

基于决策树算法的桥梁应变传感器故障识别

杨恒信¹, 张启伟¹, 王长海², 梁 才²

(1. 同济大学 土木工程学院, 上海市 200092; 2. 广西交通设计集团有限公司, 广西 南宁 530022)

摘要: 针对桥梁健康监测系统中传感器故障经常触发报警并影响后续分析的问题, 提出了一种基于决策树算法的传感器故障识别方法。根据应变传感器常见故障类型, 分别建立了偏置、漂移、卡死等故障的数学模型, 并提出分别采用对称位置传感器曼哈顿距离的平均值、前后半段数据平均值之差的绝对值、极差等作为各种故障的数据特征; 给出了数据特征提取时最优信号时长的确定方法, 并构建了以数据特征为输入、采用决策树算法进行机器学习的传感器故障识别模型。通过某斜拉桥应变监测数据的应用, 验证模型对故障类型的识别能力; 结果显示, 该模型对于故障数据的识别精度达到92.7%。该方法构建的模型复杂度低、可解释性强, 为桥梁健康监测系统的传感器故障诊断提供了可行的方法。

关键词: 桥梁健康监测; 传感器数据特征; 决策树; 故障识别; 识别精度

中图分类号: U446.2; TU997

文献标志码: B

文章编号: 1009-7716(2025)12-0268-09

Fault Identification of Bridge Strain Sensor Based on Decision Tree Algorithm

YANG Hengxin¹, ZHANG Qiwei¹, WANG Changhai², LIANG Cai²

(1. College of Civil Engineering, Tongji University, Shanghai 200092, China; 2. Guangxi Communications Design Group Co., Ltd., Nanning 530022, China)

Abstract: To address the issue of frequent false alarms caused by sensor fault in bridge health monitoring system that affects subsequent analysis, a sensor fault identification method based on decision tree algorithm is proposed. According to the common fault types of strain sensors, the mathematical models for biases, drifts and stuck faults are established respectively. It is proposed to respectively adopt the average value of the Manhattan distance of the symmetrical position sensor, the absolute value of the difference between the average values of the front and back halves of the data, the range and the others as the data characteristics of various faults. A method of determining the optimal signal duration for data feature extraction is given, and a sensor fault identification model is constructed, which takes data features as input and uses the decision tree algorithm for machine learning. The capability of the model to identify fault types is verified through the application of strain monitoring data of a cable-stayed bridge. The results show that the identification accuracy of this model for fault data reaches 92.7%. The model constructed by this method has low complexity and strong interpretability, which provides a feasible approach for sensor fault diagnosis in bridge health monitoring systems.

Keywords: bridge health monitoring; sensor data features; decision tree; fault identification; identification accuracy

0 引言

桥梁健康监测是为满足桥梁安全运营的需要发展起来的监测技术^[1], 由传感器系统、数据采集与传输系统、数据处理与储存系统、评估与决策系统等组成^[2], 其中传感器系统处于最前端, 其采集到的监测

数据是后续对桥梁进行预警和评估的基础, 重要性不言而喻。因此, 传感器系统采集到的数据的可靠性决定了后续能否有效地对桥梁预警和评估。然而, 在实际工程应用中, 传感器长期暴露于风、雨、温度、车辆荷载等复杂环境中, 同时受到电子元件老化、信号干扰等因素的影响, 其性能难免会下降, 测量的数据出现失真, 使用失真的数据进行数据处理会导致错误的预警, 有研究显示现有的桥梁健康监测系统的预警大部分起因是传感器故障而非真实结构损伤引起, 这些虚假预警会增加不必要的运维成本。因此, 研究传感器系统故障导致的异常数据的

收稿日期: 2025-06-11

基金项目: 广西重点研发计划(桂科AB23026153)

作者简介: 杨恒信(2000—), 男, 硕士, 从事桥梁工程教学研究工作。

通信作者: 张启伟(1966—), 男, 博士, 教授, 研究员, 从事桥梁工程教学研究工作。电子邮箱: zhangqw@tongji.edu.cn

识别方法很有必要。

目前传感器故障识别方法主要分为三类:基于模型的方法、基于知识的方法和基于数据驱动的方法。

基于模型的方法通过建立传感器系统的输入输出数学模型,利用残差分析检测实际输出与模型预测值之间的偏差,从而实现对传感器故障状态的识别。谈娟等^[3]对飞行控制系统进行故障建模,结合广义卡尔曼滤波方法提出了一种改进的残差检测算法进行在线学习并评估系统的故障状态。Kougiatsos等^[4]使用非线性微分代数估计器结合自适应阈值生成的残差实现故障诊断。其理论严谨、逻辑清晰,并通过仿真实验验证了其有效性,但对模型精度依赖性强;由于桥梁结构复杂、非线性、输入信号难以获取等原因导致无法建立精确的传感器模型,使得该方法在桥梁健康监测实际工程应用中面临较大挑战。

基于知识的方法利用专家经验和规则库构建诊断逻辑实现传感器故障识别。程曦等^[5]提出了基于故障树理论和灰色关联法的降雨量传感器故障诊断方法,实现了降雨量传感器的高效故障诊断。Kaluder等^[6]开发了一种模糊专家离线系统用于配电网中的故障诊断,可以诊断继电器和断路器的正确作、错误作和故障作,还可以识别和量化故障位置。其无需精确模型,可解释性高,但依赖专家知识和预定义的规则,在小规模问题上适用性较好,而桥梁健康监测传感器系统较为复杂,很难建立完整的知识库。

基于数据驱动的方法通过分析传感器采集的历史数据构建模型,无需先验模型,适应复杂非线性系统,逐渐成为研究的前沿趋势。刘张浩等^[7]提出了基于互相关分析方法的传感器故障诊断理论和算法。Bao Yuequan等^[8]提出了基于堆栈式自编码器深度神经网络的健康监测异常数据诊断方法。

然而,现有大多数基于数据驱动的传感器故障识别方法仍存在许多局限:并未有效结合传感器故障类型和数据特征,难以从根本上提升故障识别精度;需要大量的训练样本和计算资源用来构建模型,不利于实际应用的推广;建立的模型普遍缺乏明确的物理含义,其“黑箱”特性使得故障识别过程难以解释,降低了工程技术人员对识别结果的信任度。因此,本文提出了一种基于决策树的传感器故障识别方法,该方法深入分析了应变传感器故障类型和

数据特征,构建的传感器故障识别模型复杂度低,并可以直观地展示出故障类型的判别路径。这一方法在实现故障识别精度显著提升的同时,增强了故障识别过程的可解释性,也为实际应用提供了有价值的指导。

1 传感器故障数学模型及数据特征

实际工程中,常见传感器故障有以下七种类型:偏置、漂移、卡死、精度下降、增益、缺失和离群值^[9]。各种故障类型在传感器数据上具有独特的特征表现。因此,传感器故障识别的任务是:获取传感器数据,提取预先定义的特征,并基于这些特征判断传感器的故障状态及其故障类型。其中,在预先定义传感器数据特征时,需要根据各种传感器故障类型找到其独特的特征,确保所选特征能够区分各类故障。

下面将介绍正常传感器数据和各种传感器故障类型的数学模型及本文提出的数据特征。

(1) 正常

传感器采集到的数据是一个时间序列,记为 $x(t) = \{x_1, x_2, \dots, x_n\}$,设 $x_r(t)$ 表示为桥梁被测响应的真实值,由于传感器测量时受到随机测量噪声的影响,导致传感器正常工作时测量值为

$$x(t) = x_r(t) + \delta(t) \quad (1)$$

式中: $\delta(t)$ 为测量噪声,现有研究一般使用白噪声模拟测量噪声,即 $\delta(t) \sim N(0, \sigma^2)$; σ 为白噪声的标准差。

(2) 偏置

传感器偏置故障是传感器实测值与正常值相差一个固定的常数,数学模型为

$$x(t) = x_r(t) + d + \delta(t) \quad (2)$$

式中: d 为常数。

实际情况下,当传感器发生偏置故障时,往往无法直接获取正常值。而由于对称位置安装的传感器监测数据之间存在很强的相关性^[10],偏置系数 d 可以使用相同时刻下对称位置的两个传感器测量值之差进行识别。在一段时间内,相同时刻下对称位置的两个传感器测量值之差的平均值会趋于稳定,因此本文使用对称位置的两个传感器数据的曼哈顿距离的平均值作为偏置故障的数据特征,曼哈顿距离的平均值的计算方法见式(3)。

设 $x_1(t) = \{x_{11}, x_{21}, \dots, x_{n1}\}$ 和 $x_2(t) = \{x_{12}, x_{22}, \dots, x_{n2}\}$

为对称位置的两个传感器采集的数据,则它们之间的曼哈顿距离的平均值 $D(x_1(t), x_2(t))$ 为

$$D(x_1(t), x_2(t)) = \frac{1}{n} \sum_{i=1}^n |x_{i1} - x_{i2}| \quad (3)$$

(3) 漂移

传感器漂移是传感器实测值与正常值的差随时间线性变化,数学模型为

$$x(t) = x_r(t) + a + bt + \delta(t) \quad (4)$$

式中: a 和 b 是常数。

从式(4)可知,传感器漂移数据相当于在正常值上叠加一组随时间线性增长的分量,传感器漂移故障数据特征的关键是能够识别出漂移系数 b 。对于适当的数据时间段,将其分为前半部分和后半部分,正常数据的前半部分和后半部分的平均值是非常接近的,这是因为对正常数据取平均后的值反映了温致效应,而温度变化是缓慢的;漂移数据由于含有线性增长分量,因此漂移数据的前半部分和后半部分的平均值之差接近于在漂移系数 b 下所取时间段的一半的时间内的漂移量。因此本文采用前后半段数据平均值之差的绝对值作为漂移故障的数据特征,计算方法见式(5)。

$$\text{设 } a_1 = \frac{2}{n} \sum_{i=1}^{n/2} x_i, a_2 = \frac{2}{n} \sum_{i=n/2+1}^n x_i, \text{ 则该段数据的前}$$

后半段数据平均值之差的绝对值 z 为

$$z = |a_2 - a_1| \quad (5)$$

(4) 卡死

传感器卡死是无论真实值如何变化,传感器实测值只表现出恒定值或恒定值加随机噪声,数学模型为

$$x(t) = c + \delta(t) \quad (6)$$

式中: c 是常数。

对比式(1)和式(6)可知,卡死数据和正常数据相比将 $x_r(t)$ 替换为常数 c , $x_r(t)$ 由温度和车辆效应组成,因此一段时间内 $x_r(t)$ 的最大值和最小值之差即极差不为零,而常数 c 的极差为零,因此本文采用极差作为卡死故障的数据特征,计算方法见式(7)。

设 $x_{\max} = \max \{x_1, x_2, \dots, x_n\}$, $x_{\min} = \min \{x_1, x_2, \dots, x_n\}$, 则极差 R

$$R = x_{\max} - x_{\min} \quad (7)$$

(5) 精度下降

传感器精度下降是传感器测量值的均值不变,但离散性变大,数学模型为

$$x(t) = x_r(t) + \delta(t) + e(t) \quad (8)$$

式中: $e(t)$ 为传感器精度下降引起的均值为零的过量随机噪声。

从式(8)可知,精度下降是在正常数据上叠加随机噪声,精度下降故障测量值相比正常测量值偏离真实值的程度更高,导致在一段时间内相邻时间戳的测量值之差(一阶差分)的绝对值之和比正常数据的一阶差分绝对值之和更大。因此本文使用差分计算绝对值之和的平均值作为精度下降故障的数据特征,差分计算方法和差分计算绝对值之和的平均值 w 分别见式(9)和式(10)。

$$y[i] = x[i+1] - x[i] \quad 0 < i \leq n-1 \quad (9)$$

$$w = \frac{1}{n-1} \sum_{i=1}^{n-1} |y_i| \quad (10)$$

(6) 增益

传感器增益是传感器测量值是正常值与某个固定常数的乘积,数学模型为

$$x(t) = G(x_r(t) + \delta(t)) \quad (11)$$

式中: G 为增益系数,为一常数。

从式(11)可以看出,传感器增益故障数据与正常数据之差为 $(G-1)(x_r(t) + \delta(t))$,导致对称位置的两个传感器数据的曼哈顿距离的平均值变大,由于传感器偏置故障已经使用了这个特征,因此需要使用其他特征进一步区分增益故障。传感器增益故障的差分计算绝对值之和的平均值为正常数据的 G 倍,因此可以将差分计算绝对值之和的平均值作为增益故障的另一个数据特征。这两个特征的计算方法前面已经详细阐述,此处不再重复。

(7) 缺失

传感器缺失是传感器在一段时间内没有采集到数据,数学模型为

$$x(t) = \{x_1, x_2, \dots, x_m\} \quad m < n \quad (12)$$

从式(12)可以看出,缺失故障发生时在指定时间段内时间戳数量小于正常数据,因此采用时间戳数量作为缺失故障的数据特征。

(8) 离群值

传感器离群值是传感器测量值明显偏离正常范围,出现异常高或异常低的值。

传感器离群值故障通常表现为一个测量值明显大于左右两个相邻的测量值,因此采用差分计算后的两个连续跳跃点作为传感器离群值故障的数据特征。计算方法如下所示。

a. 首先计算差分计算后数据的均值 μ_y 和标准差 σ_y ;

$$\mu_y = \frac{1}{n} \sum_{i=1}^{n-1} y_i \quad (13)$$

$$\sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^{n-1} (y_i - \mu_y)^2}$$

(14)

b. 然后用 n 倍 σ 法则设定阈值 Δ ;

$$\begin{aligned}\Delta_1 &= \mu_y + n\sigma_y, \\ \Delta_2 &= \mu_y - n\sigma_y,\end{aligned}$$

(15)

式中: Δ_1 为上阈值, Δ_2 为下阈值; n 为常数,通常取值范围为 3~4。

c. 将在阈值外的数据设定为跳跃点。

如果 $y[i] > \Delta_1$ 或 $y[i] < \Delta_2$, 则认为 x_{i+1} 是跳跃点。

如果 x_{i+1} 到 x_{i+2} 都为跳跃点,且 x_i 和 x_{i+3} 不是跳跃点,则认为该段数据中存在两个连续跳跃点。

以上根据每种传感器故障类型找到了各自的特征,能够区分出传感器是否发生故障。但是,为了进一步区分出传感器故障类型,还需要保证不同传感器故障类型在这些特征上的表现形式各不相同。

因此,本文根据传感器故障类型的数学模型推导了各类故障在数据特征上的具体表现形式,见表 1。

表 1 传感器故障类型在数据特征上的表现形式

故障类型	极差	时间戳数量	差分计算后的两个连续跳跃点	差分计算绝对值之和的平均值	对称位置传感器曼哈顿距离的平均值	前后半段数据平均值之差的绝对值
偏置	√	√	√	√	×	√
漂移	√	√	√	√	√	× ¹
精度下降	√	√	√	× ¹	√	√
增益	√	√	√	× ¹	×	√
缺失	√	×	√	× ²	×	√
卡死	×	√	√	× ²	×	× ²
离群值	√	√	×	√	√	√

注:故障类型的数据特征在正常数据特征范围内为√,不在正常数据特征范围内为×,大于正常数据特征范围为×¹,小于正常数据特征范围为×²。

从表 1 可以看出,不同的故障类型在这些数据特征上的表现形式是不同的,因此可以通过一系列的判定条件逐步地将各种故障类型区分开来,这和本文使用的决策树算法的思路是相符的。

2 传感器故障识别方法

2.1 传感器故障识别流程

传感器故障识别方法分为两个阶段,分别是数据特征提取和机器学习模型训练。具体流程见图 1。

模型训练阶段分为以下步骤:

- (1)将原始数据集划分为训练集和测试集;
- (2)构建机器学习模型对训练集进行学习;
- (3)初始模型构建完成后,使用测试集进行预测性能评估,如果初始模型预测精度未达到预期,进行参数调优,直至模型在测试集上的预测精度达到预期阈值;其中,精度是指分类正确的样本数占总样本数的比例。设总样本数为 m ,有 a 个样本分类错误,则

$$\text{精度} = \left(1 - \frac{a}{m}\right) \times 100\%$$

(16)

(4)使用最佳参数重新训练,使用召回率和准确率对模型进行评估。

2.2 最优信号时长的确定方法

分段处理时选取的时间过长或过短都会对特征提取部分产生影响,因此采用合适的分段时长十分重要。本文使用的最优信号时长确定方法是传感器正常数据在分段后计算得到的数据特征离散程度最小。不同的数据特征有不同的最优信号时长,因此需要综合考虑各种数据特征的离散程度确定一个最优信号时长。

常用的衡量数据离散程度的方法有标准差和变异系数,标准差和变异系数越小说明数据离散程度越小。标准差在比较均值差异大的数据集时无法准

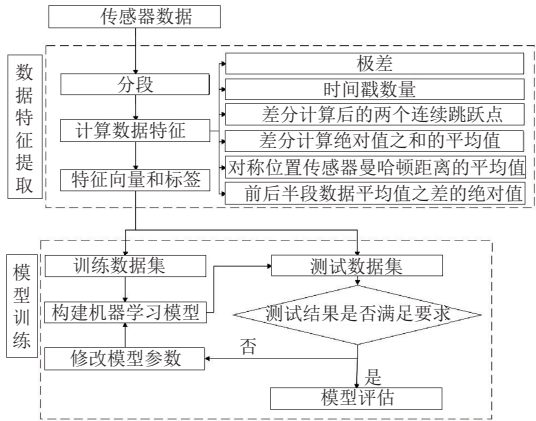


图 1 传感器故障识别流程图

在数据特征提取阶段,首先将数据进行分段,然后根据第 1 章介绍的数据特征计算方法计算出每段数据的数据特征,得到特征向量。

确反映其离散程度,而变异系数^[11](Coefficient of Variation, CV)是标准差与均值的比值,消除了均值的影响,适用于比较均值差异比较大的数据集的离散程度,变异系数的计算公式如式(17)所示, σ 和 μ 分别表示数据集的标准差和均值。当均值接近0时,微小的扰动都会对变异系数产生很大的影响。因此对于均值接近0的数据集使用标准差衡量其离散程度,对于均值远大于0的数据集使用变异系数衡量其离散程度。

$$CV = \frac{\sigma}{\mu} \times 100\% \quad (17)$$

2.3 决策树算法

由于传感器故障有许多类型,且不同的传感器故障类型在同一数据特征上的表现有可能相同。为了能够在这一多分类任务上准确地区分出各种传感器故障类型,并且确保模型具有可解释性,本文采用决策树算法对数据特征进行自主学习,结果可以直观地展示出故障类型的判别路径。

一颗决策树由一个根节点、若干个内部节点和若干个叶节点构成。根节点包含全部样本,内部节点进行特征测试,叶节点输出最终决策结果。每个节点所包含的样本集合通过特征取值划分到子节点。从根节点到每个叶节点的路径对应于一个特定的判定测试序列,反映了在给定特征条件下如何逐步决策的过程^[12]。决策树的划分方法一般基于信息增益^[13]、增益率^[14]、基尼指数^[15]。

本文采用的是ID3决策树学习算法^[16],该算法时间复杂度低,计算效率高,并且逻辑清晰简洁,易于实现代码构建。

(1)ID3采用自顶而下的方式构建决策树,分为以下三个步骤:

- a. 从根节点开始,计算所有特征的信息增益;
- b. 选择信息增益最大的特征作为当前节点的划分特征;
- c. 递归地对该特征的每个取值建立子节点,直到达到指定条件^[12]。

(2)选择信息增益最大的特征时的计算方法如下:

a. 信息熵计算

信息熵(Entropy)是衡量数据集纯度的一种指标,信息熵越大,数据越混乱。

设数据集 S 中第 i 类样本所占的比例为 p_i ,则数据集 S 的信息熵为:

$$\text{Ent}(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (18)$$

式中: c 为类别数量。

b. 信息增益计算

信息增益衡量某个特征对分类的贡献,计算公式如下所示:

$$\text{Gain}(S, A) = \text{Ent}(S) - \sum_{v=1}^V \frac{|S_v|}{|S|} \text{Ent}(S_v) \quad (19)$$

式中: A 为某个特征; V 为特征 A 的所有可能取值; S_v 为 $A=v$ 时的子数据集; $|S_v|/|S|$ 为子数据集 S_v 占整个数据集的比例。

c. 信息增益越大,则使用该特征进行划分所获得的纯度提升越大,选择信息增益最大的特征 A^* 进行划分:

$$A^* = \arg \max_A \text{Gain}(S, A) \quad (20)$$

传感器数据特征如极差、差分计算绝对值之和的平均值、对称位置传感器曼哈顿距离的平均值等均为连续特征。经典的决策树算法对于连续特征的阈值设定方法如下:首先将连续特征的值从小到大进行排列,然后用每两个相邻值的中位数作为候选阈值将连续特征值划分为两部分,再基于这些候选阈值计算样本集的信息增益,最后选取信息增益最大的候选阈值作为最终阈值。本文将阈值设为训练集中不大于经典的决策树算法确定的阈值的最大值,这样做可以使最终决策树使用的阈值都在训练集中出现过,防止训练集中传感器故障数据的故障程度过大影响传感器故障模型的识别精度。

3 实桥数据验证

为了验证上述传感器故障识别方法的可行性,首先收集实际桥梁健康监测的数据,然后构建传感器故障识别模型,最后用全时段数据验证模型的可靠性。

3.1 数据来源

本文以某斜拉桥为实际工程背景,该桥为独塔单索面四跨连续钢—混凝土组合梁斜拉桥,跨径布置50 m+2×220 m+50 m,总长540 m,半漂浮体系,上部结构采用钢—混凝土组合梁,主梁采用整体式单箱三室箱型断面。该桥健康监测系统于2020年10月开始运行,安装的传感器类型有温度、风速、挠度、位移、应变、索力、加速度等传感器。

本文以应变传感器数据作为实验数据,选取了该桥2023年1月—2024年12月两年的应变传感器

数据验证本文所提出的传感器故障识别方法。该桥在主梁跨中断面布设了光纤光栅应变传感器,采样频率为 20 Hz。本文使用的是位于底板的编号为

SEW0508 传感器的数据,其横桥向对称位置传感器编号为 SEW0518,图 2 所示为 SEW0508 的正常传感器数据及故障传感器数据。

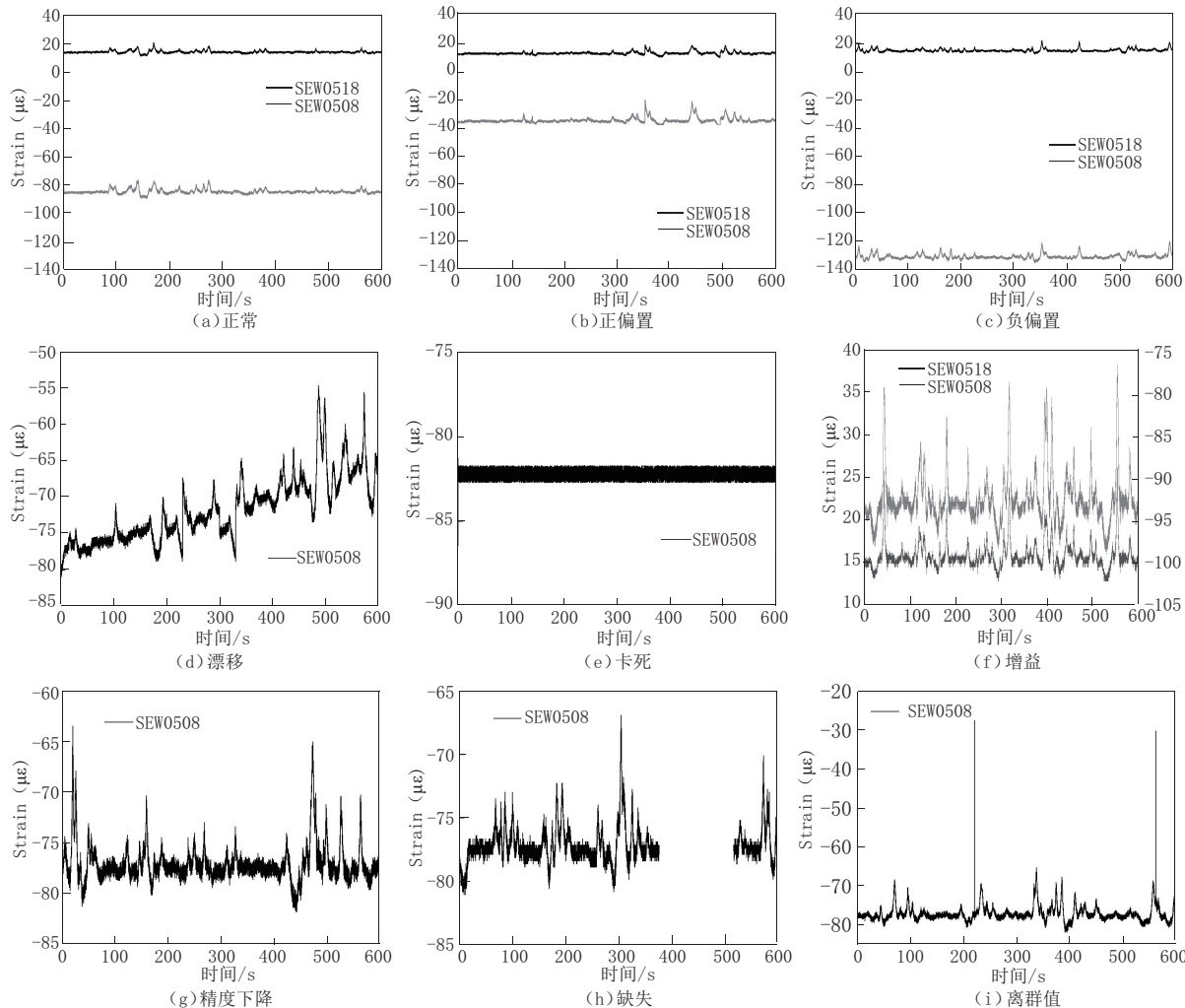


图 2 正常传感器数据及各类故障传感器数据

3.2 数据特征提取

(1) 最优信号时长的确定

采用不同信号时长计算两年的传感器正常数据的数据特征的离散程度的结果见图 3,差分计算绝对值之和的平均值和前后半段数据平均值之差的绝对值的均值接近于 0,使用标准差衡量其离散程度;对称位置传感器曼哈顿距离的平均值使用变异系数衡量其离散程度。从图 3 中可以看出,差分计算绝对值之和的平均值的标准差随着信号时长的增大不断减小,在信号时长大于 8 min 后趋于稳定;前后半段数据平均值之差的绝对值的标准差随着分段时长的增大先减小后增大,在 10 min 标准差最小;信号时长对于对称位置传感器曼哈顿距离的平均值的变异系数影响不大;综合考虑后最优信号时长确定为 10 min。将 2 年的数据按 10 min 分段后,共计 105 264

段数据,然后对这些数据进行人工标注。

(2) 计算数据特征

根据第 1 章提出的传感器数据特征计算方法,对各段数据进行了特征向量的提取。图 4 和图 5 分别展示了不同传感器故障类型 100 段数据的前后半段数据平均值之差的绝对值和差分计算绝对值之和的平均值,从图中可以看出正常数据的这两个特征较为稳定,在均值附近小范围波动。从图 4 可以明显看出,传感器漂移的前后半段数据平均值之差的绝对值显著高于其他类型的传感器数据。这一结果表明,可以在传感器漂移与其他类型数据的前后半段数据平均值之差的绝对值之间设定一个阈值:当测得的前后半段数据平均值之差的绝对值超过该阈值时,可判定为传感器漂移,而低于该阈值则认为认为是其他传感器数据。从图 5 的结果来看,传

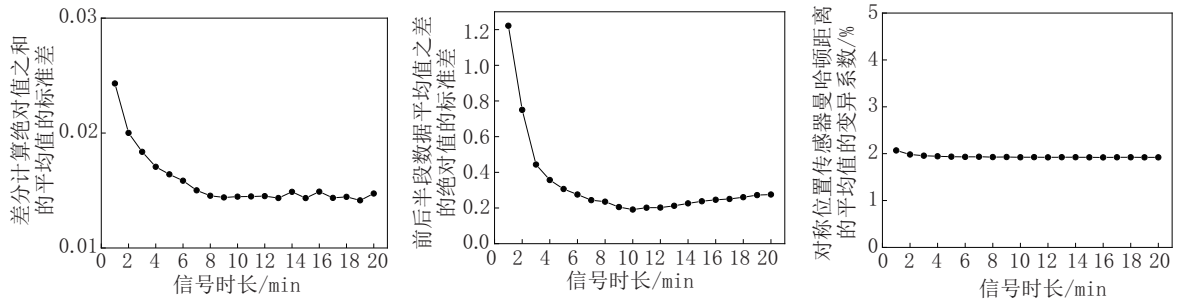


图3 分段时长的数据特征的离散程度图

传感器增益和精度下降的差分计算绝对值之和的平均值明显高于其他传感器数据。因此,针对这一特征,也可以设定一个相应的阈值:若差分计算绝对值之和的平均值超出该阈值,则判断为传感器增益

或精度下降,反之则归为其他传感器数据。决策树学习的方式是基于划分的,上述方法为基于决策树算法进行传感器故障识别提供了有效的理论依据。

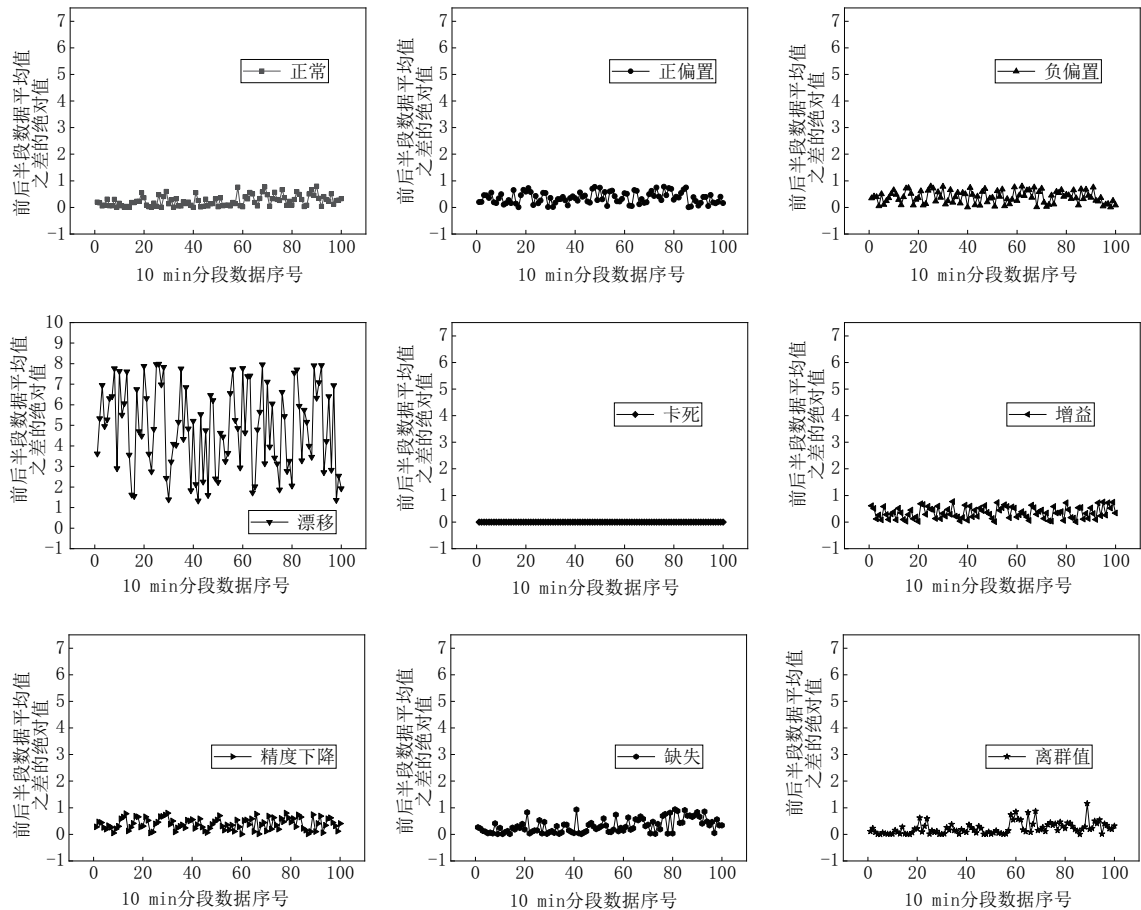


图4 各种传感器故障类型数据的前后半段数据平均值之差的绝对值

3.3 决策树模型训练

将分段数据的特征向量和标签作为样本进行决策树模型训练,每种传感器故障类型各取500个样本,然后采用分层抽样策略将总体样本按4:1的比例作为训练集和测试集。

决策树模型的性能和树的深度有关,本文取树的深度为1~10分别训练模型,得到的树的深度和训练集、测试集精度的关系见图6。从图6中可以看出,随着树的深度的增长,精度逐渐变高,当树的深

度大于等于7时,精度接近于100%,因此认为对于本桥的应用可以将树的深度设为7。

当树的深度为7时,训练后的传感器故障识别决策树可视化模型见图7。根据这个模型得到的测试集分类结果混淆矩阵见图8,图中右侧是召回率,下侧是准确率,右下角为该模型在测试集上的识别精度。从图8可以看出,决策树模型在测试集上的识别结果表现良好,精度为96.3%。由于部分正偏置、负偏置、漂移、增益、精度下降故障未被模型正确识

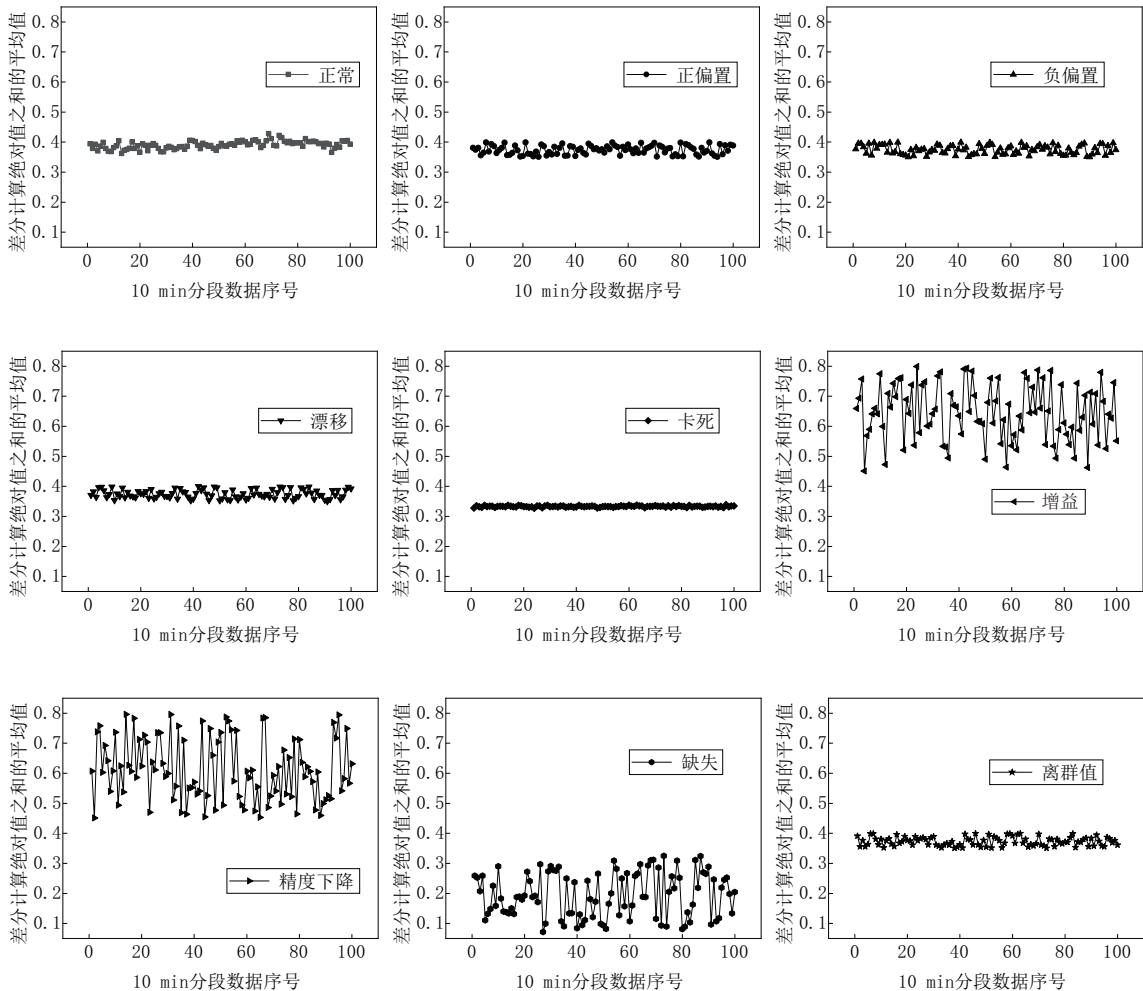


图5 各种传感器故障类型数据的差分计算绝对值之和的平均值

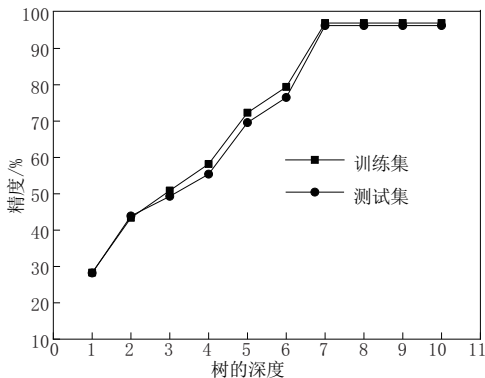


图6 树的深度和精度的关系图

别,其召回率分别为91.0%、94.0%、92.0%、97.0%和93.0%;且大多被误分类为正常,导致正常类别的准确率比较低,为78.7%;其余类别的准确率均大于94.9%。因此可以认为该决策树模型在测试集上展现出较高的整体识别性能。

3.4 全时段数据测试

为了进一步考察模型的性能,采用上文训练后的传感器故障识别模型对2023年1月—2024年12月2年(除训练数据外)的数据进行识别,与人工标记

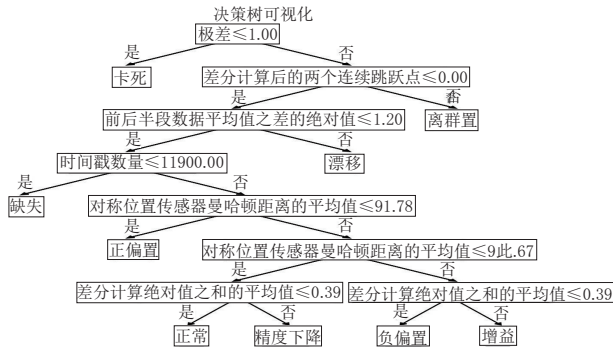


图7 传感器故障识别决策树可视化模型

的故障类型进行对比。数据的模型识别结果和人工标记故障类型的混淆矩阵见图9。由于实际传感器数据中正常数据占大多数,图9不能体现出模型对于传感器故障数据的识别能力,因此针对性绘制了仅针对故障样本的混淆矩阵,见图10,以更清晰地呈现模型在故障识别任务中的性能表现。

从图9可以看出,决策树模型对两年测试数据的识别精度达到99.4%;图10右下角的数据显示,该模型对于传感器故障数据的识别精度为92.7%;正常的召回率和准确率都很高,分别为99.8%和99.7%;由

正常	100	0	0	0	0	0	0	0	0	100%
正偏置	9	91	0	0	0	0	0	0	0	91.0%
负偏置	6	0	94	0	0	0	0	0	0	94.0%
漂移	5	1	2	92	0	0	0	0	0	92.0%
卡死	0	0	0	0	100	0	0	0	0	100%
增益	0	0	3	0	0	97	0	0	0	97.0%
精度下降	7	0	0	0	0	0	93	0	0	93.0%
缺失	0	0	0	0	0	0	0	100	0	100%
离群值	0	0	0	0	0	0	0	0	100	100%
	78.7%	98.9%	94.9%	100%	100%	100%	100%	100%	100%	96.3%
	正常	正偏置	负偏置	漂移	卡死	增益	精度下降	缺失	离群值	

图8 混淆矩阵

正常	94708	0	0	85	10	0	46	0	20	99.8%
正偏置	78	655	0	30	0	0	0	0	0	85.8%
负偏置	36	0	772	60	0	0	0	0	0	88.9%
漂移	35	15	15	698	0	0	0	0	0	91.5%
卡死	20	0	0	0	659	0	0	0	0	97.1%
增益	0	0	28	0	0	653	30	0	0	91.8%
精度下降	46	0	0	0	0	0	559	0	0	92.4%
缺失	11	0	0	0	0	0	0	563	0	98.1%
离群值	25	0	0	0	0	0	0	0	907	97.3%
	99.7%	97.8%	94.7%	80.0%	98.5%	100%	88.0%	100%	97.8%	99.4%
	正常	正偏置	负偏置	漂移	卡死	增益	精度下降	缺失	离群值	

图9 模型识别结果和人工标记的混淆矩阵

正偏置	78	655	0	30	0	0	0	0	0	
负偏置	36	0	772	60	0	0	0	0	0	
漂移	35	15	15	698	0	0	0	0	0	
卡死	20	0	0	0	659	0	0	0	0	
增益	0	0	28	0	0	653	30	0	0	
精度下降	46	0	0	0	0	0	559	0	0	
缺失	11	0	0	0	0	0	0	563	0	
离群值	25	0	0	0	0	0	0	0	907	
	正常	正偏置	负偏置	漂移	卡死	增益	精度下降	缺失	离群值	92.7%

图10 实际故障数据的混淆矩阵

于部分偏置故障被识别为正常,导致其召回率低于其他故障类型;部分漂移故障被识别为正常或偏置,并且其他故障类型也会被识别为漂移,因此漂移的召回率和准确率都较低,分别为91.5%和80.0%。综上所述,本文所提出的方法具有很好的传感器故障识别能力。另外,对比图9和图8发现,全时段数据(除训练数据外)测试结果比训练数据的测试集结果的精度高,这是由于用于训练的各种类型数据的样本数量都是相同的,而实际情况下正常数据占大多

数;这也反映了本文的传感器故障识别模型很少将正常数据识别为故障数据,虚假预警率低。

4 结 语

本文结合传感器故障类型,选取了多种传感器数据特征,提出了基于决策树算法的传感器故障识别方法,通过实测数据验证了其可靠性,主要得到以下结论:

(1)该方法利用极差、时间戳数量、差分计算后的两个连续跳跃点、差分计算绝对值之和的平均值、对称位置传感器曼哈顿距离的平均值,以及前后半段数据平均值之差的绝对值作为传感器数据特征。经实桥数据验证,该方法对于故障数据的识别精度达到92.7%。

(2)提出的基于决策树算法的传感器故障识别方法能够有效识别应变传感器数据的故障类型,并可以通过可视化决策路径直观展示故障判别过程。

(3)本文主要针对应变传感器数据,但所提出的方法同样适用于位移、挠度等传感器的故障识别。

需要指出的是,本文所提出的传感器故障识别方法在整体上具有较高的识别精度,但对于偏置、漂移、增益故障的识别仍存在比较多的误判,因此对于这些故障类型的数据特征的优化选择仍存在进一步研究的空间。

参考文献:

- [1] 张启伟.大型桥梁健康监测概念与监测系统设计[J].同济大学学报(自然科学版),2001(1):65-69.
- [2] 谢强,薛松涛.土木工程结构健康监测的研究状况与进展[J].中国科学基金,2001(5):31-34.
- [3] 谈娟,陈欣,曹东.基于模型参数辨识的机载传感器故障诊断[J].电光与控制,2020,27(5):80-84.
- [4] Nikos Kougiatsos, Rudy R Negenborn, Vasso Reppa.Distributed model-based sensor fault diagnosis of marine fuel engines[J]. IFAC-PapersOnLine, 2022, 55(6): 347-353.
- [5] 程曦,漆随平,李志乾,等.降雨量传感器的故障诊断方法研究[J].传感技术学报,2021,34(1):89-95.
- [6] Kaluder S, Fekete K, Jozsa L, et al. Fault diagnosis and identification in the distribution network using the fuzzy expert system[J].Eksplotacja i Niezawodność, 2018, 20(4): 621-629.
- [7] 刘张浩,秦小平,李力,等.基于互相关分析的桥梁应变传感器故障诊断[J].公路交通技术,2022,38(4):77-83.
- [8] Bao Yuequan, Tang Zhiyi, Li Hui, et al. Computer vision and deep learning - based data anomaly detection method for structural health monitoring[J].Structural Health Monitoring, 2019, 18(2): 401-421.
- [9] 许翔.基于大数据分析的悬索桥状态评估及动态预警方法研究[D].南京:东南大学,2019.
- [10] Xu X, Huang Q, Ren Y, et al. Sensor fault diagnosis for bridge