

基于集成深度随机森林算法的智能电厂 设备健康评估方法

庄保乾, 韩路, 李晓虎, 高社民, 刘少阳
(新疆准东特变能源有限责任公司, 新疆昌吉 831700)

摘要: 准确地评估电厂设备健康状态, 对电厂安全稳定生产、提高设备运行安全性具有十分重要的意义; 针对当前电厂设备健康评估方法存在预测精度不高的问题, 提出了一种基于集成深度随机森林算法的智能电厂设备健康评估方法; 详细介绍电厂设备健康评估系统结构, 且分析了健康评估数据结构与影响因素; 将设备评估分为6类不同的层级, 使得设备健康状态分析更方便; 结合深度学习与集成学习技术, 提出了集成深度随机森林算法; 通过仿真实验分析验证了提出方法的有效性; 结果表明, 所提方法对电厂设备健康评估准确度达到97%, 与其他评估方法相比, 文章提出的算法具有最高的健康评估准确度。

关键词: 设备健康评估; 深度随机森林; 集成学习; 集成深度森林

Health Status Assessment Method of Intelligent Power Plant Equipment Based on Integrated Deep Random Forest Algorithm

ZHUANG Baoqian, HAN Lu, LI Xiaohu, GAO Shemin, LIU Shaoyang
(Xinjiang Zhundong TBEA TBEA Energy Co., Ltd., Changji 831700, China)

Abstract: To accurately assess the health status of power plant equipment, it is of great significance to ensure the safe and stable production of power plants and improve the safety of equipment operation. Aimed at low prediction accuracy in current power plant equipment health assessment methods, an intelligent power plant equipment health assessment method based on the integrated deep random forest algorithm is proposed. Firstly, the power plant equipment health assessment system structure is introduced in detail, and the health assessment data structure and influencing factors are analyzed. Secondly, the equipment evaluation is divided into six different levels, which makes the equipment health analysis more convenient. Then, combined with deep learning and integrated learning technology, an integrated deep random forest algorithm is proposed. Finally, simulation experiments verify the effectiveness of the proposed method. The results show that the evaluation model accuracy of the proposed method can reach up to 97%, and the proposed algorithm has the highest accuracy in health assessment compared with other evaluation methods.

Keywords: equipment health assessment; deep random forest; integrated learning; integrated deep forest

0 引言

电力工业作为国民经济的先行行业, 是推动和实现中国式现代化的动力之源, 也是引领中国经济走进新时代的重要力量^[1]。随着我国经济的迅速发展, 电力资源需求不断扩大, 使得电力系统容量增大、电网规模扩大。为了提高电力系统的供能效率, 电厂电力系统自动化规模增大, 电力设备现代化程度不断提高, 电厂电力系统安全运行的难度增加, 导致电厂设备故障造成的损失概率越来越大^[2]。当前, 我国电厂设备老化严重、新机组故障诊断难, 使得电厂设备健康评估比较困难。定期的状态监测数据, 分析故障根源, 确定设备的健康状态, 构建精确的电厂设备健康评估方法, 对电厂安全稳定生产、提高设备运行安全性

具有十分重要的意义, 且对提高电厂定点巡检智能化具有深远的影响^[3]。

随着电力需求的不断提高, 电厂设备检修从最初的故障检修发展到计划检修, 但是渐渐不适应现代电力系统的运行需求^[4]。状态检修(CBM, condition based maintenance)^[5]根据当前设备的工作状态, 利用精确的设备状态检测技术, 通过对比历史状态数据, 识别出故障的早期征兆, 对故障情况做出判断, 给出设备的健康状态, 并确定检修的部位以及时机^[6]。当前, 电厂设备健康评估方法包括基于模型的状态预测方法、基于知识的故障预测方法和基于数据驱动的健康评估方法^[7]。基于模型的设备状态预测方法通过构建系统数学模型, 描述设备状态情况。高磊等

收稿日期: 2024-07-12; 修回日期: 2024-08-24。

基金项目: 新疆准东特变能源有限责任公司北一智慧电厂项目(TBEA-TCNY-ZTJG(2021)-GCFW-2021-003-01)。

作者简介: 庄保乾(1986-), 男, 大学本科, 工程师。

通讯作者: 韩路(1970-), 男, 大学本科, 高级工程师。

引用格式: 庄保乾, 韩路, 李晓虎, 等. 基于集成深度随机森林算法的智能电厂设备健康评估方法[J]. 计算机测量与控制, 2024, 32(8): 322-328.

人^[8]提出了一种面向多元信息模型的设备状态预测方法,能够客观高效地反映设备状态^[9]。该类方法状态预测实时性好,但是数据模型比较复杂,构造难度较大。基于知识的故障预测方法不需要建立精确的故障预测模型,而是利用专家经验知识对识别故障与健康状态进行定性分析。董理科等人^[10]提出了一种基于知识图谱技术的故障预测方法,实现了电力变压器运行的健康预测。徐凤建等人^[11]构建了一个用于设备故障预测的专家系统,实现了基于知识推理的设备故障预测。此类方法对专家的领域知识具有较强的依赖性,构建的故障预测模型缺乏客观性。基于数据驱动的健康评估方法为了避免数学模型和专家先验知识的局限性,利用采集的数据对隐藏的设备故障信息进行学习,从而实现设备的健康状态预测。蒋金元^[12]结合深度学习方法,提出了一种基于数据分析的故障预测模型方法。贺川双等人^[13]提出了一种基于主成分分析方法的电力设备运行状态的评价模型,通过实验证明该方法具有一定的可靠性和实用性。由上述文献分析,当前的故障诊断与健康状态预测模型存在以下不足:1)当前故障诊断方法与健康状态预测方法局限于专家先验知识,缺乏客观性;2)单一基于数据分析的设备健康状态预测方法性能泛化性弱,预测精度有待提高;3)电厂设备产生的历史数据存在随机性、非平稳性、高波动性以及周期性等特征,健康评估模型的输入未能考虑到这些特性,使得预测精度较差。

随机森林(RF, random forest)作为一种基于机器学习的预测方法,已经广泛应用于故障诊断和健康状态预测,且预测效果较为出色^[14]。随着电力能源需求的提高与电力设备朝智能化发展,电厂设备面临着故障诊断与健康评估影响因素复杂多变等问题,单一的随机森林评估模型已无法满足智能故障诊断与健康评估的要求。针对当前基于RF的故障诊断与健康评估模型存在预测误差较高、稳定性较差、效率低下等缺陷,本文结合集成学习技术与深度学习技术,提出了基于集成深度随机森林的电厂设备健康评估方法。本文主要贡献在于:1)在电厂设备健康评估系统结构基础上,分析了健康评估数据结构与影响因素,且定义了6类不同的电厂设备健康层级;2)结合深度学习与集成学习技术,提出了集成深度随机森林算法;3)通过电厂锅炉爆管故障与运行数据集验证了本文方法具有更高的评估准确度。

1 电厂设备健康评估系统结构与数据分析

1.1 电厂设备健康评估系统结构

本文提出的设备健康评估方法是建立在软件总线平台上的基础上,通过收集设备实时运行的状态数据,利用总线将数据传输到数据处理模块与模型构建模块,从而达到故障诊断和健康评估效果^[15]。电厂设备健康评估系统结构如图1所示。

如图1所示,电厂设备健康评估系统建立在软件总线平台上,通过软件总线将本文提出的方法与后端数据库发

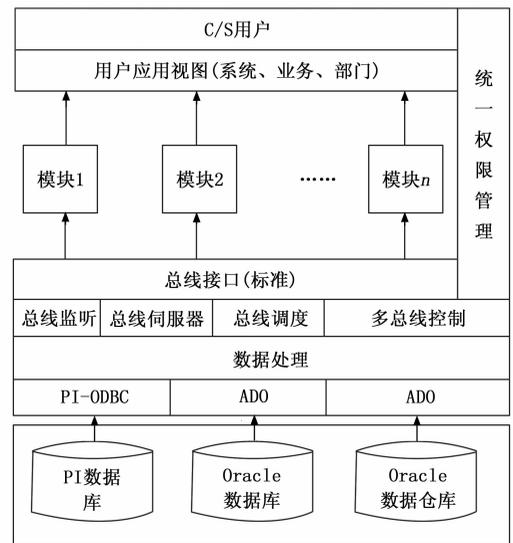


图1 电厂设备健康评估系统结构图

生交互。PI数据库利用PI-ODBC接口方式与总线连接,并且将电厂设备实时运行数据传送到电厂设备健康评估模型;Oracle数据库主要通过ADO接口方式对前端模块的业务数据进行管理;Oracle数据仓库通过ADO接口方式存放PI数据库和Oracle数据库处理后的数据。

1.2 电厂设备健康评估数据分析

1.2.1 电厂健康评估系统设备选取

电厂安全事故发生概率最高的设备是过热器和再热器。过热器和再热器是电厂最典型的热力设备,通过对其运行监测数据进行分析与学习,获取电厂设备的故障发生规律与健康状态信息,实现预测性电厂设备维修,达到电厂安全运行的效果^[16]。

大型电厂事故主要由锅炉事故造成,造成各类电站锅炉爆管事故的主要原因是过热器、再热器爆管。因此,研究和避免过热器、再热器爆管问题^[17]已成为确保电厂稳定运行、提高经济效益的关键环节。

1.2.2 数据结构分析

为了提高电厂点检智能化,针对电厂锅炉过热器、再热器状态检修问题,本小节通过确定检修主题来设计数据结构与内容。

本文涉及的数据主要来源于实时系统的实时数据库和电厂管理系统的管理数据,需要从两套独立的数据结构构建电厂设备健康评估学习训练数据。实时系统数据库与生产管理数据库通过设备编号、编码、命名对照、属性度量等方面的设置,使不同数据库的数据重构为结构一致的数据,存储在主题数据仓库,利用总线传输到应用程序服务器或者不同客户端。

本文涉及的数据是具有时间属性的设备状态信息,数据结构一般包括时间观测点、数据字段名称、数据类型、故障类型、故障内容、属性量化值等。为了简化数据输入,提高健康评估精度,去掉冗余字段,将设备故障信息、运

行参数信息整理成待分析数据,通过总线输入到应用程序服务器端口^[18]。

1.2.3 健康评估因素分析

根据文献 [18],本文分析了典型的电厂锅炉爆管故障造成因素。影响锅炉爆管故障的影响因素主要包括左引风机开度(X1)、左转向室烟温(X2)、过热器一级减温水流量(X3)、过热器三级左减温水流量(X4)、发电机负荷(X5)、过热器二级左减温水流量(X6)、给水压力(X7)等。

各故障影响因素的取值范围设置如下:1)出现故障情况下,左引风机开度(X1)的取值范围为 56.63~93;2)左转向室烟温(X2)的取值范围为 0~501.83;3)过热器一级减温水流量(X3)的取值范围为 0.9~58.1;4)过热器三级左减温水流量(X4)的取值范围为 0.28~157.32;5)故障情况下,发电机负荷(X5)的取值范围为 156.1~260.1;6)故障情况下,过热器二级左减温水流量(X6)的取值范围为 0~16.9;7)故障情况下,给水压力(X7)的取值范围为 0~14.22。

根据影响因素的取值,可以确定设备故障情况,通过故障程度与类型对设备的健康状态进行评估。

2 电厂设备健康等级划分

为了便于电厂设备健康状态分析,本文将设备评估分为 6 类不同的层级,各层级介绍如表 1 所示。

表 1 设备健康状态评估等级表

健康等级	健康指数值	设备状态
健康状态	[5, 6)	设备功能齐全,性能优异
亚健康状态	[4, 5)	设备功能齐全,性能轻微下降,堪用
注意状态	[3, 4)	设备某些特征发生异常波动,但不影响稳定运行
轻微状态	[2, 3)	设备功能短期正常,性能轻度缺陷
异常状态	[1, 2)	设备发生一定危害性的故障,性能异常
紧急状态	[0, 1)	设备发生较高危害性的故障,无法完成基本功能

在表 1 中,健康状态等级的量化采用设备健康指数,其具体的定义详见文献 [19],一般用于从定量角度判定设备的健康程度,且可以用来评估设备的剩余使用寿命。健康状态评估等级定义与取值介绍如下:

1) 健康状态。当电厂设备健康指数取值范围为 [5, 6) 时,设备处于“健康状态”,能够完成正常的功能任务,各特征指标值处于标准值附近,设备功能齐全,性能优异。

2) 亚健康状态。当电厂设备健康指数取值范围为 [4, 5) 时,设备处于“亚健康状态”。电厂设备可以正常完成所需的功能,个别特征发生轻微波动,对外部环境的适应能力有所下降。设备功能比较齐全,性能轻微下降。

3) 注意状态。当电厂设备健康指数取值范围为 [3, 4) 时,设备处于“注意状态”。电厂设备在运行过程中,某些特征发生异常波动,但不影响稳定运行,注意在短期内对异常情况进行监测,判断是否有继续恶化的趋势。

4) 轻微状态。当电厂设备健康指数取值范围为 [2, 3) 时,设备处于“轻微状态”,设备存在轻微异常故障,短期内不影响设备的稳定运行,需要加强对设备的健康监测,并实施计划检修。

5) 异常状态。当电厂设备健康指数取值范围为 [1, 2) 时,设备处于“异常状态”,设备发生一定危害性的故障,设备的稳定性运行受到一定程度的影响,需要停机局部检修。

6) 紧急状态。当电厂设备健康指数取值范围为 [0, 1) 时,设备处于“紧急状态”,设备发生较高危害性的故障,无法完成基本功能,需要停机进行大修。

3 集成深度随机森林

为了提高随机森林算法的泛化性与预测精度,结合集成学习和深度学习技术,本节提出了一种基于集成深度随机森林分类算法。

3.1 随机森林

随机森林算法(RF, random forest)是一种基于决策树的集成学习算法,其算法原理是基于 Bagging 集成算法和随机子空间方法^[14]。RF 算法利用待分类样本 X,训练产生决策树,并将所有决策树所有结果通过汇总投票方式确定预测分类结果,其具体的步骤如下:

1) 随机抽取样本,构造不同子数据集,并训练决策树;

2) 随机选择 m 个属性,采用信息增益策略来选择最优的属性进行决策树节点分裂,训练形成决策树;

3) 按照 1)~2) 建立不同的决策树,集成构建随机森林模型。

3.2 深度随机森林

深度随机森林算法(gcForest)是由南京大学周志华教授提出,通过级联的方式堆叠多层随机森林,以获得更好的特征表示和学习性能^[20]。gcForest 算法使用级联的森林结构来进行表征学习,需要很少的训练数据,且基本调节参数简单。深度随机森林算法包含多粒度扫描阶段和级联森林阶段等两个关键的步骤。

3.2.1 多粒度扫描阶段

为了提高级联森林阶段的预测效果,gcForest 算法利用多粒度扫描阶段策略最大限度地去挖掘数据特征的顺序关系。gcForest 算法多粒度扫描阶段结构如图 2 所示。gcForest 算法通过设置滑动窗口的大小和构建顺序特征向量,增强级联森林阶段分类效果。

图 2 给出了深度随机森林算法的多粒度扫描过程,具体的步骤如下:

1) 输入 200 维特征向量的样本,50 维的滑动窗口以 1 为步长获取 151 个 50 维的向量;

2) 将 151 个 50 维向量输入到 2 种不同类型的森林模型进行分类处理,从而得到 151 个 2 维向量;

3) 将分类向量按顺序组成 1 个 604 维的特征向量,将其输入到级联森林结构。

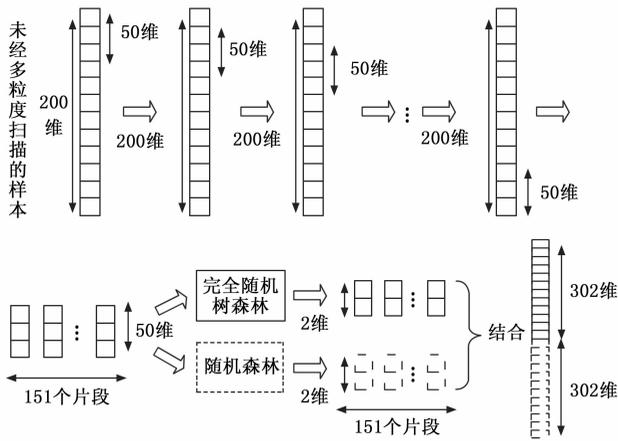


图 2 多粒度扫描阶段

3.2.2 级联森林阶段

为了增加学习算法的特征学习能力、提高算法的预测分类准确度, gcForest 算法采用级联森林阶段对多粒度扫描阶段处理后的特征向量进行逐层分析处理。在级联森林阶段, 每级都将上一级的特征向量与多粒度扫描阶段处理特征向量拼接为输入, 经过两种随机森林模型处理后传递到下一层级, gcForest 算法级联森林阶段结构如图 3 所示。

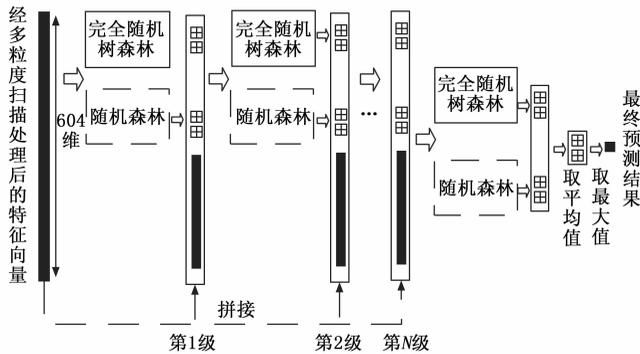


图 3 级联森林阶段

图 3 给出了深度随机森林算法的多粒度扫描过程, 具体的步骤如下:

- 1) 输入经多粒度扫描处理后的 604 维特征向量, 经过完全随机森林与随机森林处理后得到 2 个 2 维特征向量 (增强特征向量);
- 2) 将增强特征向量与 604 维的特征向量拼接为新的具有增强特征的特征向量, 输入到下一层级;
- 3) 依照 1) 和 2), 将数据传递到级联森林的最后一级;
- 4) 将最后一级产生的分类向量进行平均处理, 取最大值对应类别的分类结果。

3.2.3 深度随机森林算法

结合多粒度扫描阶段和级联森林阶段, 深度随机森林算法的整体流程如下:

- 1) 数据预处理。剔除无效的数据点, 利用插值法对缺失的数据进行填补。将数据集进行划分, 分为训练集与测

试集;

- 2) 利用训练集对深度随机森林进行学习训练。设置 gcForest 算法的超参数, 将预测处理后的数据集输入到多粒度扫描阶段, 得到特征的向量按顺序输入到级联随机森林阶段, 从而构建深度随机森林模型;

- 3) 利用测试集对深度随机森林模型进行测试验证。测试集样本按顺序输入到多粒度扫描阶段和级联随机森林阶段, 将产生的分类向量汇总平均得到分类预测结果。

3.3 集成深度随机森林算法

为了使训练过程不易发生过拟合现象, 提高弱分类器精度, 结合 AdaBoost 技术, 基于深度随机森林算法, 构建集成深度随机森林方法。假设电厂点检设备健康数据训练样本集 $X = [x_{ij}]_{n \times d}$, 其中 n 为样本数, d 为健康评估因素维度; 基评估器 gcForest, 记为 $f_{gcForest}(\cdot)$; 基评估器个数为 T 。具体的算法描述如下:

- 1) 初始化样本的权值分布 $D_1 = (\omega_{1,1}, \omega_{1,2}, \dots, \omega_{1,n})$, 每个样本的权值计算如式 (1) 所示:

$$\omega_{1,i} = \frac{1}{n}, i = 1, 2, \dots, n \quad (1)$$

- 2) 对于迭代轮次 $t = 1, 2, \dots, T$, 使用具有当前分布 D_t 的训练样本基分类器 $h_t = f_{gcForest}(X, D_t)$;

- 3) 计算基分类器 h_t 在训练样本集上的预测误差如式 (2) 和式 (3) 所示:

$$\epsilon_t = \sum_{i=1}^m \omega_{t,i} e_{t,i} \quad (2)$$

$$e_{t,i} = \begin{cases} 1 & y_i \neq h_t(x_i) \\ 0 & y_i = h_t(x_i) \end{cases} \quad (3)$$

其中: $e_{t,i}$ 为第 i 样本在 t 个基分类器上的误差, $e_{t,i} = 1$ 表示监督信号 $y_i \neq h_t(x_i)$ 时误差为 1, $e_{t,i} = 0$ 表示监督信号 $y_i = h_t(x_i)$ 时误差为 0。

- 4) 计算基分类器 h_t 的权重系数 a_t , 如式 (4) 所示:

$$a_t = 0.5 \lg \frac{\epsilon_t}{1 - \epsilon_t} \quad (4)$$

- 5) 更新训练样本集的样本分布 D_{t+1} , 直到迭代轮次达到最大值, 如式 (5) 所示:

$$\omega_{t+1,i} = \omega_{t,i} e^{-a_t y_i^t} \quad (5)$$

- 6) 线性组合 T 个基分类器, 最终得到强分类器——集成深度随机森林 (gcForest with AdaBoost, gcForest-Ada), 如式 (6) 所示:

$$f_{gcForest-ada}(X) = \text{round} \left(\sum_{t=1}^T \left(\ln \frac{1}{a_t} \right) G(X) \right) \quad (6)$$

其中: $G(X)$ 为所有 $a_t h_t(X)$ 的中位数, $\text{round}(\cdot)$ 表示取整。

4 基于集成深度随机森林的电厂设备健康评估方法

结合 AdaBoost 与深度随机森林算法, 本节提出了一种基于集成深度随机森林算法的电厂设备健康评估模型方法。该评估模型主要是以电厂设备健康影响特征因素为输入, 以设备健康等级为输出, 挖掘健康影响特征因素与健康等

级之间的映射关系。基于集成深度随机森林算法的电厂设备健康评估模型如图 4 所示。具体的步骤如下：

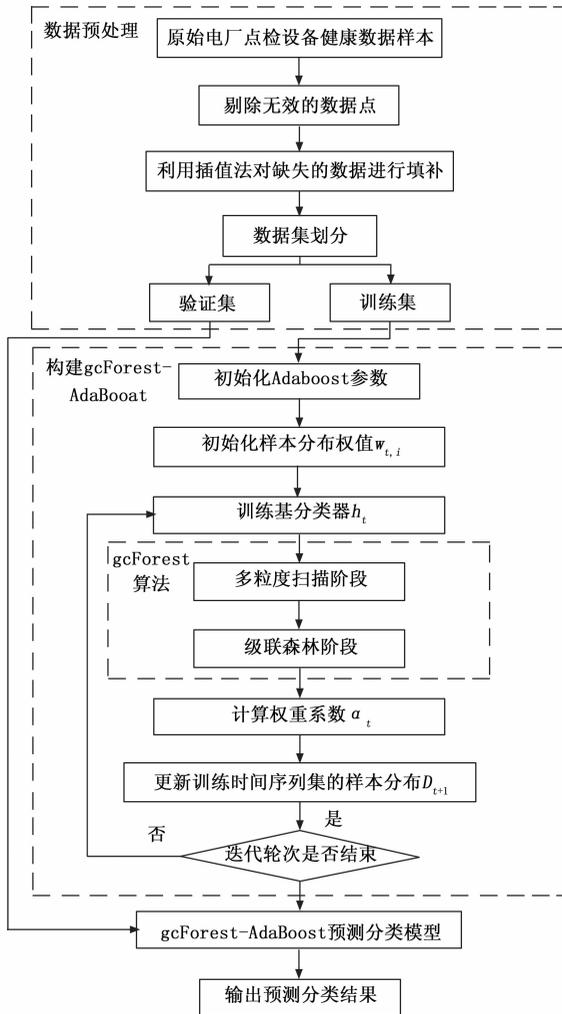


图 4 电厂设备健康评估流程图

步骤 1：获取电厂点检设备健康数据样本。利用电厂点检设备健康数据 X ，构建训练样本集 $\{X, Y\}$ ；对获取后的样本进行预处理，剔除无效的数据点，利用插值法对缺失的数据进行填补；将数据集进行划分，分为训练集与测试集；

步骤 2：初始化 AdaBoost 参数。初始化弱分类器每一个深度随机森林算法的参数；设置弱分类器个数 T ；初始化训练样本的分布权重 D_1 ；

步骤 3：训练弱分类器 gcForest。

1) 输入电厂点检设备健康数据 X ，利用多粒度扫描阶段策略最大限度地挖掘数据特征的顺序关系；

2) 输入经多粒度扫描处理后的特征向量，采用级联森林阶段对多粒度扫描阶段处理后的特征向量进行逐层分析处理，将最后一级产生的分类向量进行平均处理，取最大值对应类别的分类结果；

步骤 4：计算基分类器的权重系数 α_t 以及更新样本分布 D_{t+1} ；训练弱分类器直到迭代轮次结束，输出强分类器 gc-

Forest-Ada；

步骤 5：输入电厂点检设备健康数据测试集，通过步骤 1 的预处理方法，利用训练好的强分类器对当前健康状态进行评估，输出相应的设备健康状态等级。

5 实验分析

为了验证本文提出电厂设备健康状态评估方法的预测分类性能，本节选择文献 [19] 的电厂锅炉爆管故障与运行数据集对提出的算法的预测分类结果进行分析与讨论。

5.1 仿真环境设置

本文使用 Matlab 2021a 进行程序编写，测试环境是 Windows 10 系统，处理器为 AMD Ryzen 9 5900HX with Radeon Graphics，内存为 16.0 GB。实验数据集选取电厂锅炉爆管故障与运行数据集，其中 600 组数据集为训练集，240 组数据集为测试集，数据集具体的介绍参考文献 [19] (原始数据介绍)。本文提出的电厂设备健康评估方法与对比方法具体参数设置如表 2 所示，其中 gcForest 算法决策树数量与 gcForest-AdaBoost 算法决策树数量和分类器数量由 5.2 节分析给出。

表 2 设备健康评估方法参数设置

评估模型	方法参数设置
模糊层次分析法 ^[19]	无参数设置
BP 神经网络 ^[20]	隐层网络节点数为 10
BP-AdaBoost ^[20]	隐层网络节点数为 10, 弱分类器个数为 3
RF ^[14]	树数量为 100
RF-AdaBoost ^[21]	树数量为 100, 弱分类器个数为 3
gcForest ^[20]	多粒度扫描阶段森林数量为 2, 级联森林阶段森林数量为 2, 决策树数量由 5.2 节给出
gcForest-AdaBoost	多粒度扫描阶段与级联森林阶段森林数量同上, 决策树数量与分类器数量由 5.2 节给出

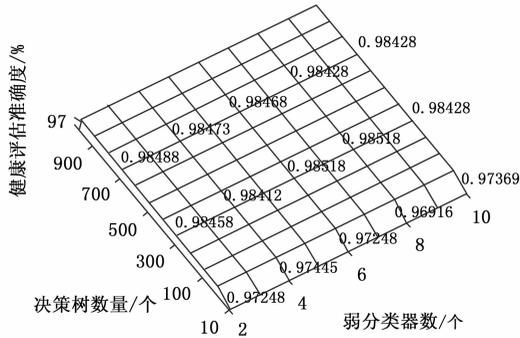
5.2 参数设置分析

为了分析深度随机森林算法与 AdaBoost 方法对模型性能影响，本文分别对深度随机森林算法决策树数量与 AdaBoost 技术基分类器数量进行分析。训练 gcForest-AdaBoost 模型的样本数量为 600，测试样本数 240。

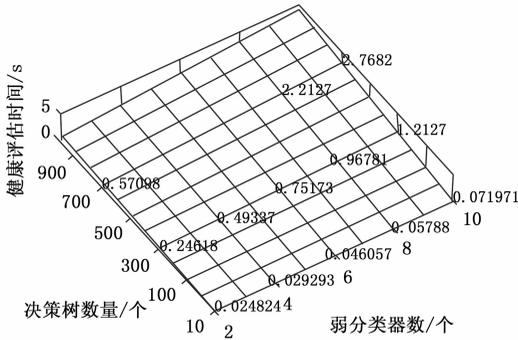
图 5 给出了不同决策树与基分类器对基于集成深度随机森林算法电厂设备健康状态评估模型性能的影响。从图 5 (a) 可以看出，弱分类器个数对 gcForest-AdaBoost 模型设备健康状态评估精度影响不大，树数量大于 50 之后对 gcForest-AdaBoost 模型设备健康状态评估精度影响不大。从图 5 (b) 可以看出，弱分类器个数越多，基于 gcForest-AdaBoost 算法的设备健康状态评估模型的评估时间越长；树数量个数越多，基于 gcForest-AdaBoost 算法的设备健康状态评估模型的评估时间越长。综上所述，本文选用树数量为 100，弱分类器个数为 3。

5.3 算法性能比较分析

为了验证 gcForest-AdaBoost 算法的高效性，本文与模



(a) 健康评估准确度影响



(b) 健康评估时间分析

图 5 不同决策树与基分类器对 gcForest-AdaBoost 评估模型性能影响分析

糊层次分析法、BP 神经网络、集成神经网络 (BP-AdaBoost)、随机森林 (RF)、集成随机森林 (RF-Ada)、深度随机森林 (gcForest) 进行比较。各个算法的参数设置如表 2 所示。gcForest-AdaBoost 算法选用树数量为 100, 弱分类器个数为 3。每个算法模型训练 20 次, 统计 20 次的评价结果, 包括设备健康评估准确度、准确度标准差以及平均样本评估时间。

表 3 给出了各算法的水质量评估精度统计结果, 其中最值标黑加粗。从表 3 可以看出, 在准确度均值与标准差

指标方面, gcForest-AdaBoost 性能最好; 在时间方面, RF 算法用时更少; 在准确度均值方面, 模糊层次分析法结果最差; 在标准差方面, RF 和 RF-AdaBoost 结果最差; 在预测时间方面, BP-AdaBoost 结果最长。gcForest-AdaBoost 的准确度均值、标准差分别为 0.97、 2.20×10^{-16} , 评估时间为 7.97×10^{-6} s, 精度较高, 实时性较好。

表 3 各模型 20 次预测结果比较

编号	算法	准确度均值	准确度标准差	t/s
1	模糊层次分析法	0.69	2.28×10^{-16}	2.17×10^{-5}
2	BP 神经网络	0.79	2.64×10^{-16}	2.78×10^{-5}
3	BP-AdaBoost	0.80	2.29×10^{-16}	2.89×10^{-5}
4	RF	0.85	3.42×10^{-16}	7.00×10^{-6}
5	RF-AdaBoost	0.89	3.42×10^{-16}	7.32×10^{-6}
6	gcForest	0.92	2.28×10^{-16}	7.57×10^{-6}
7	gcForest-AdaBoost	0.97	2.20×10^{-16}	7.97×10^{-6}

为了进一步测试差异发生情况, 采用 Iman-Davenport 测试。Iman-Davenport 测试是根据 $(k-1)$ 和 $(k-1)(N-1)$ 自由度的 F 分布进行的:

$$F_F^2 = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2} \quad (7)$$

其中: k 和 N 代表比较算法个数与测试函数的个数。在本节中, $k = 7$ 和 $N = 8$ 。根据上面公式, 可以得到 Iman-Davenport 测试结果如表 4 中所示。为了发现所有算法的差异, 基于 Nemenyi 测试的临界差异 (CD, critical difference) 被使用。CD 的计算如下:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \quad (8)$$

其中: q_α 为临界值, 一般取值为 2.445 6。因此, CD 值为 4.175 7。3 个方面基于 Friedman 测试与 Nemenyi 测试的比较结果如图 6 所示。如图 6 (a) 所示, 尽管基于 gcForest-AdaBoost 算法的设备健康评估模型在精度均值方面排名第一, 但是 gcForest-AdaBoost、gcForest、RF-AdaBoost、RF 与 BP 之间的性能差别也不大。如图 6 (b) 所示, 在标

表 4 各模型 20 次预测结果比较

测试项目	模糊层次分析法	BP	BP-AdaBoost	RF	RF-AdaBoost	gcForest	gcForest-AdaBoost
精度均值排名 ^a	6.25	5.75	4.38	4.13	3.75	1.50	1
精度方差排名 ^b	3.75	5.13	4.38	7	7	3.75	1.5
评估时间排名 ^c	4.63	6	7	1.75	2.13	3.75	4.38

- a. 1) Friedman 测试的显著概率 2.66×10^{-5} 。卡方分布值 30.89。
- 2) 6×42 自由度的 F 分布值 12.64。
- 3) Iman-Davenport 测试的显著概率 1.90×10^{-3} 。
- b. 1) Friedman 测试的显著概率 2.90×10^{-7} 。卡方分布值 41.00。
- 2) 6×42 自由度的 F 分布值 41.00。
- 3) Iman-Davenport 测试的显著概率 4.17×10^{-5} 。
- c. 1) Friedman 测试的显著概率 5.3020×10^{-8} 。卡方分布值 44.73。
- 2) 6×42 自由度的 F 分布值 95.67。
- 3) Iman-Davenport 测试的显著概率 2.33×10^{-6} 。

准差方面, 基于 gcForest-AdaBoost 方法的设备健康评估模型获得第一, 但是 gcForest-AdaBoost、gcForest、模糊层次分析法、BP 与 BP-AdaBoost 之间的性能差别不大。图 6 (c) 中, 在耗时方面 RF 算法排名第一, 但是在 RF、RF-AdaBoost、模糊层次分析法、gcForest-AdaBoost 与 gcForest 之间的显著差别不大。总之, 基于 gcForest-AdaBoost 算法的电厂设备健康评估模型除了时间之外, 其他方面表现出最好的性能。

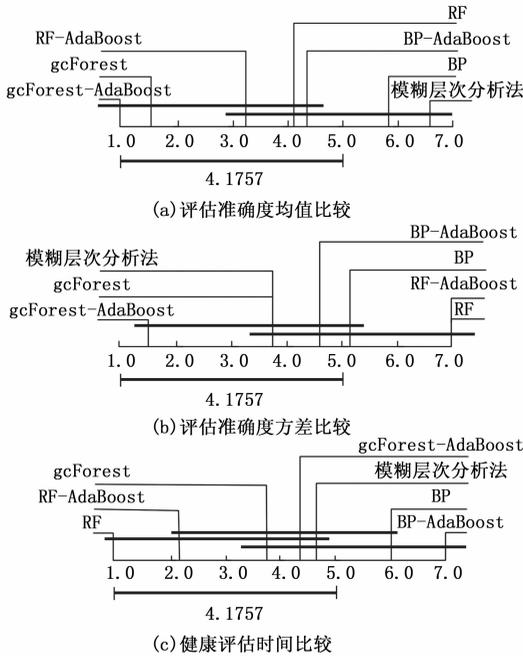


图 6 基于 Nemenyi 测试的算法比较

6 结束语

针对当前电厂设备健康状态存在预测分类精度不高的问题, 本文提出了一种基于集成深度随机森林算法的智能电厂设备健康评估方法。该方法在电厂设备健康评估系统结构基础上, 分析了健康评估数据结构与影响因素; 根据设备故障程度, 将设备评估分为 6 类不同的层级; 结合深度学习与集成学习技术, 提出了集成深度随机森林算法, 从而构建基于集成深度随机森林算法的智能电厂设备健康评估方法。通过仿真, 得出以下结论:

1) 采用集成学习与深度学习技术改进随机森林, 通过比较 gcForest、RF、RF-AdaBoost 评估模型, 验证了 gcForest-AdaBoost 能够提高评估准确度;

2) 基于 gcForest-AdaBoost 算法的健康状态评估模型不仅在预测评估准确度上好于其他预测评估模型, 且评估稳定性是最好的, 具有鲁棒性;

3) 本文提出的评估模型在评估预测时间中表现不好, 进一步提高 gcForest-AdaBoost 模型的评估时间是下一步研究重点。

参考文献:

- [1] 赵保卫. 对发电设备状态检修开展情况的分析与建议 [J]. 电力学报, 2006 (1): 104-107.
- [2] 刘歆一. 基于数据挖掘的电厂设备健康状态评价系统的研究 [D]. 北京: 华北电力大学, 2008.
- [3] 成国良. 火电厂设备健康监测与远程故障诊断 [D]. 石家庄: 石家庄铁道大学, 2015.
- [4] 高洪福. 优化组合模型在电厂设备状态预测中的研究与应用 [D]. 上海: 中国石油大学 (华东), 2016.
- [5] WANG Y F, HE R, TIAN Z G. Opportunistic condition-based maintenance optimization for electrical distribution systems [J]. Reliability Engineering and System Safety, 2023, 236 (3): 109261.
- [6] MARCO K, BRAM D J. Condition-based maintenance and production speed optimization under limited maintenance capacity [J]. Computers & Industrial Engineering, 2023, 179 (1): 109155.
- [7] 卢伟明. 数据挖掘技术在电力设备故障诊断中的应用 [J]. 现代工业经济和信息化, 2016, 6 (21): 77-78.
- [8] 高磊, 刘鑫, 曹卫强. 一种面向多元信息模型的设备状态预测及评价方法研究 [J]. 北京汽车, 2018 (1): 30-32.
- [9] 高磊. 基于多元信息模型的设备状态预测及评价方法研究 [D]. 合肥: 合肥工业大学, 2017.
- [10] 董理科, 白鹭, 武娜, 等. 基于知识图谱的电力变压器故障预测方法研究 [J]. 高压电器, 2022, 58 (11): 151-159.
- [11] 徐凤建, 车建国, 许荣. 基于知识推理的雷达装备故障预测专家系统 [J]. 电子产品可靠性与环境试验, 2010, 28 (5): 29-32.
- [12] 蒋金元. 基于深度学习的滚动轴承故障诊断与预测方法研究 [D]. 株洲: 湖南工业大学, 2021.
- [13] 贺川双, 杜修明, 严英杰, 等. 基于数据挖掘和主成分分析的电力设备状态评价 [J]. 高压电器, 2017, 53 (12): 34-41.
- [14] 杨荣利, 曹军, 张延启, 等. 基于分类算法的汽轮机故障诊断 [J]. 电子设计工程, 2023, 31 (9): 68-72.
- [15] 余笑依. 大数据挖掘技术在输变电设备故障诊断中的应用探析 [J]. 通讯世界, 2017 (22): 151-152.
- [16] 李喜奇. 设备状态检测与传感器故障诊断技术研究 [D]. 哈尔滨: 哈尔滨工程大学, 2005.
- [17] 王鹏辉, 董鹏. 对流受热面高温积灰导致锅炉爆管案例分析及防范措施 [J]. 锅炉技术, 2021, 52 (5): 60-64.
- [18] 姚雪骏. 某超临界锅炉末级过热器爆管原因分析及防范措施 [J]. 吉林电力, 2021, 49 (6): 51-54.
- [19] 郑森. 配用电设备健康状态智能评价关键技术研究及实现 [D]. 北京: 中国石油大学, 2021.
- [20] 邢书豪, 孙文慧, 颜勇, 等. 基于改进随机森林算法的电力系统短期负荷预测模型 [J]. 青岛大学学报 (工程技术版), 2019, 34 (3): 7-10.
- [21] 魏曼曼, 路皓翔, 杨辉华. 基于 RF-AdaBoost 模型的血液种属鉴别算法 [J]. 高等学校化学学报, 2020, 41 (1): 94-101.