

基于核近邻非参数回归的航材消耗预测

薛永亮,陈振林

(海军航空大学,山东 烟台 264001)

摘要:针对某型直升机服役时间短、机型较新、备件采购周期长等特点,使用核近邻非参数回归对平稳型航材建立消耗预测模型。对比多种回归方法,证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

中图分类号:V267;E926

文献标志码:A

航材保障能力是形成作战能力、遂行多样化军事任务的物质基础,航材供应的及时性影响着航空部队战斗力的生成。航材预测是航材精细化保障的重要工作,已有较多学者进行了研究^[1-5]。对某型直升机航材消耗建立模型时,该机型较新、服役时间短,难以获取足够数据。同时,该机型备件国产化水平低,大部分要依靠进口,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。本文参考多种分类、预测方法,针对上述问题,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],在预测的过程中,不对模型做先验假设,按照算法约定探究数据自身规律,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]。根据航材消耗序列转化为消耗变异系数与序列突变系数^[14],将航材分为消耗平稳型航材、突变型航材、离散型航材。本文针对平稳型航材波动性、离散性较小的消耗特点,使用核近邻非参数回归建立消耗预测模型。

1 核近邻非参数回归

核近邻估计是最经典的非参数回归算法,由 Nadariyu 和 Watson 提出,它的核心算法为在估计 $f(x_0)$ 时,给予接近观测点 x_0 更高的权重。同时,给予远离观测点更低的权重^[15]。令:

$$z_i = (x_i - x_0)/h。 \quad (1)$$

式(1)中: z_i 表示第 i 个观测值 x 与聚焦点 x_0 之间存在正负区别的距离; h 表示估计带宽,通过调整核函数的宽度来控制非参数回归曲线的光滑程度。

当带宽 h 过大时,得到的回归函数会失去数据曲线的一些细节,导致回归估计过平滑。当带宽 h 过小

时,回归估计过于密切拟合数据,导致回归曲线粗糙,并且容易受到异常因素影响^[16]。因此,需要在光滑度与拟合度之间做有效的平衡,其方法就是选取适合的带宽 h 。定义一个核函数 $k(z)$,将最大的权重靠近焦点 x_0 的观测,而后随着 $|z|$ 的增长令权重对称、平滑降低。得到权重^[15]:

$$w_i = k\left[\frac{x_i - x_0}{h}\right]。 \quad (2)$$

计算可得:

$$\hat{f}(x_0) = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}, \quad (2)$$

式中, y_i 表示第 i 个月份航材消耗数。

2 基于可变带宽的核近邻非参数回归航材消耗模型

2.1 核函数选择

在核近邻估计算法中,常用的核函数包括:三角核函数、余弦核函数、beta族核函数、高斯核函数等^[16]。

核函数的选择依据是根据距离分配各个样本点不同贡献密度的密度值。文献[17]中指出,在核近邻回归中,核函数的选取对于回归模型拟合效果影响较小。但通过航材消耗预测实验,由于数据特点多样,对于不同类型的航材消耗数据,选用不同的核函数有较大影响,针对平稳型航材,利用 Quartic 核函数,回归曲线能较好拟合数据。Quartic 核表达式为:

$$k(t) = \begin{cases} \frac{15}{16}(1-t^2)^2 & |t| \leq 1 \\ 0 & \text{其他} \end{cases}。 \quad (4)$$

2.2 带宽选择

带宽 h 对于曲线拟合效果有较大影响,当 h 趋近于 0 时, $\frac{x-x_i}{h} \rightarrow \infty$, 使 $k\left(\frac{x-x_i}{h}\right)$ 趋近于 0, $f(x) \rightarrow y$, 拟合的曲线与样本数据接近, 曲线不光滑; 当 h 趋近于 ∞ 时, $\frac{x-x_i}{h} \rightarrow 0$, 使得 $k\left(\frac{x-x_i}{h}\right)$ 趋近于 1, $f(x) \rightarrow \frac{1}{n} \sum y_i = \bar{y}$, 拟合的曲线为水平直线。

当带宽 h 相对较小时, 曲线拟合较好, 但曲线的光滑性较差, 峰值谷值突出, 随机性影响较大, 出现过度拟合; 当带宽 h 相对较大时, 曲线的光滑度很高但平均化作用过于明显, 导致拟合不足, 无法描述曲线趋势性变化, 预测效果较差^[18]。

基于核近邻回归对平稳型航材建立消耗模型, 针对两端极化较大、边界带误差较大的问题, 设计带宽 h 为可变带宽, 在两端带宽较小, 使得曲线与实际值较为贴近, 在序列中端利用经典推论。例如, 最优拇指带宽或交叉验证最优带宽等。

利用高斯核函数, 将对称轴选在序列中间值, 其数学表达式为:

$$h = h_0 \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\left(\frac{x_i}{N/2} - 1\right)^2}{2}\right) \quad (5)$$

式(5)中: h_0 是拇指带宽或最优带宽计算结果; N 为样本容量。

设拇指带宽计算得出 $h_0 = 0.5$, 取某型航材 6 年消耗记录, 以月份为单位, $N = 72$, 可变带宽最大值最小值之比为 1.6, 为进一步强化带宽在序列两端的拟合效果, 将带宽公式调整为:

$$h = h_0 \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-1 \times \left(\frac{x_i}{N/2} - 1\right)^2\right) \quad (6)$$

带宽变化更为明显, 在序列两端带宽较小, 曲线拟合时, 更加接近样本数值。

3 仿真对比试验

由于多种因素, 某型直升机在役数量波动较大。为方便研究, 本文将机群消耗数量转变为单机消耗数。同时, 根据调研结果, 可以假定该机型每年训练量、训练环境相近, 因而可以假设其服役强度一致。某型航材 72 个月份消耗数据序列见表 1。

表 1 航材消耗序列表

Tab.1 Sequence table of material consumption

| 月份 | 数量 | 月份 | 数量 | 月份 | 数量 | 月份 | 数量 | 月份 | 数量 | 月份 | 数量 |
|----|------|----|------|----|------|----|------|----|------|----|------|
| 1 | 0.14 | 13 | 1.60 | 25 | 1.53 | 37 | 2.00 | 49 | 2.67 | 61 | 3.60 |
| 2 | 0.55 | 14 | 0.35 | 26 | 2.11 | 38 | 2.00 | 50 | 2.63 | 62 | 4.00 |
| 3 | 0.81 | 15 | 1.47 | 27 | 2.11 | 39 | 2.60 | 51 | 2.63 | 63 | 3.00 |
| 4 | 0.81 | 16 | 2.42 | 28 | 2.11 | 40 | 2.60 | 52 | 3.57 | 64 | 3.00 |
| 5 | 0.81 | 17 | 3.22 | 29 | 1.31 | 41 | 3.14 | 53 | 3.43 | 65 | 3.00 |
| 6 | 1.29 | 18 | 2.88 | 30 | 2.20 | 42 | 2.77 | 54 | 3.13 | 66 | 3.00 |
| 7 | 1.75 | 19 | 2.58 | 31 | 2.73 | 43 | 2.69 | 55 | 3.88 | 67 | 4.00 |
| 8 | 2.00 | 20 | 2.00 | 32 | 2.18 | 44 | 2.25 | 56 | 3.44 | 68 | 4.60 |
| 9 | 1.45 | 21 | 1.33 | 33 | 1.63 | 45 | 2.75 | 57 | 2.78 | 69 | 3.75 |
| 10 | 1.53 | 22 | 1.33 | 34 | 2.20 | 46 | 2.91 | 58 | 2.50 | 70 | 4.00 |
| 11 | 1.00 | 23 | 0.00 | 35 | 2.79 | 47 | 3.18 | 59 | 2.86 | 71 | 4.00 |
| 12 | 1.00 | 24 | 1.21 | 36 | 2.40 | 48 | 2.20 | 60 | 3.50 | 72 | 4.33 |

由表 1 可见, 该型航材消耗有一定波动, 整体呈上升趋势, 数据离散程度相对较小, 消耗序列突变不大。

为做充分对比, 分别利用拇指法, 则带宽与可变带宽非参数回归(Non-Parametric Regression)对某型平稳型航材建立消耗预测模型, 并对比支持向量机回归(Support Vector Regression)、最小二乘法(Least Squares)3 阶、4 阶等方法。为有效、精确评价模型预测效果, 利用以下 2 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]:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (7)$$

2) 均方根差 RMSE, 其数学表达式为^[10]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

式(7)、(8)中: y_i 为实际值; \hat{y}_i 为预测值。

仿真实验对比多种预测模型, 预测结果见表 2。利用 MAPE 与 MRSE 对预测效果进行对比, 如表 3 所示。利用基于可变带宽的核近邻非参数回归模型, 对比原回归方法, 仿真结果如图 1 所示。

表2 预测结果详表

Tab.2 Table of prediction results

| 月份 | 真实值 | NPR | LS(3) | LS(4) | 改进NPR | SVM |
|----|------|------|-------|-------|-------|------|
| 1 | 4.60 | 4.88 | 4.05 | 4.21 | 3.45 | 4.69 |
| 2 | 5.00 | 4.89 | 4.09 | 4.32 | 3.79 | 4.71 |
| 3 | 5.68 | 4.92 | 4.13 | 4.40 | 2.95 | 4.72 |
| 4 | 3.75 | 4.93 | 4.17 | 4.49 | 2.95 | 4.73 |
| 5 | 4.50 | 4.95 | 4.02 | 4.59 | 2.95 | 4.75 |
| 6 | 4.66 | 4.97 | 4.24 | 4.69 | 2.95 | 4.77 |
| 7 | 5.00 | 4.99 | 4.26 | 4.79 | 3.79 | 4.78 |
| 8 | 4.00 | 5.01 | 4.28 | 4.90 | 4.30 | 4.80 |
| 9 | 5.33 | 5.03 | 4.30 | 5.02 | 3.58 | 4.83 |
| 10 | 5.00 | 5.05 | 4.32 | 5.13 | 3.79 | 4.85 |
| 11 | 1.75 | 5.07 | 4.32 | 5.25 | 3.79 | 4.86 |
| 12 | 4.81 | 5.09 | 4.33 | 5.37 | 4.07 | 4.86 |

表3 预测结果评价表

Tab.3 Comparison table of prediction results

| 模型 | MAPE | MRSE |
|--------|---------|---------|
| NPR | 0.248 1 | 1.103 2 |
| LSM(3) | 0.261 4 | 1.166 2 |
| LSM(4) | 0.246 9 | 1.038 7 |
| SVM | 0.342 0 | 1.499 0 |
| 改进NPR | 0.228 6 | 1.027 9 |

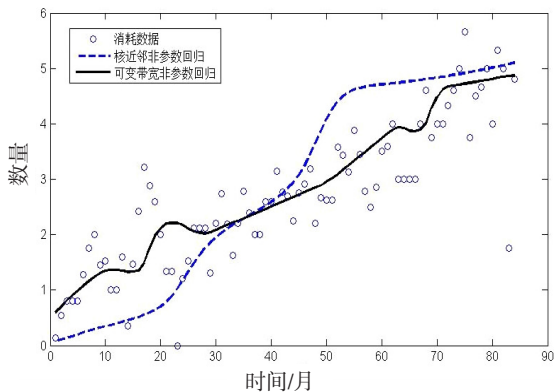


图1 非参数回归对比图

Fig.1 Contrast figure of nonparametric regression

由图1可以看出,核邻近非参数回归在序列两端偏差较大,受起始点位置影响较大。利用可变带宽,在两端带宽较小,使得回归曲线接近数据样本,对数据信息更为敏感,较好地对比原方法进行了优化。

利用改进的核邻近非参数回归模型对比参数回归模型3阶、4阶最小二乘算法,仿真结果见图2。

对于平稳型航材消耗数据,4阶最小二乘、非参数回归模型都能较好拟合消耗规律,预测航材趋势性变化,而3阶最小二乘曲线过于光滑,拟合效果不佳。

对比支持向量机回归,仿真结果见图3。

利用支持向量机回归建立模型时,消耗序列呈上升趋势,因而通过逐项训练对比,即前一数据为后一数据的训练集,得出拟合函数,再用后12项数据进行

预测,仿真效果不佳,未对航材趋势性变化做出有效预测。

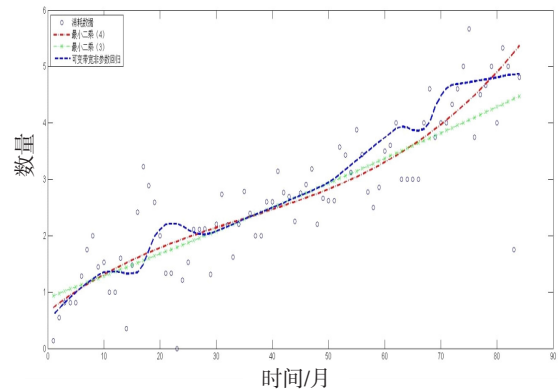


图2 回归算法对比图

Fig.2 Comparison figure of regression algorithm

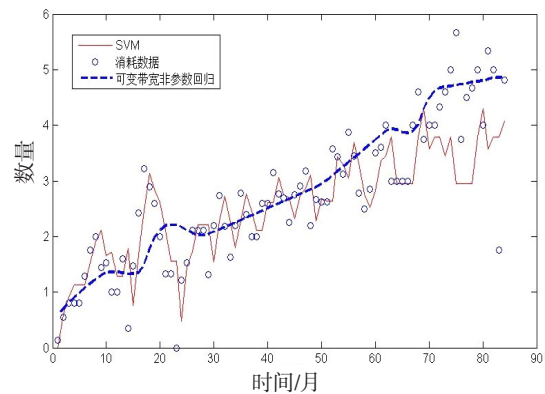


图3 支持向量机回归对比图

Fig.3 Comparison figure of support vector machine regression

4 结论

由于航材消耗相对平稳、呈上升趋势,使用核邻近非参数算法出现边界带误差较大、中位数据权重较大的问题。针对这一问题,基于高斯函数建立可变带,在两端带宽较小,使得曲线与实际值拟合更为紧密,减小中部数据权重较大的问题,对比多种预测模型,证明基于可变带宽的核邻近非参数回归对于平稳型消耗航材预测效果较好。

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(7): 105-109.
 LIU XU, ZHOU LI HUA, ZHU CHEN. Prediction of air material requirement based on PCA-SVM[J]. Ship Electronic Engineering, 2017, 37(7): 105-109. (in Chinese)
 [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(2): 243-247.
- WANG YA, WANG CHAOFENG. Demand forecast analysis of expendable aeronautical materials spare parts based on constraint scheduling[J]. Science Technology and Engineering, 2019, 19(2): 243-247. (in Chinese)
- [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(5): 1354-1360.
- LI WENQIANG, DUAN ZHENYUN, ZHAO WENHUI. Drones air-materials demands predictions method based on partial least squares model[J]. Systems Engineering Theory & Practice, 2018, 38(5): 1354-1360. (in Chinese)
- [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(5): 86-90.
- CHEN BO, XU CHANGKAI, REN JIACHENG. Consumption prediction of carried aviation material based on grey neural network[J]. Command Information System and Technology, 2018, 9(5): 86-90. (in Chinese)
- [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(9): 96-99.
- HAN YU, ZHANG ZUOGANG, ZHANG HAIJUN. Demand prediction of carried aviation materials for shipborne helicopters[J]. Ship Electronic Engineering, 2017, 37(9): 96-99. (in Chinese)
- [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(1): 26-34.
- LONG WEIFANG, YE XUGUO. Nonparametric adaptive estimation of linear regression models with the unknown error distribution[J]. Journal of Northwest Normal University: Natural Science, 2019, 55(1): 26-34. (in Chinese)
- [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(1): 20-24, 46.
- HUA ZULIN, HAN AIQIU. Improved stressor-response model of total phosphorus in Taihu Lake based on nonparametric regression[J]. Water Resources Protection, 2019, 35(1): 20-24, 46. (in Chinese)
- [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(1): 59-61.
- LI CHUNYU. Consistency of weighted kernel estimator of nonparametric regression functions under α -mixing errors[J]. Journal of Guangxi University for Nationalities: Natural Science, 2019, 25(1): 59-61. (in Chinese)
- [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(6): 1086-1094.
- LI QING, YANG QINGLONG. Single-index nonparametric adjusting of semi-parametric option pricing model[J]. Journal of Applied Statistics and Management, 2018, 37(6): 1086-1094. (in Chinese)
- [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(2): 156-162.
- DAI XIUJU, SHU ZHIBIAO. Implied volatility forecast based on nonparametric regression model[J]. Journal of Fuzhou University: Natural Science Edition, 2018, 46(2): 156-162. (in Chinese)
- [11] 夏丽丽, 田茂再. 零一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(2): 235-246.
- XIA LILI, TIAN MAOZAI. Nonparametric statistical analysis of zero-and-one inflated poisson regression models and its application[J]. Journal of Applied Statistics and Management, 2019, 38(2): 235-246. (in Chinese)
- [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(7): 89-94, 100.
- XU LIANG, SUN TAO, XU JIAN, et al. Auxiliary decision-making based on parallel mode multi-contingency for static voltage stability[J]. Electric Power Automation Equipment, 2015, 35(7): 89-94, 100. (in Chinese)
- [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(1): 1-9.
- YAO HAIXIANG, LI ZHONGFEI. Expected utility maximization optimal portfolio selection based on nonparametric estimation framework[J]. Chinese Journal of Management Science, 2014, 22(1): 1-9. (in Chinese)
- [14] 薛永亮, 陈振林. 基于消耗波动性聚类的航材分类研究[J]. 系统工程与电子技术, 2019, 41(12): 2802-2806.
- XUE YONGLIANG, CHEN ZHENLIN. Aviation material classification model based on consumption volatility clustering analysis[J]. Journal of System Engineering and Electronics, 2019, 41(12): 2802-2806. (in Chinese)
- [15] 孙志华, 尹俊平, 陈菲菲. 非参数与半参数统计[M]. 北京: 清华大学出版社, 2016: 23-31.

- SUN ZHIHUA, YIN JUNPING, CHEN FEIFEI. Nonparametric and semiparametric statistics[M]. Beijing: Tsinghua University Press, 2016: 23-31. (in Chinese)
- [16] 约翰·福克斯. 非参数回归: 平滑散点图[M]. 上海: 格致出版社, 上海人民出版社, 2015: 65-81.
- JOHN FOX. Nonparametric simple regression: smoothing scatterplots[M]. Shanghai: Gezhi Publishing House, Shanghai People's Publishing House, 2015: 65-81. (in Chinese)
- [17] CHAGNY G, ROCHE A. Adaptive estimation in the functional nonparametric regression model[J]. Journal of Multivariate Analysis, 2016, 146(4): 23-31.
- [18] SHEN HAIPENG, BROWN LAWRENCE, ZHAO LINDA, et al. Statistical analysis of a telephone call center: a queueing-science perspective[J]. Journal of the American Statistical Association, 2005, 100(469): 36-50.

Prediction of Aviation Material Consumption Based on Nonparametric Regression of Nuclear Nearest Neighbor

XUE Yongliang, CHEN Zhenlin

(Naval Aviation University, Yantai Shandong 264001, China)

Abstract: According to the characteristics of a certain type of cited helicopter, such as short service time, relatively new model and long spare parts procurement cycle, the non-parametric regression algorithm was used to establish the aviation material consumption model. Based on the classification of aviation materials, the consumption prediction model of stationary aviation materials was established by using nonparametric regression of nuclear nearest neighbor. By comparing various regression methods, it is proved that the model based on non-parametric regression has a good effect on interval prediction.

Key words: aviation material consumption; prediction; nonparametric regression

(上接第515页)

Design and Implement of Simulating Complete UAV Test General ATS

WANG Yiping, WEN Tianzhu, LIU Yong

(Naval Aviation University, Yantai Shandong 264001, China)

Abstract: How to test airborne equipment of UAV efficiently and precisely is becoming more and more urgent. A new auto test system (ATS) method, which onboard computer of UAV was used as part of ATS and test resource was used to simulate input and output signals and functions of airborne equipment was put forward by this paper. Data communications between onboard computer and other equipment of UAV could be developed to keep ATS with real work environment. And UAV BIT function test could be guaranteed normally. Multiple UAV models could be tested by configuring different UAV onboard computer and test accessory. Infield test and maintenance ability and UAV fighting effectiveness also could be improved effectively.

Key words: auto test system; simulating complete UAV test; onboard computer