

DOI: 10.20189/j.cnki.CN/61-1527/E.202502002

基于可扩展置信规则库的工业设备故障诊断方法

冯树成, 贺 维, 马 宁, 周国辉*

(哈尔滨师范大学 计算机科学与信息工程学院, 黑龙江 哈尔滨 150025)

摘 要: 针对工业设备故障诊断中常面临的样本量大、参数繁杂等问题, 提出了一种新型可扩展置信规则库模型 (Extended Belief Rule Base, EBRB), 该模型具有良好的可解释性和合理性, 便于技术人员根据模型输出进行可靠的决策。首先, 运用极致梯度提升方法对大量特征进行重要性评估, 选取最具诊断价值的特征, 有效降低问题维度; 然后, 在推理阶段, 引入投影协方差矩阵自适应策略优化方法进行参数优化, 从而提升模型性能; 最后, 在输出阶段, 采用基于阈值的方式限制激活规则数量, 并使用改进的证据推理引擎, 以提高推理效率和准确性。结果表明: 提出的 EBRB 在多个故障诊断任务中表现优异, 在滚动体故障诊断中, 准确率达 98.15%, 与 K 最邻近模型并列第一; 在内圈故障诊断中, 准确率为 90.74%, 优于所有其他对比模型; 在外圈故障诊断中, 准确率为 88.89%, 同样高于其他对比模型。实验结果验证了该模型在处理复杂工业设备故障诊断任务中的高准确性和鲁棒性, 为工业设备故障诊断领域提供了一种新的有效解决方案。

关键词: 可扩展置信规则库; 故障诊断; 特征选择; 参数优化

中图分类号: TP391

文献标志码: A

文章编号: 2097-2741(2025)02-013-09

Fault Diagnosis Method for Industrial Equipment Based on Extended Belief Rule Base

FENG Shucheng, HE Wei, MA Ning, ZHOU Guohui*

(College of Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China)

ABSTRACT: To address the problems of large sample sizes and complex parameters in industrial equipment fault diagnosis, a novel EBRB (Extended Belief Rule Base) model was proposed. This model had excellent interpretability and rationality, making it easy for technical personnel to make reliable decisions based on model outputs. First, the model employed the eXtreme Gradient Boosting method to assess the importance of numerous features and selected the most diagnostically-relevant ones, effectively reducing the dimensionality of the problem. Then, in the inference

收稿日期: 2024-11-14

基金项目: 火箭军工程大学重点实验室开放基金 (AAIE-2023-0102)

第一作者: 冯树成 (2000—), 男, 硕士研究生, 主要从事置信规则库研究。E-mail: feng0328sc@163.com

***通信作者:** 周国辉 (1974—), 男, 教授, 主要从事证据推理与人工智能研究。E-mail: zhoughuohui@hrhnu.edu.cn

引用格式: 冯树成, 贺维, 马宁, 等. 基于可扩展置信规则库的工业设备故障诊断方法[J]. 火箭军工程大学学报, 2025, 39 (2): 13-21. FENG Shucheng, HE Wei, MA Ning, et al. Fault diagnosis method for industrial equipment based on extended belief rule base [J]. Journal of Rocket Force University of Engineering, 2025, 39 (2): 13-21.

stage, the projection covariance matrix adaptive evolutionary strategy method was introduced to optimize the parameters, thereby improving the model's performance. Finally, in the output stage, a threshold-based approach was adopted to limit the number of activated rules, and an improved evidential reasoning engine was utilized to enhance inference efficiency and accuracy. The experimental results indicate that the proposed EBRB model performs exceptionally well in multiple fault diagnosis tasks. In the fault diagnosis of rolling elements, it achieves an accuracy of 98.15%, tying for first place with the K-nearest neighbors model. In the fault diagnosis of the inner circles, the model attains an accuracy of 90.74%, surpassing all the other comparative models. In the fault diagnosis of the outer circles, the model achieves an accuracy of 88.89%, which exceeds that of other comparative models. The results verify the high accuracy and robustness of the model in complex industrial equipment fault diagnosis tasks, thereby providing a new effective solution in the field of industrial equipment fault diagnosis.

KEYWORDS: extended belief rule base; fault diagnosis; feature selection; parameter optimization

现代工业设备的组成和结构愈加繁杂,各部件之间的关联也愈加紧密^[1],这些特性不仅使设备易受多方面因素影响而频繁发生故障^[2],还大幅增加了故障诊断的难度。系统复杂性导致的运行异常往往难以及时察觉^[3],其根源也难以准确快速定位。这不仅可能干扰正常生产计划,造成巨大经济损失^[4],更有可能因诊断不及时或不准确而引发严重事故,威胁生产安全^[5]。鉴于此,提出一种可靠、精确的工业设备故障诊断方法变得尤为关键。

在工业设备故障诊断领域,学者们进行了广泛深入研究并取得了显著成果^[6]。根据建模机制,可将这些方法分为3类。一是数据驱动模型,通过数据输入建立模型以检测故障。例如,胡芑庆等^[7]提出了结合经验模态分解(Empirical Mode Decomposition, EMD)和深度卷积神经网络(Deep Convolutional Neural Network, DCNN)的创新方法,实现了对行星齿轮箱状态的智能分类。二是知识驱动模型,基于设备运行原理进行故障诊断。例如,董家祥等^[8]在通用机械设备故障诊断模型之上引入了知识图谱,提高了机械故障诊断的准确性。三是知识与数据双驱动模型,将专家知识与数据相互结合,既具有数据驱动模型的高准确性,又兼具知识驱动模型的强可解释性。例如,邓丰等^[9]通过数据驱动与知识驱动的优势互补,实现了对弱非线性、存在畸变偏移和不规则畸变的高阻接地故障灵敏检测。

数据驱动模型虽准确度高,但缺乏可解释性,难以帮助工程师理解故障根源;知识驱动模型虽

可解释性强,但准确率往往不足。数据驱动与知识驱动相结合的混合模型结合了数据驱动模型准确率高和知识驱动模型可解释强的优点。因此,本文研究重点是利用混合模型解决工业设备故障诊断方面的难题。

可扩展置信规则库(Extended Belief Rule Base, EBRB)^[10]作为一种知识与数据相结合的双驱动模型,在数据驱动模型中融入专家知识,使得先行属性扩展为置信分布,从而能够更好地表示知识的不确定性,已成功应用于传感器的活动识别^[11]、环境治理成本预测^[12]、环境投资预测^[13]、碳排放预测^[14]、甲状腺结节诊断^[15]等领域。然而,在实际工业设备故障诊断过程中,EBRB可能会遇到以下3个问题:首先,工业设备的初始参数往往众多,EBRB模型该如何定义自己的前提属性^[16];其次,当规则库中规则数量过多时,必须找到有效的方法来解决推理效率低下的问题^[17];最后,还需要探索如何在EBRB模型的基础上进一步提高模型的精度^[18]。

针对上述3个问题,本文在EBRB模型的基础上进行了改进。首先,通过极端梯度提升(eXtreme Gradient Boosting, XGBoost)方法评估特征属性,筛选出重要属性来定义EBRB的前提属性;其次,采用基于阈值的方法限制激活规则的数量,并使用改进的证据推理(Evidence Reasoning, ER)引擎,提高推理效率;最后,利用投影协方差矩阵自适应策略(Projection Covariance Matrix Adaptive Evolutionary Strategy, P-CMA-ES)优化了EBRB模型参数,提升模型的准确性。

本文所提改良的EBRB模型的创新性主要体现在2个方面: 一是与原有的EBRB模型相比, 通过特征选择、限制激活规则数量、参数优化等方式, 提高了模型的准确性; 二是改进的EBRB不仅具有较好的准确性, 同时由于引入了专家系统, 更具有可解释性, 能有效实现工业设备故障诊断。

1 问题描述

在使用EBRB构建工业设备故障诊断模型时, 需要解决以下3个问题。

问题1: 在建立原始EBRB模型时, 如何有效定义EBRB的前提属性。EBRB是一种数据与知识双驱动的混合模型, 具有一定的可解释性。然而, 过多的特征会使模型变得非常复杂, 难以解释每个特征对最终决策的影响, 降低了模型的可解释性。且大多数情况下, 并非所有的特征都对预测有帮助, 无关特征的引入可能会给模型带来额外的噪声, 降低模型的准确性, 可以描述为

$$\begin{aligned} A = \{a_1, a_2, \dots, a_M\} \rightarrow A' = \{a_1, a_2, \dots, a_{N'}\}, \\ N' \leq M' \end{aligned} \quad (1)$$

式中: A 为原始属性集; A' 为筛选后的属性集。 M' 与 N' 分别为原始属性集与筛选后的属性集中的属性个数; $a_i (i \leq M')$ 为第 i 个属性。

问题2: 如何避免规则库规则数目过大导致的推理效率低下。EBRB模型会为先行属性设置信念分布生成规则, 当训练集样本数量相对较少时, 这种一一对应的关系并不会造成明显问题。但是, 随着训练集规模的增长, 规则库的规模会急剧膨胀, 导致模型推理效率严重下降。一方面, 这种规则库膨胀问题会增加计算复杂度和存储需求; 另一方面, 过大的规则库可能会引入冗余信息, 反而降低模型的泛化能力和整体性能。因此, 为了提高EBRB模型的应用效果, 亟需开发有效的策略来控制和优化规则库的规模, 在保持模型优势的同时, 以提高其推理效率和实用性, 可以描述为

$$\begin{aligned} \Theta = \{R_1, R_2, \dots, R_{j'}\} \rightarrow \Theta' = \\ \{R_1, R_2, \dots, R_{k'}\}, k' \leq j' \end{aligned} \quad (2)$$

式中: Θ 为原始EBRB中待融合的规则集合; Θ' 为使用某种方法后待融合的规则集合; j' 为原始规则库总数; k' 为优化后的规则库总数; $R_{i'} (i' \leq k')$ 为第 i' 条规则。显然, Θ' 中的规则数量少于 Θ 中

的规则数量。

问题3: 如何进一步提高模型的精度。在原始EBRB中, 规则权重与属性权重等参数主要依赖专家知识进行分配。然而, 随着数据集规模的不断扩大, 专家难以为每条规则和属性精确分配权重, 这可能导致模型准确性和有效性显著下降。根据EBRB的原理, 最终结果与规则权重、属性权重之间存在着强烈的因果关系。因此, 如何在大规模数据环境下科学、高效地确定这些权重, 成为提升EBRB模型性能迫切需要解决的问题。开发一种能够自动化或半自动化分配权重的方法, 不仅能够减轻专家的负担, 还能够更好地捕捉数据中的隐含模式, 从而提高模型的整体性能和适应性, 可以描述为

$$P = (x, y, \gamma) \rightarrow P_{\text{best}} = \text{optimize}(x, y, \gamma') \quad (3)$$

式中: P 为原始参数集; P_{best} 为优化后的参数集; x 为数据集; y 为评估结果; γ 为原始参数; γ' 为优化后的参数; $\text{optimize}(\cdot)$ 为优化函数。

2 可扩展置信规则库模型

针对EBRB模型面临的3个问题, 本文提出了一套多层次的优化策略。

2.1 模型整体结构

本文所提出的EBRB优化模型框架如图1所示, 该框架包含了特征选择过程、推理过程以及训练过程。在特征选择过程中, 先采用XGBoost方法对特征进行选择。在推理过程中, 进行参数优化, 先将数据库划分为训练数据集和验证数据集; 然后采用P-CMA-ES方法对训练数据进行优化。输出过程中, 先在规则融合之前采用基于阈值的方法限制激活规则的数量, 然后使用改进后的ER引擎缩短推理时间。

2.2 模型特征选择

本文使用XGBoost特征重要性分析作为特征选择方法^[19], 该方法在许多机器学习和数据挖掘竞赛中被广泛使用。

使用XGBoost方法, 一个有 n 个样本, M 个特征值的数据集定义为

$$\hat{y}_i = \sum_{t=1}^T f_t(x_i), f_t \in F \quad (4)$$

式中: $\hat{y}_i$ 为预测值; x_i 为输入样本; T 为决策树的总数; f_t 为决策树的结构; F 为树空间; 目标是最

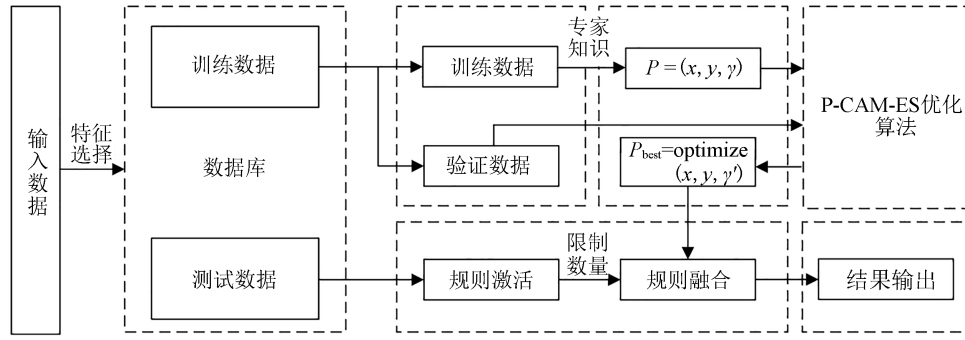


图 1 改良后的 EBRB 模型整体结构

Fig. 1 Architecture of improved EBRB model

小化 O_{bj}

$$O_{bj} = \sum_{i=1}^n L(y_i, \hat{y}) + \sum_{i=1}^l \Omega(f_i) \quad (5)$$

式中: O_{bj} 为优化目标; $L(\cdot)$ 为损失函数; $\Omega(f_i)$ 为正则化项, 是 L 在第 t 次迭代时的二阶泰勒级数。损失函数为

$$L^{(t)} = \sum_{i=1}^k [L(y_i, y_i^{t-1} + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x)) + \Omega(f_i)] \quad (6)$$

式中: $L^{(t)}$ 为 L 的第 t 次迭代; g_i 和 h_i 分别为一阶和二阶梯度。增益函数为式 (7), 作为确定最优分裂节点的度量

$$g_{ain} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \eta} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \eta} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \eta} \right] - \gamma \quad (7)$$

式中: g_{ain} 代表增益函数; I_L 和 I_R 分别为分割后左右节点的样本; $I = I_L \cup I_R$; η 和 γ 为惩罚参数。增益表示树的每个分裂的增益得分, 最终的特征重要性得分由平均增益计算。平均增益为所有树的总增益除以每个特征的分割总数, 特征重要性得分越高, 相应的特征越重要。在本文的 EBRB 规则库建立过程中, 选取最重要的 2 个参数进行规则库构建。

2.3 模型构建

在完成特征选择后, 根据选择的特征建立 EBRB 规则库, 第 k 条扩展信念规则为

$$R_k: \text{if } U_1 \text{ is } \{ (A_{1,1}, \alpha_{1,1}^k), (A_{1,2}, \alpha_{1,2}^k), \dots, (A_{1,J_1}, \alpha_{1,J_1}^k) \} \wedge \dots \wedge U_M \text{ is } \{ (A_{M,1}, \alpha_{M,1}^k), (A_{M,2}, \alpha_{M,2}^k), \dots, (A_{M,J_M}, \alpha_{M,J_M}^k) \} \\ \text{then } \{ (D_1, \beta_1^k), (D_2, \beta_2^k), \dots, (D_N, \beta_N^k) \}$$

$$\text{s.t. } 0 \leq \alpha_{i,j}^k, \beta_j^k \leq 1, \sum_{j=1}^N \beta_j^k \leq 1,$$

$$\sum_{j=1}^{J_i} \alpha_{i,j}^k \leq 1, i \in \{1, 2, \dots, M\} \quad (8)$$

式中: $U_i (i = 1, 2, \dots, M)$ 为第 i 个前件属性; M 为表前件属性的数量; $A_{i,j} (j = 1, 2, \dots, J_i)$ 为第 i 个前件属性的第 j 个参考值; J_i 为参考值的数量; $\alpha_{i,j}^k (k = 1, 2, \dots, L)$ 为对应参考值的匹配度; ($k = 1, 2, \dots, L$) 中 L 为规则库中规则的数目; $D_n (n = 1, 2, \dots, N)$ 为后件属性的第 n 个参考值; β_n^k 为第 k 条规则的规则后件的第 n 个参考值的置信度。其中, $\sum_{n=1}^N \beta_n^k = 1$ 代表第 k 条规则是完整的, 否则是不完整的。

2.4 模型推理

在设定完相关参数后, 先输入待评估数据, 然后进行 EBRB 模型的输出, 步骤如下。

Step1 由专家确定前件参考值 $A_{i,j}$ 以及后件参考值 D_n 后, 将 $x^k = (x_1^k, x_2^k, \dots, x_M^k)$ 转为如下的置信分布

$$E(x_i^k) = \{ (A_{i,j}, \alpha_{i,j}^k), j = 1, 2, \dots, J_i, i = 1, 2, \dots, M \} \quad (9)$$

式中: $u(A_{i,j})$ 为 $A_{i,j}$ 的效用值, 且需满足 $u(A_{i,j}) < x_i^k \leq u(A_{i,j+1}) (j = 1, 2, \dots, J_i - 1)$ 。 $\alpha_{i,j}^k$ 的计算步骤为

$$\begin{cases} \alpha_{i,j}^k = \frac{u(A_{i,j+1}) - x_i^k}{u(A_{i,j+1}) - u(A_{i,j})}, u(A_{i,j}) \leq x_i^k \leq u(A_{i,j+1}) \\ \alpha_{i,j+1}^k = 1 - \alpha_{i,j}^k, u(A_{i,j}) \leq x_i^k \leq u(A_{i,j+1}) \\ \alpha_{i,t}^k = 0, t = 1, \dots, j-1, j+2, \dots, J_i \end{cases} \quad (10)$$

Step2 计算激活权重。激活权重是一个与规则匹配度 S_i^k 、规则权重 θ_k 、属性权重 δ_i 相关的变量, 具体计算为

$$S_i^k = 1 - d_i^k = 1 - \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2}$$

$$\omega_k = \frac{\theta_k \times \prod_{i=1}^M (S_i^k)^{\delta_i}}{\sum_{l=1}^L [\theta_l \times \prod_{i=1}^M (S_i^l)^{\delta_i}]}, \bar{\delta}_i = \frac{\delta_i}{\max_{i=1,2,\dots,M} \{\delta_i\}} \quad (11)$$

Step3 规则融合。对于满足 $0 \leq \omega_k \leq 1$ 的规则, 使用规则推理算法进行融合, 生成结果 $S(x)$ 满足

$$S(x) = (D_j, \beta_j), j = 1, 2, \dots, N \quad (12)$$

$$\beta_n = [\mu \times \prod_{k=1}^L (\omega_k \beta_n^k + 1 - \omega_k \sum_{i=1}^N \beta_i^k) - \prod_{k=1}^L (1 - \omega_k \sum_{i=1}^N \beta_i^k)] / [1 - \mu \times \prod_{k=1}^L (1 - \omega_k)] \quad (13)$$

$$\mu = [\sum_{j=1}^N \prod_{k=1}^L (\omega_k \beta_j^k + 1 - \omega_k \sum_{i=1}^N \beta_i^k) - (N-1) \prod_{k=1}^L (1 - \omega_k \sum_{i=1}^N \beta_i^k)]^{-1} \quad (14)$$

对于分类问题

$$f(x) = D_j, j = \arg \max_{n=1,2,\dots,N} \hat{\beta}_n \quad (15)$$

针对传统的规则融合在样本数量过多时会产生推理效率低下的问题, 本文方法进行2处优化。首先, 对于满足 $0 \leq \omega_k \leq 1$ 的规则, 使用参数 λ 控制激活规则的数量。对于规则数量较少的规则库, 通过设置 λ 来仅激活前30%权重最高的规则; 对于规则数量较多的规则库, λ 直接用于控制待融合规则的具体数量。其次, 对于传统EBRB分类系统中的输出如式(15)所示。其中, $\beta_n (n=1, 2, \dots, N)$ 值的求法如式(13)所示, 假设 β_a 为最大值、 β_b 为次最大值, 结合式(13)、(15)可以得到

$$\beta_a - \beta_b = [\prod_{k=1}^L (\omega_k \beta_a^k + 1 - \omega_k \sum_{i=1}^N \beta_i^k) - \prod_{k=1}^L (\omega_k \beta_b^k + 1 - \omega_k \sum_{i=1}^N \beta_i^k)] \times \mu / [1 - \mu \times \prod_{k=1}^L (1 - \omega_k)] \quad (16)$$

由此可见, 最后的输出实际上是比较项 $\prod_{k=1}^L (\omega_k \beta_n^k + 1 - \omega_k \sum_{i=1}^N \beta_i^k)$ 的大小。因此, 在最后的ER推理引擎输出时, 仅计算项 $\prod_{k=1}^L (\omega_k \beta_n^k + 1 - \omega_k \sum_{i=1}^N \beta_i^k)$ 对结果有影响。

2.5 参数优化

传统EBRB模型中, 规则权重 θ_k 、 δ_i 以及后件

置信度 β_n^k 都是由专家给出。然而, 在实际过程中, 这些参数会对最后的推理结果产生重要影响。因此, 需要对这些参数进行优化。本研究将这些参数的集合定义为 P , $P = \{\theta_k, \delta_i, \beta_n^k\}$, 为了最小化目标函数, 本文设目标函数为 $\xi(P)$, $\xi(P) = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{i,k} \log P(\hat{y}_i = k)$ 。其中, $y_{i,k}$ 为0, 1变量。取值为1时表示第 i 个样本属于第 k 类 (K 为类别总数; N 为样本数); 取值为0时则表示预测错误。在执行优化过程中, 需要满足以下约束条件

$$\begin{cases} 0 \leq \theta_k \leq 1 \\ 0 \leq \delta_i \leq 1 \\ 0 \leq \beta_n^k \leq 1 \\ \sum_{n=1}^N \beta_n^k = 1 \end{cases} \quad (17)$$

在本文提出的模型中使用P-CMA-ES方法进行参数优化^[20]。与其他算法相比, 如随机搜索^[20], 梯度下降^[22]以及贝叶斯优化^[23], P-CMA-ES可以在可行域内生成解, 这意味着解总是严格满足等式约束, 步骤如下。

Step1 参数初始化。设定初始参数集合 P^0 , 其中 $P = \{\theta_1, \theta_2, \dots, \theta_L, \beta_{1,1}, \beta_{1,2}, \dots, \beta_{L,N}, \delta_1, \delta_2, \dots, \delta_M\}$ 。

Step2 采样。执行抽样操作以生成总体

$$P_i^{g+1} \sim w^g + \sigma^g N(0, C^g), i = 1, 2, \dots, \lambda \quad (18)$$

式中: P_i^{g+1} 为第 $g+1$ 个后代中的第 i 个解; w 为总体的平均值; σ 为步长; N 为正态分布; C^g 为第 g 代的协方差矩阵。

Step3 预测。执行投影操作以满足约束

$$P_i^{g+1}(1 + n_e \times (j-1) : n_e \times j) = P_i^{g+1}(1 + n_e \times (j-1) : n_e \times j) - A_e^T \times (A_e \times A_e^T)^{-1} \times P_i^{g+1}(1 + n_e \times (j-1) : n_e \times j) \times A_e \quad (19)$$

超平面可以表示为 $P_i^{g+1}(1 + n_e \times (j-1) : n_e \times j) = 1$

式中: n_e 为解 P_i^g 中等式约束的变量数量; $j = 1, 2, \dots, N+1$ 为解 P_i^g 中等式约束的数量; $A_e = (1, \dots, 1)_{1 \times N}$ 为参数向量。

Step4 选择。执行选择操作以更新平均值

$w^{g+1} = \sum_{i=1}^{\tau} h_i P_{EA}^{g+1}$, P_{EA}^{g+1} 为第 $g+1$ 代中 λ 个解中的第 i 个解; τ 为后代种群规模。

Step5 调整操作。执行自适应操作来更新协方差矩阵。

上述优化过程递归运行, 直至获得最优解

P_{optimal} , 然后就可以建立最优的 EBRB 模型。

3 案例研究

3.1 实验定义

轴承作为工业机械中不可或缺的核心配套部件, 在设备运行中发挥着关键作用。尽管其结构尺寸相对较小, 却直接影响着整个机械系统的工作状态^[24]。值得注意的是, 轴承也是旋转机械中最易发生故障的组件之一, 统计数据显示约 40% 的旋转机械故障源于轴承^[25]。本文采用的是美国西储大学轴承数据中心轴承数据集, 采用其荷载模式 0 (1 797 r/min) 下的 3 种数据集, 分别对应着 3 种故障, 分别为内圈 (IR) 故障、外圈 (OR) 故障以及滚动体 (B) 故障, 每个数据集均包含 3 个关键特征和 3 个故障程度类别。特征值由 3 个关键测量点获得: 基座加速度数据 (base accelerometer data, BA)、驱动端加速度数据 (drive end accelerometer data, DE) 以及风扇端加速度数据 (fan end accelerometer data, FE)。故障程度则通过不同的划痕深度来定义, 分别为 0.017 78、0.035 56 和 0.053 54 cm, 在本研究中将其简称为故障一、故障二和故障三。原始数据由 12 kHz 采样率的传感器采集而来, 为了降低数据维度并提高信噪比, 对原始数据进行预处理。步骤如下: 将每 1 000 条连续的传感器数据点通过计算均值融合为 1 条数据; 对于不足 1 000 条的剩余数据, 采用均值方法进行融合, 实验中各数据集的基本信息如表 1 所示。

表 1 轴承故障实验数据集

Table 1 Experimental dataset of bearing failure 个

故障类型	总样本	训练样本	测试样本
内圈故障一	122	86	36
内圈故障二	122	86	36
内圈故障三	123	87	36
外圈故障一	122	86	36
外圈故障二	122	86	36
外圈故障三	123	87	36
滚动体故障一	123	87	36
滚动体故障二	122	86	36
滚动体故障三	122	86	36

3.2 实验结果分析

在预处理结束之后, 使用 XGBoost 方法, 通

过进行特征重要性分析的方式, 对每种故障模式进行评估, 结果如表 2 所示。

表 2 轴承属性重要性评估

Table 2 Importance evaluation of bearing features

属性	BA	DE	FE
B	0.29	1.00	0.39
IR	0.66	1.00	0.92
OR	0.66	1.00	0.85

根据评估结果, 选取 DE 与 FE 特征作为 EBRB 规则库建立的属性, 建立 EBRB 规则库。然后, 将建立好的 EBRB 规则库按照 7:3 的比例分为训练数据集和测试数据集。这样的比例有利于准确捕获模式和特征, 并充分评估学习系统的泛化能力和性能。然后, 在训练数据库中, 再次按照 7:3 的比例分割数据库, 通过这 30% 的验证集优化其余 70% 训练数据生成的规则库的参数, 然后输入待评估数据, 使用所提出的改进的 ER 推理引擎评估结果。

该方法在数据集上的效果如图 2 所示。其中, 标签 0、1、2 分别代表故障一、二、三。

3.3 对比试验

对于本次故障诊断任务, 有 3 个属性 3 个类别, 分别通过特征选择、规则约简和参数优化 3 个改进方法解决此问题。为了验证本实验所提出模型的有效性, 将其与 5 个著名的分类方法进行比较, 这些方法包括决策树以及决策树的 2 种集成模型 (Bagging Tree、Boosting Tree)、K 近邻算法 (K-Nearest Neighbor, KNN) 以及支持向量机 (Support Vector Machine, SVM), 比较指标为准确率 and 加权 F1 评分。其中, 准确率是正确分类的样本个数占总样本个数的比例, 而加权 F1 评分是包含精确率 P_i 与召回率 R_i 的综合评分, 其计算过程为

$$P_i = \frac{T_{P_i}}{T_{P_i} + F_{P_i}}$$

$$R_i = \frac{T_{P_i}}{T_{P_i} + F_{N_i}}$$

$$F_i = \frac{2R_i P_i}{P_i + R_i} \quad (20)$$

$$W_i = \frac{c_i}{C}$$

$$F_{\text{weight}} = \sum_{i=1}^N F_i W_i$$

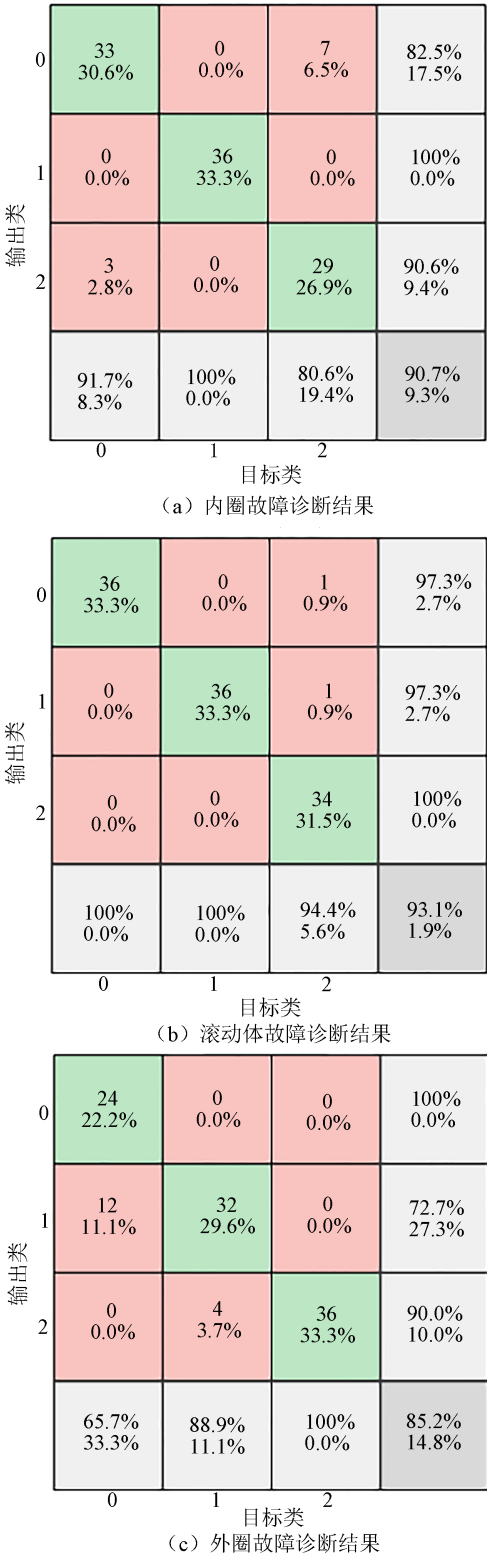


图 2 轴承故障诊断实验结果

Fig. 2 Experimental results of bearing failure diagnosis

式中： T_{Pi} 、 F_{Pi} 与 F_{Ni} 分别为某一类预测当中，被模型预测为正类的正样本、被模型预测为正类的负样本和被模型预测为负类的正样本； c_i 为第 i 样本的数量； C 为样本总数， P_i 、 R_i 、 W_i 、 F_i 均为中间

变量； F_{weight} 为加权 F1 评分。不同模型的故障诊断性能对比如表 3 所示。可以看出，本文所提模型的性能在滚动体故障诊断中与 KNN 并列第一，在内圈故障与外圈故障诊断中，也均优于其他模型。综合 3 个实验结果可以看出，本文所提模型相比其他模型具有明显优势，且由于专家知识的引入，更容易帮助专家认识故障的本质。

表 3 不同模型的故障诊断性能对比实验

Table 3 Comparative experiments of failure diagnosis performame of different models

数据集	模型	准确率	加权 F1 评分
滚动体故障	本文所提模型	98.15	0.981 3
	Bagging Tree	97.22	0.972 1
	Boosting Tree	97.22	0.972 3
	决策树	96.30	0.962 8
	KNN	98.15	0.981 3
	SVM	93.63	0.935 2
内圈故障	本文所提模型	90.74	0.907 1
	Bagging Tree	87.04	0.868 4
	Boosting Tree	87.04	0.868 4
	决策树	85.19	0.850 3
	KNN	87.96	0.877 5
	SVM	77.27	0.773 8
外圈故障	本文所提模型	88.89	0.888 0
	Bagging Tree	84.26	0.841 5
	Boosting Tree	79.63	0.787 5
	决策树	78.70	0.774 8
	KNN	87.96	0.879 4
	SVM	80.00	0.788 5

4 结论

为显著提升故障诊断的准确性和效率，本文提出了一种基于 EBRB 的工业设备故障诊断方法，通过引入 XGBoost 进行特征选择，并采用 PCMA-ES 进行模型参数优化。实验结果表明，所提出的 EBRB 模型在多个故障诊断任务中表现优异。在滚动体故障诊断中，模型准确率达 98.15%，在内圈故障与外圈故障诊断中，准确率分别为 90.74% 和 88.89%，说明该 EBRB 模型在处理复杂的工业设备故障时具备较高的准确性和鲁棒性。同时，

通过对特征的有效筛选和参数的智能优化,降低了模型的复杂性,提高了可解释性,满足了技术人员在实际应用中的需求。然而,本文也存在一定的局限性。首先,获取专家知识以指导模型训练和验证的过程较为困难,而专家知识对于理解数据背后的物理意义和识别潜在的故障模式至关重要,缺乏这种深层次的理解可能会阻碍模型性能的进一步提升;此外,尽管本研究的方法已经在多种故障类型上展现出了良好的性能,但面对更加复杂或未知类型的故障时,模型的表现仍有待检验。未来的研究可以集中在以下几个方面:进一步完善模型,以应对更多类型的故障;探索如何在动态环境中实时更新模型参数;结合其他机器学习方法,增强模型的普适性。通过这些努力,可以为工业设备的安全和可靠运行提供更为坚实的保障。

参考文献

- [1] 周玉蓉, 张巧灵, 于广增, 等. 基于声信号的工业设备故障诊断研究综述[J]. 计算机工程与应用, 2023, 59(7): 51-63.
ZHOU Y R, ZHANG Q L, YU G Z, et al. Review of acoustic signal-based industrial equipment fault diagnosis[J]. Computer Engineering and Applications, 2023, 59(7): 51-63.
- [2] WEN L, LI X Y, GAO L, et al. A new convolutional neural network-based data-driven fault diagnosis method[J]. IEEE Transactions on Industrial Electronics, 2018, 65(7): 5990-5998.
- [3] ZHANG L, XIA H, CHA B. Deep learning based fault diagnosis for chemical process with statistical feature fusion[C]//2023 42nd Chinese Control Conference(CCC). Piscataway, NJ, USA: IEEE, 2023: 8515-8520.
- [4] JARDINE A K S, LIN D M, BANJEVIC D. A review on machinery diagnostics and prognostics implementing condition-based maintenance[J]. Mechanical Systems and Signal Processing, 2006, 20(7): 1483-1510.
- [5] LEE J, NI J, DJURDJANOVIC D, et al. Intelligent prognostics tools and e-maintenance[J]. Computers in Industry, 2006, 57(6): 476-489.
- [6] LEI Y G, YANG B, JIANG X W, et al. Applications of machine learning to machine fault diagnosis: a review and roadmap[J]. Mechanical Systems and Signal Processing, 2020, 138: 106587.
- [7] 胡芑庆, 陈徽鹏, 程哲, 等. 基于经验模态分解和深度卷积神经网络的行星齿轮箱故障诊断方法[J]. 机械工程学报, 2019, 55(7): 9-18.
- HU N Q, CHEN H P, CHENG Z, et al. Fault diagnosis for planetary gearbox based on EMD and deep convolutional neural networks[J]. Journal of Mechanical Engineering, 2019, 55(7): 9-18.
- [8] 董家祥, 翟纪宇, 马昕, 等. 知识驱动的机械设备故障诊断[J]. 计算机科学, 2023, 50(5): 82-92.
DONG J X, ZHAI J Y, MA X, et al. Mechanical equipment fault diagnosis driven by knowledge[J]. Computer Science, 2023, 50(5): 82-92.
- [9] 邓丰, 陈依林, 曾哲, 等. 数据-知识联合驱动的配电网高阻接地故障检测方法[J]. 中国电机工程学报, 2024, 44(24): 9618-9633.
DENG F, CHEN Y L, ZENG Z, et al. Combined data-knowledge driven detection method for high impedance faults in distribution networks[J]. Proceedings of the CSEE, 2024, 44(24): 9618-9633.
- [10] LIU J, MARTINEZ L, CALZADA A, et al. A novel belief rule base representation, generation and its inference methodology[J]. Knowledge-Based Systems, 2013, 53: 129-141.
- [11] CALZADA A, LIU J, NUGENT C D, et al. Sensor-based activity recognition using extended belief rule-based inference methodology[C]//Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ, USA: IEEE, 2014: 2694-2697.
- [12] YANG L H, WANG S H, YE F F, et al. Environmental investment prediction using extended belief rule-based system and evidential reasoning rule[J]. Journal of Cleaner Production, 2021, 289: 125661.
- [13] WANG Y M, YE F F, YANG L H. Extended belief rule based system with joint learning for environmental governance cost prediction[J]. Ecological Indicators, 2020, 111: 106070.
- [14] YE F F, YANG L H, LU H T, et al. A novel data-driven decision model based on extended belief rule base to predict China's carbon emissions[J]. Journal of Environmental Management, 2022, 318: 115547.
- [15] FU C, HOU B B, XUE M, et al. Extended belief rule-based system with accurate rule weights and efficient rule activation for diagnosis of thyroid nodules[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(1): 251-263.
- [16] YANG L H, YE F F, WANG Y M, et al. Ex-

- tended belief rule-based system using bi-level joint optimization for environmental investment forecasting[J]. *Applied Soft Computing*, 2023, 140: 110275.
- [17] ZHANG A, GAO F, YANG M, et al. A new rule reduction and training method for extended belief rule base based on DBSCAN algorithm[J]. *International Journal of Approximate Reasoning*, 2020, 119: 20-39.
- [18] HUANG H Y, LIN Y Q, SU Q, et al. An ensemble approach for extended belief rule-based systems with parameter optimization[J]. *International Journal of Computational Intelligence Systems*, 2019, 12(2): 1371-1381.
- [19] QIU Y G, ZHOU J, KHANDELWAL M, et al. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration[J]. *Engineering with Computers*, 2022, 38(5): 4145-4162.
- [20] HU G X, HE W, SUN C, et al. Hierarchical belief rule-based model for imbalanced multi-classification[J]. *Expert Systems with Applications*, 2023, 216: 119451.
- [21] MENON L, ANTENEODO C. Random search with resetting in heterogeneous environments[J]. *Physical Review E*, 2024, 110(5-1): 054111.
- [22] ZHOU X J, ZHAO C N, HUANG Y Q, et al. Improved fractional-order gradient descent method based on multilayer perceptron[J]. *Neural Networks*, 2025, 183: 106970.
- [23] MAHBOUBI N, XIE J Y, HUANG B. Point-by-point transfer learning for Bayesian optimization: An accelerated search strategy[J]. *Computers & Chemical Engineering*, 2025, 194: 108952.
- [24] QIU M Q, ZHAO Z B. Bearing fault diagnosis using a novel coding-statistic feature combined with NNC[J]. *Journal of Vibroengineering*, 2022, 24(5): 848-861.
- [25] PENG X J, ZHANG B K, GAO D. Research on fault diagnosis method of rolling bearing based on 2DCNN[C]//2020 Chinese Control And Decision Conference (CCDC). Piscataway, NJ, USA: IEEE, 2020: 693-697.

(编辑: 夏菁)