



基于深度学习的风洞天平测力试验数据异常检测方法研究

张靖 孙文举 尼文斌 魏巍 张江 杨武兵 李清勇

Study on deep learning-based anomaly detection method for wind tunnel balance force data

ZHANG Jing, SUN Wenju, NI Wenbin, WEI Wei, ZHANG Jiang, YANG Wubing, LI Qingyong

引用本文:

张靖, 孙文举, 尼文斌, 等. 基于深度学习的风洞天平测力试验数据异常检测方法研究[J]. 实验流体力学, 2022, 36(6): 67-73. DOI: 10.11729/sytlx20210061
ZHANG Jing, SUN Wenju, NI Wenbin, et al. Study on deep learning-based anomaly detection method for wind tunnel balance force data[J]. Journal of Experiments in Fluid Mechanics / J Exper Fluid Mech, 2022, 36(6): 67-73. DOI: 10.11729/sytlx20210061

在线阅读 View online: <https://doi.org/10.11729/sytlx20210061>

您可能感兴趣的其他文章

Articles you may be interested in

基于减小轴向力测量干扰的高精度测力天平研制

Development of a six-component wind tunnel balance with lower interference on axial force measurement

实验流体力学. 2018, 32(6): 61-67 <https://doi.org/10.11729/sytlx20180082>

旋转导弹风洞动态测力试验技术研究

Research on test technology of dynamic force measurement of rotating missile in wind tunnel

实验流体力学. 2020, 34(4): 81-86 <https://doi.org/10.11729/sytlx20190119>

基于CVAE的超高速碰撞碎片云运动过程的快速预测技术

Evolution prediction of HVI debris based on CVAE model

实验流体力学. 2021, 35(5): 40-46 <https://doi.org/10.11729/sytlx20200058>

多风洞共享的通用型试验数据处理系统

General data processing system for multiple wind tunnels

实验流体力学. 2017, 31(4): 64-70 <https://doi.org/10.11729/sytlx20170051>

风洞自由飞试验中气动参数辨识准确度评价方法研究

Research on accuracy assessment method of aerodynamic parameters identified from wind tunnel free-flight test data

实验流体力学. 2017, 31(1): 39-46 <https://doi.org/10.11729/sytlx20160148>

CHN-T1标模大型低速风洞试验结果相关性分析

Correlation analysis of large low speed wind tunnel test on CHN-T1 calibration model

实验流体力学. 2019, 33(3): 106-111 <https://doi.org/10.11729/sytlx20180046>

地址: 四川省绵阳市二环路南段8号11信箱9分箱

电话: 0816-2463376

Email: sytlx@163.com

网址: <http://www.sytlx.com>



关注微信公众号

获得更多资讯

基于深度学习的风洞天平测力试验数据 异常检测方法研究

张靖¹, 孙文举¹, 尼文斌², 魏巍², 张江², 杨武兵², 李清勇^{1,*}

1. 北京交通大学 智慧高铁系统前沿科学中心, 北京 100044 2. 中国航天空气动力技术研究院, 北京 100074

摘要: 风洞天平测力试验数据异常检测有助于分析试验异常原因、排查设备故障、改进试验方案, 为解决目前人工检测时间成本高、效率低的问题, 提出一种基于深度学习的异常检测框架。首先针对异常零样本问题, 对风洞试验常见的异常类型及规律进行总结; 然后为解决不同车次数据维度不同的问题, 提出基于统计特征的标准化特征表示方案; 最后利用神经网络学习异常特征, 完成异常检测。试验结果表明: 基于深度学习的异常检测方法对风洞异常数据检测的准确率和检出率分别达到了 81.7% 和 72.6%, 能够较好地识别孤立跳点异常和多点异常。

关键词: 风洞试验; 天平测力; 异常检测; 异常规则; 深度学习

中图分类号: V211.74; TP391 文献标识码: A

Study on deep learning-based anomaly detection method for wind tunnel balance force data

ZHANG Jing¹, SUN Wenju¹, NI Wenbin², WEI Wei²,
ZHANG Jiang², YANG Wubing², LI Qingyong^{1,*}

1. Frontiers Science Center for Smart High-speed Rail System, Beijing Jiaotong University, Beijing 100044, China

2. China Academy of Aerospace Aerodynamics, Beijing 100074, China

Abstract: Anomaly detection for wind tunnel balance force data is beneficial to analyze anomaly reasons, improve test schema and troubleshoot equipment problems. To solve the high time cost and low-efficiency problems of the manual detection method, a deep learning-based anomaly detection method is proposed. To solve the problem of no abnormal data, we summarize the most common abnormal types in the wind tunnel test. For the problem that the dimensions of data in different experiments are different, a standardization scheme based on statistical characteristics is proposed. Finally, a deep learning model is utilized to learn abnormal features and detect abnormal data. Experimental results show that our deep learning-based anomaly detection method can achieve 81.7% accuracy and 72.6% recall, and has a good detection performance for isolated jump points and multipoint anomalies.

Keywords: wind tunnel test; balance force; anomaly detection; anomaly rule; deep learning

0 引 言

飞行器的气动外形是决定飞行器气动性能的重要因素。目前, 国内外进行空气动力试验的方法主

要有数值模拟、风洞试验和飞行试验等。其中, 风洞试验作为一种操作便捷、准确率较高且成本可接受的试验方式, 广泛应用于航空航天、机械制造等诸多领域。然而, 作为一种实物试验, 风洞试验会受风洞

收稿日期: 2021-06-16; 修回日期: 2021-07-28; 录用日期: 2021-08-06

基金项目: 中央高校基本科研业务费专项(2022JBQY007)

* 通信作者 E-mail: liqy@bjtu.edu.cn

引用格式: 张靖, 孙文举, 尼文斌, 等. 基于深度学习的风洞天平测力试验数据异常检测方法研究 [J]. 实验流体力学, 2022, 36(6): 67-73.

ZHANG J, SUN W J, NI W B, et al. Study on deep learning-based anomaly detection method for wind tunnel balance force data[J]. Journal of Experiments in Fluid Mechanics, 2022, 36(6): 67-73.

设备故障、试验人员操作失误等诸多因素影响产生异常数据。其中,风洞设备故障和人员操作失误是导致风洞异常数据的主要原因,据统计,二者导致的风洞异常数据分别占异常数据总数的70%和25%。针对风洞天平测力试验异常数据,目前皆采用人工检测的方式。这种方式虽然具有较高的准确率和可靠性,但十分依赖分析人员的专业知识和处理经验,且时间成本较高。为了弥补人工检测方式的不足,本文提出将人工智能技术用于风洞试验天平测力数据监测,实现自动化异常识别,提高风洞试验的整体效率。

国内外已有很多将人工智能技术引入风洞试验领域的研究。谢艳等^[1]尝试在风洞常规试验中构建起大数据采集、收集和分析处理平台,并利用大数据较强的洞察能力,助力风洞试验中的疑难问题分析。Belmahdi等^[2]借助人工智能中的模糊逻辑推理系统,对风洞试验中的高斯山峰前、中、后的风散速度进行建模,预测风洞中的风速。美国NASA Lanley研究中心在亚声速风洞中安装了神经网络数据分析系统,该系统可以高效地执行模型结构优化、流动参数研究等诸多任务^[3]。Vlachas等^[4]借助深度学习中的长短时记忆模型(LSTM)来预测高维混沌系统的演变,并取得了比高斯过程模型更优的效果。Mifsud等^[5]利用不连续本征正交分解(POD)和受限最小二乘法来融合CFD模拟数据和稀疏的风洞试验数据。这些工作证明了人工智能与风洞试验相结合的可行性,但根据调研,目前还没有利用人工智能技术对风洞天平测力试验数据进行异常检测的先例。

在人工智能领域中,异常检测问题是一个经典问题,相关技术广泛应用于欺诈检测^[6]、故障诊断^[7]、医疗辅助^[8]等领域。传统的异常检测算法包括基于统计特征的方法、基于近邻的方法、基于聚类的方法和基于分类的方法等。例如Barnett等^[9]提出将符合某个特定分布或概率模型的数据视为正常数据,否则视为异常数据。Schölkopf等^[10]利用单类支持向量机学习正常样本与异常样本之间的最佳超平面,提出了单类支持向量机算法(One-Class Support Vector Machine, OC-SVM)。Liu等^[11]利用异常样本相对密度低、易被分割孤立的特点提出了基于规则树结构的孤立森林算法(Isolation Forest, IF)。传统方法的优点是运行时间快、可解释性高,但大部分方法依赖于数据的特征设计,无法进行自主学习。深度学习方法引入了神经网络。Akçay等^[12]利用正

常数据训练自编码重构模型,将测试时重构误差较大的样本视为异常样本。Ruff等^[13]提出深度支持向量数据描述算法(Deep Support Vector Data Description, Deep SVDD),引入神经网络完成特征学习和映射,并用软边界损失约束训练过程。深度学习方法的优点是准确率高、特征自学习,缺点则是可解释性和鲁棒性较差。

不同于普通的异常检测任务,本文要解决的风洞天平测力试验数据异常检测有两大难点:1)无真实异常数据;2)不同车次的迎角个数不同导致数据维度不同。为解决这2个难点,本文提出一种基于深度学习的异常检测框架,通过针对风洞异常数据的仿真策略来解决异常零样本问题。首先,设计了基于统计特征的标准化特征表示,将数据处理为统一尺度;然后,利用神经网络提取深度特征,增强对数据的判断能力。以此为基础的检测系统已部署于实际生产环境,为现场分析人员进行异常检测提供帮助。

1 风洞天平测力试验异常数据仿真

对风洞异常数据进行分析有利于及时发现异常原因、排查设备问题、改进试验方案。但在实际的生产环境中,异常数据产生频率极低,且通常被视为垃圾数据而丢弃,这给异常检测模型的训练和验证带来了挑战。为解决异常数据零样本的问题,归纳了常见的风洞数据异常类型,并根据专家知识形成了异常数据规则,以此来生成异常数据。具体而言,根据异常的主要特征,将异常类型分为3类:孤立跳点异常、多点异常和整组试验异常。

1) 孤立跳点异常

导致风洞试验出现孤立跳点的原因包括天平虚焊、采集系统故障、模型内有碎屑等。此类异常表现为在某车次的试验数据中,某个迎角下的力分量值脱离真实值,形成跳点。

2) 多点异常

天平内外引线不绝缘、激波返回、模型碰到支杆、测压模块异常、杂质冲击等都可能造成多点异常。此类异常多发生于法向力数据,表现形式为:在某车次的试验数据中,连续或离散的多个(≥ 2)迎角下的力分量值产生异常。

3) 整组试验异常

导致整组试验异常的原因有很多,如天平零飘、迎角不准、总压管故障、迎角机构异常、模型加工问

题、模型抖动、模型安装、压力滞后等。此类异常表现为某试验车次所有数据发生了偏移，特征不明显，识别难度较高。

以上 3 类异常在法向力 (N) 和轴向力 (A) 上都有体现，本文选择最常见的 6 种异常作为检测目标。表 1 记录了这些异常的基本信息。其中， $N_{\text{normal},\alpha_i}$ 和 $N_{\text{abnormal},\alpha_i}$ 分别表示法向力正常或异常曲线在迎角为

α_i 时的值， $A_{\text{normal},\alpha_i}$ 和 $N_{\text{abnormal},\alpha_i}$ 分别表示轴向力正常或异常曲线在迎角为 α_i 时的值。 a 为异常变化率，取值在 0 到 1 之间随机变化。 $k_{N,\text{normal}}$ 和 $k_{N,\text{abnormal}}$ 分别表示法向力正常或异常曲线的斜率。在生成异常数据时，本文将以正常数据为基础，按照随机的原则选择表 1 中的异常类型，并根据异常规则对正常数据进行变化。

表 1 风洞天平测力试验数据异常信息
Table 1 The information of wind tunnel balance force anomaly data

异常类型	序号	异常名称	异常规律
孤立跳点异常	1	A 孤立跳点	$ A_{\text{abnormal},\alpha_i} - A_{\text{normal},\alpha_i} \geq A_{\text{normal},\alpha_i} \times a$
	2	N 孤立跳点	$ N_{\text{abnormal},\alpha_i} - N_{\text{normal},\alpha_i} \geq N_{\text{normal},\alpha_i} \times a$
多点异常	3	N 模型碰支杆	$ N_{\text{abnormal},\alpha \geq 20^\circ} = c, c \text{ 为一常数}$
	4	N 多个跳点	$ N_{\text{abnormal},\alpha_i} - N_{\text{normal},\alpha_j} \geq N_{\text{normal},\alpha_j} \times a, \alpha_i \text{ 值有多个}$
整组试验异常	5	N 斜率异常	$ k_{N,\text{abnormal}} - k_{N,\text{normal}} \geq k_{N,\text{normal}} \times a$
	6	A 整体偏大	$A_{\text{abnormal},\alpha_i} > A_{\text{normal},\alpha_i}$

2 模型方法

2.1 问题定义

在本文中，以 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 表示风洞天平测力试验数据集，其中 \mathbf{x}_i 表示第 i 个车次的数据， n 表示车次的总数量。在每一个车次的风洞试验中，试验人员通常固定其他参数，仅改变迎角 α ，获取气动模型在不同姿态时的受力变化情况。因此，每个车次的数据 \mathbf{x}_i 又可以表示为：

$$\mathbf{x}_i = \{(\alpha_1^i, N_1^i, A_1^i), \dots, (\alpha_{z_i}^i, N_{z_i}^i, A_{z_i}^i)\} \quad (1)$$

式中： α_1^i 表示第 i 个车次的第 1 个迎角； N_1^i 和 A_1^i 分别表示气动模型在第 1 个迎角时受到的法向力和轴向力； z_i 表示第 i 个车次的迎角总数。需要注意的是，不

同车次在风洞试验时使用的迎角个数及具体数值可能并不相同，也就是说 $\exists i, j$ 使得 $i \neq j \wedge z_i \neq z_j$ 。

本模型的任务目标是利用真实的风洞试验数据和少量的仿真异常数据，训练一个异常数据判别器 f 。当给定某个车次的试验数据 \mathbf{x}_i 时， f 需要判断该车次的数据是否存在异常，表示为：

$$y_i = f(\mathbf{x}_i) \quad (2)$$

式中： $y_i \in \{-1, 1\}$ 表示第 i 个车次的异常检测结果，1 表示异常，-1 表示正常。

2.2 总体流程

图 1 展示了本文所使用的异常检测算法的整体框架。首先，给定真实的正常数据和异常规则，生成部分异常数据，为后续模型的训练和验证提供数据

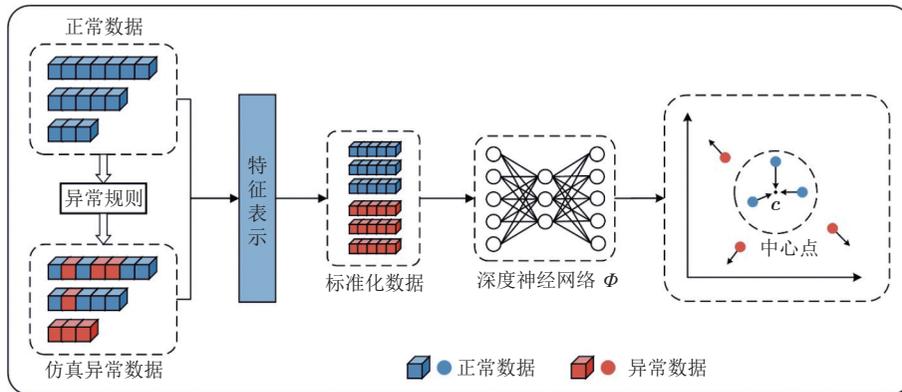


图 1 基于深度学习的异常检测算法框架图

Fig. 1 The framework of our deep learning-based anomaly detection method

基础。然后,为解决不同车次迎角个数可能不同而导致的尺度不同的问题,设计了针对风洞天平测力数据的特征表示,从而将所有的车次数据转化为等尺度的标准化数据。在完成标准化处理之后,所有样本将通过一个深度神经网络映射到一个特征空间中。在此空间中,要求真实的正常样本尽可能靠近中心点,而仿真异常样本尽可能远离中心点,并使用基于距离的损失对神经网络进行训练。

在测试阶段,给定某个车次的数据,对其同样进行标准化特征处理,并输入到训练好的模型中。测试数据将被映射到特征空间中,其与中心点的距离将作为该测试数据的异常分数并与一个异常阈值进行比较。若大于该阈值,则判定为异常;否则为正常。

2.3 标准化特征表示

不同车次迎角个数不同,导致无法用相同尺度(长度)的向量表示所有的试验数据,也就无法将其输入到全连接神经网络中。因此,设计了标准化特征表示模块,对尺度不一的车次数据进行标准化处理。具体而言,从统计学角度选取 13 个特征来描述每个车次数据的变化趋势,其中包括 7 个法向力特征和 6 个轴向力特征。具体设计如下:

1) 法向力特征

法向力特征包括法向力曲线的均值、方差、突变率、大迎角变化率和曲线分斜率的均值、方差、突变率。均值和方差为常见的统计特征;突变率为曲线所有迎角下突变率的最大值,其中,迎角为 α_j 时的突变率 δ 为该迎角下的力分量值与期望值的相对偏差:

$$\delta = \frac{|N_{\alpha_j} - (N_{\alpha_{j-1}} + N_{\alpha_{j+1}})/2|}{(N_{\alpha_{j-1}} + N_{\alpha_{j+1}})/2} \quad (3)$$

曲线分斜率是指曲线中相邻两个迎角之间的斜率。若某车次数据共有 11 个迎角,则可以计算得到 10 个分斜率。以 k_i 表示第 i 个分斜率,则分斜率的突变率为 $\max |k_i - k_{i+1}|$,利用此特征可以检测斜率的突变,对孤立跳点异常检测非常有效。

2) 轴向力特征

轴向力特征包括轴向力曲线的均值、方差、突变率和曲线分斜率的均值、方差、突变率。这些特征的计算方式与法向力特征中的同名特征相同。

给定某个车次的试验数据,计算该车次数据的上述特征,并按照固定的顺序将其拼接成维度为 13 的特征向量,这个过程可以表示为:

$$\hat{\mathbf{x}}_i = T(\mathbf{x}_i) \quad (4)$$

式中: T 表示特征计算及拼接过程; $\hat{\mathbf{x}}_i$ 表示标准化特征向量,维度为 13。

2.4 深度异常检测方法

由于异常数据的多样性和复杂性,简单利用基于统计的特征表示无法实现理想的异常检测功能。因此,受深度神经网络强大的特征提取能力启发,选择使用一种基于深度学习的异常检测方法来进行更深入的特征学习。

具体而言,在训练阶段,输入第 i 个车次的标准化特征向量 $\hat{\mathbf{x}}_i$,使用一个 3 层的全连接神经网络将其映射到一个特征空间中,并得到该车次的嵌入特征表示。此过程表示为:

$$\tilde{\mathbf{x}}_i = \Phi(\hat{\mathbf{x}}_i; \mathbf{W}) \quad (5)$$

式中: Φ 为神经网络所实现的特征映射函数; \mathbf{W} 为神经网络的参数; $\tilde{\mathbf{x}}_i$ 为 $\hat{\mathbf{x}}_i$ 通过映射之后在目标特征空间中的特征表示。

将所有的训练数据映射到同一个特征空间后,真实的正常数据应尽可能靠近给定的中心点,而仿真异常数据应尽可能远离中心点。为实现这一目标,可通过下式计算每次迭代的损失:

$$\min_{\mathbf{W}} \frac{1}{n+m} \sum_{i=1}^n \|\Phi(\hat{\mathbf{x}}_i; \mathbf{W}) - \mathbf{c}\|^2 + \frac{\eta}{n+m} \sum_{j=1}^m \|\Phi(\hat{\mathbf{x}}_j; \mathbf{W}) - \mathbf{c}\|^{-2} + \lambda \|\mathbf{W}\|_F^2 \quad (6)$$

式中: \mathbf{c} 为中心点,由随机初始化获得并固定不变; n 为真实风洞试验数据个数; m 为仿真异常数据个数; η 为异常数据损失权重; λ 为正则化项的权重。损失函数第一项为正常样本惩罚项,要求正常样本与中心点的欧氏距离要尽可能小;第二项为异常样本损失项,为异常样本与中心点距离的倒数,要求异常样本尽可能远离中心点;最后一项为正则化项(下标 F 表示 Frobenius 范数),约束神经网络参数 \mathbf{W} 的复杂度,避免参数过于精细而导致对训练数据过拟合。每一次迭代计算得到的损失将通过链式法则进行反向传播,通过随机梯度下降法^[14]对神经网络中的参数进行更新,直至模型收敛。

在测试阶段,对于每一个测试样本 \mathbf{x} ,利用训练好的神经网络对其进行映射,并将该样本与中心点 \mathbf{c} 的距离作为异常分数,表示为:

$$s(x) = \|\Phi[T(\mathbf{x}); \mathbf{W}^*] - \mathbf{c}\|^2 \quad (7)$$

式中: $s(x)$ 为该测试样本的异常分数; \mathbf{W}^* 为训练完成的模型参数。若 $s(x)$ 大于阈值 σ ,则判定该样本为

异常；否则为正常。异常阈值 σ 是一个超参数，通过对所有数据的异常分数进行排序后，再根据数据集的异常比例进行确定。

3 试 验

3.1 数据集

试验使用的数据包括真实正常数据和仿真异常数据。真实正常数据取自于某风洞在 2017—2018 年进行的真实风洞试验，包含了 120 多个不同型号气动模型的共 11675 个车次的数据。仿真异常数据则通过对正常数据施加表 1 中的异常规则获得，共生成了 3000 个车次的数据。正常数据与异常数据共同组成了试验数据集，并按照 3:1:1 的比例划分为训练集、验证集和测试集，分别用于模型训练、超参数调优和性能测试。为验证模型对数据的依赖性，按随机原则制作了 5 份不同的数据集，最后用模型在这些数据集上的均值和方差来验证方法的稳定性。

3.2 参数设置

深度神经网络共有 3 层，神经元个数分别为 128、256 和 128，激活函数选择 ReLU^[15] 激活函数。模型训练设置的迭代次数为 300，学习率为 0.0001，正则化权重 λ 设置为 0.00001。

本文异常检测方法中共包含 2 个超参数：异常数据损失权重 η 和异常阈值 σ 。前者的取值范围为 [0.1, 0.2, ..., 2.0]，通过交叉验证试验确定。后者与数据集的异常比例有关，按照由高到低对所有数据的异常分数进行排序，并按照异常比例确定异常阈值。

3.3 评价指标

在本任务中，将异常作为正类、正常作为负类，表 2 为异常检测混淆矩阵。其中，TP 表示正确地将异常判定异常，FN 表示错误地将异常判定为正常，FP 表示错误地将正常判定为异常，TN 表示正确地将正常判定为正常。

表 2 异常检测混淆矩阵
Table 2 Anomaly detection confusion matrix

预测	实际	
	异常	正常
异常	TP	FP
正常	FN	TN

在本任务中，选择准确率($P_{\text{precision}}$)、召回率(P_{recall})以及 F1 分数(F1-score)作为评价指标。准确

率 $P_{\text{precision}}$ 表示预测为异常的数据中实际为异常的比例：

$$P_{\text{precision}} = \frac{TP}{TP + FP} \quad (8)$$

召回率 P_{recall} 表示被检测出的异常数据占总异常数据的比例：

$$P_{\text{recall}} = \frac{TP}{TP + FN} \quad (9)$$

F1 分数为准确率和召回率的调和平均数，表示模型在准确率和召回率上的整体性能：

$$F1 = \frac{2 \times P_{\text{precision}} \times P_{\text{recall}}}{P_{\text{precision}} + P_{\text{recall}}} \quad (10)$$

3.4 结果及分析

我们选择了 3 种经典的异常检测方法作为对比方法，分别是：单类支持向量机算法(OC-SVM)、孤立森林算法(IF)和深度支持向量数据描述算法(Deep SVDD)。其中，单类支持向量机算法和孤立森林算法为传统方法，深度支持向量数据描述算法为深度学习方法，三者的基本原理已在引言部分进行了简要介绍。需要说明的是，由于在测试阶段正常数据不可知，无法直接利用表 1 中的异常规则进行异常检测。在试验过程中，所有方法使用相同的标准化特征表示。表 3 展示了不同方法在风洞天平测力试验数据集上的异常检测结果(5 个数据集上检测结果的均值及标准差)。

表 3 风洞天平测力试验数据异常检测结果
Table 3 The performance of anomaly detection methods for wind tunnel test data

方法	准确率/%	召回率/%	F1分数/%
OC-SVM	51.1 ± 1.9	68.5 ± 3.3	58.5 ± 2.4
IF	42.8 ± 1.1	52.7 ± 2.2	47.2 ± 1.0
Deep SVDD	29.5 ± 3.8	35.5 ± 9.5	30.4 ± 4.5
Current work	81.7 ± 1.7	72.6 ± 2.4	76.9 ± 1.4

在准确率、召回率和 F1 分数这 3 个指标上，本文方法的性能全面优于另外 3 种方法，比性能次优的方法分别高 30.6%、4.1% 和 18.4%。在模型稳定性方面，孤立森林算法表现最为稳定，对数据依赖性较低；本文的方法和单类支持向量机算法次之；深度支持向量数据描述算法方差最大，稳定性最差。除此之外，观察到 2 种传统方法的性能优于深度支持向量数据描述算法，但弱于本文方法。这是因为深度

支持向量数据描述算法虽然利用了神经网络来学习样本特征,但没有异常监督信息,导致神经网络无法学习到正常数据与异常数据之间的辨别性特征,因此效果不如直接使用规则化统计特征表示的传统方法。与3种对比方法相比,本文方法以基于统计的标准化特征为基础,以仿真异常数据作为监督信息,充分结合了基础统计特征和深度区分特征,因此检测效果更好。

图2为训练过程中的损失收敛曲线。可以看到:在经过约30次迭代后模型基本完成收敛,最后的损失值收敛于0.23左右。

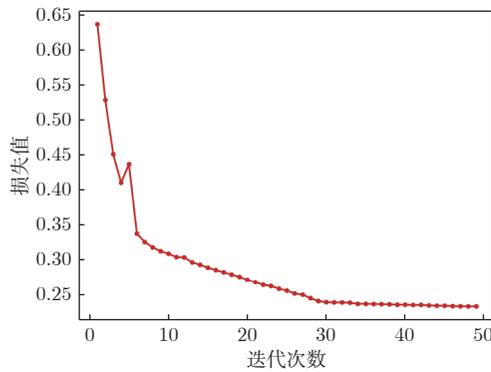


图2 训练过程损失收敛曲线

Fig. 2 Loss convergence curve in the training period

图3展示了4种异常检测方法在风洞天平测力试验数据集上的准确率-召回率曲线。从图中可以看出:本文的方法在整体性能上最优,单类支持向量机算法次之,另外2种方法性能较差。

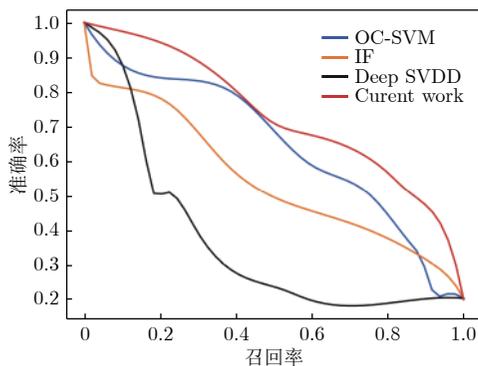


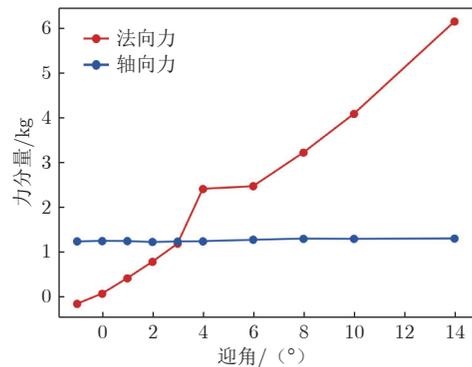
图3 异常检测方法准确率-召回率曲线图

Fig. 3 The precision-recall curve of anomaly detection methods

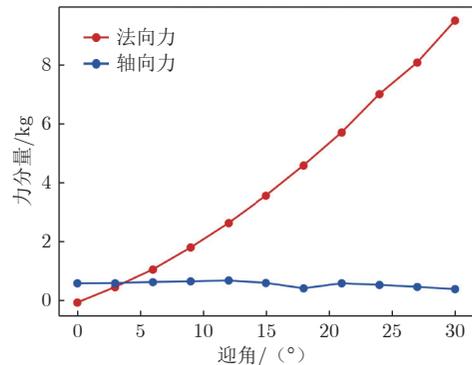
表4展示了表1中各类异常的具体检测召回率。图4为本文方法的具体检测实例,包括正确检测(异常数据被检测为异常,即TP)、错误检测(正常数据被检测为异常,即FP)以及漏检(异常数据被检测为正常,即FN)的车次数据样例。

表4 各种类异常识别率
Table 4 The detection results of all anomaly types

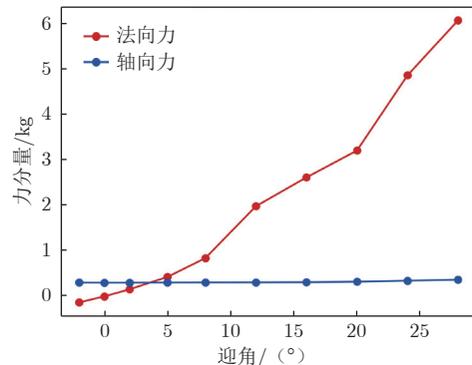
异常类型	召回率/%
1	77.6
2	89.9
3	89.7
4	95.0
5	38.0
6	63.3



(a) 正确检测样例(法向力异常数据被检测为异常)



(b) 错误检测样例(轴向力正常数据被检测为异常)



(c) 漏检样例(法向力异常数据被检测为正常)

图4 异常检测实例

Fig. 4 Anomaly detection examples

检测结果显示本文的方法对孤立跳点异常和多点异常识别效果较好,对整组试验异常识别效果较

差。这是因为相较于另外 2 类异常, 整组试验异常特征不明显。整组数据发生偏移后, 虽然与该车次的正常值差距较大, 但可能与其他车次的分布十分相似, 导致模型难以对此类异常进行识别。在实际的生产环境中, 整组试验异常本身就是一种难以检测的异常。为了识别这一类异常, 风洞数据分析人员通常需要结合气动模型具体外形数据和同模型不同车次数据进行分析处理。

4 结 论

1) 针对风洞天平测力试验数据, 基于标准化特征表示和深度学习的异常检测方法性能优于单类支持向量机算法、孤立森林算法和深度支持向量数据描述算法。

2) 深度异常检测模型能够快速识别大部分异常类型(孤立跳点异常、多点异常), 但对异常特征不明显的整组试验异常识别效果较差。在实际生产环境中, 能够帮助风洞数据分析人员进行数据初筛, 提高检测效率。

下一步将拓展模型可检测的风洞异常类型, 从特征学习、对比学习角度入手, 提高模型的精度, 进一步探索人工智能在风洞试验中的应用。

参考文献:

- [1] 谢艳, 李平, 蒋鸿. 大数据分析技术在风洞试验中的应用[J]. *空气动力学学报*, 2019, 37(6):1004-1009.
XIE Y, LI P, JIANG H. Application of big data analytics approach in wind tunnel test[J]. *Acta Aerodynamica Sinica*, 2019, 37(6):1004-1009.
doi:10.7638/kqdlxxb-2018.0105
- [2] BELMAHDI M, ZEGADI R, BOUHARATI S, et al. Modeling of air flow in wind tunnel using artificial intelligence techniques[J]. *Wulfenia*, 2013, 20(6):94-101.
- [3] 竹朝霞, 惠增宏, 金承信. 虚拟仪器技术在风洞测控智能化中的应用[J]. *实验技术与管理*, 2006, 23(9):76-79.
ZHU Z X, HUI Z H, JIN C X. Virtual instrument application for intelligent system in wind tunnel test[J]. *Experimental Technology and Management*, 2006, 23(9):76-79.
doi:10.3969/j.issn.1002-4956.2006.09.026
- [4] VLACHAS P R, BYEON W, WAN Z Y, et al. Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks[J]. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2018, 474(2213):20170844.
doi:10.1098/rspa.2017.0844
- [5] MIFSUD M, VENDL A, HANSEN L U, et al. Fusing wind-tunnel measurements and CFD data using constrained gappy proper orthogonal decomposition[J]. *Aerospace Science and Technology*, 2019, 86:312-326.
doi:10.1016/j.ast.2018.12.036
- [6] YU W F, WANG N. Research on credit card fraud detection model based on distance sum[C]//Proc of the 2009 International Joint Conference on Artificial Intelligence, 2009:353-356.
doi:10.1109/IJCAI.2009.146
- [7] PURARJOMANDLANGRUDI A, GHAPANCHI A H, ESMALIFALAK M. A data mining approach for fault diagnosis: an application of anomaly detection algorithm[J]. *Measurement*, 2014, 55:343-352.
doi:10.1016/j.measurement.2014.05.029
- [8] BAUDER R A, KHOSHGOFTAAR T M. Multivariate outlier detection in medicare claims payments applying probabilistic programming methods[J]. *Health Services and Outcomes Research Methodology*, 2017, 17(3):256-289.
doi:10.1007/s10742-017-0172-1
- [9] BARNETT V, LEWIS T. Outliers in statistical data[M]. 3rd ed. Chichester: Wiley & Sons, 1994.
- [10] SCHÖLKOPF B, PLATT J C, SHAWE-TAYLOR J, et al. Estimating the support of a high-dimensional distribution[J]. *Neural Computation*, 2001, 13(7):1443-1471.
doi:10.1162/089976601750264965
- [11] LIU F T, TING K M, ZHOU Z H. Isolation forest[C]//Proc of the 2008 Eighth IEEE International Conference on Data Mining, 2008:413-422.
doi:10.1109/ICDM.2008.17
- [12] AKCAY S, ATAPOUR-ABARGHOUEI A, BRECKON T P. GANomaly: semi-supervised anomaly detection via adversarial training[M]//Computer Vision - ACCV 2018. Cham: Springer International Publishing, 2019:622-637.
doi:10.1007/978-3-030-20893-6_39
- [13] RUFF L, VANDERMEULEN R, GÖERNITZ N, et al. Deep one-class classification[C]//International Conference on Machine Learning, 2018:4393-4402.
- [14] ROBBINS H, MONRO S. A stochastic approximation method[J]. *The Annals of Mathematical Statistics*, 1951, 22(3):400-407.
doi:10.1214/aoms/1177729586
- [15] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//Proc of the Proceedings of the 27th International Conference on International Conference on Machine Learning, 2010:807-814.

作者简介:



张靖(1995—),男,山东潍坊人,博士研究生。研究方向:机器学习、计算机视觉。通信地址:北京市海淀区上园村3号北京交通大学(100044)。
E-mail:j_zhang@bjtu.edu.cn

(编辑:宋嘉慧)