

doi: 10.11823/j.issn.1674-5795.2023.02.02

起落架载荷标定模型的集成学习方法

倪天琦, 田永卫, 王勇, 齐贺阳, 辛育霞, 刘家旭

(航空工业北京长城计量测试技术研究所, 北京 100095)

摘要: 为解决飞机起落架载荷标定实验使用线性回归建立标定方程结果不理想的问题, 考虑到实验中起落架压缩行程和应变片布片位置等因素对标定载荷的非线性影响, 运用特征融合、集成学习理论, 通过使用 AdaBoost 和 XGBoost 非线性回归方法, 构建起落架载荷标定模型。首先, 通过起落架载荷标定实验获取实验数据, 使用主成分分析方法建立输入特征矩阵; 其次, 构建起落架载荷标定模型, 将起落架三向加载载荷分别作为标签向量, 训练集和测试集根据随机取样原则划分, 使用 AdaBoost 和 XGBoost 两种方法训练标定模型; 最后, 在测试集中对载荷进行拟合预测, 并使用均方根误差、平均绝对误差、决定系数、耗时4个评价指标对模型进行评估。实验结果显示, 与广泛使用的最小二乘法相比, XGBoost 方法建立的标定模型能够更好地拟合加载载荷, 在不考虑时效性的场景下 XGBoost 算法更具优势。研究结果对提高飞机起落架载荷实测准确性以及飞机结构健康监测的进一步研究具有重要价值。

关键词: 飞机起落架; 载荷标定; 非线性回归; 集成学习

中图分类号: TB9; V226

文献标志码: A

文章编号: 1674-5795(2023)02-0010-09

Ensemble learning method for landing gear load calibration model

NI Tianqi, TIAN Yongwei, WANG Yong, QI Heyang, XIN Yuxia, LIU Jiaxu

(Changcheng Institute of Metrology & Measurement, Beijing 100095, China)

Abstract: In order to solve the problem that the calibration equation established by linear regression is not ideal for the aircraft landing gear load calibration experiment, considering the nonlinear influence of factors such as landing gear compression stroke and strain gauge position on the calibration load in the experiment, the landing gear load calibration model is constructed by using AdaBoost and XGBoost nonlinear regression methods using feature fusion and ensemble learning theory. Firstly, experimental data are obtained through landing gear load calibration experiments, and the input feature matrix is established using principal component analysis method. Then, a landing gear load calibration model is constructed, using the three directional loading loads of the landing gear as label vectors. The training and testing sets are divided according to the random sampling principle, and the calibration model is trained using AdaBoost and XGBoost methods. Finally, the load is fitted and predicted in the testing set, and the model is evaluated using four evaluation indicators: root mean square error, average absolute error, determination coefficient and time. The experimental results show that compared with the widely used least squares method, the calibration model established by XGBoost method can bet-

收稿日期: 2023-04-10; 修回日期: 2023-04-21

基金项目: 航空工业计量所创新基金(ZC02102282)

引用格式: 倪天琦, 田永卫, 王勇, 等. 起落架载荷标定模型的集成学习方法[J]. 计测技术, 2023, 43(2): 10-18.

Citation: NI T Q, TIAN Y W, WANG Y, et al. Ensemble learning method for landing gear load calibration model[J]. Metrology & Measurement Technology, 2023, 43(2):10-18.



ter fit the loading load. XGBoost algorithm is more advantageous in scenarios without considering timeliness. The research results have important value for improving the accuracy of aircraft landing gear load measurement and further research on aircraft Structural health monitoring.

Key words: landing gear; load calibration; nonlinear regression; ensemble learning

0 引言

起落架是在滑行、起飞和着陆阶段提供悬挂支撑的系统,是飞机的重要受载结构,起落架的载荷谱编制对飞机结构健康监测有着重要意义。对起落架进行载荷标定实验是进行载荷谱编制的关键一步,在标定实验中需要建立加载载荷和应变码值之间的对应关系,即载荷标定模型。

国内外诸多专家学者对飞机载荷标定技术进行了一系列研究。郑甲宏^[1]等人研究了一种直升机载荷标定回归变量的优选方法,在使用多元线性方程进行载荷标定方程的确立时,考虑不同变量的选择对方程准确度的影响,采用逐个回归法将变量依次引入,对已选入的变量逐个检验,进而得到最优的自变量参数组合,该方法有效降低了标定方程的误差。黄祎丰^[2]等人提出了一种支柱式起落架载荷标定数据工程处理方法,分别使用2种标定数据回归方法:最小二乘法和反向回归法,保证了飞机起落架载荷谱实测准确性。Wang Jee-Ray^[3]等人使用灰色关联度法,通过选择具有代表性的变量构建传统载荷标定方程,并在仿真实验中验证了此方法的有效性。

目前国内外的载荷标定模型普遍采用多元线性回归方法确定,并且针对多元线性回归问题提出了不同改进方法,但由于在起落架受载过程中缓冲支柱压缩行程会影响传力,叠加应变片粘贴角度、加载载荷输出、信号传输误差等因素影响,在标定实验过程中加载载荷与应变码值之间呈现非线性关系^[4-6]。为确保起落架载荷标定模型与实测数据的准确可靠,对加载载荷和应变码值建立非线性载荷标定模型具有重要意义。

基于国内外现状,本文提出采用集成学习的2种方法分别构建非线性起落架载荷标定回归模型,并与当前载荷标定工程实践中较为常用的多元线性回归—最小二乘法进行对比实验。

1 起落架载荷标定实验

1.1 起落架载荷标定任务切面

飞机载荷实测项目中普遍采用应变计测量法测量飞机具体结构的受力情况。通过对飞机实际飞行过程中的受力分析,设计应变片布片位置。应变电桥桥路采用惠斯通电桥,惠斯通电桥可以将微小应变信号放大两倍,便于后续数据分析与处理。

对起落架施加载荷,针对某型号飞机,以左起落架载荷标定实验为例,考虑起落架在飞行过程中的实际受力情况,分别设计7个任务切面,具体位置如图1所示。标定模型的标签向量为左起落架法向载荷 P_z ,左起落架轮芯逆航向载荷 P_y ,左起落架逆航向接地点载荷 F_y 以及左起落架侧向载荷 F_x 。标定模型的输入特征有应变码值和起落架缓冲支柱压缩行程,其中应变码值包括起落架法向、航向、侧向三向载荷在7个任务切面的应变电桥输出码值和起落架撑杆处的应变电桥输出码值。

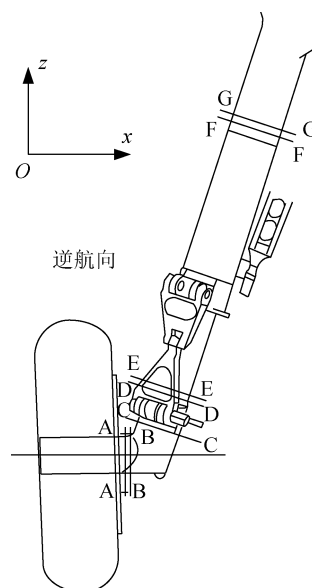


图1 逆航向起落架任务切面示意图

Fig.1 Schematic diagram of reverse heading landing gear mission section

上述载荷的单位为 kN, 压缩行程参数的单位为 mm, 应变码值为信号采集系统输出的应变电桥电信号编码值, 为无量纲数。

1.2 起落架载荷标定加载工况设计

起落架载荷标定实验的硬件设施由标定工装、液压加载装置、数据采集器等组成。对起落架施加载荷的工况设计考虑起落架实际工作状态, 模拟起落架实际受力情况, 对起落架缓冲支柱压缩行程设计为4种情况: 0, 50, 158, 180 mm。在这4种压缩行程下设计13种工况分别进行标定实验, 包括5种单向加载工况、4种双向加载工况、4种三向加载工况。工况1至工况5分别为法向、逆航向、航向、左侧向、右侧向单向加载工况; 工况6为法向和逆航向同时加载; 工况7为法向和航向同时加载; 工况8为法向和左侧向同时加载; 工况9为法向和右侧向同时加载; 工况10为法向、逆航向、左侧向同时加载; 工况11为法向、逆航向、右侧向同时加载; 工况12为法向、航向、左侧向同时加载; 工况13为法向、航向、右侧向同时加载。所施加载荷按照大小逐级加载, 根据不同工况设计下的载荷配比不同, 单向载荷、双向载荷和三向载荷的加载分辨力在5, 3, 1.5 kN之间不等。

1.3 输入特征矩阵的构建

在建立起落架载荷标定模型的过程中, 输入特征矩阵由12个应变测试参数特征和1个起落架缓冲支柱压缩行程参数组成。面对标定载荷 y , 普遍使用2种方法进行回归变量的选取: 一种为将全部应变测试参数特征作为输入特征矩阵进行计算^[7-8], 得到多元线性回归方程; 另一种为专家经验选元法回归^[9], 即根据结构实际受力分析以及力学基础选择相关应变测试参数, 进而纳入输入特征矩阵。两种方法各有优劣: 将全部应变测试参数作为输入特征矩阵虽然节省时间成本, 但是会把对标定载荷不敏感的远端测试参数也考虑进来, 增加标定方程的误差; 而第二种方法需要具备载荷谱测试专业知识的专家, 增加的选元步骤也会引入新的人为误差^[10]。为了提升载荷实测数据处理效率, 减少不同特征之间可能存在的重复或冗余信息的影响, 本文拟采用特征融合方法——主成分分析法 (Principle Component Analysis, PCA) 构建输入特征矩阵。

PCA通过计算输入特征矩阵的协方差矩阵, 采用方差最大化的方法, 对输入特征矩阵进行分解, 是一种无监督的线性变换方法。PCA寻找到高维输入特征矩阵的最大方差方向后, 取前几维较大方差方向进行重新投影以得到一个新的特征子空间。PCA的原理可以用图2进行解释, 假设有一组数据的空间分布如图所示, 通过对这组数据的协方差矩阵进行计算, 得到两个最大方差方向如图2中PC1和PC2所示, 在新的特征空间投影也是以PC1和PC2为基准。

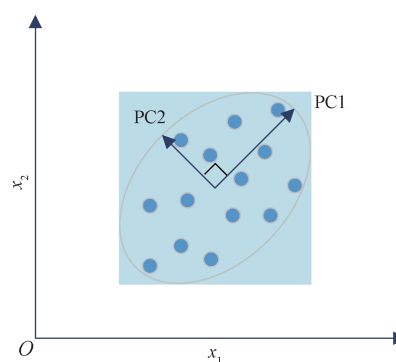


图2 主成分分析示意图

Fig.2 Diagram of principal component analysis

通过构建 $d \times k$ 维的变换矩阵 W , 把输入特征矩阵 x 映射到 k 维特征子空间 z , 其中 $d \geq k$, $\mathbb{R}$ 为时域空间, 则

$$x = [x_1, x_2, \dots, x_d], x \in \mathbb{R}^d \quad (1)$$

$$\downarrow xW, W \in \mathbb{R}^{d \times k} \quad (2)$$

$$z = [z_1, z_2, \dots, z_k], z \in \mathbb{R}^k \quad (3)$$

在PCA降维算法中, 输入特征矩阵所转换到的新特征子空间中的每个坐标轴彼此正交, 所体现的主成分为输入特征矩阵的协方差矩阵中的特征值, 通过对特征值进行排序, 选取前 k 个特征值进行新特征子空间的投影映射, 并且要求各个主成分之间彼此互不相关。

使用主成分分析法对输入特征矩阵进行特征融合, 实验发现当特征维数为6维时模型效果最好。表1为部分原始实验数据, 应变1~应变12的应变码值为信号采集系统输出的应变电桥电信号编码值, 为无量纲数。表2列出了使用主成分分析法得到的六维特征, 在新的空间中各个特征不具有物理意义, 为无量纲数。

表1 部分原始标定实验数据

Tab.1 Partial original calibration experimental data

法向载荷/kN	逆航向载荷/kN	侧向载荷/kN	压缩行程/mm	应变1	应变2	应变3	应变4	应变5	...	应变12
0.00	8.99	0.00	0	17	-111	30	946	169		1 296
14.46	-5.97	-5.97	0	-179	760	529	-645	-501		-1 154
15.13	8.95	0.00	50	93	577	277	916	-499		1 578
20.12	11.95	-0.00	50	133	742	348	1 208	-643		1 536
25.03	15.03	-0.03	158	196	1 015	512	1 751	-1 056		2 567
30.08	18.05	-0.03	158	320	1 250	646	2 050	-1 153		3 117
35.12	-13.96	-0.06	158	147	1 655	546	-1 830	-1 335		-1 085
40.13	23.99	-0.06	180	279	1 637	786	2 737	-1 649		3 939
45.12	27.00	-0.06	180	349	1 827	1 827	3 019	-1 859		4 503
54.17	15.31	-0.07	180	319	2 329	1 036	1 642	-2 207		3 196

表2 使用主成分分析法后的输入特征矩阵

Tab.2 Features after dimensionality reduction using principal component analysis

特征1	特征2	特征3	特征4	特征5	特征6
336.656	1 402.256	1 811.873	-1 538.391	103.507	22.219
1 439.102	1 928.764	2 119.468	-1 586.197	108.724	16.411
2 537.675	2 467.948	2 413.714	-1 588.252	231.412	54.513
...	...	...	...	...	...
8 745.782	-3 915.313	132.869	1 456.923	-199.436	-266.584
10 178.694	-4 395.555	24.160	1 799.599	-219.598	-219.598
11 570.452	-4 939.394	-67.077	2 155.709	-219.243	-272.799

2 集成学习非线性回归模型

建立起落架载荷标定模型时,输入特征为起落架测试切面的应变参数以及缓冲支柱压缩行程,与加载载荷之间具有复杂的非线性关系,无法用简单的线性模型进行描述,使用非线性回归方法能够建立更准确的数学模型。而仅使用一种非线性回归方法建立的载荷标定模型不具有较强的泛化性能,容易产生过拟合。集成学习构建多个具有差异化的弱回归模型,运用优化策略得到一个强回归模型,在保证弱回归模型多样性的同时,得到的结果比单个回归模型的结果更好。

比较简单的模型被称为弱学习器。集成学习的思想就是将一系列弱学习器组合成一个预测效果更好的复杂模型(或称为强学习器),如图3所示。建立非线性回归模型的过程中,常用的集成学

习方法包括 AdaBoost 和 XGBoost 等方法。

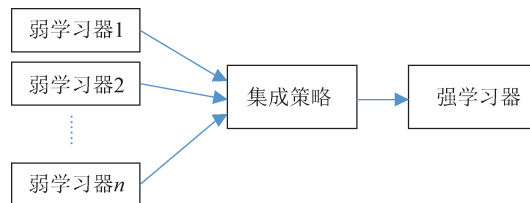


图3 集成学习思路示意图

Fig.3 Schematic diagram of ensemble learning ideas

2.1 AdaBoost算法

最初的集成学习算法由 Schapire 于 1989 年提出,并进行了试验和理论性的证明^[11]。在此之后, Freund 研究出一种更高效的自适应提升集成算法^[12],被称为 Adaboost,其作为集成学习算法中最为典型的算法,在多种数据挖掘的应用场景下展现出十分优秀的能力^[13-14]。

面对起落架载荷标定实验的载荷-码值数据集 D , 由 m 个样本组成

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\} \quad (4)$$

式中: $\mathbf{x}_i \in R^d$, y_i 为样本 $\mathbf{x}_i$ 的标签值。AdaBoost 回归算法的具体步骤如下:

①初始化权重。记初始状态下的数据及样本分布为 $Dist_1$, 对每一个样本 $\mathbf{x}_i$ 的权重初始化为 $1/m$, 则 $Dist_1(\mathbf{x}_i) = 1/m$ 。 $Dist_1$ 分布用于第一个弱分类器 h_1 的训练, 其他同理。

②循环进行 T 轮迭代, 记每一轮迭代中弱学习器的编号为 t , 且 $t \in \{1, 2, 3, \dots, T\}$ 。以该步骤作为循环体, 循环体中的步骤进一步细分为:

在样本分布为 $Dist_t(\mathbf{x})$ 的基础上, 在数据集 D 上训练弱学习器 h_t ; 计算学习器 h_t 在训练集 D 上的最大误差 E_t 为

$$E_t = \max |y_i - h_t(\mathbf{x}_i)|, i = 1, 2, \dots, m \quad (5)$$

式中: $h_t(\mathbf{x}_i)$ 为弱学习器 h_t 对样本 $\mathbf{x}_i$ 的预测结果, y_i 为样本 $\mathbf{x}_i$ 的目标数值。

根据求得的 h_t 的最大误差 E_t , 计算 h_t 对每个样本的相对误差为

$$e_{it} = \frac{(y_i - h_t(\mathbf{x}_i))^2}{E_t^2}, i = 1, 2, \dots, m \quad (6)$$

根据求得的样本相对误差 e_{it} , 计算出当前弱学习器 h_t 的误差率为

$$e_t = \sum_{i=1}^m Dist_t(\mathbf{x}_i) e_{it} \quad (7)$$

对于样本 $\mathbf{x}_i$, 更新权重的计算公式为

$$w_i = \frac{e_t}{1 - e_t} \quad (8)$$

更新数据集样本的权重分布

$$Dist_{t+1}(\mathbf{x}_i) = \frac{Dist_t(\mathbf{x}_i)}{Z_t} \omega_i^{1-e_{it}} \quad (9)$$

式中: Z_t 为归一化因子, 计算公式为

$$Z_t = \sum_{i=1}^m Dist_t(\mathbf{x}_i) \omega_i^{1-e_{it}} \quad (10)$$

令迭代次数 t 加 1, 返回循环体初始步骤。

$$Obj^{(t)} = \sum_{i=1}^N L(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \approx \sum_{i=1}^N \left[L(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(X_i) + \frac{1}{2} h_i f_i^2(X_i) \right] + \Omega(f_t) + \text{const} \quad (14)$$

式中: g_i 和 h_i 分别为损失函数 L 在第 i 个样本点上关于第二个变量的一阶和二阶偏导数; $\hat{y}_i^{(t-1)}$ 为经

结束 T 轮迭代, 最终得到强学习器为

$$H(x) = \sum_{i=1}^m \ln \left(\frac{1}{\omega_i} \right) f(x) = \left[\sum_{i=1}^m \ln \left(\frac{1}{\omega_i} \right) \right] f(x) \quad (11)$$

式中: $f(x)$ 为所有 $\omega_i h_i(x)$ ($t = 1, 2, \dots, T$) 的中位数, 即所有弱学习器的加权输出结果的中位数。

2.2 XGBoost 算法

XGBoost 算法是 2016 年华盛顿大学陈天奇教授提出的一种在 GBDT 算法基础上优化的算法, 其主要思想是通过不断地残差拟合, 生成多个弱学习器, 最后将生成的弱学习器累加起来得到一个强学习器, 并使得损失函数最小化^[15]。XGBoost 算法在优化过程中将损失函数进行二级泰勒展开, 引入二阶导数信息, 与 GBDT 算法只使用一阶导数相比, 模型在训练过程中的收敛速度更快。此外, XGBoost 还通过在损失函数中添加正则化项来抑制模型复杂度, 防止过拟合现象发生^[16]。与传统的梯度提升算法相比, XGBoost 算法可以实现更加快速的梯度提升, 被广泛应用于各种机器学习任务中。

XGBoost 算法使用回归树作为弱学习器, 算法模型主要由以下几个方面构成:

1) 规定算法的目标函数

XGBoost 算法通过对目标函数增加正则项, 对每棵树的复杂度进行惩罚来防止模型的过拟合现象。XGBoost 算法的目标函数为

$$Obj = \sum_{i=1}^N L(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(f_t) \quad (12)$$

$$\Omega(f_t) = \gamma M + \frac{1}{2} \lambda \sum_{m=1}^M c_m^2 \quad (13)$$

式中: $L(y, \hat{y})$ 为损失函数; $\Omega(f_t)$ 为刻画树复杂度的正则项; T 为迭代次数; f_t 为第 t 个 CART 回归树; M 为该树的叶子节点的个数; c_m 为第 m 个叶子节点上的输出值; γ 和 λ 为惩罚系数。

在对目标函数进行极小化训练时, 使用损失函数的二阶泰勒展开近似原损失函数为

过 $t-1$ 次迭代后强学习器的预测值; const 为与 f_1 无关的常数部分。

在具体训练过程中, 可将树 f_i 拆分成结构部分 q 和输出部分 $\vec{c}$ 为

$$f_i(X) = \vec{c}_{q(X)}, \vec{c} \in R^M, q: R^K \rightarrow \{1, 2, \dots, M\} \quad (15)$$

式中: q 为一个映射, 它将每个样本点 $X \in R^K$ 映射到某一个叶子节点上; $\vec{c} = (c_1, c_2, \dots, c_M)$ 为一个 M 维向量, 其分量 c_m 对应第 m 个叶子节点上的输出值, 符号 $\vec{c}_m$ 为取其第 m 个分量。

2) 确定树的最优结构

在已经划分的叶子节点基础上寻找新分割点, 达到目标函数的最小化, 新的分割点需要依据分割增益定义, 分割增益 $Gain$ 定义为

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (16)$$

式中: $G_L^2/(H_L + \lambda)$ 为引入分割后左子树的分数; $G_R^2/(H_R + \lambda)$ 为右子树的分数; $(G_L + G_R)^2/(H_L + H_R + \lambda)$ 为不引入分割情况下的分数; γ 为引入新分割对模型复杂度的惩罚系数。同 GBDT 算法一样, 在对叶节点进行分割时可以采用贪心算法, 计算所有候选特征和其所有可能的切分点对应的增益 $Gain$, 选取增益最大的特征和切分点进行分割, 确定树的结构 q 。

3) 最小化目标函数, 得到最优输出值向量

在树的结构 q 确定后, 计算该结构下对应的最优输出值 $\vec{c}$ 的表达式。将 f_i 和 Ω 的表达式代入近似的目标函数式 (12) 中, 忽略与 f_i 无关的常数部分后得到的目标函数是 M 个相互独立的变量 c_m 的二次函数之和。通过求导可以求得极小点和目标函数的极小值, 这 M 个极小值点构成了该结构 q 所对应的最优输出值向量 $\vec{c}$ 为

$$\begin{aligned} c_m^* &= -\frac{G_m}{H_m + \lambda} \\ Obj^{(t)*} &= -\frac{1}{2} \sum_{m=1}^M \frac{G_m^2}{H_m + \lambda} + \gamma M \end{aligned} \quad (17)$$

2.3 AdaBoost 和 XGBoost 的比较分析

AdaBoost 与 XGBoost 的主要差别在于: AdaBoost 每轮学习的弱学习器通过改变样本的权值, 关注上轮回归误差较大的样本的权值, 以逐步减少在训练集上的误差率, 算法最终生成的强学习

器是多个弱学习器线性组合; 而 XGBoost 每轮学习的弱学习器是通过改变输出值, 每轮拟合的值为真实值与已有的加法模型的残差。AdaBoost 能够有效降低误差, 在泛化性能非常弱的学习器上构建很强的集成, 缺点是对噪声敏感。XGBoost 对损失函数进行了二阶泰勒展开, 关注到了损失函数的二阶导数信息, 能够快速收敛, 并且在目标函数中添加正则化项, 降低弱学习器的复杂程度, 避免模型产生过拟合问题。

3 实验与结果分析

3.1 模型的建立

以起落架载荷标定实验中的左主起落架法向载荷 P_z 为例, 选取实验过程中的左主起落架侧向 AFG 切面、航向 CFG 切面、法向 ABDE 切面、左主起落架撑杆和左主起落架缓冲支柱压缩行程等应变测试参数作为数据集, 使用主成分分析法进行特征融合处理, 并构建载荷标定模型的输入特征矩阵。其中左主起落架法向载荷 P_z 作为回归模型中的标签变量。在对实验数据进行基本的预处理操作后, 数据集共有 2 793 条有效数据。

将数据集按照 6 : 4 的比例随机分为训练集和测试集, 划分好的数据集统计信息如表 3 所示。从训练集与测试集的统计学信息中可以得出, 在最大最小值、平均值、偏度和峰度 4 个维度下两个数据集相似, 数据集的划分较为合理。

表 3 训练集和测试集左起垂向载荷的统计学信息
Tab.3 Statistical information on the left vertical load of training and testing sets

数据集	样本数	最大~最小值	平均值	偏度	峰度
训练集	1 675	60.25 ~ -0.03	27.34	-0.14	-0.78
测试集	1 118	60.25 ~ -0.02	26.56	-0.09	-0.81

图 4~图 6 为部分应变参数与法向载荷的数据散点图。可以看到, 不同应变参数与法向载荷的关系较为明显, 应变参数 4 和应变参数 9 都与法向载荷具有一定的非线性关系, 而应变参数 12 的线性关系更突出。

3.2 模型调参

使用 AdBoost 和 XGBoost 两种非线性集成学习

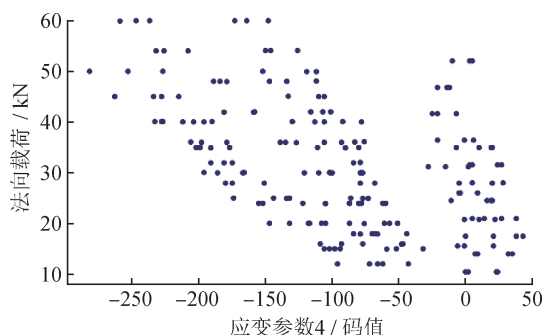


图4 法向载荷与应变参数4的数据散点图

Fig.4 Scatter plot of data for normal load and strain parameter 4

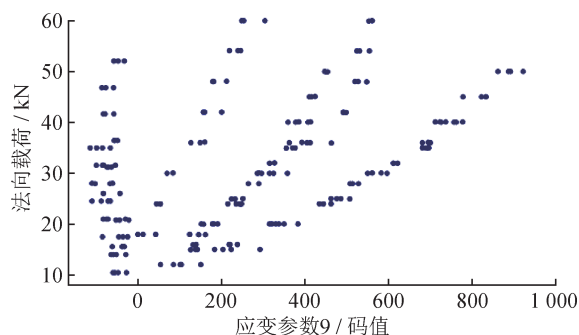


图5 法向载荷与应变参数9的数据散点图

Fig.5 Scatter plot of data for normal load and strain parameter 9

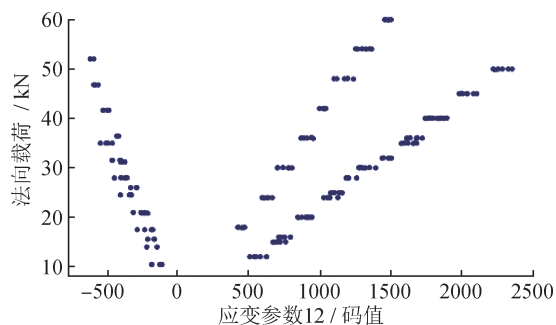


图6 法向载荷与应变参数12的数据散点图

Fig.6 Scatter plot of data for normal load and strain parameter 12

算法对标签变量左起落架法向载荷构建回归模型,为了衡量所使用模型的回归性能,采用广泛应用的线性回归方法—最小二乘法进行对比实验。

调节参数数值是模型构建的一个重要过程,一个模型往往通过指定算法和相应的参数得以表征。AdBoost算法是在决策树模型的基础上进行迭代更新的,即弱学习器为决策树,所涉及的参数主要有最大迭代次数 $n_estimators$, 学习率 $learning_rate$ 和损失函数 $loss$ 。XGBoost算法可以指定弱学习器类型,作为其中一个参数 $booster$, 其他主要参数有最大迭代次数 $n_estimators$, 学习率 $learn-$

ing_rate , 指定叶节点进行分支所需的损失减少的最小值 $gamma$, L1 正则化权重项 $alpha$, 树的最大深度 max_depth , 指定叶节点中最小的样本权重和 min_child_weight 。模型参数如表4所示。

表4 AdBoost和XGBoost的模型参数

Tab.4 Model parameters for AdBoost and XGBoost

模型	参数	参数值
AdBoost	$n_estimators$	60
	$learning_rate$	0.3
	$loss$	'linear'
XGBoost	$booster$	'gbtree'
	$n_estimators$	50
	$learning_rate$	0.1
	$gamma$	0
	max_depth	8
	min_child_weight	1

3.3 结果分析

按表4的模型参数对训练集数据进行训练,将训练得到的模型使用测试集数据进行测试,得到的实验结果如表5所示。使用均方根误差 $RMSE$ 、平均绝对误差 MAE 、决定系数 R^2 、耗时4个评价指标对模型进行定量度量,其中 $RMSE$ 和 MAE 的值越小,表示模型拟合效果越好, R^2 越接近于1,表示回归拟合值越接近于真实值,效果越好。

表5 载荷标定模型的评价指标比较

Tab.5 Comparison of evaluation indicators for load calibration models

对比模型	$RMSE$	MAE	R^2	耗时 / ms
最小二乘法	1.283	0.941	0.993	6.02
AdaBoost	0.779	0.605	0.997	217
XGBoost	0.463	0.258	0.999	86.3

从表5中可以看出,在测试集中3个载荷标定模型的 R^2 均大于0.99,模型拟合效果较好,但非线性回归模型AdaBoost和XGBoost的拟合效果要更加优于线性回归方法最小二乘法,其中XGBoost模型的前3个评价指标均优于AdaBoost和最小二乘法,结果表现最好。另外在耗时方面,最小二乘法的耗时6.02 ms,时间最短,XGBoost模型所用

时间为 86.3 ms, 牺牲了一定的效率来提升准确率, AdaBoost 模型的耗时最长, 用时为 217 ms, 但是模型效果不如 XGBoost 理想。

由于起落架载荷标定模型使用了多个解释变量, 所以无法在二维图中可视化回归效果, 但可以通过绘制残差与预测值来判断回归模型。残差图是判断回归模型常用的图形工具。本文所用三种方法的残差对比见图 7。在最理想的情况下, 残差均刚好为零, 但在现实实验中, 残差越接近中心线模型拟合效果越好, 同时考察回归模型的误差是否呈随机分布。从图 7 中可以看出三种模型的残差均随机分布在中心线附近。最小二乘法残差分布范围偏大, 残差值在 $[-4, 4]$ 之间; AdaBoost 的残差值分布相较于最小二乘法更为收敛; XGBoost 的残差值分布最集中在中心线零线, 表现效果最好。

图 8~图 10 给出了最小二乘法、AdaBoost、XGBoost 三种方法在测试集中预测值与真实值的差异。XGBoost 方法优于 AdaBoost 和最小二乘法, 能较好地拟合样本数据。

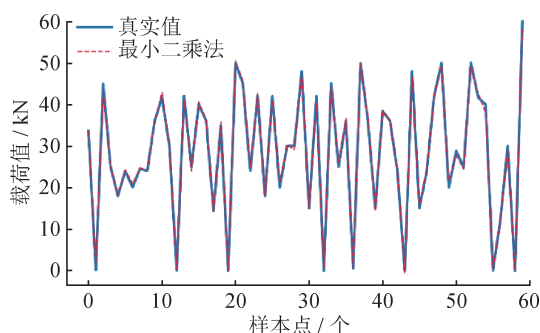


图 8 最小二乘法回归模型

Fig.8 Least squares regression model

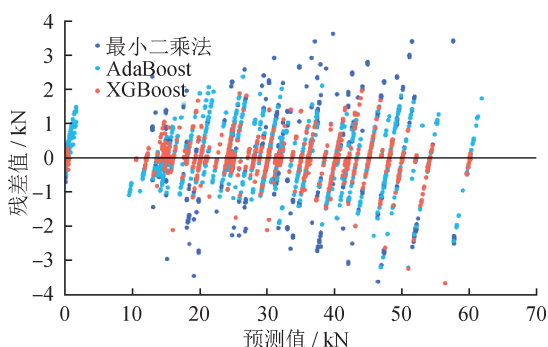


图 7 三种模型的残差图

Fig.7 Residual plots of three models

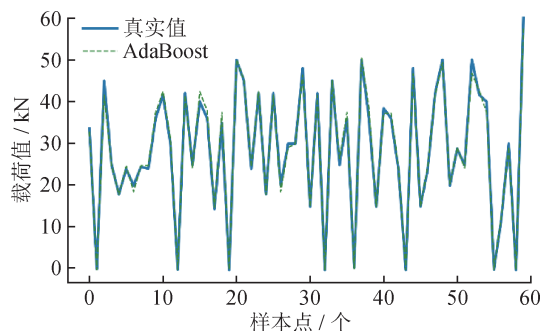


图 9 AdaBoost 回归模型

Fig.9 AdaBoost regression model

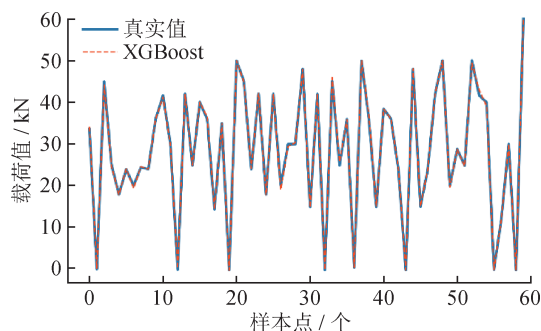


图 10 XGBoost 回归模型

Fig.10 XGBoost regression model

4 结束语

研究了起落架载荷标定模型的构建方法, 介绍了 AdaBoost 和 XGBoost 算法的基本原理, 以某型号飞机左起落架垂向载荷的标定数据为研究对象, 对输入特征矩阵进行主成分分析法特征融合处理, 建立起落架载荷标定模型, 并与最小二乘法进行对比。

对于本文中的数据样本, 测试集的 MAE 能够降低到 0.258, RMSE 能够降低到 0.463, 决定系数能够达到 0.999。相比于最小二乘法和 AdaBoost 算法, XGBoost 能够识别相关应变码值、起落架压缩行程与左起落架垂向载荷之间的耦合关系, 对于样本回归具有较好的效果, 但其运行算法所需时间较最小二乘法更长, 在不考虑时效性的场景下 XGBoost 算法更具优势。

参考文献

- [1] 郑甲宏, 王泽峰, 沈雳. 直升机载荷标定方程回归变量的优选方法[J]. 科学技术与工程, 2017, 17 (24): 299-303.
ZHENG J H, WANG Z F, SHENG L. Optimization method for regression variables of helicopter load calibration

- tion equation [J]. Science Technology and Engineering, 2017, 17(24): 299-303. (in Chinese)
- [2] 黄祎丰, 裘进浩, 刘克格. 支柱式起落架载荷标定数据工程处理方法[J]. 振动、测试与诊断, 2020, 40(2): 248-254, 416.
- HUANG Y F, QIU J H, LIU K G. Engineering treatment methods for load calibration data of telescopic land gear [J]. Journal of Vibration, Measurement & Diagnosis, 2020, 40(2): 248-254, 416. (in Chinese)
- [3] WANG J R, JAU J S, LEE J H, et al. Apply grey relational grade in optimal airplane landing gear loads calibration [C]// International Conference on Intelligent & Advanced Systems. IEEE, 2008.
- [4] 侯乔乔, 张清勇. 一种改进的载荷标定试验数据回归算法[J]. 工程与试验, 2018, 58(4): 22-24.
- HOU Q Q, ZHANG Q Y. An improved regression algorithm for load calibration test data [J]. Engineering & test, 2018, 58(4): 22-24. (in Chinese)
- [5] 齐婵颖, 李育超, 岳健, 等. 基于飞行载荷的大数据实测方法研究[J]. 计算机测量与控制, 2017, 25(7): 66-69.
- QI C Y, LI Y C, YUE J, et al. Research on big data test technology of flight load [J]. Computer Measurement & Control, 2017, 25(7): 66-69. (in Chinese)
- [6] 周强, 易晖. 光纤光栅传感器在直升机旋翼桨叶载荷标定上的应用[J]. 电子技术与软件工程, 2020(3): 94-96.
- ZHOU Q, YI H. Application of fiber Bragg grating sensors in load calibration of helicopter rotors and blades [J]. Electronic Technology & Software Engineering, 2020(3): 94-96. (in Chinese)
- [7] 韩聪玲, 刘克格, 刘彦鹏. 非线性回归算法在飞机载荷标定试验中的应用[J]. 工程与试验, 2022, 62(4): 31-35.
- HAN C L, LIU K G, LIU Y P. Application of nonlinear regression algorithm in aircraft load calibration test [J]. Engineering & Test, 2022, 62(4): 31-35. (in Chinese)
- [8] 田立勇, 戴渤海, 王启铭. 基于采煤机摇臂销轴多应变数据融合的煤岩识别方法[J]. 煤炭学报, 2020, 45(3): 1203-1210.
- TIAN L Y, DAI B H, WANG Q M. Coal rock identification method based on multi strain data fusion of coal mining machine rocker pin shaft [J]. Journal of China Coal Society, 2020, 45(3): 1203-1210. (in Chinese)
- [9] 段垚奇, 刘克格, 赵丽娜, 等. 飞机载荷标定试验数据的多元回归选元方法[J]. 振动、测试与诊断, 2015(1): 178-182, 197.
- DUAN Y Q, LIU K G, ZHAO L N. Multiple regression method for selecting elements of aircraft load calibration test data [J]. Journal of Vibration, Measurement & Diagnosis, 2015(1): 178-182, 197. (in Chinese)
- [10] 刘彦鹏, 刘克格, 张书明, 等. 有起落架布置的双三角翼载荷测试技术研究[J]. 振动、测试与诊断, 2016, 36(5): 1003-1008, 1031.
- LIU Y P, LIU K G, ZHANG S M, et al. Research on load test technology of double delta wing with landing gear arrangement [J]. Journal of Vibration, Measurement & Diagnosis, 2016, 36(5): 1003-1008, 1031. (in Chinese)
- [11] SCHAPIRE R E. The strength of weak learnability [J]. Proceedings of the Second Annual Workshop on Computational Learning Theory, 1989, 5(2): 197-227.
- [12] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Computational Learning Theory, 1995, 904: 23-37.
- [13] WYNER A J, OLSON M, BLEICH J, et al. Explaining the success of AdaBoost and random forests as interpolating classifiers [J]. Journal of Machine Learning Research, 2017, 18: 1-33.
- [14] TIEN B D, HO T C, PRADHAN B, et al. GIS-based modeling of rainfall-induced landslides using data mining-based functional trees classifier with AdaBoost, Bagging, and MultiBoost ensemble frameworks [J]. Environmental Earth Sciences, 2016, 75(14): 1-22.
- [15] CHEN T, GUESTRIN C, XGBoost. A scalable tree boosting system [C]// KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 785-794.
- [16] 倪天琦. 集合预报洋面阵风订正研究[D]. 天津: 天津大学, 2020.
- NI T Q. Research on correction of ensemble forecast of ocean gust [D]. Tianjin: Tianjin University, 2020. (in Chinese)

(本文编辑: 田艳玲)



第一作者: 倪天琦 (1996—), 女, 助理工程师, 硕士, 主要研究方向为动力学测试与诊断的算法应用、软件开发。