

基于核函数的三坐标空间测量误差估计

张月梅,张梅,高陈媛

安徽大学

摘要: 由于三坐标测量过程中涉及多种测量误差因素,因此在快速测量过程中对误差特性进行详细分析就比较复杂。本文以移动桥式三坐标测量机为研究对象,运用基于非参数回归模型中的核函数估计法对手动三坐标测量机的空间测量误差进行了多元非线性建模和预测。通过对采集数据进行回归拟合,并与参数建模的拟合结果进行对比分析后可知,非参数回归模型中的核函数估计法具有更好的预测效果。

关键词: 三坐标测量机,误差建模,非参数回归,核函数估计

中图分类号: TG806;TH71

文献标志码: A

Error Estimation of Coordinate Measuring Machine Based on Kernel Function

Zhang Yuemei, Zhang Mei, Gao Chenyuan

Abstract: Since the error is affected by various factors in the process of CMM measuring, the detailed error characteristic analysis in fast measurement can be complicated. Taking a moving bridge CMM as the research object, and based on the kernel function estimation in non-parametric regression, the multivariable nonlinear modeling and estimation of manual CMM measuring errors are carried out. By comparing the fitting and regression results from the data acquisition and the parametric modeling, the kernel function estimation method shows a better result.

Keywords: CMM; error modeling; non-parametric regression; kernel function estimation

1 引言

由于三坐标测量机测量过程中涉及到多种测量误差因素,测量误差与空间坐标的关系更多地表现为非稳定、非线性关系,且不知测量误差的总体分布形式,因此很难用具体准确的数学表达式来描述。由于经典的参数建模法要求被分析的数据遵从某一已知特定分布如正态分布,一旦这种假定条件不成立,其推断的正确性就不存在。

核估计法属于非参数建模,在抽取样本对总体

进行估计时不必依赖于样本所属的总体分布形式,对模型的限制很少。比较两种建模方法的回归拟合可知,核估计法比偏最小二乘法有更好的拟合效果,非参数建模比参数建模更能准确预测三坐标测量机动态误差。为此,本文采用核估计法对三坐标测量机的动态误差进行估计。

2 核函数估计理论

2.1 核估计简介

核估计也称为局部加权平均估计,利用加权函数进行拟合函数估计,是一种非参数估计方法。对于一组关于 X 和 Y 的观测数据 $\{(x_i, y_i)\}_{i=1}^n$, 假设它们存在如下关系: $y_i = m(x_i) + \varepsilon_i, i = 1, \dots, n$ 。当

基金项目: 安徽省教育厅重点项目(KJ2016A037);安徽大学2015大学生科研训练计划
收稿日期: 2017年4月

[9] Zhang Xu, Ma Zi, Hu Yang, et al. A novel robot system for surface inspection and diameter measurement of large size pipes. [C]//The IEEE International Conference on Industrial Informatics(INDIN 2008), 2008:1717-1721.

[10] Zhang Liancun, Zhang Yating, Zhang Guoyu, et al. Research of optoelectronic non-contact inspecting method for dimension of inner diameter[J]. Journal of University of Electronic Science and Technology of China, 2006, 35(5):829-832.

[11] 苏丽梅, 马宏, 白素萍. 管道内径光电检测技术研究[J]. 长春理工大学学报, 2005, 28(1):30-31.

[12] 邢书剑, 曲兴华, 王晟, 等. 采用高精度自动定心机构的大尺寸内径测量[J]. 仪器仪表学报, 2012, 33(8):1681-1687.

[13] 赵士磊, 曲兴华, 邢书剑, 等. 激光非接触式大尺寸内径自动测量系统[J]. 红外与激光工程, 2012, 41(8):2185-2191.

第一作者: 赵海峰, 博士, 副教授, 南京信息职业技术学院机电学院, 210046 南京市

First Author: Zhao Haifeng, Doctor, Associate Professor, School of Microelectronics, Nanjing College of Information Technology, Nanjing 210046, China

解释变量为确定变量时,假设随机误差项的数学期望为零,即 $E(u_i) = 0$,此时因变量的数学期望 $E(Y_i) = m(X_i)$;当解释变量为随机变量时,假定解释变量与随机误差项独立,假定随机误差项的条件数学期望为零,即 $E(\frac{Y_i}{X_i}) = m(X_i)$ 。一般来说,可以认定 $m(x)$ 为 Y 在 $X = x$ 条件下的期望值。

通常核估计的目的在于确定 $m(x)$ 的最佳形式。在非参数化估计中,对 $m(x)$ 的形式不做任何假设,主要从样本中获取分布信息。若 $\{X^1, X^2, \dots, X^N\}$ 为随机变量 x 的 N 个样本,随机变量 x 的密度函数记为 $f(x)$,则核函数估计 $f(x;h)$ 定义为

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K(\frac{x - X_i}{h}) \quad (1)$$

式中, h 为窗宽,满足条件 $\lim_{n \rightarrow \infty} h = 0$ 和 $\lim_{n \rightarrow \infty} Nh = 0$,其作用是调节核函数密度的光滑程度; $K(\cdot)$ 为核函数,是一个权函数。

核函数估计是一个加权平均。核函数估计的好坏依赖于核函数和窗宽 h 的选取。常用核函数如表 1 所示。

表 1 常用核函数

名称	表达式
Uniform	$\frac{1}{2}(u \leq 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} (u \leq 1)$
Epanechnikov	$\frac{3}{4}(1 - u^2) (u \leq 1)$
Biweight	$\frac{15}{16}(1 - u^2)^2 (u \leq 1)$
Triweight	$\frac{35}{32}(1 - u^2)^3 (u \leq 1)$

2.2 核函数窗宽选择

核函数估计 $\hat{f}_n(x)$ 的偏差和方差分别定义为

$$Bias \{ \hat{f}_n(x) \} = E[\hat{f}_n(x)] - f(x) \quad (2)$$

$$Var \{ \hat{f}_n(x) \} = E[\hat{f}_n(x)]^2 - [E\hat{f}_n(x)]^2 \quad (3)$$

窗宽的选择对估计值 $\hat{f}_n(x)$ 的影响很大。窗宽 h 越小,估计结果越接近针状分布;窗宽 h 越大,估计结果越平滑,但会光滑掉 f 的一些重要特征。MISE 准则旨在避免了这两种极端情况发生。MISE 准则要求最小化以下值

$$MISE \{ \hat{f}_n(x) \} = E[\hat{f}_n(x) - f(x)]^2 \quad (4)$$

将 $MISE \{ \hat{f}_n(x) \}$ 简写为 $MISE(h)$ 以表示它依赖于窗宽 h 。基于最小化 MISE 准则的最优窗宽定义为

$$h_{opt} = \operatorname{argmin}_h MISE(h) \quad (5)$$

根据式(4)可以将 MISE 写成

$$MISE(h) = \int Bias^2 \{ \hat{f}_n(x) \} + \int Var \{ \hat{f}_n(x) \} \quad (6)$$

对偏差和方差进行泰勒展开可得

$$MISE(h) = \frac{R(K)}{nh} + \frac{h^4 \mu_K^2 R(f'')}{4} + o(\frac{1}{nh}) + o(h^4) \quad (7)$$

当拥有大量样本时,以下定义的渐进 MISE (AMISE) 可以作为一个好的近似

$$AMISE(h) = C_V R(K) n^{-1} h^{-1} + \frac{1}{4} C_B \mu_K^2 h^4 \quad (8)$$

式中, C_V 为常量; $C_B = R(f'')$, 均是与核函数 $K(\cdot)$ 无关的量。

对上式求一阶导数,然后令其为 0,可得到基于最小化 AMISE(h) 准则的最佳窗宽为

$$h_{opt} = (\frac{C_V}{4C_B})^{1/5} (\frac{R(K)}{\mu_K^2})^{1/5} n^{-1/5} \quad (9)$$

2.3 核函数选择

在 AMISE 准则下将式(9)式带入式(8)式中,可得

$$AMISE_{opt} = n^{-4/5} (C_V)^{4/5} (C_B)^{1/5} (4^{1/5} + 4^{-4/5}) [R^2(K) \mu_K]^2 \quad (10)$$

当确定最佳窗宽后,由于 C_V 和 C_B 为常量,所以要使 $AMISE(h)$ 最小化,则需令 $R^2(K) \mu_K$ 最小。 $R^2(K) \mu_K$ 与核函数的关系如下

$$R^2(K) \mu_K = \left[\int K^2(u) du \right]^2 \int u^2 K(u) du \quad (11)$$

所以最优的核函数是使式(11)达到最小的核函数。

3 非参数回归建模

非参数回归建模流程如图 1 所示。

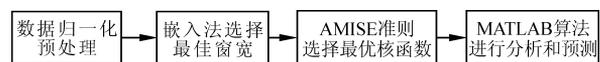


图 1 非参数回归建模流程

3.1 模型数据及量化

共采集了 480 组试验数据。在同一试验样本的基础上随机选取 5% (即 24 组) 数据作为预测样本,其余数据为建模样本。对训练集和测试集进行归一化预处理,归一化映射如下

$$f: x \rightarrow y = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (12)$$

式中, $x, y \in R^n$; $x_{min} = \min(x)$; $x_{max} = \max(x)$ 。

归一化的效果是使原始数据被规范到 $[1, 2]$ 范围内。

在 MATLAB 中, mapminmax 函数可以实现上述归一化,其常用的函数接口如下

$$[y, ps] = \text{map min max}(x, y_{\min}, y_{\max}) \quad (13)$$

3.2 窗宽选择

在式(9)中, $C_B = R(f'')$ 仍然未知, 一般的做法是用嵌入法 (Plug-in) 嵌入 $R(f'')$ 一个恰当估计。在此采用 Silverman 提出的一种初等方法, rule of thumb 拇指法则, 即根据经验的方法。如果事先知道 f 来自正态总体且具有均值 u 及方差 σ^2 , 那么

$$R(f'') = \sigma^{-5} R(\varphi'') = \frac{3}{8\sqrt{\pi}} \sigma^{-5} \quad (14)$$

式(14)中 $\varphi(\cdot)$ 表示标准正态分布的密度函数, 未知参数可用样本标准差 $\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ 。由此得到窗宽为

$$\hat{h}_{rot} = \left(\frac{4}{3n}\right)^{1/5} \hat{\sigma} \quad (15)$$

通过 MATLAB 编程计算得出最佳窗宽 $h = 0.5$ 。

3.3 通过 AMISE 准则选择核函数

通过对式(11)积分求解, 得到使 AMISE 最优的核函数为

$$K(u) = \frac{d(d+2)}{2S_d} (1 - u_1^2 - u_2^2 - \dots - u_d^2) \quad (16)$$

式中, $S_d = 2\pi^{d/2} / \Gamma(d/2)$, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ 。

所以一元核估计的最优函数是 $K(u) = 0.75(1 - u^2)$, 即 Epanechnikov 核函数。

3.4 MATLAB 分析与预测

在三坐标测量机空间误差建模研究中, 以测量坐标值 x, y, z 为自变量, 标准量块的测量误差 e 为因变量, 通过嵌入法选择最佳窗宽 h , 由 AMISE 准则选择最优核函数。

通过核估计公式可以得到三坐标测量机在测量空间内不同测量点产生的测量误差的预测值为

$$\hat{m}(X) = \frac{7.5 \sum_{i=1}^n [1 - (\frac{x_i - x}{h})^2 - (\frac{y_i - y}{h})^2 - (\frac{z_i - z}{h})^2] e_i}{S_3 \sum_{i=1}^n [1 - (\frac{x_i - x}{h})^2 - (\frac{y_i - y}{h})^2 - (\frac{z_i - z}{h})^2]} \quad (17)$$

式中, $X = (x, y, z)$; $S_3 = 2\pi^{3/2} \Gamma(3/2)$, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ 。

根据式(17), 通过 MATLAB 编程实现公式算法, 得到预测样本的预测值, 见图 2。计算出均方误差 $MSE = 0.00634942\text{cm}$, 相关系数 $R = 97.3449\%$ 。

在偏最小二乘法的 MATLAB 算法中, 拟合曲线与实际测得数据间的均方误差为 0.03069cm 。与偏最小二乘法相比, 核函数估计减小了均方误差, 提高

了曲线拟合的精确度, 更适合进行空间坐标测量误差曲线的拟合。

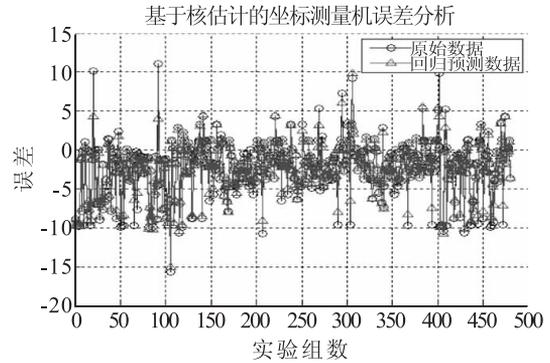


图 2 核估计预测结果

4 结语

本文针对空间测量误差和各影响因素间可能存在着复杂的非线性关系, 采用核函数估计的非参数回归方法对三坐标空间测量误差进行了建模分析, 同时与偏最小二乘法作对比。

由两种建模方法的回归拟合比较可知, 核估计法比偏最小二乘法有更好的拟合效果。在手动三坐标测量机空间测量误差的后续研究中, 如果在对总体分布类型难以知晓、参数估计中的假设得不到满足时, 可以考虑选择合适的非参数回归模型法。

参考文献

- [1] 费叶泰, 赵静, 王宏涛, 等. 三坐标测量机动态误差研究分析[J]. 仪器仪表学报, 2004, 25(1): 773-776.
- [2] 郭照庄, 霍东升, 孙月芳. 密度核估计中窗宽选择的一种新方法[J]. 佳木斯大学学报, 2008, 26(3): 401-403.
- [3] 杨永生, 张宗杰. 基于核函数和带宽的海杂波概率密度函数估计[J]. 探测与控制学报, 2010, 32(5): 38-41.
- [4] 王文圣, 丁晶. 基于核估计的多变量非参数随机模型初步研究[J]. 水利学报, 2003(2): 9-14.
- [5] 刘敬伟, 胡爽, 康进, 等. 倒向随机微分方程中非参数估计的核函数选择[J]. 统计与信息论坛, 2011, 26(11): 8-12.
- [6] 高采文, 甘华来. 增长曲线模型的非参数估计[J]. 应用概率统计, 2013, 29(6): 655-665.

第一作者: 张月梅, 本科生, 安徽大学电气工程与自动化学院, 230601 合肥市

First Author: Zhang Yuemei, Undergraduate, School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China