

doi: 10.7690/bgzd.2022.02.007

# 基于集成学习的设备流场性能分析与应用

肖乾柯, 赵智聪, 刘治红

(中国兵器装备集团自动化研究所有限公司智能制造事业部, 四川 绵阳 621000)

**摘要:** 为解决某高速设备试验中流场性能实时分析的问题, 对该设备的数据采集与分析技术展开研究。阐述某高速试验设备数据特点和性能分析的需求, 对采集的数据进行可信度校验; 针对其各核心系统模块中不平衡样本数据, 基于 Bagging 集成学习对各个核心系统进行建模分析, 实现对该设备的流场性能分析。实际应用结果表明: 该方法能完成对核心系统的流场性能分析, 提高对该设备的智能管控水平。

**关键词:** 集成学习; 流场; 可信度; 性能分析

**中图分类号:** TP181 **文献标志码:** A

## Performance Analysis and Application of Equipment Flow Field Based on Ensemble Learning

Xiao Qianke, Zhao Zhicong, Liu Zhihong

*(Department of Intelligent Manufacture, Automation Research Institute Co., Ltd. of China South Industries Group Corporation, Mianyang 621000, China)*

**Abstract:** In order to solve the problem of real-time analysis of flow field performance in a high-speed equipment test, the data acquisition and analysis technology of the equipment was studied. The data characteristics and performance analysis requirements of a high-speed test equipment were described, and the credibility of the collected data was verified. Aiming at the unbalanced sample data in each core system module, the modeling analysis of each core system was carried out based on bagging ensemble learning, and the flow field performance analysis of the equipment was realized. The practical application results show that the method completes the flow field performance analysis of the core system and improves the intelligent management and control level of the equipment.

**Keywords:** ensemble learning; flow field; credibility; performance analysis

### 0 引言

近年来全球制造业都朝着信息化、智能化的方向发展<sup>[1]</sup>, 期望通过设备的信息化、智能化升级优化实际生产中的管理问题<sup>[2]</sup>, 通过设备中的数据构建算法模型, 智能化地辅助管理者作出决策。对某高速试验设备管理而言, 同样希望通过对数据的分析, 搭建智能化的机器学习算法模型库, 为其试验中压力和控制数据的实时分析和智能化提供辅助决策, 提高试验设备的管理效率, 减少对专家人员的依赖。

集成学习是机器学习中非常重要的热点分支, 通常由多个弱分类器构成一个强分类器, 进而提高分类器的性能<sup>[3]</sup>。集成学习的特点是强分类器可以克服各个弱分类器自身的不足(如各弱分类器间算法以及适用范围的差异), 通过分治与集成的思想将问题分解后再集成, 保证了学习效率, 这一特点在数据融合、小样本数据集等领域有显著效果<sup>[4]</sup>。集

成学习有多种集成算法: Bagging、Boosting 以及 Stacking。在 Bagging 中, 从整体数据集中采取有放回抽样得到  $N$  个数据集, 在每个数据集上分别学习出一个模型, 最终结果利用  $N$  个模型的输出得到, 分类采用模型预测投票的方式, 回归采用模型预测平均值的方式<sup>[5]</sup>。Boosting 也是学习一系列弱分类器, 组合为一个强分类器, 是可以用来减小监督学习中偏差的算法<sup>[4]</sup>。Stacking 方法是先训练多个不同的模型, 随后将各个模型的输出为输入来训练一个模型, 以得到一个最终的输出<sup>[6]</sup>。

集成学习算法在设备性能分析上已有诸多应用, 相比传统的性能分析算法显著地提高了准确度。文献[7]研究了集成学习在基于特征集和数据集之间的性能差异, 表明了性能依赖于数据集的特性。文献[6]设计了以长短记忆神经网络和 XGBoost 作为基学习器的集成学习算法, 完成了对负荷的预测。文献[8]提出了一种权重自适应调整的集成学习多

收稿日期: 2021-11-19; 修回日期: 2021-12-14

作者简介: 肖乾柯(1996—), 男, 四川人, 硕士, 从事智能制造、数据挖掘研究。E-mail: xiaoqianke1996@163.com。

分类器算法，其分析结果表明提高了分类的性能。文献[9]在半监督学习的情感分类方法基础上提出了基于一致性标签的集成学习方法，该方法在小规模标注样本的情况下有效地降低了对未标注样本的误判率。文献[10]通过 Boosting 重采样集成方法对原始遥感数据进行亚像元估算，其效果优于传统单一的 CART 算法。文献[11]根据各品牌的技术参数从多方面分析了钻机性能与施工的关系。文献[12]使用变权分析方法对电动汽车充电设备的性能进行了综合分析，其建立的综合评价指标体系在性能分析之外还突出了薄弱环节的作用。综上所述，国内外对集成学习算法和设备性能分析都做了较多研究，但针对高速试验设备流场性能的分析研究和应用较少。

笔者从某高速试验设备的流场性能分析需求出发，结合机器学习与工业物联网相关技术，对设备中核心系统流场性能分析技术进行研究，通过对该设备核心系统性能监测系统应用验证，实现了其试验过程的实时自动化流场性能分析，促进了对设备试验过程中的流场性能智能管控，减少了对专家人员的依赖，提升了对设备的自动化智能管理水平。本文中流场性能检测系统使用的是以感知机为弱分类器的 Bagging 集成学习算法。

## 1 设备性能分析

### 1.1 现状

某企业的高速试验设备，对于试验运行过程的智能管控水平较低，主要有以下问题：

1) 试验过程中采集数据量大，但与设备核心子系统流场性能相关的压力传感器数据中有大量未挖掘的数据特征，未充分利用数据，无法结合试验情况进行自动分析和问题的快速定位。

2) 现有的流场性能分析依赖于专家手工计算，没有通过程序将性能指标进行规则化，无法自动实时地进行流场性能分析。

3) 数据采集频率高，压力测点多，且传感器原始数据存在可信度问题；现有的历史数据规律对照计算的工作量大，且无法进行实时数据可信度计算。

### 1.2 实时性能分析需求

该高速试验设备传统的喷管段流场稳定性评估方法存在诸多问题：1) 无法实时进行流场性能分析，需要试验后经过有空气动力学数据处理经验的专家进行分析；2) 其设备试验控制系统的马赫数稳

定标志位后存在部分时间段流场不稳定的情况，该情况目前只通过控制系统进行判断，存在流场性能判断不准确的问题；3) 面临多种试验条件下喷管段的流场稳定性情况复杂。

该设备在试验运行过程中，其控制系统判定流场稳定后，核心系统实际出现流场不稳定和发生异常均为低概率事件，导致训练数据出现严重不均衡，有标签异常数据极为短缺；因此，流场不稳定性预警与异常检测问题为典型的不平衡样本数据分析与数据挖掘问题。在处理该设备的流场稳定性分析时，使用集成学习算法进行流场稳定性分析，使用感知机模型作为基分类器对流场稳定性进行分类预测，减少不平衡数据的影响。

## 2 关键分析技术研究

### 2.1 传感器数据校验

由于设备试验过程中的传感器数据受振动影响，传感器测值在某些试验过程中测值会抖动。高速试验设备对数据的可信度要求极高，因此压力传感器数据须进行可信度校验。对压力传感器进行基于一致性度量的可信度校验流程如下。

计算压力传感器  $\sigma_i$  的自支持度算子  $\sigma_i(k)$ <sup>[13-15]</sup>：

$$\sigma_i(k) = \begin{cases} 1 - |y_i(k) - y_{\text{med}}(k)| / \partial(k), & \text{else} \\ 0, & \text{if } |y_i(k) - y_{\text{med}}(k)| \geq \partial(k) \end{cases} \quad (1)$$

其中： $y_i(k)$ 为  $k$  时刻压力传感器观测结果； $y_{\text{med}}(k)$ 为  $k$  时刻压力传感器数据  $y_1(k), y_2(k), \dots, y_N(k)$  的序列中值。

计算压力传感器  $\sigma_i$  和压力传感器  $\sigma_j$  的互支持度算子  $\sigma_{ij}(k)$ <sup>[14-15]</sup>如下：

$$\sigma_{ij}(k) = \begin{cases} 1 - |y_i(k) - y_j(k)| / \partial(k), & \text{else} \\ 0, & \text{if } |y_i(k) - y_{\text{med}}(k)| \geq \partial(k) \end{cases} \quad (2)$$

其中  $\partial(k)$  为压力传感器  $i$  的支持度因子，通过基于绝对偏差中值 (median absolute deviation, MAD) 估计的鲁棒标准差计算获得，具体如下：

$$\partial(k) = c * \text{med}_i |y_i(k) - \text{med}_j(y_j(k))| \quad (3)$$

其中  $c$  为常数，通过历史数据结合专家知识经验计算得出。

计算压力传感器  $\sigma_i$  的观测值和其他  $(n-1)$  个压力传感器观测值的一致性度量  $R_i(k)$ <sup>[15]</sup>如下：

$$R_i(k) = \sum_{j=1, j \neq i}^N \frac{\sigma_i(k) * \sigma_j(k) * \sigma_{ij}(k)}{n-1} \quad (4)$$

计算压力传感器  $\sigma_i$  的鲁棒性度量如下：

$$\tau_i(k) = \exp - \frac{\sum_i^{L_R} (R_i(t) - \text{med}_i(R_i(k)))^2 / L_R}{(\text{med}_i(R_i(k)))^2} \quad (5)$$

其中： $R_i(k)$ 为压力传感器  $i$  一致性度量数据结果； $L_R$  为一致性度量数据队列长度； $\text{med}_i(R_i(k))$ 为一致性度量结果队列中值。

计算得到压力传感器可信度评价  $R_{\tau_i}$  如下：

$$R_{\tau_i} = R_i(k) * \tau_i(k) \quad (6)$$

经过上述传感器数据可信度评价加权流程后，每个压力传感器的一致性度量结果综合了相邻时间内与其他压力传感器的相关程度，有效避免了压力传感器出现传感器自身不稳定观测值后在后续数据融合分析中的异常影响。

### 2.2 基于感知机模型的弱分类器

感知机原理：设定模型的待训练分类超平面为  $\omega * t + b = 0$ ，其中  $\omega = [\omega_1, \omega_2, \dots, \omega_n]$  为  $n$  维向量，称为权重， $n$  为从压力传感器原始数据中提取出用于分类的特征数据的个数， $b$  为偏移量，输入向量为  $n$  维向量  $t = [x_1, x_2, \dots, x_n]$ 。在每一次迭代中，计算每一个压力传感器数据样本点  $[x_1^k, x_2^k, \dots, x_n^k]$  在待训练直线上所得到的值，按照所选取符号函数作为激活函数对每一个样本点进行分类，若所得值大于 0 则标记为 +1，反之小于 0 则标记为 -1，若等于 0 则标记为 0。感知机模型的参数更新公式如下

$$\omega_i = \omega_i + \omega_i; \quad (7)$$

$$\omega_i = \eta * (l - o) * x_i; \quad (8)$$

$$b = b + \eta * (l - o) \quad (9)$$

针对设备核心系统的流场性能分析感知机的输入数据首先需要通过数据特征提取，减少处理维度与模型训练时间。感知机模型训练与检测流程如图

1 所示。

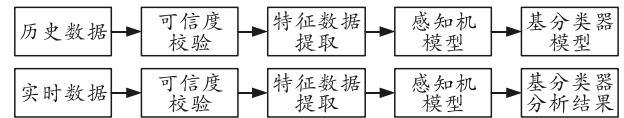


图 1 感知机模型训练与检测流程

### 2.3 基于 Bagging 的集成性能分析

为解决设备不平衡数据分析的难点，在该设备核心系统进行流场性能分析时使用基于欠采样的集成学习方案，使用 Bagging 集成基础分类器的分类结果，综合投票得到最终性能分析结果。使用的方案流程如下：

1) 构建数据集：对样本进行特征选取并挑选出训练集后，将多样本类训练数据集根据小样本类数据集的大小随机分为  $m$  个子集，使得这  $m$  个子集的大小与小样本类数据集的大小基本相同，随后将这  $m$  个子集分别与少样本类数据组合得到  $m$  个数据集，经过上述训练集分割后， $m$  个训练集的各样本分布均衡，适用于感知机模型。

2) 弱分类器：将感知机模型作为弱分类器。针对不同的分类目标需求将数据进行分割，感知机作为基本的弱分类器独自进行训练，形成互相独立的分类器。

3) 集成：根据感知机在相互独立的数据集上训练得到独立的  $m$  个基分类器。在对新样本进行分类时，在  $m$  个感知机模型上进行数据分类，最终通过预设分类器权重的方式计算得到压力传感器的最终性能分析结果，原理为：

$$f(x) = 1 / M \sum_{m=1}^M f_m(x) \quad (10)$$

设备的流场性能监测流程如图 2 所示。

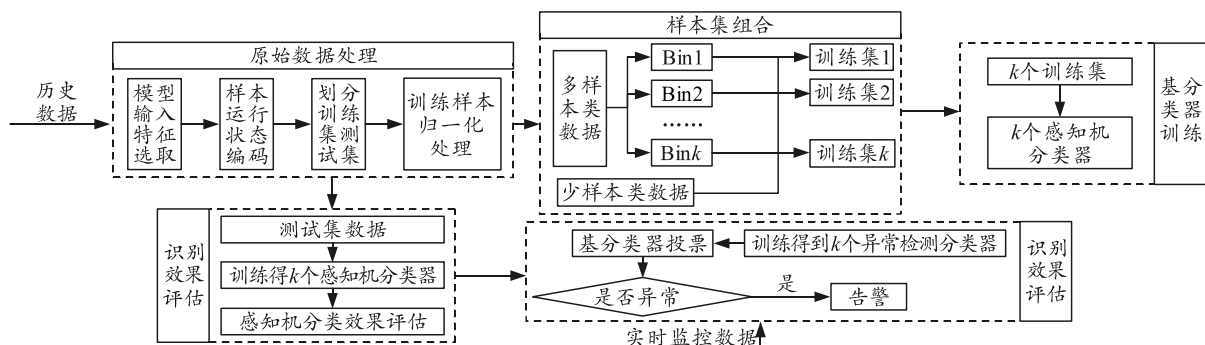


图 2 集成学习的性能检测流程

## 3 应用实现

### 3.1 技术路线

试验设备流场性能监测系统使用基于 Windows

平台的 .NET5 框架，使用 C# 编程语言在 Visual Studio 2019 开发环境上进行数据采集程序和数据分析程序开发。前端和客户端界面分别由 Vue 和 WPF 进行开发，后台服务接口统一使用 WebApi，实时

数据使用 MQTT 进行数据分发。整个系统结构如图 3 所示。

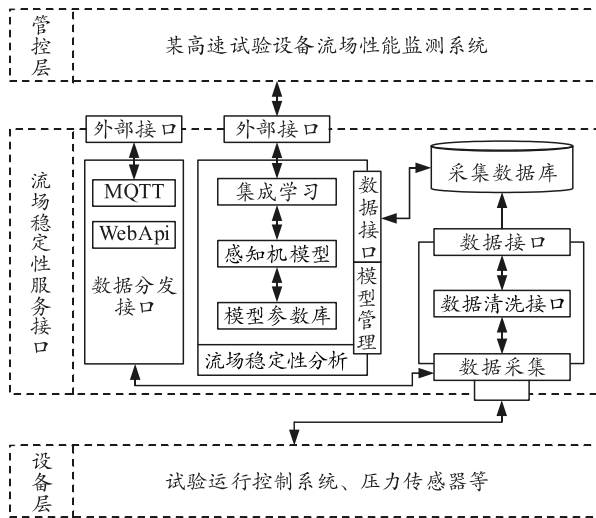


图 3 流场性能监测系统结构

该高速设备试验过程中的数据采集与数据分析是实现设备智能管控的基础，采集与分析主要

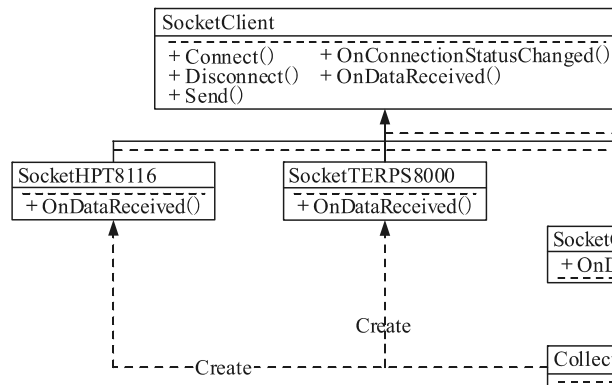


图 5 基于 TCP/IP 的数据采集

2) 性能分析。

研究分析设备核心子系统的历史数据，结合专家经验知识规则进行数据集分割，基于 C#语言 MathNet 库建立感知机模型计算类，通过建立集成学习模型类完成 Bagging 集成。

基于 C#编程语言的流场性能分析程序对设备核心系统的试验过程状态信息定义如表 1 所示。

表 1 核心系统状态

序号	核心系统状态码	说明
1	Boot	设备上电
2	Run	运行(试验中)
3	Idle	空闲(未试验)
4	FlowStable	流场稳定性
5	Fault	故障
6	Error	报警

3.2 验证系统实现

高速试验设备性能监测系统运行，实现了压力

包括：

1) 压力传感器数据采集。

监测系统测点信号分别由 HPT8116 多通道数字压力测量仪、CYG1204F 差压传感器、CPT6180 绝压传感器、TERPS8000 大气压力传感器和 BWS5000 倾角传感器采集。其中压力测量仪采用网口通信，信号直接使用 TCP/IP 协议读取，其他传感器都支持 RS485 通信，经过串口服务器转接后也使用 TCP/IP 协议读取。部分传感器接线拓扑图如图 4 所示。采集与分析程序设备管理的类关系如图 5 所示。

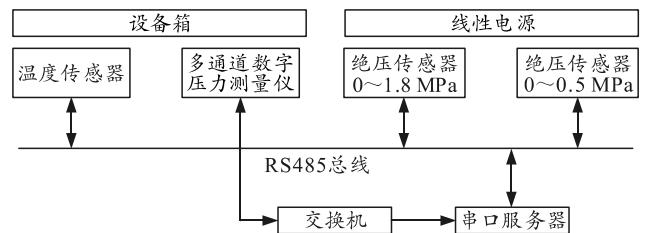


图 4 部分传感器接线

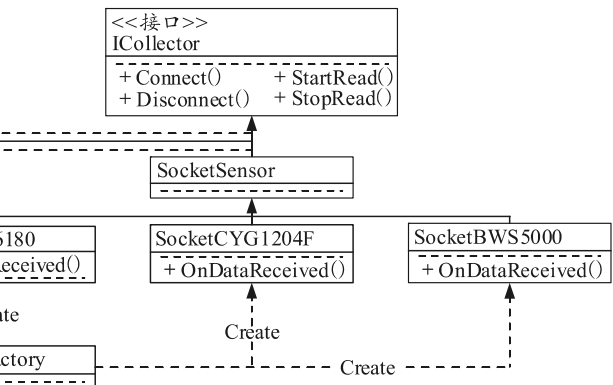


图 5 基于 TCP/IP 的数据采集

传感器数据可信度校验功能和该高速试验设备稳定段及喷管段等核心系统的流场性能分析功能，如图 6 所示。

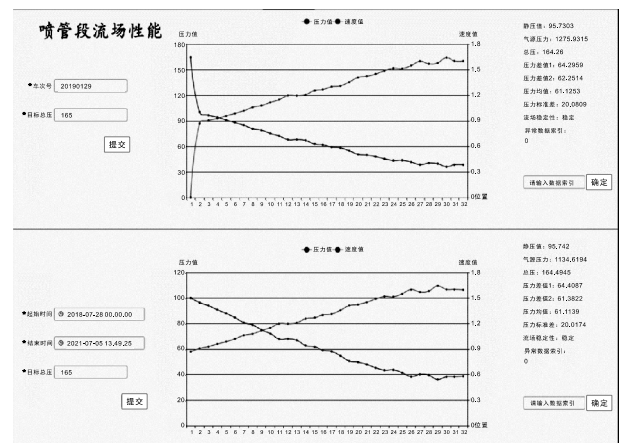


图 6 性能监测系统部分效果