



基于 XGBoost 的多源气动数据融合建模

林枫 海春龙 梅立泉

Multi-source aerodynamic data fusion modeling with XGBoost

LIN Feng, HAI Chunlong, MEI Liqun

引用本文:

林枫, 海春龙, 梅立泉. 基于 XGBoost 的多源气动数据融合建模[J]. 空气动力学学报, 2024, 42(7): 27–34. DOI: 10.7638/kqdlxxb-2023.0066

LIN Feng, HAI Chunlong, MEI Liqun. Multi-source aerodynamic data fusion modeling with XGBoost[J]. *Acta Aerodynamica Sinica*, 2024, 42(7): 27–34. DOI: 10.7638/kqdlxxb-2023.0066

在线阅读 View online: <https://doi.org/10.7638/kqdlxxb-2023.0066>

您可能感兴趣的其他文章

Articles you may be interested in

飞行仿真气动数据机器学习建模方法

Machine learning method for aerodynamic modeling based on flight simulation data

空气动力学学报. 2019, 37(3): 488–497 <https://doi.org/10.7638/kqdlxxb-2019.0024>

工程翼型气动特性数据挖掘与建模

Modeling and data mining of engineering airfoil aerodynamic characteristics

空气动力学学报. 2021, 39(6): 175–183 <https://doi.org/10.7638/kqdlxxb-2021.0219>

机器学习方法在气动特性建模中的应用

Applications of machine learning for aerodynamic characteristics modeling

空气动力学学报. 2019, 37(3): 470–479 <https://doi.org/10.7638/kqdlxxb-2019.0033>

月地高速再入返回器气动辨识数据偏差分析

Deviation analysis on aerodynamic identification data for lunar–earth high–speed reentry capsule

空气动力学学报. 2018, 36(5): 871–877 <https://doi.org/10.7638/kqdlxxb-2018.0140>

嵌入式大气数据传感系统的模糊逻辑建模方法

Fuzzy logic algorithm for flush air data sensing system

空气动力学学报. 2019, 37(3): 412–418 <https://doi.org/10.7638/kqdlxxb-2016.0140>

基于实验数据同化的湍流模型常数标定: 含滤网蒸汽阀门通流特性数值预测

Calibration of turbulence model constants using measurement data assimilation: prediction of steam valve flow characteristics with filter

空气动力学学报. 2021, 39(2): 12–22 <https://doi.org/10.7638/kqdlxxb-2020.0183>

地址: 四川省绵阳市二环路南段8号11信箱9分箱

电话: 0816-2463375

Email: kqdlxxb@163.com



关注微信公众号

获得更多资讯信息

文章编号: 0258-1825(2024)07-0027-08

基于 XGBoost 的多源气动数据融合建模

林 枫, 海春龙, 梅立泉*

(西安交通大学 数学与统计学院, 西安 710049)

摘 要: 飞行试验和计算流体力学(CFD)是气动数据获取的两种常用手段, 其中, 通过飞行试验获取的气动数据精度高, 但成本高昂, 而通过 CFD 获取气动数据成本低, 数据量丰富。因此, 为了以尽可能低的成本获取精度更高的气动数据, 本文利用数据融合方法, 提出了一种可以融合大量的由 CFD 获取的气动数据与少量由飞行试验获取的气动数据的嫁接 XGBoost (eXtreme Gradient Boosting)集成模型框架。首先, 利用充足的 CFD 气动数据建模, 作为融合框架中的低精度模型; 然后, 将少量飞行试验参数输入低精度模型得到相应的气动输出; 最后, 结合飞行试验气动数据的其他特征进行二次建模, 实现对真实气动参数的预测。为了证明所提算法的有效性, 本文设置了相关的对比实验, 结果表明: 1) 多源气动数据融合建模较单源气动数据建模具有更高的预测精度; 2) 集成学习模型比传统的机器学习模型具有更强的泛化性能, 其中 XGBoost 模型泛化能力最强, 能够实现对气动参数的准确预测。

关键词: 多源气动数据; XGBoost; 数据融合; 集成学习

中图分类号: V211.3 文献标识码: A doi: 10.7638/kqdlxxb-2023.0066

Multi-source aerodynamic data fusion modeling with XGBoost

LIN Feng, HAI Chunlong, MEI Liqun*

(School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Flight tests and CFD (Computational Fluid Dynamics) are frequently used to obtain aerodynamic data. The aerodynamic data obtained by flight tests are usually highly accurate but expensive, while those obtained by CFD are relatively less accurate, but cheap and abundant in data. Therefore, to obtain aerodynamic parameters with higher accuracy at the lowest possible cost, this paper proposes a grafted XGBoost (eXtreme Gradient Boosting) integrated model framework, combining aerodynamic data obtained by CFD and flight tests for fusion modeling. First, a large amount of CFD aerodynamic data are used to establish a low-precision model in the fusion modeling framework regarding the trend of aerodynamic data. Then, a small number of design variables of flight tests are put into the low-precision model to obtain the corresponding aerodynamic output. Finally, the secondary modeling is carried out by integrating the information from other flight test aerodynamic data characteristics to predict the exact aerodynamic output. In order to verify the effectiveness of the proposed algorithm, relevant comparative experiments are conducted. The results show that: 1) Multi-source aerodynamic data fusion modeling has higher prediction accuracy than its single-source counterparts; 2) The integrated learning model has a better generalization performance than traditional machine learning models, among which the XGBoost model has the best generalization ability and can realize the accurate prediction of aerodynamic parameters.

Keywords: multi-source aerodynamic data; XGBoost; data fusion; ensemble learning

收稿日期: 2023-05-08; 修订日期: 2023-06-30; 录用日期: 2023-06-05; 网络出版时间: 2023-08-15

作者简介: 林枫(1998—), 男, 江西新余人, 硕士研究生, 研究方向: 气动数据建模与数据挖掘. E-mail: llinmaple@163.com

通信作者: 梅立泉*(1969—), 男, 教授, 研究方向: 偏微分方程数值解与计算物理. E-mail: lqmei@mail.xjtu.edu.cn

引用格式: 林枫, 海春龙, 梅立泉. 基于 XGBoost 的多源气动数据融合建模[J]. 空气动力学学报, 2024, 42(7): 27–34.

LIN F, HAI C L, MEI L Q. Multi-source aerodynamic data fusion modeling with XGBoost[J]. Acta Aerodynamica Sinica, 2024, 42(7): 27–34(in Chinese). doi: 10.7638/kqdlxxb-2023.0066

0 引言

目前空气动力学的研究和数据获取,主要依赖3种技术手段:飞行试验、风洞试验和数值模拟。其中,飞行试验可以获得真实的气动数据,但成本十分高昂,所以获取的试验数据十分有限^[1];风洞试验可以获得较为精确的气动数据,但其成本依旧过高,且与真实的气动参数存在一定的误差^[2];数值模拟通过求解物理方程获得气动参数结果^[3],极大节省了获取成本,但其精度有限,尤其在高超声速条件下,计算结果往往会出现失真,影响了其可用性^[4]。因此,如何在保证数据准确性的前提下,降低试验和计算的成本,是空气动力学领域备受关注的研究方向^[5]。

近些年来,数据融合技术被广泛应用于多源气动数据的建模中。He等^[6]通过深度神经网络(deep neural network, DNN)算法融合多源气动数据,提出了一种低成本生成更多高质量气动数据的有效方法。Lin等^[7]针对多保真度(multi-fidelity, MF)数据融合问题,提出了一种融合高保真度(high-fidelity, HF)和低保真度(low-fidelity, LF)响应以及相应梯度的贝叶斯高斯过程的回归模型,降低了建模复杂度,提高了模型的预测精度。Zhang等^[8]提出了一种基于协同代理的数据融合技术,证明了多个保真度水平的数据集可以显著提高飞行性能分析的精度。Pham等^[9]提出了一种基于多保真度数据建模的三层次Kriging法,证明了所提方法比现有的层次Kriging法有更好的预测性能。Kou和Zhang^[10]提出了一种基于数据融合和机器学习等多种技术的非定常多保真度气动建模框架,该方法仅需使用少量高保真度数据和低保真度模型,就可以准确再现高保真度结果。王旭等^[11]基于飞行试验数据与风洞试验数据,提出了一种基于随机森林^[12]的数据融合框架,并证明该模型可以提升气动数据的预测精度。邓晨等^[13]通过对气动数据融合算法的研究,发现基于融合算法的模型有更好的预测效果。陈海等^[14]提出了一种基于卷积神经网络(convolutional neural network, CNN)的图形化预测方法,该方法能够基于翼型图像准确预测出翼型气动系数。王文正等^[15]利用K-means聚类与最近邻算法等数据挖掘技术,建立了基于飞行试验数据的气动模型,通过与地面试验结果的比较分析,验证了辨识结果的一致性和可信度,证明了该方法的可行性与有效性。

在已有的基于多源气动数据融合建模的研究中,大多都使用单一模型进行建模分析,使用集成学习思想的研究较少,因此本文考虑将XGBoost(eXtreme

Gradient Boosting)集成算法用于气动参数的预测。为了提升预测精度,本文提出了一种融合CFD气动数据与少量飞行试验数据的嫁接XGBoost模型框架。该框架实现了对真实气动参数有效且合理的预测,为以更少的成本生成更准确的气动数据提供了一种可行方案。

1 CFD气动数据与飞行试验数据

以某对称性飞行器模型的飞行试验数据作为本文研究的实验数据。飞行试验总时长约400 s,收集了目标飞行器飞行时轴向力系数 C_x 、法向力系数 C_y 以及侧向力系数 C_z 等3个气动参数随马赫数 Ma 、迎角 α 、侧滑角 β 、俯仰控制量 δ_z 、偏航控制量 δ_y 以及滚转控制量 δ_x 等6个变量变化的数值结果。而CFD气动数据是通过目标飞行器的几何模型进行数值计算获取的,计算时的 Ma 、 α 、 β 、 δ_z 、 δ_y 以及 δ_x 等6个变量与目标飞行器飞行状态下的保持一致。飞行试验气动数据与CFD气动数据的对比如图1~图3所示。

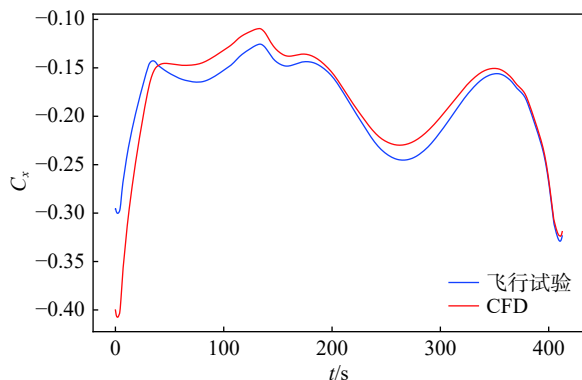
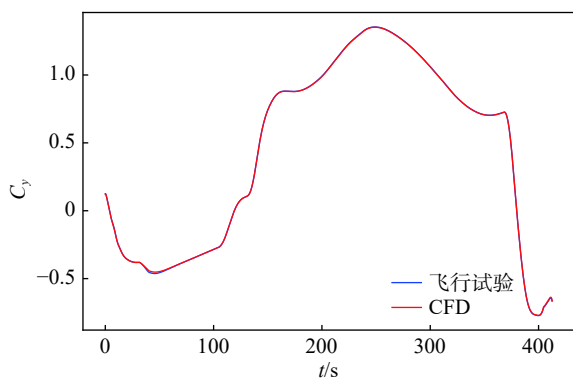
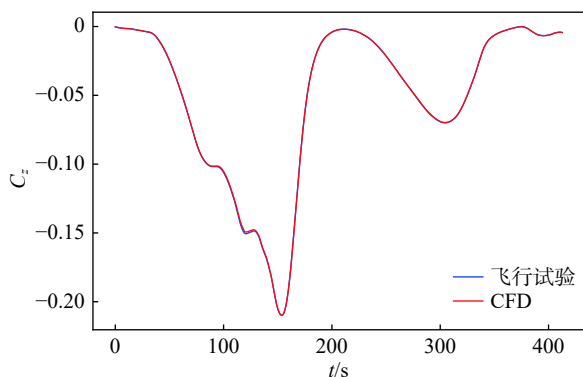
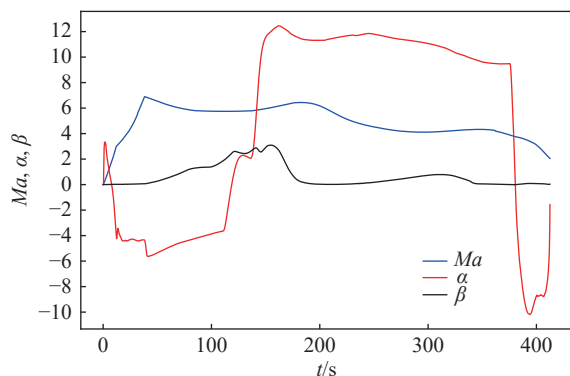


图1 CFD与飞行试验的 C_x 对比图

Fig. 1 Comparison of C_x between CFD and flight test

由图1~图3可以看出,CFD气动数据与飞行试验数据在 C_y 和 C_z 两个数值上的变化趋势一致,且差异极小,但在 C_x 上存在显著差异,说明由CFD数值计算得到的气动数据对飞行试验的 C_x 的拟合效果还有待提升。因此,本文将着力于研究应用融合模型对轴向力系数 C_x 的预测效果。

此外,如图4所示,在飞行过程中, α 的变化幅度最大,而 Ma 与 β 的变化则相对平滑,未出现较大波动。 α 的大幅波动主要在3个时段:飞行器启动阶段、100~150 s飞行时间段以及飞行结束阶段。这说明在这3个时段飞行器受到的干扰较大,从而导致迎角出现了较大的波动。由于传感器灵敏度与精度的

图 2 CFD 与飞行试验的 C_y 对比图Fig. 2 Comparison of C_y between CFD and flight test图 3 CFD 与飞行试验的 C_z 对比图Fig. 3 Comparison of C_z between CFD and flight test图 4 飞行过程中 Ma 、 α 以及 β 的变化曲线图Fig. 4 Variations of Ma , α and β during flight

限制, 相关结果可能会存在一定的误差。

2 集成学习

不同学习器对数据集的偏差和噪声有不同的敏感性, 所以单个学习器在不同的初始参数设置下训练的模型可能不是最优的。而集成学习可以将多个弱学习器集成为一个强学习器, 将它们组合起来能够降低整体模型的方差和偏差^[16], 提高模型的准确性和稳定性, 从而提高模型的泛化性能。

XGBoost 模型是目前具有代表性的集成模型之

一, 其在工业领域有着广泛的应用^[17-21], 模型本身具有很强的鲁棒性和稳健性, 于是, 本文考虑将其应用于轴向力系数 C_x 的预测中。

2.1 XGBoost 算法

XGBoost 算法是一种高效、灵活、可扩展的梯度提升决策树集成算法, 由 Chen 和 Guestrin^[22] 于 2014 年提出。算法在进行决策树集成时采用了 Boosting 的建模思想, 逐一建立多棵回归决策树, 并且下一棵树的建立方式依赖于上一棵树的评估结果, 即每棵决策树的输出是相互关联的, 每次加入新的决策树需要考虑已有决策树的误差, 从而不断迭代地提高模型的预测精度。在所有决策树建好之后, 对于回归问题, 是将所有决策树的预测值相加求平均得到最终的预测结果; 而对于分类问题, XGBoost 模型最后的输出结果则是基于每棵树的输出概率分布进行加权平均, 通常采用的方法是选取概率最大的类别作为最终的分类结果, 其模型架构如图 5 所示。

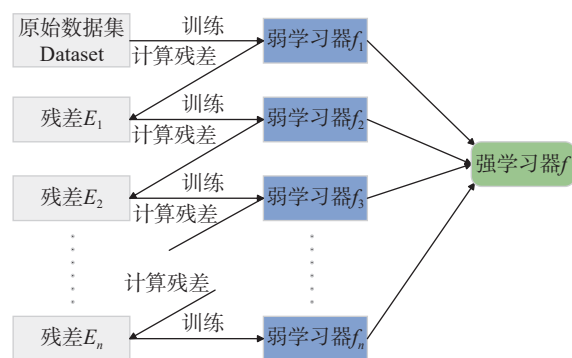


图 5 XGBoost 模型架构图

Fig. 5 Architecture of XGBoost

XGBoost 的建模流程如下:

1) 初始化: 将训练数据中所有样本的预测值都设为 0, 作为模型的初始值, 用来训练第一棵决策树, 并将其作为初始的弱分类器 f_1 。

2) 计算残差: 根据当前模型的输出结果 $\hat{y}_i$, 计算每个样本的残差, 即当前样本真实值 y_i 与当前模型的预测值 $\hat{y}_i$ 之间的误差 E_i 。

3) 计算梯度和 Hessian 矩阵: 通过计算残差的一阶梯度 g_i 和二阶导数 h_i , 分别得到每个样本的梯度 G_i 和 Hessian 矩阵 H_i 。

4) 计算损失函数: 将当前模型的输出结果 $\hat{y}_i$ 和样本的真实值 y_i 代入损失函数中计算误差, 并根据误差计算模型的损失函数。

$$L^{(t)} = \sum_{j=1}^T \left[G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2 \right] + \gamma T \quad (1)$$

其中, $L^{(t)}$ 为第 t 棵决策树的损失函数, ω_j 为第 j 个叶

子结点的权重, T 为模型叶子节点的数目, 叶子节点越少模型越简单, λ 与 γ 为模型的正则项系数, 用于控制模型的复杂度。

5) 计算新的决策树: 根据样本的梯度 G_i 、Hessian 矩阵 H_i , 以及当前模型的损失函数, 生成一棵新的决策树 f_i , 并将其加入现有决策树集合 $\{f_1, f_2, \dots, f_{i-1}\}$ 中。

6) 加权更新: 为了防止过拟合, XGBoost 采用加权更新的方式, 即为每个样本分配一个权重, 根据其误差大小来调整权重, 使误差大的样本得到更高的权重, 从而使得模型更多关注难以预测的样本。

7) 模型集成: 将新生成的决策树 f_i 和现有的决策树集合 $\{f_1, f_2, \dots, f_{i-1}\}$ 进行集成, 得到新的模型输出。

8) 迭代更新: 重复执行步骤 2~7, 直到模型误差达到预设阈值或者达到最大迭代次数为止。

2.2 多源气动数据融合的 XGBoost 模型

CFD 气动数据与飞行试验气动数据本质上是流体控制方程的解的不同精度近似, 因此使用 CFD 气动数据建立的模型, 可以为真实气动数据建模提供信息。为了充分利用这种信息以达到更好的数据融合效果, 设计了一种可以同时结合两种数据源的 XGBoost 嫁接集成建模方法, 该方法旨在利用低精度模型的趋势信息对建模过程进行引导, 同时结合少量飞行试验气动数据作为高精度模型的输入, 以达到更好的结果, 最终实现对真实气动参数的预测, 图 6 给出了该方法的模型架构图。

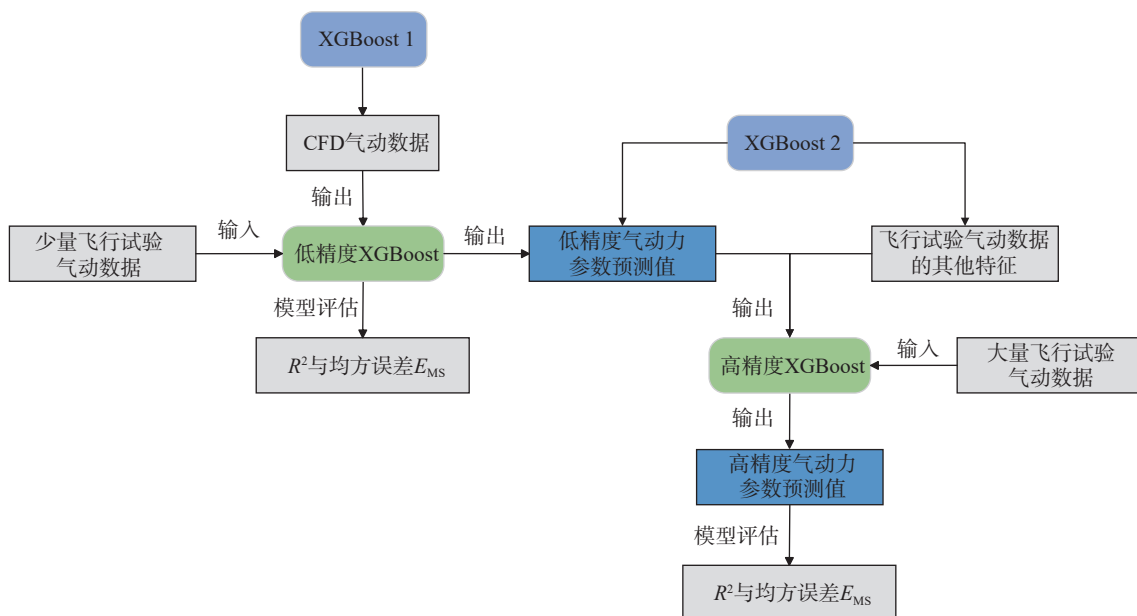


图 6 多源气动数据融合的 XGBoost 嫁接模型架构图

Fig. 6 Architecture of the XGBoost grafting model for multi-source aerodynamic data fusion

该方法的具体实现步骤如下:

1) 以 CFD 气动数据的 Ma 、 α 、 β 以及 δ_z 、 δ_y 、 δ_x 这 6 个参数作为低精度 XGBoost 模型的输入, 以 CFD 气动数据的 C_x 作为输出, 建立低精度的 XGBoost 模型。

2) 为了结合 CFD 气动数据的信息对 C_x 进行预测, 将少量飞行试验的 6 个参数输入低精度的 XGBoost 模型, 从而得到输出的低精度轴向力系数 $C_{x,low}$, 并将其作为后续高精度 XGBoost 模型构建的特征之一。

3) 结合飞行试验气动数据的 6 个特征参数和 $C_{x,low}$ 作为高精度 XGBoost 模型的输入进行二次建模, 实现多源气动数据融合的目的, 最终输出飞行试验气动数据的 C_x 。

4) 为了评估模型的泛化能力, 将大量的飞行试

验气动数据输入高精度 XGBoost 模型得到高精度 C_x 的预测值, 通过计算真实值与预测值之间的 R^2 (拟合优度) 评估模型拟合数据的能力, 均方误差 E_{MS} 用于衡量模型的预测误差。

3 基于 XGBoost 的多源气动数据融合建模

基于决策树的 Bagging 和 Boosting 等集成学习模型建立最优决策树是通过寻找最优特征分裂点实现的, 因此树模型是阶跃的, 对于特征参数取值的大小并不敏感, 所以树模型样本点的数值缩放不影响分裂点的位置, 对树模型的结构也不造成影响, 因此不需要对数据进行归一化或标准化处理。

3.1 基于 XGBoost 的多源气动数据融合建模过程

首先, 将 CFD 气动数据随机划分成了训练集与测试集, 划分比例为 9:1, 训练集用于低精度 XGBoost 模型的训练, 测试集用于评估模型对 CFD 气动数据的拟合程度和泛化性能。

在 XGBoost 模型中, 特征重要性的计算是通过特征在所有决策树中的分裂次数和分裂增益来评估的, 通过特征重要性可以了解每个特征对模型预测结果的影响程度。由图 7 的特征重要性柱状图可以看到, Ma 对于低精度 XGBoost 模型的贡献分是最大的, 表明 Ma 对模型预测结果产生的影响最大, δ_z 和 α 次之。

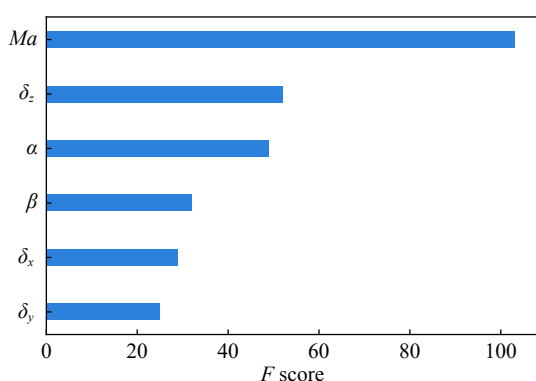


图 7 低精度 XGBoost 的特征重要性

Fig. 7 Feature importance of low-precision XGBoost

其次, 为了分析低精度 XGBoost 模型对 CFD 气动数据的拟合效果与泛化性能, 绘制了图 8 所示的散点回归图, 其中横轴代表 C_x 预测值, 纵轴代表 C_x 的 CFD 真实数值解, 蓝色点为低精度 XGBoost 模型在测试集上的 C_x 预测值。可以看到, 几乎所有测试数据的预测值都分布在了 $y = x$ 这条直线上, 这说明模型的预测值与 CFD 数值解几乎相等。模型在训练集与测试集上的 R^2 分别为 0.9987 和 0.9963, 均方误

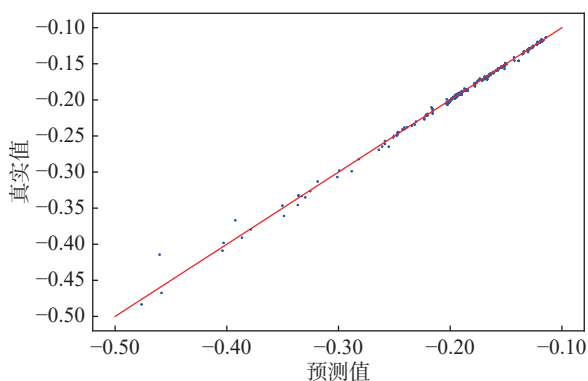


图 8 低精度 XGBoost 的预测值与真实值的散点回归图

Fig. 8 Scatter regression of predicted values from low-precision XGBoost

差 E_{MS} 分别为 3.47×10^{-6} 和 1.08×10^{-5} , 说明没有出现拟合现象, 低精度 XGBoost 对于 CFD 气动数据的拟合效果和泛化性能较好。

低精度 XGBoost 模型建立之后, 结合两种气动数据的信息进行融合建模, 以实现飞行试验气动数据的高效预测。首先, 将飞行试验气动数据按训练集与测试集 1:9 的比例进行随机划分, 仅用 1/10 的飞行试验数据就能更好地体现高精度数据稀疏的特点以及数据融合建模的意义。为了与多源气动数据融合建模保持一致, 单源气动数据的建模仅使用飞行试验数据, 按同比例对数据进行了随机划分。此外, 在模型训练时均采用了五折交叉验证^[23], 降低模型出现过拟合的风险。

为了分析模型训练的收敛情况, 绘制了模型在训练集与验证集上的损失变化曲线图, 如图 9、图 10 所示, 其中横坐标为模型的迭代次数, 纵坐标为均方根误差 E_{RMS} 。由图 9 可以看到, 单源 XGBoost 模型在迭代 20 次左右时, 训练集和验证集的 E_{RMS} 均降至 0 左右, 但训练集和验证集之间的损失还存在一定的差距, 说明模型存在过拟合的风险。而由图 10 可以看到, 高精度 XGBoost 模型同样在训练大约 20 次时, 训练集与验证集的 E_{RMS} 都降至 0 附近, 但与单源气动数据建模的结果相比, 训练集与验证集的差异更小, 模型出现过拟合的风险更小。

为了解每个特征对单源 XGBoost 模型预测结果的影响程度, 绘制了特征重要性图如图 11 所示。由图可以看到, 单源气动数据建模时, β 与 Ma 对模型的贡献最大, 其次是 δ_y , 且 3 个特征的贡献分相差不大, 说明 3 个特征对模型预测结果的影响程度几乎一致。与单源气动数据建模不同, 由图 12 所示的高精度 XGBoost 模型的特征重要性图可以看到, 低精度 XGBoost 模型输出的轴向力系数 $C_{x,low}$ 对模型的贡献

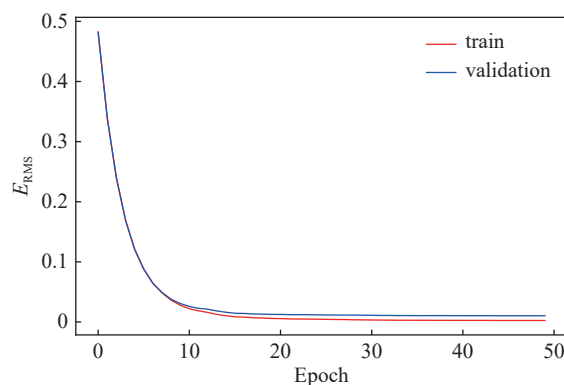


图 9 单源 XGBoost 在训练集与验证集上的损失变化曲线图

Fig. 9 Iterative loss variations of single-source XGBoost for the training and verification sets

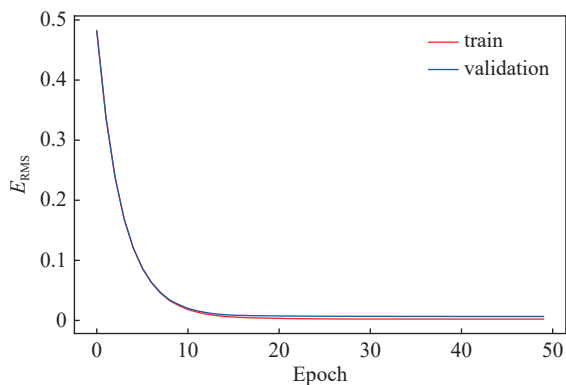


图 10 高精度 XGBoost 在训练集与验证集上的损失变化曲线图

Fig. 10 Iterative loss variations of high-precision XGBoost for the training and verification sets

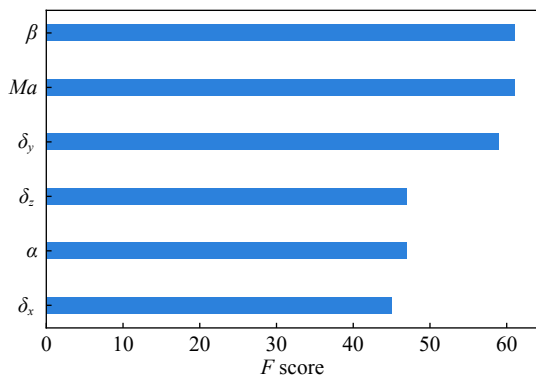


图 11 单源 XGBoost 的特征重要性

Fig. 11 Feature importance of single-source XGBoost

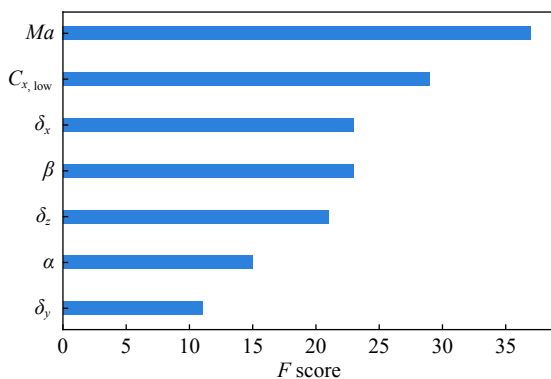


图 12 高精度 XGBoost 的特征重要性

Fig. 12 Feature importance of high-precision XGBoost

仅小于 Ma , 比其他特征的贡献分都要高, 表示其对模型预测结果产生了较大影响, 对模型的构建也起到了关键性的作用, 说明 CFD 气动数据可以为高精度 XGBoost 模型带来较大的信息增益。

3.2 结果分析与讨论

本节展示并描述了两部分的实验结果: 第一部分是多源气动数据建模与单源气动数据建模的结果对

比; 第二部分是传统机器学习算法引入本文所提的嫁接模型框架进行的多源气动数据融合建模的结果。

3.2.1 单源气动数据与多源气动数据融合建模的结果对比

在模型完成训练之后, 为了分析模型的拟合程度以及预测精度, 将飞行试验气动数据的测试集输入 XGBoost 模型进行测试, 绘制了图 13、图 14 所示的散点回归图。

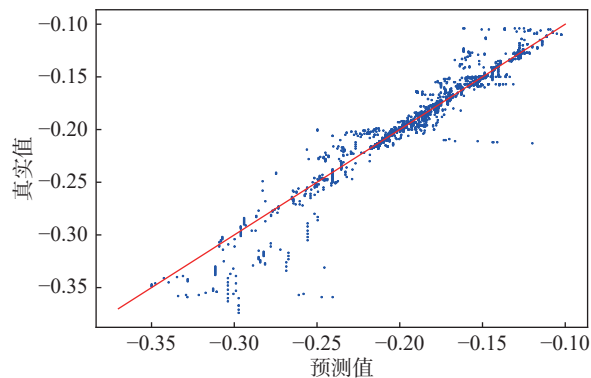


图 13 单源 XGBoost 的预测值与真实值的散点回归图

Fig. 13 Scatter regression of predicted values from single-source XGBoost

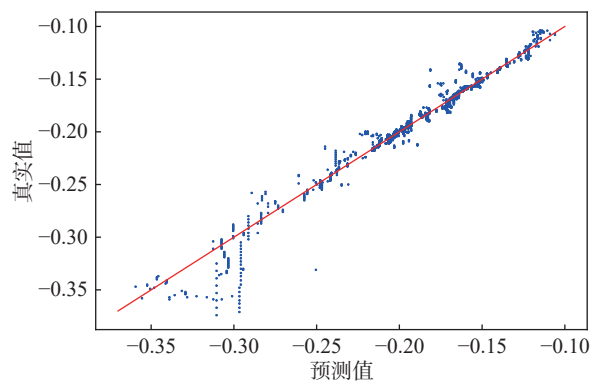


图 14 高精度 XGBoost 的预测值与真实值的散点回归图

Fig. 14 Scatter regression of predicted values from high-precision XGBoost

图 13 中的结果表明, 单源 XGBoost 模型在测试集上的大部分预测值与真实值出现了较大偏差, 数据点的分布较为分散, 模型预测效果不理想。而由图 14 中高精度 XGBoost 模型在测试集上的预测结果可以看到, 针对轴向力系数 C_x , 除少量数值较小 (小于 -0.3) 的数据点的预测误差较大外, 其余点的预测值与真实值几乎均匀地分布在 $y=x$ 这条直线上。回顾图 1 可以发现, C_x 数值较小的这部分点分布在飞行器启动与结束阶段, 飞行器在这两个阶段的飞行稳定性较弱, 受到的干扰噪声较大, 从而导致了模型在这附近的预测精度较低。在飞行器的稳定飞行阶段, 模

型对于 C_x 的预测值与真实值几乎相等, 可见多源 XGBoost 模型的预测效果是比较理想的。

为了进一步分析模型对 C_x 的预测效果, 将 C_x 的预测值、真实值与 CFD 的数值解进行了可视化分析, 如图 15、图 16 所示。可以看到 CFD 数值解只有少部分与试验真实值重合, 大部分与试验真实值存在较大差异。图 15 中单源 XGBoost 模型得到的预测值曲线波动较大, 未能很好地逼近真实值曲线, 而由图 16 可以看到高精度 XGBoost 模型的 C_x 预测值几乎与真实值重合, 说明基于 XGBoost 的多源气动数据融合模型较为精确地预测了轴向力系数 C_x , 模型的预测效果与泛化性能比单源气动数据模型更为理想。

3.2.2 传统机器学习算法的多源气动数据融合建模

为了验证 XGBoost 模型的优越性, 本文还使用线性回归、支持向量回归、随机森林进行了多源气动数据的嫁接学习。为了减轻量纲或数值对模型结果的影响, 建模前已对数据进行了标准化处理。采用不同算法在 CFD 气动数据和飞行试验气动数据上的预测效果如表 1 所示。

表 1 中的结果显示, 无论是在 CFD 气动数据还是飞行试验气动数据上, XGBoost 模型都有着更强的拟合能力和更高的泛化性能, 在飞行试验气动数据测试集上的 E_{MS} 达到了 1×10^{-5} , 比其他模型都有着更高的预测精度。此外, 由表 1 还可以看出, 随机森林与 XGBoost 模型在飞行试验气动数据测试集上的 E_{MS} 分别为 1.498×10^{-4} 与 4.404×10^{-5} , 而支持向量机

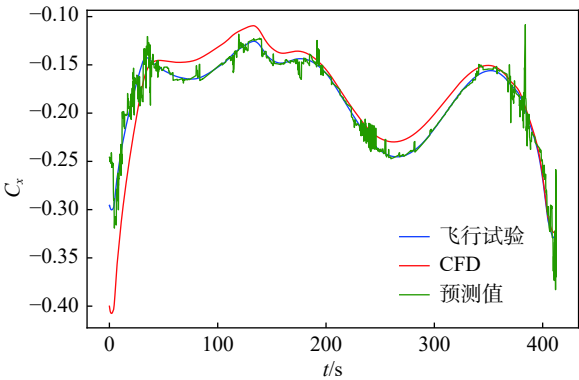


图 15 单源气动数据建模结果

Fig. 15 Results of single-source aerodynamic data modeling

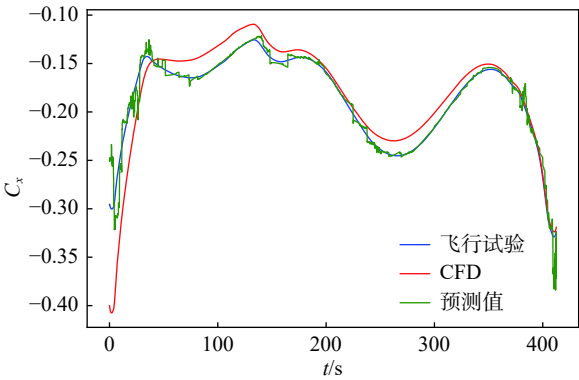


图 16 多源气动数据融合结果

Fig. 16 Results of multi-source aerodynamic data fusion

与线性回归模型在测试集上的 E_{MS} 为 3.977×10^{-2} 与 2.091×10^{-1} , 这说明集成模型要比传统机器学习算法的泛化能力更强, 可以得到更准确的预测结果。

表 1 不同模型在 CFD 气动数据以及飞行试验气动数据上的预测结果

Table 1 MSE and R-square of CFD and flight test aerodynamic data obtained by different models

模型	CFD 气动数据				飞行试验气动数据			
	训练集 E_{MS}	训练集 R^2	测试集 E_{MS}	测试集 R^2	训练集 E_{MS}	训练集 R^2	测试集 E_{MS}	测试集 R^2
随机森林	1.487×10^{-5}	0.994 48	2.039×10^{-5}	0.993 05	9.979×10^{-5}	0.936 54	1.498×10^{-4}	0.911 31
多源 XGBoost	3.470×10^{-6}	0.998 71	1.079×10^{-5}	0.996 32	5.170×10^{-6}	0.996 71	4.404×10^{-5}	0.973 93
支持向量机	2.806×10^{-2}	0.971 67	2.329×10^{-2}	0.978 45	2.110×10^{-2}	0.977 48	3.977×10^{-2}	0.960 50
线性回归	1.644×10^{-1}	0.834 07	1.693×10^{-1}	0.843 26	1.925×10^{-1}	0.794 59	2.091×10^{-1}	0.792 34
单源 XGBoost	—	—	—	—	6.150×10^{-6}	0.996 09	1.070×10^{-4}	0.936 66

4 结 论

本文提出了一种基于 XGBoost 的多源气动数据融合建模框架, 并将其应用于 CFD 气动数据与飞行试验气动数据, 很好地实现了对飞行试验气动数据轴向力系数 C_x 的预测。主要得出以下结论:

1) 结合 CFD 气动数据与飞行试验气动数据的嫁接 XGBoost 模型可以实现气动数据的精准预测, 预测误差约为 10^{-5} 量级, 模型的泛化能力较强。

2) 相较于单源气动数据的建模, 多源气动数据融合建模的效果更好, 对气动数据的预测更准确, 模型的泛化性能更强。

3) 相较于传统的机器学习模型, 基于集成学习的数据融合模型的预测误差更小, 泛化能力更强。

后续研究中可以改进的地方主要包括以下几个方面:

1) 数据点位置选取方面, 可以考虑多选取一些

在飞行器启动阶段与结束阶段的数据点进行训练。

2) 模型选择与调参方面, 可以尝试选取其他模型进行嫁接建模, 比如神经网络模型; 在模型的训练过程中可以采用网格搜索和贝叶斯优化等调参手段寻找最优的模型参数, 从而进一步提高模型的预测精度和泛化性能。

3) 低精度模型的建立可以考虑选取与目标飞行器飞行状态下各变量不一致的 CFD 气动数据进行建模, 增强低精度模型的外推能力, 为后续高精度模型的建立提供更多信息。

参 考 文 献:

- [1] PALACIOS F, ECONOMON T D, ALONSO J J. Large-scale aircraft design using SU2[C]// 53rd AIAA Aerospace Sciences Meeting, Kissimmee, Florida. Reston, Virginia: AIAA, 2015: 1946.
doi: 10.2514/6.2015-1946
- [2] BUSAN R C, MURPHY P C, HATKE D B, et al. Wind tunnel testing techniques for a tandem tilt-wing, distributed electric propulsion VTOL aircraft[C]// AIAA Scitech 2021 Forum, VIRTUAL EVENT. Reston, Virginia: AIAA, 2021: 1189.
doi: 10.2514/6.2021-1189
- [3] SONI B, ALIABADI S. Large-scale CFD simulations of airflow and particle deposition in lung airway[J]. *Computers & Fluids*, 2013, 88: 804–812.
doi: 10.1016/j.compfluid.2013.06.015
- [4] HU L W, XIANG Y, ZHANG J, et al. Aerodynamic data predictions based on multi-task learning[J]. *Applied Soft Computing*, 2022, 116: 108369.
doi: 10.1016/j.asoc.2021.108369
- [5] 时梓芳. 基于深度神经网络的空气动力学数据建模方法研究[J]. 成都: 电子科技大学, 2020.
SHI Z F. Research on aerodynamic data modeling method based on deep neural network[J]. *Chengdu: University of Electronic Science and Technology of China*, 2020 (in Chinese).
doi: 10.27005/d.cnki.gdzku.2020.003871
- [6] HE L, QIAN W Q, ZHAO T, et al. Multi-fidelity aerodynamic data fusion with a deep neural network modeling method[J]. *Entropy*, 2020, 22(9): 1022.
doi: 10.3390/e22091022
- [7] LIN Q, HU D W, HU J X, et al. A screening-based gradient-enhanced Gaussian process regression model for multi-fidelity data fusion[J]. *Advanced Engineering Informatics*, 2021, 50: 101437.
doi: 10.1016/j.aei.2021.101437
- [8] ZHANG M M, BARTOLI N, JUNGO A, et al. Enhancing the handling qualities analysis by collaborative aerodynamics surrogate modelling and aero-data fusion[J]. *Progress in Aerospace Sciences*, 2020, 119: 100647.
doi: 10.1016/j.paerosci.2020.100647
- [9] PHAM V, KIM M, TYAN M, et al. Numerical experience with variable-fidelity metamodeling for aerodynamic data fusion problems[J]. *Journal of Defense Acquisition and Technology*, 2019, 1(1): 1–8.
doi: 10.33530/jdaat.2019.1.1.1
- [10] KOU J Q, ZHANG W W. Multi-fidelity modeling framework for nonlinear unsteady aerodynamics of airfoils[J]. *Applied Mathematical Modelling*, 2019, 76: 832–855.
doi: 10.1016/j.apm.2019.06.034
- [11] 王旭, 宁晨伽, 王文正, 等. 面向飞行试验的多源气动数据智能融合方法[J]. *空气动力学学报*, 2023, 41(2): 12–20, I0001.
WANG X, NING C J, WANG W Z, et al. Intelligent fusion method of multi-source aerodynamic data for flight tests[J]. *Acta Aerodynamica Sinica*, 2023, 41(2): 12–20, I0001 (in Chinese).
doi: 10.7638/kqdlxxb-2021.0428
- [12] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
doi: 10.1023/A:1010933404324
- [13] 邓晨, 陈功, 王文正, 等. 基于不确定度和气动模型的气动数据融合算法[J]. *空气动力学学报*, 2022, 40(4): 117–123.
DENG C, CHEN G, WANG W Z, et al. Aerodynamic data fusion algorithms based on aerodynamic model and uncertainly[J]. *Acta Aerodynamica Sinica*, 2022, 40(4): 117–123 (in Chinese).
doi: 10.7638/kqdlxxb2020-0151
- [14] 陈海, 钱炜祺, 何磊. 基于深度学习的翼型气动系数预测[J]. *空气动力学学报*, 2018, 36(2): 294–299.
CHEN H, QIAN W Q, HE L. Aerodynamic coefficient prediction of airfoils based on deep learning[J]. *Acta Aerodynamica Sinica*, 2018, 36(2): 294–299 (in Chinese).
doi: 10.7638/kqdlxxb-2017.0098
- [15] 王文正, 郑鹏鹏, 陈功, 等. 数据挖掘技术在飞行试验数据分析和气动参数辨识中的应用研究[J]. *空气动力学学报*, 2016, 34(6): 778–782.
WANG W Z, ZHENG K P, CHEN G, et al. Application research of data-mining method to flight test data analysis and aerodynamic parameter identification[J]. *Acta Aerodynamica Sinica*, 2016, 34(6): 778–782 (in Chinese).
doi: 10.7638/kqdlxxb-2015.0032
- [16] WEBB G I, ZHENG Z. Multistrategy ensemble learning: reducing error by combining ensemble learning techniques[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2004, 16(8): 980–991.
doi: 10.1109/TKDE.2004.29
- [17] DU P F, ZHANG S, ALPHONES A, et al. Faster deployment for indoor visible light positioning using xgboost algorithms in industrial Internet-of-things[C]// IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, Toronto, ON, Canada. IEEE, 2021: 1–7.
doi: 10.1109/IECON48115.2021.9589151
- [18] ABBASIMEHR H, PAKI R, BAHIRI A. A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models[J]. *Sustainable Computing: Informatics and Systems*, 2023, 38: 100863.
doi: 10.1016/j.suscom.2023.100863
- [19] KIANGALA S K, WANG Z H. An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment[J]. *Machine Learning With Applications*, 2021, 4: 100024.
doi: 10.1016/j.mlwa.2021.100024
- [20] WU T W, HE P J, LAN D Y, et al. Application of XGBoost for fast identification of typical industrial organic waste samples with near-infrared hyperspectral imaging[J]. *ACS ES& T Engineering*, 2023, 3(6): 841–850.
doi: 10.1021/acsestengg.2c00426
- [21] GUMUS M, KIRAN M S. Crude oil price forecasting using XGBoost[C]// 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey. IEEE, 2017: 1100–1103.
doi: 10.1109/UBMK.2017.8093500
- [22] CHEN T Q, GUESTIN C. Xgboost: A scalable tree boosting system[C]// 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. 2016: 785–794.
doi: 10.1145/2939672.2939785
- [23] BROWNE M W. Cross-validation methods[J]. *Journal of Mathematical Psychology*, 2000, 44(1): 108–132.
doi: 10.1006/jmps.1999.1279

(本文责编: 张天航)