

基于热态信息挖掘的机床主轴系统热误差建模方法研究^①

孙惠娟^{②*} 林海峰^{**} 向胜华^{**} 殷国富^{*} 方 辉^{③*}

(^{*} 四川大学制造科学与工程学院 成都 610065)

(^{**} 成都飞机工业(集团)有限责任公司 成都 610091)

摘 要 针对数控机床热误差建模问题,通过对精密卧式加工中心主轴系统热变形数据特征进行分析,提出了一种利用均值理论进行数据分类、利用粗糙集理论进行数据挖掘和利用线性回归分析构建误差模型的机床热误差建模方法。该方法首先进行机床主轴的热误差实验,利用传感器同时检测机床主轴所选各检测点的温升和主轴轴端的热变形,然后利用均值理论和粗糙集理论对获得的数据进行分类和数据挖掘精简,最后通过线性回归分析进行热误差建模。为了评价所建模型的有效性,将所建模型与 BP 神经网络模型进行了对比,结果表明这种基于热态信息挖掘的热误差建模方法在机床主轴热误差预测方面具有较高的准确性和实用价值。

关键词 数控(NC)机床,热误差,数据挖掘,建模,主轴

0 引言

高精度数控机床的热变形误差约占其总误差的 60%~70%,已成为其最大的误差源^[1],因而减小热误差的研究一直受到重视。目前减小机床误差主要用两种方法^[2],一是误差防止法,二是误差补偿法,其中误差补偿法能够降低仪器和设备的制造成本,因而应用比较广泛。误差补偿技术的实施一般分为误差测量、误差建模及误差补偿^[3]三个步骤,由于误差补偿系统的性能主要依赖于所使用的误差模型,因此误差建模成为误差补偿过程中最关键的因素之一。国内外研究人员在热误差建模方面做了大量研究,最常用的建模方法是实验建模法,即通过实验分析来建立机床温度场和热变形之间关系的热误差预测模型,例如回归分析、神经网络^[4]、灰色系统理论^[5]、模糊逻辑^[6]和自适应模糊推理系统^[7]等方法。实验建模法在实际应用中已经表现出良好的预测精度,如文献[8]基于对机床结构链的刚体运动学分析建立了热误差模型,实验证明所建模型可预测出 60% 的机床热误差。文献[9]建立的数控车削中心热误差 RBF 神经网络模型,其预测残差带宽度

为 6.92 μm 。然而,以上热误差建模方法存在以下不足:对实验数据的处理缺乏理论依据;建模方法太复杂,模型完全依赖于建模数据;适应性不强,对于一些比较复杂的加工条件很难建立实时、准确的热误差模型;需要大量的建模数据,辨识时间长;对温度传感器的安装位置很敏感。因而,热误差建模,需寻求新的方法。考虑到近年来数据挖掘^[10]技术在商业、服务、管理等领域的应用已经取得了显著的成效,并正在逐步向工业领域扩展,如文献[11]对带钢在冷轧过程中的辊系径向变形值进行数据挖掘,并在某四辊轧机轧制数据上进行应用,取得了比较清晰的变形规律和完整的变形过程,文献[12]提出了基于数据挖掘技术的通用物料清单重构过程,重构后的通用物料清单的适应性和配置能力大大提高,本研究将数据挖掘技术引入到机床主轴系统的热误差建模中,以一台精密卧式加工中心为研究对象,综合运用数据挖掘技术的均值理论、粗糙集理论、回归分析,建立了一个简单、准确、适应性强的机床主轴系统热误差模型,并通过实验进行对比分析,验证了所建模型能够有效克服目前热误差建模方法存在的弊端。

① 国家自然科学基金(51175356),国家科技重大专项(2010ZX04015-011)和四川省科技支撑计划(2011GZ0075)资助项目。

② 女,1983 年生,博士生;研究方向:数控机床精度稳定性技术;E-mail:scdx2013@163.com

③ 通讯作者,E-mail:jfh@scu.edu.cn

(收稿日期:2011-08-22)

1 基本理论

数据挖掘技术是 20 世纪 90 年代初期新崛起的一个研究领域,数据挖掘是通过综合运用统计学、粗糙集、模糊数学、机器学习和专家系统等多种学习手段和方法,从大量不完全的、有噪声的、模糊的随机数据中提取隐含的、事先不知道的有潜在用途的信息和知识的过程。

对于机床热变形实验过程中所采集的大量温度与热变形数据,存在不确定、模糊、有噪声、随机性等问题,如检测得到的数据可能存在一些偏差,单个热变形变量与多个温度变量的对应导致温度变量耦合现象,各检测点温度与热变形之间的关系及变化规律不确定等。如果直接用这些数据建模会导致严重的偏差并降低模型的鲁棒性。本文在进行热误差建模时采用了三种数据挖掘分析方法,即均值理论法、粗糙集理论法和回归理论法,其中均值理论法主要用于对所测得的温升和热变形数据进行分类预处理,在此基础上利用粗糙集理论对获得的信息进一步约简得到有用的知识元素,最后,用线性回归分析来构建热误差模型并进行回归建模。

1.1 均值理论

K-means 聚类算法是 McQueen 在 1967 年提出的,它是一种早期的聚类算法,主要用于解决分组类的问题。K-means 算法是一种很典型的基于距离的聚类算法,它采用距离作为相似性的评价指标,认为两个对象的距离越近,其相似度就越大。因此把距离较近、相似度较大的对象归为同一簇。当聚类是密集的,且类与类之间区别明显时,效果较好,对于处理大数据集,该算法具有较高的可伸缩性和高效性。其计算步骤如下:

步骤 1 首先从 n 个数据对象中任意选择 K 个对象作为初始聚类中心。

步骤 2 对于剩下的对象,根据它们与这些聚类中心的相似度(即距离),分别将它们分配给与其距离最近的聚类。

步骤 3 计算所获得的每个新聚类中心(即该聚类中所有对象的均值)。

步骤 4 重复步骤 3,直到满足停止条件为止。

1.2 粗糙集理论

粗糙集理论(rough set theory, RST)作为一种数据分析处理理论,由波兰科学家 Pawlak^[13] 在 1982 年

创立。粗糙集理论是一种处理模糊与不确定性问题的数学方法,其主要思想是基于不可分辨关系,即每个对象与一些信息相联系,且每个对象仅能用获得的信息表示,因此具有相同或相似信息的对象不能被识别,而具有不同或差别较大信息的对象可以被识别。与粗糙集理论相关的一些知识,如信息系统,下近似、上近似、负区及边界区,核与属性约简,区分矩阵,决策表等概念与算法详见文献[14]。本文在对数据进行预处理后,选用粗糙集理论来推导两种不同类型的变量(温度和热变形)之间的关系。

1.3 回归分析理论

回归分析(regression analysis, RA)是一种确定两种或两种以上变量间相互依赖的定量关系的统计分析方法。回归分析的具体步骤如下:

步骤 1 确定自变量和因变量。

步骤 2 建立回归预测模型。

步骤 3 进行相关分析。

步骤 4 检验回归预测模型。

本文在对实验数据进行约简和规则推理的基础上,得到能够反映机床温度和热变形之间关系的所有可能的决定规则集,再应用回归分析理论来对之前获得的属性组合进行相关性分析,以得到最优的热误差模型。

1.4 数据挖掘总体技术框架及挖掘过程

图 1 和图 2 分别为利用数据挖掘技术对机床热态信息进行挖掘的总体技术框架和主要过程。由图 1 和图 2 可知,热态信息挖掘分为三个模块,即聚类分析模块、粗糙集约简模块及回归分析模块。聚类分析模块主要对实验中采集到的大量原始数据进行预处理;粗糙集约简模块主要进行进一步的数据约简和知识推理;回归分析模块主要对得到的决策规则集进行相关分析和最优模型选择。

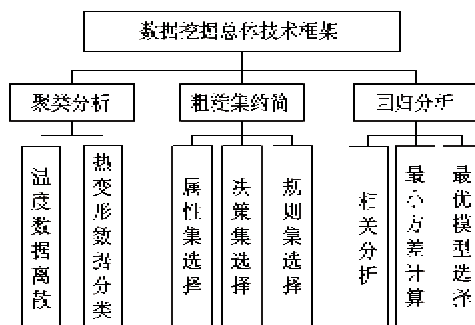


图 1 数据挖掘总体技术框架

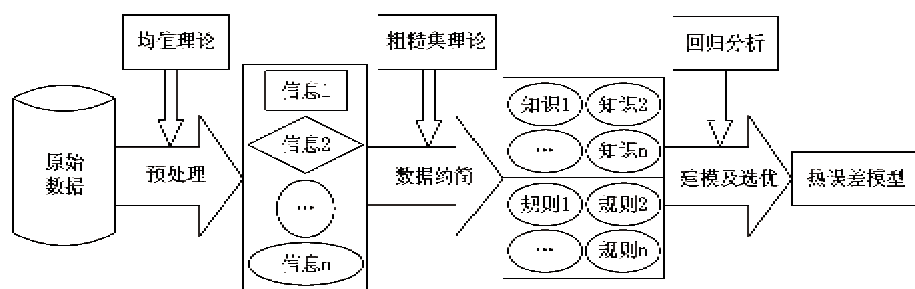


图2 数据挖掘过程

2 实验测试与结果分析

2.1 实验设备

机床的热误差实验所用的实验设备包括一台卧式加工中心(如图3所示)和传感器单元。传感器包括温度传感器和位移传感器。其中,温度传感器用于检测机床主轴、电机、变速箱等所选温度检测点处的温度变化,位移传感器用于检测机床主轴轴端Y向的热变形量。

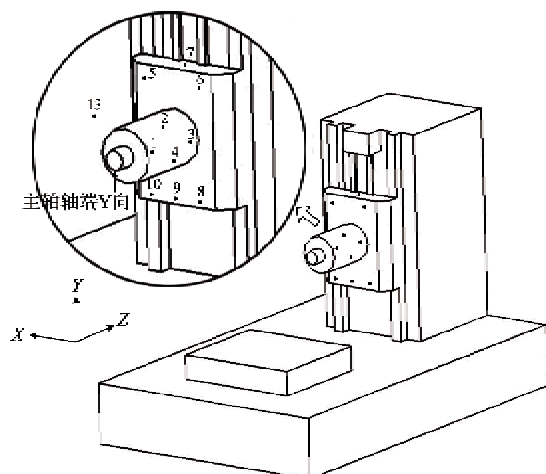


图3 机床及各热检测点分布

2.2 热检测点的选择

文献[15-17]对机床主轴系统温度场进行的分析表明,机床主轴系统的热源主要分布在主轴、主轴箱、变速箱及主轴电机上。考虑到如果选择的温度检测点太多会造成检测时间和热误差模型计算时间过长,而如果选择的检测点太少则可能会造成一些与热误差有关的重要信息的丢失,本文在实验所用的卧式加工中心主轴附近共布置了12个温度检测点,温度检测点的分布情况如表1所示,在机床上的位置如图3所示,其中第11、12号检测点在图中未标出,第13号检测点为环境温度。

表1 温度测点分布

测点编号	位置	测点编号	位置
1	主轴侧面前端	7	主轴箱顶部
2	主轴侧面中部上端	8	主轴箱前面右下端
3	主轴侧面后端	9	主轴箱前面下端中部
4	主轴侧面中部下端	10	主轴箱前面左下端
5	主轴箱前面左上端	11	主轴减速箱上端
6	主轴箱前面右上端	12	主轴电机上端

2.3 热变形实验及结果分析

热变形实验从机床的冷态开始,当机床正常开机后,主轴以3000r/min的速度空载运转,在机床正常运行过程中,利用传感器检测所选温度检测点的温升并同步检测机床主轴轴端Y向的热变形,每隔10 min采集一次温度和热位移数据,直到数据稳定为止,且同一时刻采集到的温度和热位移数据为一组样本,本实验中共得到18组样本。测得温升和热变形的检测结果如图4和图5所示。

由图4可以看出,在实验过程中,机床主轴温升最大值出现在主轴减速箱上端即第11个温度检测点处,温度变化量为15℃左右。由图5可以看出,机床主轴轴端Y方向在运转的3 h内最大热变形量为40μm左右。

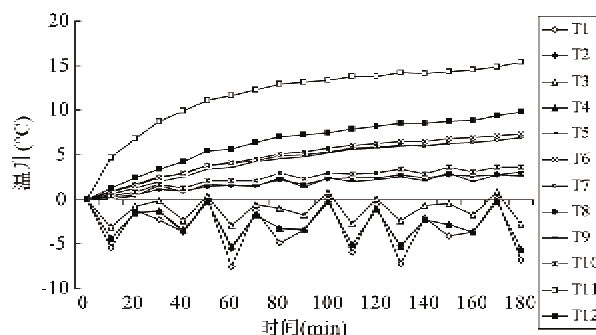


图4 机床主轴不同测点的温升曲线

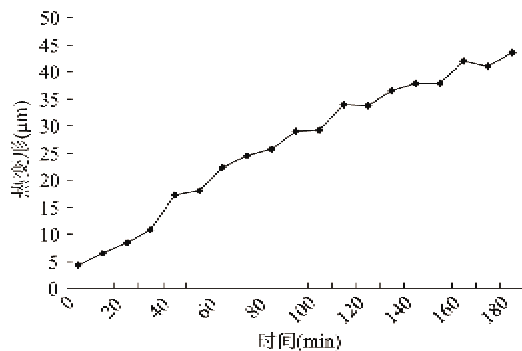


图5 机床主轴轴端Y向热变形曲线

3 热误差建模

3.1 利用 K-means 理论对数据进行预处理

机床的热变形实验共选择 12 个检测点进行温升检测,12 个检测点的温度可记为 $T_1, T_2, \dots, T_{12}$, 并且在每个检测点每隔 10 min 采样一次,共得到 18 个时间段的数据,记为 $X_1, X_2, \dots, X_{18}$ 。在应用粗糙集理论之前,需对实验数据进行预处理,首先将图 4 中的 12 个温度检测点的温升值 ΔT_i 根据其重要性程度^[18]划分为 3 或 4 个层次(如表 2 所示),并记为 P_i ,称为条件属性;将图 5 中的热变形量经过计算得到 $\Delta\delta_{y_i}$ ($\Delta\delta_{y_i} = \sqrt{(\delta_{y_i} - \bar{\delta}_{y_i})^2}$, $i = 1, 2, \dots, 18$, 其

中, δ_{y_i} 为各检测点的热变形量, $\bar{\delta}_{y_i}$ 为各检测点热变形量的平均值),记为 Q ,称为决策属性。利用 K-means 均值理论将 Q 划分为两组,其中第一组数据的中心为 17.2,代表大线性误差,记为 1;第二组数据的中心为 6.212963,代表小线性误差,记为 0。由此可以得到在不同时间段内的条件属性和决策属性的水平值,如表 3 所示。

表2 条件属性的范围

	Level1	Level2	Level3	Level4
ΔT_1	$(-\infty, -5]$	$(-5, -3]$	$(-3, -1]$	$(-1, +\infty)$
ΔT_2	$(-\infty, -5]$	$(-5, -3]$	$(-3, -1]$	$(-1, +\infty)$
ΔT_3	$(-\infty, -5]$	$(-5, -3]$	$(-3, -1]$	$(-1, +\infty)$
ΔT_4	$(-\infty, -5]$	$(-5, -3]$	$(-3, -1]$	$(-1, +\infty)$
ΔT_5	$(-\infty, 2]$	$(2, 4]$	$(4, 6]$	$(6, +\infty)$
ΔT_6	$(-\infty, 2]$	$(2, 4]$	$(4, 6]$	$(6, +\infty)$
ΔT_7	$(-\infty, 2]$	$(2, 4]$	$(4, 6]$	$(6, +\infty)$
ΔT_8	$(-\infty, 1]$	$(1, 2]$	$(2, +\infty)$	
ΔT_9	$(-\infty, 1]$	$(1, 2]$	$(2, +\infty)$	
ΔT_{10}	$(-\infty, 1]$	$(1, 2]$	$(2, +\infty)$	
ΔT_{11}	$(-\infty, 5]$	$(5, 7]$	$(7, 9]$	$(9, +\infty)$
ΔT_{12}	$(-\infty, 5]$	$(5, 7]$	$(7, 9]$	$(9, +\infty)$

表3 在不同时间段内的条件和决策属性水平值

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	Q
X_1	1	2	1	1	1	1	1	1	1	1	1	1	1
X_2	3	3	2	3	1	1	1	1	1	1	1	2	1
X_3	3	3	2	3	1	2	2	2	2	2	2	2	1
X_4	2	2	1	2	2	2	2	1	1	2	2	3	0
X_5	4	4	3	4	2	2	2	2	2	3	2	3	0
X_6	1	1	1	1	2	3	2	2	2	3	2	3	0
X_7	3	3	2	3	3	3	3	2	2	3	3	4	0
X_8	2	2	2	2	3	3	3	3	3	3	3	4	0
X_9	2	2	2	2	3	3	3	2	2	3	3	4	0
X_{10}	4	4	3	4	3	3	3	3	3	3	3	4	0
X_{11}	1	1	1	1	3	4	3	3	2	3	3	4	0
X_{12}	4	3	3	3	3	4	3	3	3	3	3	4	0
X_{13}	1	1	1	1	3	4	4	3	3	3	4	4	0
X_{14}	3	3	2	3	4	4	4	3	3	3	4	4	0
X_{15}	2	3	2	3	4	4	4	3	3	3	4	4	0
X_{16}	2	2	2	2	4	4	4	3	2	3	4	4	1
X_{17}	4	4	3	4	4	4	4	3	3	3	4	4	1
X_{18}	1	1	1	1	4	4	4	3	3	3	4	4	1

3.2 利用粗糙集理论进行信息挖掘和约简

根据 1.2 节所介绍的粗糙集理论,本文采用一

种简单的属性标准来构造区分矩阵,标准^[14]如下:

$$\begin{cases} P \in A, P(X_i) \neq P(X_j), \text{ 若 } Q_i \neq Q_j \\ 0, & \text{ 若 } Q_i = Q_j \\ 1, & P(X_i) = P(X_j), \text{ 若 } Q_i \neq Q_j \end{cases} \quad i, j = 1, 2, \dots, 18 \quad (1)$$

其中, P 为条件属性, Q 为决策属性。根据以上标准,并利用表 3 中的数据,可以构造出如表 4 所示的区分矩阵。

表 4 中,数字 1,2,...,12 分别代表属性 $P_1, P_2, \dots, P_{12}$ 。利用区分矩阵来搜寻最小的属性子集(即约简),从而生成与整个属性集 P 相同的数据划分。为了找到最小的属性集,需要构建区分函数 $f(P)$ 。区分函数 $f(P)$ 是一个布尔函数,构建方法如下:首

先对区分矩阵中的每一列进行“或”操作,对结果进行“且”操作,再通过进一步的逻辑操作得到最小的属性子集和核属性。应用此方法对本实例中的条件属性构建区分函数,结果为:

$$f(P) = (P_2 \wedge P_4) \vee (P_5 \wedge P_7) \vee (P_8 \wedge P_9) \vee P_{12} \quad (2)$$

对于公式(2)所表示的条件属性有 8 种可能的组合形式,分别为:① P_2, P_5, P_8, P_{12} ;② P_2, P_5, P_9, P_{12} ;③ P_2, P_7, P_8, P_{12} ;④ P_2, P_7, P_9, P_{12} ;⑤ P_4, P_5, P_8, P_{12} ;⑥ P_4, P_5, P_9, P_{12} ;⑦ P_4, P_7, P_8, P_{12} ;⑧ P_4, P_7, P_9, P_{12} 。

表 4 区分矩阵

	X1	X2	X3	X4	X5	...	X12	X13	X14	X15	X16	X17
X2	0											
X3	0	0										
X4	1,4~7,10~12	1~7,10~12	1~5,8,9,12									
X5	1~12	1~12	1~5,10,12	0								
⋮	⋮	⋮	⋮	⋮								
X12	1~12	1,3,5~12	1,3,5~12	0	0							
X13	2,5~12	1~12	1~12	0	0							
X14	1~12	5~12	5~12	0	0							
X15	1~12	1,5~12	1,5~12	0	0							
X16	0	0	0	3,5~12	1~8,11,12	...	1~5,7,9,11	1~5,9	1,2,4,9	2,4,9		
X17	0	0	0	1~12	5~9,11,12	...	2,4,5,7,11	1~5	1~4	1~4	0	
X18	0	0	0	1,2,4~12	1~9,11,12	...	1~5,7,11	5	1~4	1~4	0	0

3.3 回归分析及热误差建模

利用 3.2 节所得到的 8 种可能的条件属性组合作为自变量,3.1 节主轴轴端 Y 向热变形量作为因变量来建立热误差线性回归模型。首先对这两种变量进行相关性分析,由条件属性的 8 种组合计算其最大相关系数和最小方差结果如表 5 所示。

表 5 不同温升组合与热变形量的相关系数

序号	回归变量组合	相关系数	最小方差
1	P_2, P_5, P_8, P_{12}	3.23	0.086
2	P_2, P_5, P_9, P_{12}	3.127	0.080
3	P_2, P_7, P_8, P_{12}	3.221	0.085
4	P_2, P_7, P_9, P_{12}	3.118	0.079
5	P_4, P_5, P_8, P_{12}	3.239	0.076
6	P_4, P_5, P_9, P_{12}	3.136	0.077
7	P_4, P_7, P_8, P_{12}	3.23	0.083
8	P_4, P_7, P_9, P_{12}	3.127	0.083

由表 5 可以看出,在条件属性的 8 种组合中,只有第 5 组即 P_4, P_5, P_8, P_{12} 属性组合计算所得到的

相关系数(3.239)最大,最小方差(0.076)最小。因此,线性回归热误差建模选用第 5 组属性组合作为建模的温度自变量,热变形数据作为因变量进行回归建模。本文在实验过程中选取了 18 组测量数据用于建模,通过计算得到机床主轴的温度-热位移之间线性回归方程为

$$dy = -62.28 - 0.51 \times T_4 + 34.02 \times T_5 - 11.42 \times T_8 - 17.68 \times T_{12} \quad (3)$$

其中, dy 为主轴轴端 Y 向热变形量, T_4, T_5, T_8, T_{12} 分别为温升值。

4 建模结果对比分析

为了验证本文所提出的基于热态信息挖掘的热误差建模方法的有效性,将引入反向传播(back propagation, BP)神经网络模型与其进行对比。BP 神经网络是一种典型的多层网络,由输入层、隐含层和输出层组成,层间多为全互连方式,同层单元间不存在连接。将本文 3.3 节所得到的结果即 T_4, T_5 ,

T_8, T_{12} 四个温度检测点的温升值作为网络的输入, 主轴轴端 Y 向热变形量即 dy 作为网络的输出来构建机床热误差神经网络模型。将热变形实验所测得的 18 组数据分为两部分, 其中前 13 组数据作为训练样本, 后 5 组数据作为测试样本, 对所建立的神经网络进行训练。

本文通过机床的热变形实验来对两种误差建模方法进行比较: 机床主轴以 3000r/min 速度运转 180 min, 每隔 10 min 进行一次数据采集。然后分别应用 BP 神经网络模型和基于热态信息挖掘方法所建模型来预测机床的热误差变形, 预测结果及建模精度对比分别如图 6 和表 6 所示。

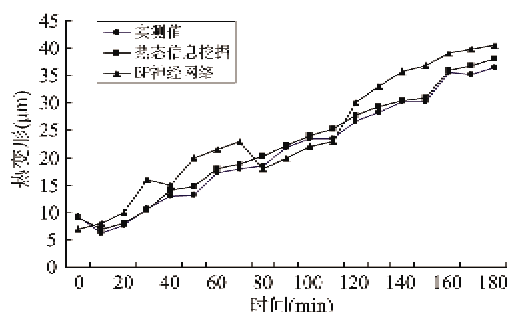


图 6 预测值和实验结果对比图

表 6 建模精度对比表

参数	热态信息挖掘	BP 神经网络
残差范围	$[-0.2, 1.9]$	$[-2.2, 5.8]$
残差平均值	0.07	0.1
残差方差	0.9836	5.4289

由图 6 和表 6 可知:

(1) 基于热态信息挖掘方法所建模型的预测结果与实测数据比较接近, 说明本建模方法具有较高的建模精度。

(2) 与 BP 神经网络相比, 基于热态信息挖掘方法所建模型的预测结果更接近于机床的实际情况, 表明本文所建立的热误差建模方法在机床热误差预测上具有优越性。

5 结 论

本文提出了一种基于热态信息挖掘的机床主轴系统热误差建模方法, 以某精密卧式加工中心为研究对象, 对机床主轴系统进行了温升和热变形误差建模。通过与其它建模方法 (BP 神经网络方法) 进行对比分析得出: 利用数据挖掘技术可以快速、高效

地从大量实验数据中挖掘出有价值的信息用于机床热误差建模; 经实验结果验证, 本文建立的热误差建模方法有较高的预测精度。下一步的研究工作是将该方法应用到数控机床的进给系统及整机的热误差补偿中, 并对其进行实验验证。

参考文献

- [1] Ramesh R, Mannan M A, Poo A N. Error compensation in machine tools—a review part II: thermal errors. *International Journal of Machine Tools and Manufacture*, 2000, 40(9): 1257-1284
- [2] 杨建国. 数控机床误差综合补偿技术及应用: [博士学位论文]. 上海: 上海交通大学机械工程学院, 1998. 14-15
- [3] Hsu Y Y, Wang S S. A new compensation method for geometry errors of five-axis machine tools. *International Journal of Machine Tools & Manufacture*, 2007, 47: 352-360
- [4] 左红, 方辉, 刘立新等. 基本热误差神经网络预测模型的机床重点热刚度辨识方法研究. *机械工程学报*, 2011, 47(11): 117-124
- [5] 李永祥, 杨建国. 灰色系统模型在机床热误差建模中的应用. *中国机械工程*, 2006, 17(23): 2439-2442
- [6] Lee J H, Yang S H. Thermal error modeling of a horizontal machining center using fuzzy logic strategy. *Journal of Manufacturing Processes*, 2001, 3(2): 120-127
- [7] Wang K C, Tseng P C, Lin K M. Thermal error modeling of a machining center using grey system theory and adaptive network-based fuzzy inference system. *JSME International Journal, Series C*, 2006, 49(4): 1179-1187
- [8] Delbressine F L M, Florussen G H J, Schijvenaars L A, et al. Modeling thermal mechanical behavior of multi-axis machine tools. *Precision engineering*, 2006, 30: 47-53
- [9] 张宏韬, 杨建国. RBF 网络在线建模方法在热误差实时补偿技术中的应用. *上海交通大学学报*, 2009, 43(5): 807-810
- [10] 蔡自兴, 徐光祐. 人工智能及其应用. 北京: 清华大学出版社, 2004. 21-22
- [11] 刘涛, 王益群, 薛志勇. 带钢冷轧过程辊系统径向变形的数据挖掘与预报. *机械工程学报*, 2011, 47(6): 69-79
- [12] 朱海平, 王忠浩, 张国军等. 基于数据挖掘的通用物料清单重构方法研究. *计算机集成制造系统*, 2008, 14(2): 315-320
- [13] Nanda S, Majumdar S. Fuzzy rough sets. *Fuzzy Sets and Systems*, 1992, 2(11): 157-160
- [14] 菅利荣. 面向不确定性决策的杂合粗糙集方法及其应用. 北京: 科学出版社, 2008. 12-27

- [15] Yang J G, Yuan J X, Ni J. Thermal error model analysis and robust modeling for compensation on a CNC turning center. *International Journal of Machine Tools & Manufacture*, 1999, 39(9): 1367-1381
- [16] 蹇小龙, 杨建国, 关贺等. 温度测点优化在机床主轴热误差建模中的应用. *航空精密制造技术*, 2003, 39(4): 33-36
- [17] 张海燕, 曾黄麟, 唐建芳. 基于粗集方法的机床热补偿误差的温度测点优化. *机床与液压*, 2010, 38(1): 39-41
- [18] 刘凌霞. 基于粗糙集理论属性重要性的离散化算法. *广西轻工业*, 2007, 10: 75-76

A thermal error modeling method for the spindle system of the machine tools based on thermal information mining

Sun Huijuan^{*}, Lin Haifeng^{**}, Xiang Shenghua^{**}, Yin Guofu^{*}, Fang Hui^{*}

(^{*} School of Manufacturing Science and Engineering, Sichuan University, Chengdu 610065)

(^{**} Chengdu Aircraft Industrial(Group) Co. Ltd., Chengdu 610091)

Abstract

Aiming at the thermal error modeling of computerised numerical control (CNC) machine tools, a new modeling method, characterized by classifying data using the *K*-means theory, mining data using the rough set theory, and establishing an error model with the linear regression analysis technique, is proposed by analyzing the characteristics of the thermal deformation data of a high accuracy horizontal machining center spindle system. The method consists of three steps: Firstly, the temperature ascent at the selected check points and the thermal deformation at the spindle nose are measured with sensors through the thermal error experiment on a machine tool spindle system; then, the obtained data is classified by the *K*-means method, mined and reduced by the rough-set method; and finally, a thermal error model is established by means of linear regression analysis. To evaluate the performance of the proposed method, a back propagation (BP) neural network is introduced for comparison. The experimental results reveal that the proposed thermal error modeling method based on thermal information mining is highly accurate and effective in predicting the thermal behavior of a spindle of machine tools.

Key words: numerical control (NC) machine tool, thermal error, data mining, modeling, spindle