

基于非参数回归的航材消耗预测模型研究

陈振林, 薛永亮

引用本文:

陈振林, 薛永亮. 基于非参数回归的航材消耗预测模型研究[J]. [兵器装备工程学报](#), 2020, 41(6): 132–135.

您可能感兴趣的其他文章

1. ANFIS在航材需求量预测中的应用

引用本文: 康正亮, 胥凯晖, 杨传起. ANFIS在航材需求量预测中的应用[J]. [兵器装备工程学报](#), 2015, 36(1): 84–87.

2. 基于灰色多项式回归组合模型的维修器材需求预测方法研究

引用本文: 张磊, 于战果, 李世民. 基于灰色多项式回归组合模型的维修器材需求预测方法研究[J]. [兵器装备工程学报](#), 2019, 40(1): 179–183.

3. 战时航母编队导弹消耗预测模型研究

引用本文: 邹强, 王城超, 王栋, 等. 战时航母编队导弹消耗预测模型研究[J]. [兵器装备工程学报](#), 2018, 39(2): 90–94,118.

4. 基于改进EMD的排气温度裕度预测

引用本文: 戴邵武, 陈强强, 丁宇. 基于改进EMD的排气温度裕度预测[J]. [兵器装备工程学报](#), 2020, 41(1): 157–162.

5. 基于EOQ下我国航材存储模型的实例研究

引用本文: 苏涛, 曹文静, 赵敏. 基于EOQ下我国航材存储模型的实例研究[J]. [兵器装备工程学报](#), 2015, 36(4): 64–67.

基于非参数回归的航材消耗预测模型研究

陈振林, 薛永亮

(海军航空大学 岸防兵学院, 山东 烟台 264001)

摘要:针对某型引俄直升机,服役时间短、机型较新、备件采购周期长等特点,使用非参数回归算法建立了航材消耗模型。非参数回归分析是一种不设置先验知识的回归模型,通过每个数据点计算权重,使得回归曲线延伸性能较好,能较好适应该型直升机航材消耗预测模型缺乏先验知识、预测周期长等特点。通过对消耗数据建模仿真,并对比多种回归方法,结果表明基于非参数回归的航材消耗模型对区间预测具有较好的效果。

关键词:航材消耗;预测;非参数回归;局部多项式回归

本文引用格式:陈振林,薛永亮.基于非参数回归的航材消耗预测模型研究[J].兵器装备工程学报,2020,41(06):132-135.

Citation format:CHEN Zhenlin, XUE Yongliang. Research on Prediction Model of Aviation Material Consumption Based on Nonparametric Regression[J]. Journal of Ordnance Equipment Engineering, 2020, 41(06): 132-135.

中图分类号:V267;E926

文献标识码:A

文章编号:2096-2304(2020)06-0132-04

Research on Prediction Model of Aviation Material Consumption Based on Nonparametric Regression

CHEN Zhenlin, XUE Yongliang

(Navy Aviation University Coast Dedence College, Yantai 264001, China)

Abstract: Prediction of aviation materials has become an important guarantee for accurate guarantee of aviation materials. In view of the characteristics of a certain type of Russian-piloted helicopter, such as short service time, relatively new type and long spare parts procurement cycle, non-parametric regression analysis was used to establish the aviation materials consumption model. By modeling and simulating the consumption data and comparing various regression methods, it is proved that the aircraft material consumption model based on non-parametric regression has a good effect on interval prediction.

Key words: aviation material consumption; prediction; nonparametric regression; local polynomial regression

关于航材的预测,国内外学者采用了多种方法进行试验拟合,其中包括线性回归、神经网络、支持向量机等方法^[1-5]。线性回归模型对缺少明确变化趋势、离散型数据拟合较差,但其算法固定,通过规定的算法得出唯一结果,且其延伸性能较好;支持向量机是一种优秀的机器学习预测算法,能较好解决局部最优、过度学习等问题,对小样本信息预测效果较好,但其对于较大步幅预测,由于缺少样本支撑,经常拟合效果较差。

对某型引俄直升机航材消耗建立模型时,需要考虑两个特点:一是机型较新、服役时间短,难以获取足够数据,不能划分训练集;二是机型备件国产化水平低,大部分要依靠进口,补给难度较大且易受多种因素影响,对该机型航材消耗,不仅要预测 $T+1$ 时刻,还要预测 T 之后较长步幅的区间。基于上述问题,本文使用非参数回归算法建立模型。非参数回归是一种不对模型参数做任何假设的回归算法^[6],仅规定一些一般性条件,近年来该算法在多学科运用较为广泛^[7-13]。

收稿日期:2019-07-19;修回日期:2019-11-24

作者简介:陈振林(1969—),男,教授,主要从事雷达应用研究,E-mail:bigtotty@126.com。

通讯作者:薛永亮(1990—),男,硕士研究生,主要从事航材保障研究,E-mail:402062810@qq.com。

1 非参数回归分析

回归分析是应用最广的统计分析方法,其一般模型数学表达式为:

$$Y = m(X) + U \quad (1)$$

式(1)中: X 预测变量; Y 是 X 的响应变量; U 是随机误差,且满足 $E(U|X) = 0, E(U^2|X) = \sigma^2(X)$; $m(\cdot)$ 是光滑函数^[14]。若 $m(\cdot)$ 模型已知,参数未知即为参数回归。对于模型的估计,需要一定的先验知识做出强假设,否则会有较大偏差。当 $m(\cdot)$ 模型未知,该回归函数即为非参数回归,在分析时不做任何参数假定。

对该型直升机,预测时可以根据相似机型基础上修正,计算出足够量的先验知识,利用贝叶斯大数定律预测^[15-17]。但是类比推测的过程中不可避免会产生误差,并且使用参数回归形式一旦固定,经常拟合效果差。在这种情况下,当数据样本容量足够,使用非参数回归具有一定的可行性,且应用范围更广,性能更稳健^[18]。使用非参数回归建模,不需要对数据样本做先验估计,仅依靠数据自身规律进行拟合,通过每个数据计算权重,使得回归曲线具有整体性,对长步幅区间的预测效果较好^[6],能较好解决该机型样本数据少、预测长区间等问题。

非参数回归有多种算法,最经典的是局部核回归,对 $m(\cdot)$ 估计的数学表达式为:

$$m(X) = \frac{\sum_{i=1}^n Y_i k\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n k\left(\frac{X_i - x}{h}\right)} \quad (2)$$

式(2)中: $k(\cdot)$ 是核函数; h 是带宽; X_i, Y_i 分别为预测变量与响应变量。如式(2)所示,平滑函数仅与 X, Y 相关,通过选取适当的核函数与带宽,建立预测模型。

2 基于非参数回归的航材消耗预测

2.1 航材消耗序列

选取某型引俄直升机一种关键航材,统计 8 年的航材消耗数据。在 8 年时间内,该型飞机由于各种原因服役数量存在波动,在本文中,转换为单机月平均消耗件数,方便预测模型的建立。其消耗序列如图 1 所示,可以初步看出,消耗序列较为离散,有缓慢上升趋势。利用非参数回归模型,对数据样本不做先验假设,依据数据样本自身特点进行回归运算。

2.2 局部多项式回归算法

局部核回归存在边界带估计偏差较大的缺点,对其回归算法进行改进,建立局部多项式回归^[7]。其主要方法是在 x 的一个邻域内,用多阶多项式去估计光滑函数 $m(\cdot)$,然后进行核函数加权,求出 $m(\cdot)$ 及各阶系数。 p 阶局部多项式估计是求下式最小值的解,即:

$$M = \sum_{i=1}^n (Y_i - \beta_0 - \dots - \beta_p (X_i - x)^p)^2 K\left(\frac{X_i - x}{h}\right) \quad (3)$$

式(3)中: Y_i 为响应变量即为每个月单机消耗量; X_i 为预测变量即为月份。设 $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^T$ 使式(3)达到最小值 $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ 的值,则 β_0 即回归函数 $m(x)$ 的估计, β_s 是 $\frac{1}{s!} m_s(x)$ 的估计, $m_s(x)$ 是 $m(x)$ 的 s 阶导数,可以看出,当 $p=0$ 时,即为局部核估计^[19]。令

$$w(x) = \text{diag}\left(K\left(\frac{X_1 - x}{h}\right), \dots, K\left(\frac{X_n - x}{h}\right)\right) \quad (4)$$

$$Y = (Y_1, Y_2, \dots, Y_n)' \quad (5)$$

$$\beta = (\beta_0, \beta_1, \dots, \beta_p)' \quad (6)$$

$$X_x = \begin{pmatrix} 1, (X_1 - x), \dots, (X_1 - x)^p \\ 1, (X_2 - x), \dots, (X_2 - x)^p \\ \vdots \\ 1, (X_n - x), \dots, (X_n - x)^p \end{pmatrix} \quad (7)$$

将式(3)转化为:

$$M = (Y - X_x \beta)^T W_x (Y - X_x \beta) \quad (8)$$

即转化为求式(4)的最小二乘问题。

取 $p=1$,有:

$$m_2(x) = (X_x^T W_x X_x)^{-1} (X_x^T W_x Y) \quad (9)$$

经计算可得

$$m_2(x) = \frac{1}{n} \sum_{i=1}^n \frac{(S_2(x, h) - S_1(x, h) \times (X_i - x)) \times K_h \times Y_i}{S_2(x, h) \times S_0(x, h) - S_1(x, h)^2} \quad (10)$$

式(10)中: $S_0(x, h) = \frac{1}{n} \sum_{i=1}^n K_h$; $S_1(x, h) = \frac{1}{n} \sum_{i=1}^n K_h \times (X_i - x)^T$; $S_2(x, h) = \frac{1}{n} \sum_{i=1}^n K_h \times (X_i - x) \times (X_i - x)^T$; $K_h(\cdot)$ 是核函数; X_i, Y_i 分别代表预测变量与响应变量; h 代表带宽。由式(6)~式(9)可以看出,在选定核函数后,关于 x 的光滑函数需要确定带宽 h 。

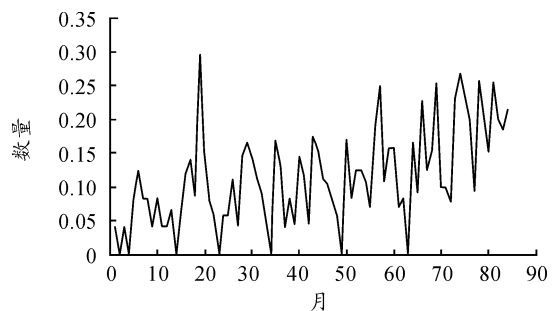


图 1 航材消耗序列图

2.3 模型选择

核函数有多种选择,通常有 Triangle、Epanechnikov、Quartic、Triweight、Gaussian 等核函数,有学者指出当带宽 h 合适时,核函数的选择对于回归结果影响较小^[20-22]。针对上述数据较为离散、变化趋势不明显的问题,使用高次方函数精

确性会有一定提高^[23],使用 Triweight 函数作为备选函数,其数学表达式为:

$$K(x) = \frac{35}{32}(1 - x^2)^3 \quad (11)$$

在试验中发现,使用 Triweight 核函数,在序列两端偏差较大,高次幂对权重影响较大,对该核函数改造为:

$$K(x) = (1 - 3x^2 + 0.5x^4)(1 - 0.2x^2) \quad (12)$$

由此能有效降低高次幂在两端的影响,一定程度消除偏差。

对于带宽 h 的选择优化问题,有了较多的研究,较多学者使用交叉验证选取^[20-24]。交叉验证是通过构建均方误差最小的估计量来确定最优带宽,通过剩余数据量来预测响应变量,但对于预测外延有一定误差,本文使用经典拇指法则^[6],其数学表达式为:

$$h = 1.06\sigma n^{-0.2} \quad (13)$$

式(13)中: σ 为样本方差; n 为样本个数。

3 仿真实验

为有效对比非参数回归模型准确性,使用多种方法对消耗数据进行建模。使用前 72 个数据建立模型,预测 73-84 数据,与真实数据做对比。

3.1 数据检验

数据检验是回归建模的基础工作,对数据的平稳性做出检验,若不满足要求,在回归拟合前需要进行偏差处理。使用 Eviews 进行单位根检验,检验结果如图 2 所示。

Augmented Dickey-Fuller Unit Root Test on SERIES01		
Null Hypothesis: SERIES01 has a unit root		
Exogenous: Constant		
Lag Length: 0 (Automatic - based on SIC, maxlag=11)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.476568	0.0000
Test critical values:		
1% level	-3.525618	
5% level	-2.902953	
10% level	-2.588902	

图 2 平稳性检验结果

通过 ADF (Augmented Dickey-Fuller) 检验 T 值满足要求显著性检验, P 值为 0, 认为消耗序列平稳, 无需进行差分运算^[15]。

3.2 试验对比

分别使用最小二乘法 (Least Square Method) 2 阶、3 阶、4 阶, 自回归滑动平均模型 (Autoregressive Moving Average Model)、非参数回归 (Non-parametric Regression) 等方法建模, 仿真曲线如图 3 所示。

由图 3 可以看出, 自回归滑动平均模型对于长步距预测效果较差, 最小二乘法拟合曲线由于序列自身波动性较大而缺少明显趋势, 拟合过于平滑, 但 3 阶算法预测结果与实际消耗趋势相同。非参数回归由于两端高次幂影响偏差较大, 通过改进后模型有了较好优化, 曲线拟合与实际消耗情况

相近。

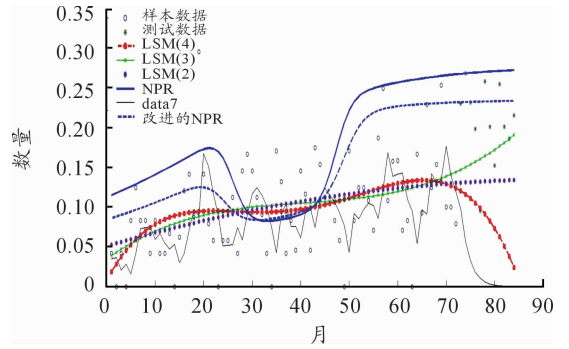


图 3 回归仿真曲线

支持向量机 (Support Vector Machine) 对小样本预测较为理想, 多名学者在此基础上进行了改进, 取得了一定的成果^[25]。为充分对比, 本文使用 SVM 对上述仿真进行建模。在上文建模仿真中, 73 至 84 数据设为未知, 因此在设计 SVM 试验时, 设置 [1, 60] 数据为训练集、[61, 72] 数据为标准集, 对 [61-84] 数据预测拟合, 其推导原理及试验过程在此不做描述, 仿真结果如图 4 所示。

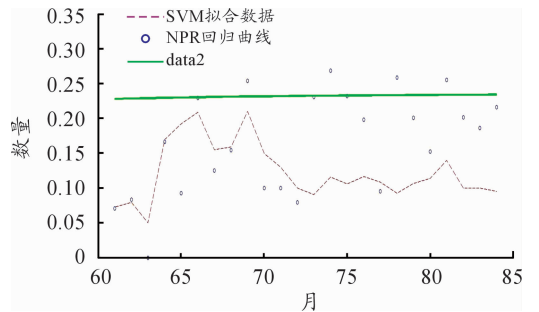


图 4 SVM 仿真曲线

由图 4 可以看出, 由于前期数据支撑, 对于近邻阶段预测效果较好, 61~72 仿真数据拟合较为理想, 对比回归曲线更加接近数据样本。但 74 之后数据拟合效果较差, 对于消耗数据上升趋势未做有效预测分析。

为有效评价模型拟合效果, 使用以下两个指标分析对比不同模型的预测准确性。

1) 均值百分比误差 MAPE。其数学表达式为^[10]:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (14)$$

2) 均方根差 RMSE。其数学表达式为^[10]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

式(15)中: y_i 为实际值; \hat{y}_i 为预测值。

对其后 12 项数据预测结果如表 1 所示。由表 1 可以看出, 最小二乘 3 阶模型与非参数局部多项式回归预测效果接近, 而改进后非参数局部多项式回归预测效果明显较好。

表1 预测效果对比

模型	MAPE	MRSE
LSM(2)	0.396 4	0.093 1
LSM(3)	0.281 1	0.067 9
LSM(4)	0.587 1	0.142 6
ARMA	0.912 7	0.199 0
SVM	0.475 9	0.115 3
NPR	0.285 6	0.070 1
改进 NPR	0.261 3	0.053 6

4 结论

基于非参数回归算法建立航材消耗预测模型,在建模时不对变量作强假设推定,通过样本数据自身进行拟合,使得模型具有自适应性。利用局部多项式回归算法建立模型,并根据数据特点修正模型,仿真实验表明,对于多种回归算法,基于非参数回归的航材消耗预测模型拟合效果更好、预测误差更小,具有较高的准确性。

参考文献:

- [1] 刘旭,周丽华,朱臣.基于PCA-SVM的航材需求预测方法研究[J].舰船电子工程,2017,37(07):105-109.
- [2] 汪娅,王超峰.基于约束调度的消耗性航材备件需求预测分析[J].科学技术与工程,2019,19(02):243-247.
- [3] 李文强,段振云,赵文辉.基于偏最小二乘模型的无人机航材需求预测方法[J].系统工程理论与实践,2018,38(05):1354-1360.
- [4] 陈博,徐常凯,任佳成.基于灰色神经网络的携行航材消耗预测[J].指挥信息系统与技术,2018,9(05):86-90.
- [5] 韩玉,张作刚,张海军.基于粗糙集和神经网络的舰载直升机航材消耗预测研究[J].舰船电子工程,2017,37(09):96-99.
- [6] 龙伟芳,叶绪国.未知误差分布下线性回归模型的非参数自适应估计[J].西北师范大学学报(自然科学版),2019,55(01):26-34.
- [7] 华祖林,韩爱秋.基于非参数回归改进的太湖总磷压力-响应模型[J].水资源保护,2019,35(01):20-4,46.
- [8] 李春雨. α -混合误差下非参数回归函数加权核估计的相合性[J].广西民族大学学报(自然科学版),2019,25(01):59-61.
- [9] 李庆,杨青龙.模型指导的单指标非参数期权定价[J].数理统计与管理,2018,37(06):1086-1094.
- [10] 戴秀菊,舒志彪.基于非参数核回归模型的隐含波动率预测[J].福州大学学报(自然科学版),2018,46(02):156-162.
- [11] 夏丽丽,田茂再.零-膨胀泊松回归模型的非参数统计分析及其应用[J].数理统计与管理,2019,38(02):235-246.
- [12] 许梁,孙涛,徐箭,等.基于函数型非参数回归模型的中

长期日负荷曲线预测[J].电力自动化设备,2015,35(07):89-94,100.

- [13] 姚海祥,李仲飞.基于非参数估计框架的期望效用最大化最优投资组合[J].中国管理科学,2014,22(01):1-9.
- [14] ASSAF A G, TSIONAS M. Non-parametric regression for hypothesis testing in hospitality and tourism research[J]. International Journal of Hospitality Management, 2019, 76(2):125-139.
- [15] CHAGNY G, ROCHE A. Adaptive estimation in the functional nonparametric regression model[J]. Journal of Multivariate Analysis, 2016, 146(4):23-31.
- [16] GIORDANO F, PARRELLA M L. Bias-corrected inference for multivariate nonparametric regression: Model selection and oracle property[J]. Journal of Multivariate Analysis, 2016, 143(1):165-181.
- [17] HLÁVKA Z, HUŠKOVÁ M, MEINTANIS S G. Tests for independence in non-parametric heteroscedastic regression models[J]. Journal of Multivariate Analysis, 2011, 102(4):215-232.
- [18] HU Y, YANG Y, WANG C, et al. Imputation in nonparametric quantile regression with complex data[J]. Statistics and Probability Letters, 2017, 127(4):111-135.
- [19] KRIEF J M. Direct instrumental nonparametric estimation of inverse regression functions[J]. Journal of Econometrics, 2017, 39(1):54-69.
- [20] MANGILI F. A prior near-ignorance Gaussian process model for nonparametric regression[J]. International Journal of Approximate Reasoning, 2016, 78(1):89-103.
- [21] MOREIRA C, UÑA-ÁLVAREZ J D, MEIRA-MACHADO L. Nonparametric regression with doubly truncated data[J]. Computational Statistics and Data Analysis, 2016, 93(3):129-141.
- [22] PARK C G, KIM I, LEE Y S. Error variance estimation via least squares for small sample nonparametric regression[J]. Journal of Statistical Planning and Inference, 2012, 142(8):239-251.
- [23] TUFFAHA H W, STRONG M, GORDON L G, et al. Efficient Value of Information Calculation Using a Nonparametric Regression Approach: An Applied Perspective[J]. Value in Health, 2016, 19(4):187-206.
- [24] WANG W, YU P. Asymptotically optimal differenced estimators of error variance in nonparametric regression[J]. Computational Statistics and Data Analysis, 2017, 105(1):63-87.
- [25] 张作刚,宋大勇,宫明慧,等.基于EEMD-SVM的不稳定需求航材消耗预测研究[J].环境技术,2018,36(05):70-73.