $i, j, l_{i,j} = 1$ 。

1.2 核主成分分析实现步骤

1) 构造样本矩阵如下:

$$X = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (12)$$

式中: x_{ij} 表示第 i 组温度数据中的第 j 个变量的值。

2) 模糊聚类和相关性分析。

3) 选择核函数,选择高斯径向基(RBF)为本文核函数:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right) \quad (13)$$

4) 求 KL 的特征值 $\lambda_1, \lambda_2, \dots, \lambda_n$ 和特征向量 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 。

5) 确定 t , 使温度特征信息的覆盖率达到 85% 以上。确定方法为求特征信息的累积覆盖率 $B_1, B_2, \dots, B_n$, 当 $B_i \geq 85\%$, 则选取 t 个主成分 $\alpha_1, \alpha_2, \dots, \alpha_n$ 。

2 温度场测量实验

以一台五轴龙门式镗铣加工中心为测试平台,如图1所示。定位误差采用激光干涉仪(分辨能力为 $0.001 \mu\text{m}$, 测量最大距离为 80 m)测量,采用 HIOKI8423 数据采集仪和安利计器磁吸式温度传感器(E-type MG-24E-GW1-ANP, 测量精度 $\pm 2.5^\circ\text{C}$) 测量温度数据,测温点布局如图2所示。机床空载运行,设定进退刀速度为 10 m/min 。

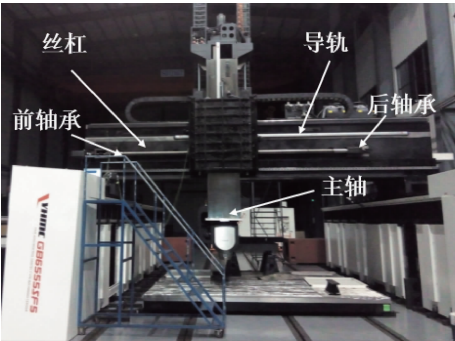


图 1 测试机床示意图

Fig.1 Schematic diagram of test machine tool

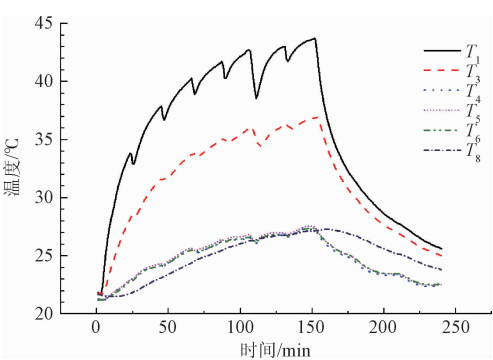


图 4 进给系统关键点温度曲线

Fig.4 Temperature curves of key points of feed system

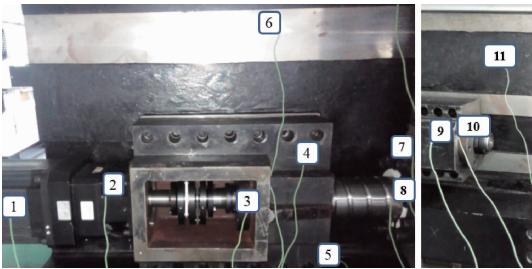


图 2 测试系统测温点布局

Fig.2 The layout of temperature measurement point on the test system

2.1 温度数据测量

由于龙门机床的体积庞大,获得全面的温度数据需要大量传感器,但考虑到成本因素,必须对传感器有针对性的布局。因此,首先借助红外成像仪拍摄进给系统的温度场分布,所得成像图如图 3 所示,由此确定测量部件为进给系统的前、后轴承、丝杠、导轨和螺母。结合传感器布置方法^[8]最终将温度传感器布置如图 2 所示。采用上述温度测量系统数据,绘得温度变化曲线如图 4 所示。

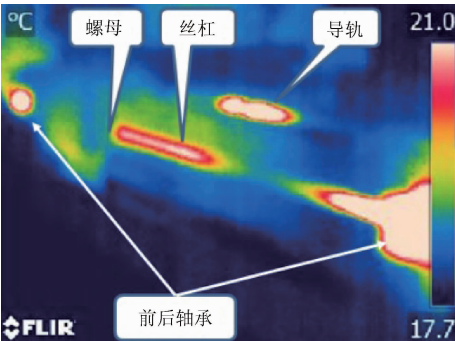


图 3 测试机床红外成像图

Fig.3 The infrared imaging of experiment machine tool

2.2 热变形数据采集

进给系统误差利用图 5 所示的激光干涉仪进行测量。考虑到激光干涉仪使用规则,均匀设置了 11 个测量点,测量点分布如图 6 所示。为了预防丝杠反向间隙对测量误差的影响,于左右末端设置了超程,测量运动的方向为正反两个方向。

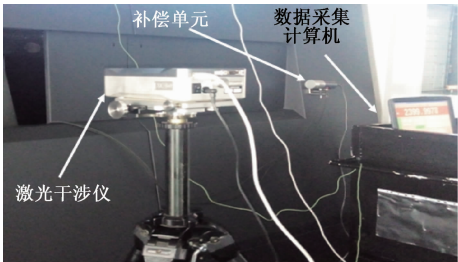


图 5 进给系统误差测量

Fig.5 Measurement of feed system position error

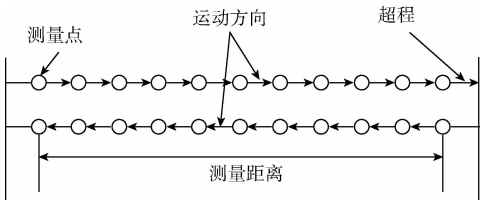


图 6 误差测量装置测点分布

Fig.6 The measurement point layout of laser interferometer

3 KPCA 实例

3.1 聚类分析

根据参考文献^[8],可得模糊聚类分组为 2,分组效果如表 1 所示。

表 1 分组结果
Table 1 The grouping result

组 1	$T_1, T_2, T_3, T_7, T_8, T_9, T_{10}$
组 2	$T_4, T_5, T_6, T_{11}, T_{12}$

3.2 特征变量计算

根据式(9), 求出第 1 组和第 2 组的特征值, 如表 2、3 所示。

表 2 组 1 特征值 λ 及其覆盖率

Table 2 Eigenvalues λ and its coverage rate of group 1

特征	λ	覆盖率/%	累计覆盖率/%
1	3.741	76.6	76.6
2	0.731	15	91.6
3	0.102	2.1	93.7
4	0.098	2.0	95.7
5	0.082	1.8	97.5
6	0.064	1.5	99
7	0.045	1	100

表 3 组 2 特征值 λ 及其覆盖率

Table 3 Eigenvalues λ and its coverage rate of group 2

特征	λ	覆盖率/%	累计覆盖率/%
1	3.779	87.5	87.5
2	0.296	6.9	94.4
3	0.103	2.4	96.8
4	0.099	2.3	99.1
5	0.041	0.9	100

从表 2 和 3 中可以发现, 组 1 中特征 1 和 2 的累积覆盖率已经超过 85%; 组 2 中特征 1 的累积覆盖率已经超过 85%。因此, 组 1 中 2 个综合特征即可表达其主要信息, 组 2 中 1 个综合特征即可表达其主要信息。

3.3 KPCA 温度综合表达式

计算组 1 中 2 个综合特征的特征向量 K_1 、 K_2 和组 2 中 1 个综合特征的特征向量 K_3 , 得:

$$\begin{aligned} K_1 &= [0.297 \quad 0.378 \quad 0.019 \quad 0.659 \quad 0.723 \\ &\quad 0.15 \quad 0.381] \\ K_2 &= [0.512 \quad 0.023 \quad 0.153 \quad 0.479 \quad 0.414 \\ &\quad 0.325 \quad 0.12] \\ K_3 &= [0.431 \quad 0.105 \quad 0.221 \quad 0.137 \quad 0.086] \end{aligned}$$

由上述特征向量可得核主成分分析温度变量综合表达式为:

$$\begin{aligned} T_a &= 0.297T_1 + 0.378T_2 + 0.019T_3 + 0.659T_7 + \\ &\quad 0.723T_8 + 0.15T_9 + 0.381T_{10} \quad (14) \\ T_b &= 0.512T_1 + 0.023T_2 + 0.153T_3 + 0.479T_7 + \\ &\quad 0.414T_8 + 0.325T_9 + 0.12T_{10} \quad (15) \end{aligned}$$

$$\begin{aligned} T_c &= 0.431T_4 + 0.105T_5 + 0.221T_6 + 0.137T_{11} + \\ &\quad 0.086T_{12} \quad (16) \end{aligned}$$

4 KPCA 模型预测效果分析

为了对比基于 PCA 和 KPCA 两种测点优化方法建立的预测模型的预测效果, 根据滚珠丝杠系统误差建模理论^[11]可得:

$$\begin{aligned} E_{KPCA} &= (1.575 + 1.32\delta_y - 2.93\delta_y^2 + \\ &\quad 1.83 \times 10^{-6}\delta_y^3 - 2.05 \times 10^{-9}\delta_y^4) + (-2.1 + \\ &\quad 0.12T_a + 0.058T_b + 0.021T_c) \times (\delta_y - \delta_0) \quad (17) \end{aligned}$$

$$\begin{aligned} E_{PCA} &= (1.783 + 1.521\delta_y - 3.32\delta_y^2 + \\ &\quad 1.341 \times 10^{-6}\delta_y^3 - 1.23 \times 10^{-9}\delta_y^4) + (-1.97 + \\ &\quad 0.87T_a + 0.106T_e + 0.015T_f) \times (\delta_y - \delta_0) \quad (18) \end{aligned}$$

式(17)和(18)分别为 KPCA 自变量优化法和 PCA 自变量优化法建立的热误差预测模型, 式(17)中 T_a 、 T_b 和 T_c 分别由式(14)~(16)得到。其中 T_e 、 T_f 和 T_g 为主成分分析后得到的温度变量综合表达式, δ_y 为主轴所在位置的 y 轴坐标, δ_0 为机床的 y 轴坐标原点。

KPCA 和 PCA 方法自变量优化得到的预测模型的预测效果如图 7、8 所示, 从图 7、8 可以发现, KPCA 方法建立的模型预测曲线更接近实际曲线; 从图 9、10 所示的模型预测值的残差也可以看出, KPCA 法建立的模型预测效果优于 PCA 模型。

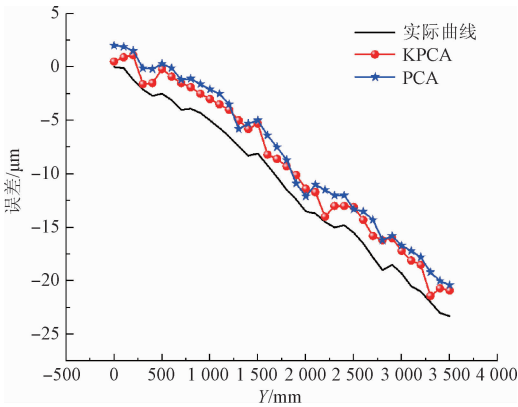


图 7 进给系统运行 30 min Y 正方向预测效果

Fig. 7 The prediction result of Y positive direction(30 min)

表 4、5 分别列出了两种试验情况下, KPCA 和 PCA 模型的部分指标对比, 其中 (variance inflation factor, VIF) 为方程膨胀因子, 它是衡量多元共线性影响的指标, 判断经验是 $0 < VIF < 10$, 则不存在多元共线性, 从表中数据可见, 两种方法都可以很好地消除多元共线性的影响。从表 4、5 可见, KPCA 模型相较于 PCA 模型, RMSE 降低了约 36%, 最大残差降低了约 29%。

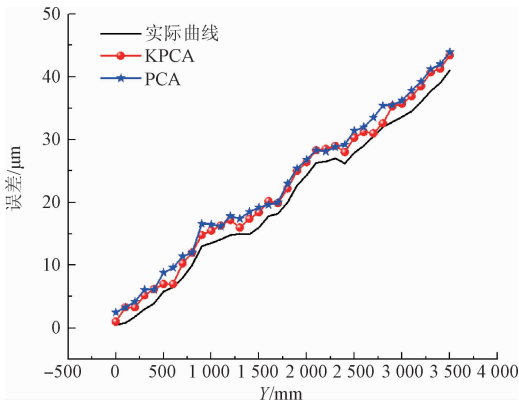


图 8 进给系统运行 120 min Y 负方向预测效果
Fig.8 The prediction result of Y negative direction(120 min)

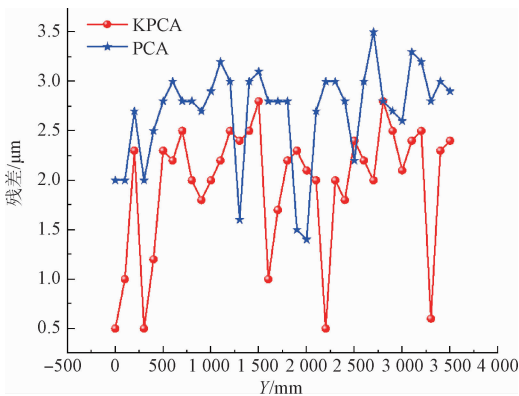


图 9 进给系统运行 30 min Y 正方向预测残差对比
Fig.9 The comparison ofresidual errors of Y positive directionprediction(30 min)

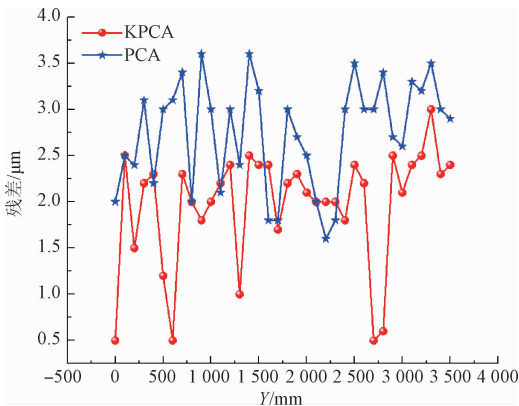


图 10 进给系统运行 120 min Y 负方向预测残差对比
Fig.10 The comparison ofresidual errors of Y negative direction prediction(120 min)

表 4 30 min 时正向,模型预测结果对比
Table 4 The prediction result of model in positive direction(30 min)

模型	$RMSE/\mu m$	Max Residual/ μm	VIF
KPCA	1.89	2.5	5.76
PCA	2.93	3.5	6.83

表 5 120 min 时负向,模型预测结果对比
Table 5 The prediction result of model in negative direction(120 min)

模型	$RMSE/\mu m$	Max Residual/ μm	VIF
KPCA	1.96	2.5	5.76
PCA	2.99	3.6	6.83

5 结 论

本文在一台数控加工中心进给系统上进行了热误差模型自变量的优化试验,基于 KPCA 自变量优化方法,结合进给系统误差建模理论建立了热误差模型,预测试验结果表明,KPCA 模型相对于 PCA 模型, $RMSE$ 降低约 36%,最大残差降低 29%。证明了 KPCA 自变量优化法能有效解决温度场的非线性对预测模型精度的影响。

提出的核主成分分析自变量优化法有效地克服了多元共线性对误差模型鲁棒性的影响,同时,提高了自变量的信息携带量;此方法选择高斯径向基作为核函数,将数据应用范围由线性扩展至非线性,可以取得很好的优化效果,该方法有望在其他机床的测点优化中进一步应用。

参考文献

- [1] RAMESH R. Error compensation in machine tools: A review Part II; Thermal errors[J]. International Journal of Machine Tools and Manufacture, 2000, 40(9): 1257-1284.
- [2] 王海同,李铁民,王立平,等. 机床热误差建模研究综述[J]. 机械工程学报, 2015,51(9): 119-128.
WANG H T, LI T M, WANG L P, et al. Review on thermal error modeling of machine tools[J]. Journal of Mechanical Engineering,2015,51(9): 119-128.
- [3] CHEN J S. A study of thermally induced machine tool errors in real cutting conditions[J]. International Journal of Machine Tools and Manufacture, 1996, 36 (12): 1401-1411.
- [4] 张毅,杨建国. 基于灰色理论预处理的神经网络机床热误差建模[J]. 机械工程学报, 2011, 47 (7): 134-139.
ZHANG Y, YANG J G. Modeling for machine tool thermal error based on grey model preprocessing neural network[J]. Journal of Mechanical Engineering, 2011, 47(7): 134-139.
- [5] 林伟青,傅建中,许亚洲,等. 基于 LS-SVM 与遗传算法的数控机床热误差辨识温度传感器优化策略[J]. 光学精密工程, 2008, 16(9): 1682-1687.

- LIN W Q, FU J ZH, XU Y ZH, et al. Optimal sensor placement for thermal error identification of NC machine tool based on LS-SVM and genetic algorithm[J]. Optics and Precision Engineering, 2008, 16(9): 1682-1687.
- [6] RAMESH R, MANNAN M A, POO A N. Support vector machines model for classification of thermal error in machine tools[J]. International Journal of Advanced Manufacturing Technology, 2002, 20(2): 114-120.
- [7] MIAO E, LIU Y, LIU H, et al. Study on the effects of changes in temperature-sensitive points on thermal error compensation model for CNC machine tool[J]. International Journal of Machine Tools and Manufacture, 2015, 97(10): 50-59.
- [8] 魏弦,高峰,李艳,等. 龙门机床进给系统热误差模型关键点优化[J]. 仪器仪表学报, 2016, 37(6): 1340-1346.
- WEI X, GAO F, LI Y, et al. Optimization of thermal critical point on gantry machine tool feeding system[J]. Chinese Journal of Scientific Instrument, 2016, 37(6): 1340-1346.
- [9] 张成新,高峰,李艳,等. 基于分段拟合的机床大尺寸工作台热误差补偿模型[J]. 机械工程学报, 2015, 51(3): 146-152.
- ZHANG CH X, GAO F, LI Y, et al. Model of thermal error compensation of large size worktable for machine tools based on piecewise fitting[J]. Journal of Mechanical Engineering, 2015, 51(3): 146-152.
- [10] 要小鹏,殷国富,李光明. 数控机床进给轴综合误差解耦建模与补偿研究[J]. 机械工程学报, 2016, 52(1): 1-9.
- YAO X P, YIN G F, LI G M, et al. Positioning error of feed axis decouple-separating modeling and compensating research for CNC machine tools[J]. Journal of Mechanical Engineering, 2016, 52(1): 1-9.
- [11] 魏弦,吴继忠,王允威. 数控机床进给系统误差综合补偿[J]. 现代制造工程, 2016(10): 1-5.
- WEI X, WU J ZH, WANG Y W. Comprehensive error compensation of CNC feed system[J]. Modern Manufacturing Engineering, 2016(10): 1-5.
- [12] HAN J, WANG L, CHENG N, et al. Thermal error modeling of machine tool based on fuzzy c-means cluster analysis and minimal-resource allocating networks[J]. The International Journal of Advanced Manufacturing Technology, 2012, 60(5-8): 463-472.
- [13] ABDULSHAHEED A M, LONGSTAFF A P, FLETCHER S. The application of ANFIS prediction models for thermal error compensation on CNC machine tools[J]. Applied Soft Computing, 2015, 27(7): 158-168.
- [14] 杨漪,姚晓栋,杨建国,等. 基于主成分分析与BP神经网络相结合的机床主轴热漂移误差建模[J]. 上海交通大学学报, 2013, 47(5): 750-753.
- YANG Y, YAO X D, YANG J G, et al. Thermo-drifting error modeling of spindle based on combination of principal component analysis and bp neural network[J]. Journal of Xi'an Jiaotong University, 2013, 47(5): 750-753.
- [15] 梁胜杰,张志华,崔立林. 主成分分析法与核主成分分析法在机械噪声数据降维中的应用比较[J]. 中国机械工程, 2011, 22(1): 80-83.
- LIANG SH J, ZHANG ZH H, CUI L L. Comparison between PCA and KPCA method in dimensional reduction of mechanical noise data[J]. China Mechanical Engineering, 2011, 22(1): 80-83.
- [16] 高绪伟. 核PCA特征提取方法及其应用研究[D]. 南京:南京航空航天大学, 2009.
- GAO X W. Feature extraction method based on KPCA and its application[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2009.

作者简介



魏弦, 2008年于电子科技大学获得硕士学位, 现为西安理工大学博士研究生, 主要研究方向为数控机床的热误差补偿。

E-mail: 110801844@qq.com

Wei Xian received M. Sc. from University of Electronic Science and Technology of China in 2008. He is now a Ph. D. candidate in Xi'an University of Technology. His main research field is thermal error compensation of CNC machine tool.