

文章编号:1671-4229(2023)04-0029-13

基于模型集成的保修期内外设备故障预测

张全新^a, 倪紫薇^a, 古筝琴琴^b, 周伟伟^a, 谭毓安^b

(北京理工大学 a. 计算机学院, b. 网络空间安全学院, 北京 100081)

摘要: 工业设备故障预测是智能制造领域的重要课题,在工业生产中被广泛应用。设备故障预测能够提前对设备进行故障预判和诊断,有利于降低设备的运营风险并提高设备的利用率。由于传统预测方法相对简单,对记录数据利用不充分,使得故障预测准确率较低,并且对于保外期的设备故障预测没有系统的解决方案。为解决上述问题,文章设计了一种基于模型集成的保内保外设备故障预测方法。通过分析挖掘预测性维护数据(包含设备的运行状态数据、发生异常和故障数据以及历史维修数据等),构建高维设备特征属性,使用灰色关联度分析提取主要特征,并结合支持向量机模型与XGBoost模型建立保内保外故障预测协同机制,实现了高、低故障风险设备在保修期内的协同预测。实验结果表明,该方法在多预测周期、多特征空间、多模型对比下的预测准确率均有提升。

关键词: 故障预测; 预测性维护; 特征工程; 支持向量机; XGBoost

中图分类号: TP 311.13

文献标志码: A

Equipment fault prediction technology within and outside the warranty based on model integration

ZHANG Quan-xin^a, NI Zi-wei^a, GU Zheng-qin-qin^b, ZHOU Wei-wei^a, TAN Yu-an^b

(a. School of Computer Science, b. School of Cyberspace Science and Technology,

Beijing Institute of Technology, Beijing 100081, China)

Abstract: The fault prediction of industrial equipment is an important subject in the field of intelligent manufacturing, which is widely used in industrial production. Equipment fault prediction can predict and diagnose equipment failures in advance, which is conducive to reducing equipment operation risk and improving equipment utilization. Because the traditional prediction method is relatively simple and the recorded data is not fully utilized, the accuracy of fault prediction is low, and there is no systematic solution for equipment fault prediction outside the warranty period. To solve the above problems, this paper designs a fault prediction method for equipment within and outside the warranty based on model integration. By analyzing and mining Predictive Maintenance data (Telemetry Time Series data, Error and Failure data, Maintenance data, etc.), constructing high dimensional equipment feature attributes, using gray correlation analysis to extract main features, and combining the support vector machine model and the XGBoost model, a cooperative mechanism for fault prediction within and outside the warranty is established. It realizes synergistic forecasts within and outside the warranty between high and low failure risk equipment. The experimental results show that the prediction accuracy of this method is improved under the comparison of multiple prediction periods, multiple feature spaces and multiple models.

Key words: fault prediction; predictive maintenance; feature engineering; SVM; XGBoost

收稿日期: 2023-06-01; 修回日期: 2023-06-21

基金项目: 国家重点研发计划资助项目(2018YFB1101402)

作者简介: 张全新(1974—),男,副教授. E-mail: zhangqx@bit.edu.cn

引文格式: 张全新, 倪紫薇, 古筝琴琴, 等. 基于模型集成的保修期内外设备故障预测[J]. 广州大学学报(自然科学版), 2023, 22(4): 29-41.

故障预测技术是故障预测与健康管理技术(Prognostics and Health Management, PHM)^[1-3]的一个分支,是从设备使用现状出发,基于工业生产中的大数据,结合信号处理和数据挖掘技术,分析设备的结构特征、历史数据和状态数据,预测和判断设备的未来故障情况。故障预测有利于辅助设备运维,通过预测和管理未知风险,进而提前对设备进行维护保养。

在大数据、云计算和 5G 技术的支撑下,故障预测技术被广泛应用于制造企业的售后服务场景中^[4-5]。对设备进行故障预测可以帮助企业售后团队和设备维修人员及时消除故障隐患,有利于企业获得成本节约和竞争优势,同时帮助用户将设备尽早恢复至稳定运行状态,以此减少维修费用,实现企业与用户双赢^[6-8]。基于产品售后市场的设备预测性维护也将成为未来发展趋势,具有广阔的市场前景^[9-10]。相关数据统计表明,截止至 2022 年底,中国 PHM 市场规模已达到 46 亿元,相比于 2017 年的 13 亿元,增长了近 254%;全球 PHM 市场规模在 2022 年已达到 94 亿元,相比于 2017 年的 33 亿元,增长了近 185%^[11]。

尽管各国对 PHM 足够重视且持续加大了投资力度,现有的故障预测技术仍然存在不可忽视的瓶颈。

在面向保内保外的售后服务方面,随着设备使用者对保修服务的要求也越来越高,制造厂商为了在市场竞争中占据优势,纷纷加大售后保修服务力度^[12],不仅要注重保修期内的服务质量,还要提高保修期外的服务质量。文献[13-14]基于保修期内外的不同失效率,提出搭建失效率衰减的模型,该方法实现了在整个设备使用期限内进行不同周期的预防性维护。文献[15]提出了带缓冲区和保修期限的预测性维护建模方法,以此降低产品维护的成本。文献[16]提出了基于效益最大化的全局最优保修期模型,通过对系统故障概率进行量化建模,结合保修期内外的维修成本,计算消费者和生产商效益同时最大化的保修期,此方法适用于保修期待确定的情况。目前考虑保修期因素的运维策略研究多数基于预防性维护,仅仅是对设备进行定期的检查和更换,但仍会出现维护不及时或过度维护的问题,需要研究更多的预测性维护策略,预测设备未来的运行状况。

传统的基于模型的故障预测方法需要科研工

作者具有一定的技术经验积累,通过经验来总结设备退化过程中的状态演变规律,构造出恰当的模型来拟合设备的退化过程,如灰色预测模型^[17]、隐 Markov 模型^[18]、主元分析^[19]和 BP 神经网络模型^[20]等。这些方法在设备单一、外界干扰较少的理想情况下能够精准预测,但针对复杂的动态系统难以建立精确的模型。此外,不同的设备环境和工作条件也会导致模型的变化,使得基于模型的故障预测方法不适用于实际生产中。

机器学习算法作为人工智能技术的核心,虽然在部分性能上不如深度学习算法,但其具有简单、高效和可解释的优点,且能够针对特定场景人为构建特征,在工业预测性维护中仍被广泛应用。其中,支持向量机模型(Support Vector Machine, SVM)具有小样本强学习能力,工业系统中故障数据为小样本数据。实验数据集为小样本,且数据获取具有一定难度,因此,具有样本强学习能力的 SVM 有实际应用的优越性。如文献[21]提出了一种基于 SVM 预测分析方法,并使用主成分分析法(Principal Component Analysis, PCA)对油气钻井进行故障预测,结果表明 PCA-SVM 方法具有优秀的性能,准确率达到了 97.99%。文献[22]提出了一种基于特征空间聚类间距离优化的 SVM 模型,从振动信号中提取多时间尺度特征,用于机轴承的故障预测和诊断。文献[23]在对汽车子系统的故障预测中,与决策树模型(Decision Tree, DT)、随机森林(Random Forest, RF)和 K-近邻(K-Nearest Neighbor, KNN)等模型相比, SVM 的预测性能最佳。

除了单一模型的预测,模型融合的方法能够进一步提升预测效果,如文献[24]提出一种基于 ARIMA 时间序列预测和极端梯度提升(Extreme Gradient Boosting, XGBoost)的预测模型,首先对振动信号进行数值预测,然后将预测的数值输入 XGBoost 模型中进行故障分类,此融合模型的效果更好。文献[25]将模型方法与数据驱动方法进行融合,实现设备未来故障时间的预测。在以上设备故障预测方法中,更多聚焦在如何进行设备本身属性的特征构建以及模型的配置优化问题上,且仅侧重于保内期的设备维护。然而,目前的研究仅对设备运行的相关状态进行特征构建,针对保内保外的售后服务场景,在考虑设备本身的运行数据外,还需考虑用户的主动维护情况,因此需要从更多角度和更高维度构建特征。面对新的智

能运维和产品后市场环境,应当针对性设计面向保修期内外的特征构建和故障预测方法。本文研究了售后服务特点及各模型特征后,提出了基于模型集成的保内保外设备故障预测方法,充分挖掘了该场景的应用潜力。

本文的主要贡献有以下两点:

(1)通过特征构建方法构建高维度特征,并利用灰色关联度分析提取出主要特征,在此基础上结合 SVM 和 XGBoost 模型进行双模型故障预测,实现了准确率较高的故障预测能力。

(2)构建 FME 模型进行 K-Means 聚类分析,判定设备的故障风险等级,结合多种预测机制,为保修期内外的工业设备故障预测均提供了一定的保障。

1 故障预测技术

基于模型集成的保内保外设备故障预测方法主要包含3个部分,分别是设备属性特征构建、基于灰色关联度分析的主特征提取,以及结合 SVM 和 XGBoost 的故障预测集成模型构建。

1.1 设备属性特征构建方法

特征工程是机器学习的重要组成部分,通过以原始数据为基础构建和提取数据的关键特征,不断提高模型的训练效果和预测性能。构建属性特征的方式主要有以下3种:

(1)原特征的分解。把一个属性分解或切分成多个属性,将原始字段细颗粒度化。如将时间相关属性划出不同的时间窗口,得到同一属性在不同时间下的特征值。

(2)特征值分析。根据特征值分析其背后含义,挖掘出关联的新特征。

(3)特征间组合。通过特征间的加减乘除,组合出新特征。

本文通过上述3种方式将原始特征拓展成更高维度,以提高预测的准确性。

1.2 灰色关联度分析

灰色关联度分析的基本思想是根据序列曲线形状之间的相似程度来确定它们之间的关联性,如果曲线越近,相关性越大,反之亦然^[26]。灰色关联度的计算包含4个步骤:①选定参考序列和比较序列。参考序列是结果特征组成的序列,比较序列是需要进行排序的属性特征组成的序列。②无量纲化处理。采用均值法将参考序列和比较序

列中的数据统一映射到一定区间中,以便在相同维度下观测不同属性的变化趋势。③公式计算灰色关联系数。④计算关联度并排序。将比较序列中各个属性的灰色关联系数取平均值,根据关联度大小由高到低进行排序。保留数值大的主要因素,删除数值小的次要因素,以此降低特征空间维数。

灰色关联度分析是根据各个因素之间发展趋势的相似或相异程度来衡量因素间的关联程度,本文使用此方法以样本数据为基础,根据各特征属性变化趋势的相似度来分析其对决策属性的影响度大小。通过对关联度进行排序,提取出主要影响因素。

在将原始特征拓展为高维度的特征后,可能会造成特征的冗余,通过灰色关联度分析结合 pred_1d 每日故障预测、高频率的 pred_1w 每周故障预测机制,以及低频率的 pred_1m 每月故障预测机制,对各影响因素进行重要性排名,以此提取主要特征,降低特征空间维数。其过程如图1所示。

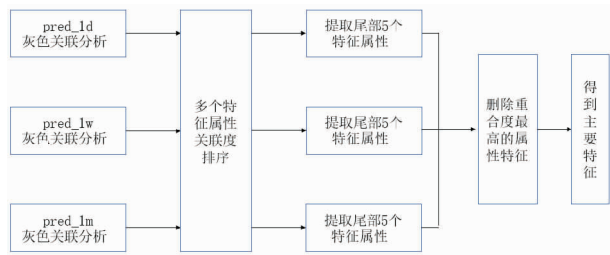


图1 灰色关联分析特征选取流程

Fig. 1 Gray correlation analysis feature selection process

1.3 结合 SVM 和 XGBoost 的故障预测集成模型构建

1.3.1 SVM 算法原理

SVM 是一般化线性分类方法^[27],通过寻找一个分类线来划分两个类别,使得两类的间隔最大,分界面为 $\omega x + b = 0$,在线性可分情况下为求解 ω 和 b 的问题。

对于线性不可分的情况,SVM 引入核函数将数据从低维空间映射到高维空间,把求解分类线的问题转化为求分类面的问题,进而有效地分类。常用的核函数有 RBF 核函数、线性核函数和多项式核函数等。SVM 能够在海量数据中高效筛选出对预测任务最有效的少数训练样本,对大型机械设备非线性高维空间故障数据的分类具有显著效果。

1.3.2 XGBoost 算法原理

XGBoost 算法是以分类回归树 (Classification And Regression Tree, CART) 为基础, 组合多个 CART 的 Boosting 集成学习方法^[28]。XGBoost 与梯度提升决策树 (Gradient Boosting Decision Tree, GBDT) 相比, 执行速度更快且模型性能更高。近年来, XGBoost 广泛应用于产品分类、消费者行为预测等领域, 并取得了很好的效果^[29]。

XGBoost 的实现是不断添加一个个新的决策树进来, 学习器以此为基础去拟合前一次的预测残差, 最终通过添加所有的决策树获得预测的最后结果。XGBoost 相比 GBDT 进行了优化, 其对损失函数进行二阶泰勒展开并引入一阶和二阶导数, 能够防止过拟合, 大大提升了模型的泛化能力, 适用于故障分类预测。

1.3.3 基于 SVM 和 XGBoost 的保内保外故障预测协同模型

利用 SVM 解决非线性精度高、速度快, 以及 XGBoost 泛化度高的特点, 以特征工程为基础结合双模型进行故障预测, 故障预测模型构建过程如图 2 所示。

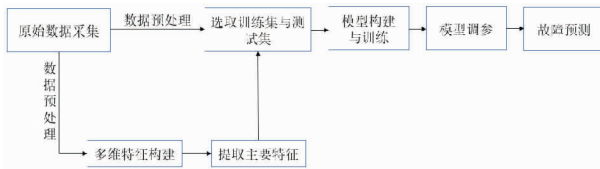


图 2 故障预测模型构建流程图

Fig. 2 Flow chart of fault prediction model construction

由于保内保外业务的差异性, 本文将两种业务场景进行融合, 研究设备在双场景下运用不同故障风险等级设备的故障预测方法, 设计基于 SVM 和 XGBoost 的保内保外故障预测协同模型, 进一步丰富售后服务的业务流程, 实现设备信息共享与服务数据反馈。

保内保外业务的主要差异在于设备信息采集的完整性, 由于保内期可免费维保, 商家与用户均重视设备健康状况, 因此, 保内期的设备跟踪数据更加完整。而保外期不在商家所承担的服务范围, 且用户可任意选择商户进行维保, 因此保外期设备数据难以追踪。本文以保修期为时间轴, 对保内和保外的设备数据进行关联, 设计高、低故障风险等级的故障预测机制, 如图 3 所示, 以此实现对设备故障情况的持续预测。

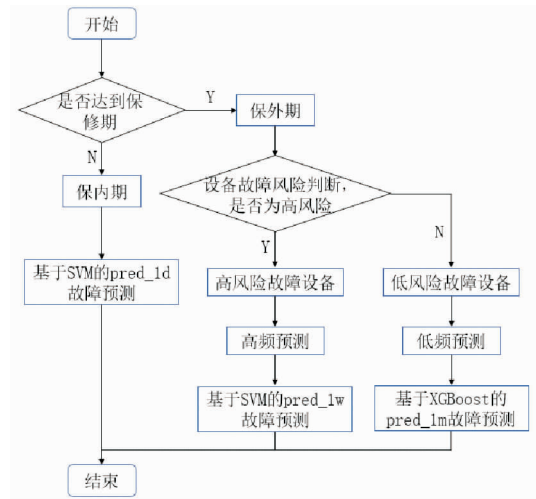


图 3 结合故障风险的保内保外预测模型

Fig. 3 Prediction model both inside and outside of warranty based on failure risks

对于保修期内的设备, 由于数据更加完整, 采取日预测机制, 即基于 SVM 模型进行 pred_ld 的每日故障预测, 通过对设备高频率监控和数据挖掘, 促进用户的维保行为, 以此丰富用户的历史活动数据, 形成制造服务数据驱动的业务挖掘模式。当设备到达保修期, 则需要对保内的全部历史数据进行分析并确定设备目前的故障风险等级。

本文分别从设备故障情况、设备维护情况和设备报错情况 3 个方面来衡量设备故障风险。构建 FME 模型进行 K-Means 聚类分析, 以此对设备故障风险等级分类。FME 模型各字段含义为: F 表示设备的平均故障时间间隔 (Mean Time Between Failures, MTBF), $MTBF = \text{总运行时间} / \text{总故障次数}$; M 表示设备的平均维护时间间隔 (Mean Time Between Maintenance, MTBM), 包含主动维护和被动维护两种, $MTBM = \text{总运行时间} / \text{总维护次数}$; E 表示设备的平均报错时间间隔 (Mean Time Between Errors, MTBE), $MTBE = \text{总运行时间} / \text{总报错次数}$ 。

K-Means 算法作为一种无监督学习算法, 处理分类问题时简单快速、算法效率高、可解释性强且其本身具有迭代功能, 可以满足低时间复杂度的要求。因此, 本文在确定设备故障风险评估指标后, 使用 FME 模型结合 K-Means 的方法进行设备故障的风险等级划分。由于本文研究的故障预测是二分类问题, 这里与其对应将设备故障风险等级划分为两种, 即高故障风险和低故障风险。接

下来,对聚类后的设备故障风险等级情况进行分析,通过对比指标属性值,总结各类别特征,从而为设备故障预测提供建设性指导和建议。若为高故障风险设备,则采用高频率的 pred_1w 预测机制进行短周期预测;若为低故障风险设备,则采用低频率的 pred_1m 预测机制进行长周期预测。

本文采用 GridSearchCV 网格搜索技术和三折交叉验证进行参数调参,以求解最优超参数组合。网格搜索技术搜索的是参数,即在指定的参数范围内,按步长依次调整参数,利用调整的参数训练学习器,从所有的参数中找到在验证集上精度最高的参数。三折交叉验证是将训练集等分为3个部分,分别记做 S1、S2 和 S3。每次选取两部分作为训练集,剩下的一部分作为测试集。对于任意模型,首先用 S1 和 S2 作为训练集进行训练,得到模型 model_1,并用 S3 作为测试集进行测试,得到准确率 accuracy_1;再用 S1 和 S3 作为训练集进行训练,得到模型 model_2,并用 S2 作为测试集进行测试,得到准确率 accuracy_2;最后用 S2 和 S3 作为训练集进行训练,得到模型 model_3,并用 S1 作为测试集进行测试,得到准确率 accuracy_3。则该模型在此训练集上的平均准确率为 $F1 = (accuracy_1 + accuracy_2 + accuracy_3)/3$ 。选取 F1 平均得分最高的超参数组合配置到模型中,以此构建出最佳超参数下的模型,为后续模型预测做准备。最后,选取高性能模型进行训练和预测。后续的试验结果表明,对于保外期的设备,若为高风险故障设备,采用基于 SVM 的 pred_1w 的高频预测更为高效准确;若为低风险故障设备,则采用基于 XGBoost 的 pred_1m 的低频预测更为有效。由此进行双模型故障预测,实现了准确率较高的故障预测能力。

2 故障预测实验与分析

2.1 实验环境配置

本实验平台的操作系统:Windows 10 21H1 版;CPU: Intel(R) Core(TM) i7 - 4790 CPU @ 3.60GHz; IDE: Pycharm2020.2.3 Professional Edition。

2.2 数据集分析

实验数据集来自 kaggle 平台的预测性维护项目,提取 100 台设备从 2015 年 1 月 1 日至 2016 年 1 月 1 日的遥测数据、报错数据、维护数据、故障数

据,以及设备元数据。

遥测数据是 100 台设备的传感器采集数据,包含日期和时间、设备 ID、每小时的平均电压值、每小时的平均转速值、每小时的平均压力值和每小时的平均振动值,共计 876 100 条。

故障数据记录 100 台设备某部件发生的故障情况,共计 761 条,其中包含设备发生故障的日期和时间、发生故障的设备 ID 以及设备的故障部件类型。其中故障部件类型有 4 种,分别为 comp1、comp2、comp3 和 comp4。

设备维护数据记录了 100 台设备的部件更换情况,包含设备主动维护和被动维护记录(被动维护记录即故障记录),共 3 286 条,其中包括维护的日期和时间、维护的设备 ID 以及设备的维护部件类型。设备的维护部件类型与故障部件类型相同,共 4 种,分别为 comp1、comp2、comp3 和 comp4。

设备报错数据记录 100 台设备一年内的报错情况,共计 3 919 条,其中包括报错的日期和时间、设备 ID,以及设备的报错类型号。报错的类型有 5 种,分别是 error1、error2、error3、error4 和 error5。

在预测性维护数据集中,遥测数据共 876 100 条,其中故障数据 761 条,正常运行数据 875 339 条。将设备故障记为正样本(1),设备正常记为负样本(0),正负样本比达到 1:1 150 数量级,比例严重失调,因此需要对负样本进行下采样。先提取出 100 台设备每天 6:00:00 时刻的数据,总计 36 600 条,此时正负样本比例 1:51。然后再对负样本随机采样,使正负样本比在 1:3 左右,以此得到样本数据集,其中包含故障数据 702 条,正常运行数据 1 996 条。最后,处理样本数据。使用 shuffle 函数将样本数据集的顺序打乱,使用 train_test_split 函数将样本数据集划分为训练集与测试集,其中测试集占比 30%。

设备元数据记录 100 台设备的基本信息,是静态数据,包含设备 ID,设备类型和设备年龄。设备类型有 4 种,分别为 model1、model2、model3 和 model4。从数据量级上来看。设备运行中每小时记录的数据共 876 100 条,其中正常数据 875 381 条,占比 99.92%;故障数据 719 条,占比 0.08%;维护数据 2 163 条,占比 0.25%。设备故障与设备正常的比例约 1:1 217,严重不平衡,在之后的数据集选取中需要对负样本进行下采样来构造训

练集与测试集。

2.3 设备属性特征构建

由于预测性维护数据中采集到的设备基本属性仅有 4 个,特征空间维度较低,不利于设备故障预测,本文通过对设备运行数据和售后历史数据进一步分析,基于原始 4 维特征构造出 24 维衍生特征。

首先,以设备运行数据为基础将设备属性特

征由 4 维扩展至 14 维。设备运行时从传感器获取的特征有 4 个,分别是 volt(电压)、rotate(转速)、pressure(压力)和 vibration(振动),这些属性均满足正态分布且属性值的范围基本确定,本文根据分布特点将每个属性划分为 6 个 bins,各 bins 的频数分布如图 4 所示。

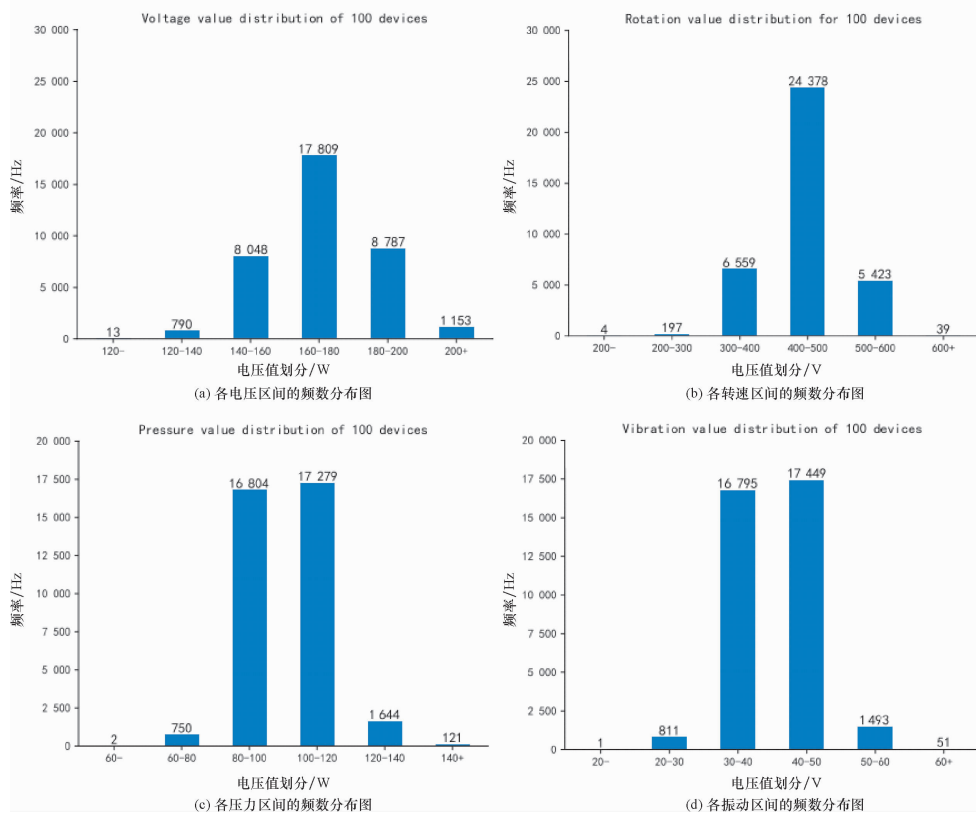


图 4 设备 4 个属性区间的分布直方图

Fig. 4 Distribution histogram of four attribute intervals of equipment

由此得到 4 个衍生特征,分别为 volt_bins、rotate_bins、pressure_bins 和 vibration_bins,特征值与属性区间的映射关系见表 1。如果属性值在区间

内,则对应区间号 i ,区间序号 i 的值根据值区间的大小,由低到高分别对应 $1, 2, 3 \dots 6, i \in R$ 。

表 1 属性区间映射

Table 1 Attribute interval mapping

volt_bins	1	2	3	4	5	6
volt	$(-\infty, 120)$	$[120, 140)$	$[140, 160)$	$[160, 180)$	$[180, 200)$	$[200, +\infty)$
rotate_bins	1	2	3	4	5	6
rotate	$(-\infty, 200)$	$[200, 300)$	$[300, 400)$	$[400, 500)$	$[500, 600)$	$[600, +\infty)$
pressure_bins	1	2	3	4	5	6
pressure	$(-\infty, 60)$	$[60, 80)$	$[80, 100)$	$[100, 120)$	$[120, 140)$	$[140, +\infty)$
vibration_bins	1	2	3	4	5	6
vibration	$(-\infty, 20)$	$[20, 30)$	$[30, 40)$	$[40, 50)$	$[50, 60)$	$[60, +\infty)$

设备运行的原始数据中,表示日期的字段为 datetime,其包含年、月、日、时、分、秒。由于采集的是设备一年的数据,且运行数据以小时为单位进行统计,因此,年份相同,分、秒字段均为 0,数据具有一定的冗余。本文对日期和时间细颗粒度化,将运行时间拆分为 6 个特征,分别是 month(月)、day(日)、quarter(季度)、week of year(一年中的星期数)、day of week(一星期中的天数)和 day of year(一年中的天数),以此从更多时间维度进行分析。

接下来以设备售后历史数据为基础将设备属性特征由 14 维扩展至 24 维。在 14 维特征构建的基础上,增加设备故障、报错、维护等相关特征。对于设备元、设备故障和维护数据,本文关注故障、维护的时间,以及发生的时间间隔,新增 4 个特征属性、4 个标签属性。4 个特征属性分别是 age、maint、m2m 和 f2f,4 个标签属性为 label_1d、label_1w、label_1m 和 is_in_maint。age 代表设备年龄,maint 代表设备当前是否进行了维护,构建方法是将设备运行数据与维护数据关联,如果当前运行数据也在维护数据表中则置 1,否则为 0;m2m 和 f2f 分别代表距离上次设备维护的时间间隔和距离上次发生故障的时间间隔,即计算两个置 1 数据的时间差。label_1d、label_1w 和 label_1m 分别代表 1 天后的设备状态、1 周后的设备状态和 1 个月后的设备状态。标签的构建方法是先确定当前运行数据中的设备状态,将设备运行数据与故障数据进行关联,如果当前运行数据在维护数据表中则置 1,不在则置 0。然后根据设备状态构建标签属性,即将设备状态前移 X ,label_1d、label_1w、label_1m 对应的预测周期 X 分别为 1、7、30 和 31。由于设备的年龄均匀分布在(0,20),本文假设设备的保修期均为 10,is_in_maint 标签代表当前设备是否在保修期内,如果设备年龄未超过保修期,即($age \leq 10$ 时)is_in_maint 标签置 1,否则 is_in_maint 标签置 0。

考虑到设备的报错情况与设备故障的相关度高,因此,新增 6 个报错相关的特征属性,包括当日总报错量 error_sum,以及当日各个报错类型的报错量 error1_num、error2_num、error3_num、error4_num 和 error5_num。最终构建的 24 维属性特征的详细情况见表 2。

表 2 设备特征构建参数

Table 2 Equipment feature building parameters

变量类型	变量	特征参数	含义
影响变量	X1	volt	设备每小时的电压均值
	X2	rotate	设备每小时的转速均值
	X3	pressure	设备每小时的压力均值
	X4	vibration	设备每小时的振动均值
	X5	volt_bins	设备电压值分区号
	X6	rotate_bins	设备转速值分区号
	X7	pressure_bins	设备压力值分区号
	X8	vibration_bins	设备振动值分区号
	X9	month	月
	X10	day	日
	X11	quarter	季度
	X12	weekofyear	一年中的星期数
	X13	dayofweek	一星期中的天数
	X14	dayofyear	一年中的天数
	X15	maint	是否进行了维护
	X16	m2m	距离上次维护的时间间隔
	X17	f2f	距离上次故障的时间间隔
	X18	age	设备年龄
	X19	error_sum	当日总报错量
	X20	error1_num	当日 error1 的报错量
	X21	error2_num	当日 error2 的报错量
	X22	error3_num	当日 error3 的报错量
	X23	error4_num	当日 error4 的报错量
	X24	error5_num	当日 error5 的报错量
决策变量	Y1	label_1d	设备的故障标签,预测 1 天后是否故障
	Y2	label_1w	设备的故障标签,预测 1 周后是否故障
	Y3	label_1m	设备的故障标签,预测 1 个月是否故障
	M	is_in_main	保内保外标签,标志设备是否在保修期之内

2.4 灰色关联度分析

经过特征构建,设备具有 28 个属性,其中包含 24 个特征属性和 4 个决策属性。由于属性较多,本文分析影响设备的主要属性,对各影响因素进行重要性排名,以此提取主要特征,降低特征空间维数。

本文选择构建的 24 个特征属性作为特征变量,对应表 2 中 X1 ~ X24,选取 Y1、Y2、Y3 作为 3 个决策变量,分别对 Y1、Y2、Y3 进行灰色关联度分

析。由于不同属性的数值范围差异大,本文采取均值法,进行无量纲化处理,以 Y_1 为例处理后的数据见表 3。

表 3 无量纲化数据
Table 3 Non-dimensional data

X_1	X_2	X_3	...	X_{24}	Y_1
1.05	1.13	1.10	...	0.00	0.00
0.84	0.99	0.93	...	0.83	0.00
0.86	1.09	0.88	...	1.67	0.00
...
1.19	1.11	1.07	...	0.00	11.24
0.80	0.69	0.89	...	0.00	11.24

然后,选取标签类别属性 Y 作为系统的参考序列,将各个特征属性参数 X 作为比较序列,分别与参考序列求差序列,然后根据灰色关联度公式,计算出灰色关联度系数。由于本文有 3 种类别属性 Y_1 、 Y_2 、 Y_3 ,需进行 3 次灰色关联度系数的计算。以 Y_1 为例,计算得到的特征属性 $X_1 \sim X_{24}$ 与决策属性 Y_1 关联系数大小见表 4。

表 4 各特征属性与决策属性的关联系数

Table 4 Relation coefficient of each characteristic attribute and decision attribute

X_1	X_2	X_3	...	X_{24}	Y_1
0.998 6	0.999 0	0.995 8	...	0.982 2	0.982 2
0.997 4	0.998 4	0.994 1	...	0.999 9	0.985 7
0.996 0	0.999 7	0.998 6	...	0.986 2	0.985 3
...
0.998 8	0.997 9	0.999 7	...	0.979 0	0.851 6
0.998 1	0.998 5	0.995 2	...	0.986 3	0.846 8

对基于 Y_1 、 Y_2 和 Y_3 的 3 个关联度结果进行综合分析,选取出主要特征。本文计算关联系数对应的平均数作为灰色关联度,并将关联度由高到低排序,得到基于 Y_1 、 Y_2 和 Y_3 的关联度,见表 5。

根据关联度,对 Y_1 、 Y_2 和 Y_3 的 3 种决策属性下的 24 个特征变量进行由高到低排序,对比 Y_1 、 Y_2 和 Y_3 的 3 个排序结果中尾部特征,计算尾部 5 个特征的重合度,删去重合度最高的属性,保留其余属性作为主要特征。基于 Y_1 的尾部特征为 X_{15} 、 X_{16} 、 X_{21} 、 X_{14} 和 X_{24} ;基于 Y_2 的尾部特征为 X_{16} 、 X_{17} 、 X_{14} 、 X_{24} 和 X_{21} ;基于 Y_3 的尾部特征为

X_{15} 、 X_{16} 、 X_{21} 、 X_{14} 和 X_{24} 。其中 X_{17} 的重合度为 1, X_{16} 的重合度为 2, X_{21} 、 X_{14} 和 X_{24} 的重合度均为 3,本文删除重合度最高的 4 个特征属性,分别为 X_{21} 、 X_{14} 、 X_{24} 与 X_{16} 。其余 20 个特征属性为主要特征属性被保留下来,分别是 vibration、rotate、volt、pressure、volt_bins、pressure_bins、rotate_bins、day、error5_num、vibration_bins、quarter、f2f、monthdayofweek、weekofyear、error4_num、error3_num、error2_num 和 error_sum,实现了 24D ~ 20D 特征提取,在保证主要特征完整的情况下减少了不必要的计算复杂度。由此,构建基于 20D-SVM 模型和基于 20D-XGBoost 模型的故障预测,最终形成制造服务数据驱动的业务挖掘模式和产品保修期内的业务关联模式的“双模式”模型。

表 5 特征属性 X 与决策属性 Y 关联度大小

Table 5 Relation between characteristic attribute X and decision attribute Y

特征变量	特征属性	Y_1 关联度	Y_2 关联度	Y_3 关联度
X_1	volt	0.997 9	0.999 1	0.998 0
X_2	rotate	0.998 0	0.999 1	0.998 2
X_3	pressure	0.997 8	0.999 0	0.997 9
X_4	vibration	0.998 3	0.999 2	0.998 4
X_5	volt_bins	0.997 6	0.998 9	0.997 8
X_6	rotate_bins	0.997 1	0.998 7	0.997 3
X_7	pressure_bins	0.997 2	0.998 7	0.997 3
X_8	vibration_bins	0.992 0	0.996 3	0.99 25
X_9	month	0.991 6	0.996 1	0.991 9
X_{10}	day	0.993 1	0.996 8	0.993 5
X_{11}	quarter	0.991 6	0.996 1	0.992 1
X_{12}	weekofyear	0.990 1	0.995 4	0.990 5
X_{13}	dayofweek	0.991 3	0.996 0	0.991 8
X_{14}	dayofyear	0.971 3	0.986 0	0.973 0
X_{15}	error_sum	0.973 2	0.986 3	0.973 6
X_{16}	error1_num	0.972 8	0.986 6	0.973 2
X_{17}	error2_num	0.973 5	0.986 5	0.973 7
X_{18}	error3_num	0.974 2	0.986 8	0.974 5
X_{19}	error4_num	0.977 4	0.988 2	0.976 0
X_{20}	error5_num	0.992 3	0.996 4	0.992 6
X_{21}	age	0.971 4	0.985 9	0.973 1
X_{22}	maint	0.979 0	0.990 0	0.980 0
X_{23}	f2f	0.980 7	0.991 0	0.982 1
X_{24}	m2m	0.971 1	0.986 0	0.972 4

2.5 设备风险等级划分

基于构建的 FME 模型,分别计算出平均故障时间间隔(F)、平均维护时间间隔(M)和平均报错时间间隔(E)3类指标。

F = 总运行时间/总故障次数,即从 2015 年 1 月 1 日至 2016 年 1 月 1 日的总小时数($366 * 24 = 8\ 784\ h$)/该设备在此时间段的故障总次数。

M = 总运行时间/总维护次数,即从 2015 年 1 月 1 日至 2016 年 1 月 1 日的总小时数($366 * 24 = 8\ 784\ h$)/该设备在此时间段的维护总次数。

E = 总运行时间/总报错次数,即从 2015 年 1 月 1 日至 2016 年 1 月 1 日的总小时数($366 * 24 = 8\ 784\ h$)/该设备在此时间段的报错总次数。

接下来通过 K-Means 聚类分析将设备故障风险等级划分为高风险和低风险两种,故提前将 k 设置为 2。本文采用随机函数来设置初始聚类中心位置,如表 6 所示为迭代 200 次后得到聚类结果,此时聚类中心点间距最大,划分效果最好。

表 6 各数据属性类别描述

Table 6 Description of each data attribute category

设备类别	高风险设备	低风险设备
聚类数	92	8
所占比例/%	92	8
F 均值	1 212.42	3 660.00
M 均值	308.40	305.29
E 均值	229.02	261.36

通过分析聚类中心、 F 均值、 M 均值和 E 均值,对聚类结果进行分析,总结出如下特征:

(1) 高故障风险设备:此类设备占比 92%,可见绝大多数的设备是不稳定的,存在很大的故障隐患。它们的平均故障间隔时间间隔为 1 212 个小时,即平均 1.7 个月出现一次故障。它们的平均报错间隔时间间隔为 229 个小时,即平均 1.3 周发生一次异常。从维护的角度来看,此类设备平均维护间隔时间间隔为 308 个小时,即平均 12.8 天进行一次维护,因此在实际业务场景中,此类设备属于高故障风险设备。

(2) 低故障风险设备:此类设备占比 8%,数量很少,说明基本“无隐患”的高稳定性设备仍是少见的。它们的平均故障间隔时间为 3 660 个小

时,即平均 5 个月出现一次故障。它们的平均报错间隔时间间隔为 261 个小时,即平均 1.6 周发生一次异常。从维护的角度来看,此类设备平均维护间隔时间间隔为 305 个小时,即平均 12.7 天进行一次维护,在实际业务场景中,此类设备属于低故障风险设备。

上述结果表明,故障风险等级不同的设备发生故障的频率不同,采用单一的预测模型难以在低成本的情况下实现对所有设备进行故障预测。因此,需要对比不同模型在不同场景下的效率,以便对于不同的设备使用不同的预测模型来提升预测效率。对于保外期的设备,若为高故障风险设备,则采用高频率的 pred_1w 预测机制进行短期预测;若为低故障风险设备,则采用低频率的 pred_1m 预测机制进行长周期预测。

2.6 三折交叉验证调参

本文 SVM 模型选取的核函数是 RGB,两个超参数分别是惩罚系数 C 和核函数的影响参数 γ ,实验中设置的 C 取值范围是 $\{32, 64, 128, 256, 512\}$, γ 范围是 $\{1e-1, 1e-2, 1e-3\}$ 。

XGBoost 模型采用 L2 正则化项,其参数设置为 $n_estimators = 50$ 、 $max_depth = 10$ 、 $learning_rate = 0.01$ 、 $objective = binary:logistic$ 、 $\gamma = 0$ 、 $min_child_weight = 1$ 、 $subsample = 0.85$ 和 $reg_lambda = 1$ 。

本文调节的超参数组合为树的个数 $n_estimators$, 树的最大深度 max_depth 和学习速率 $learning_rate$, 实验设置的 $n_estimators$ 的取值范围是 $\{50, 100, 200\}$, max_depth 的取值范围是 $\{5, 10, 15\}$, $learning_rate$ 的范围是 $\{0.06, 0.1, 0.6\}$ 。

表 7 为三折交叉验证实验结果。调参后 SVM 模型 4D 特征空间下最佳超参数为 $C = 512$, $\gamma = 0.001$; 14D 和 24D 特征空间下最佳超参数为 $C = 32$, $\gamma = 0.001$ 。XGBoost 模型在 4D 特征空间下最佳超参数为 $learning_rate = 0.6$ 、 $max_depth = 5$ 和 $n_estimators = 100$; 14D 特征空间下最佳超参数为 $learning_rate = 0.6$ 、 $max_depth = 5$ 和 $n_estimators = 50$; 24D 特征空间下最佳超参数为 $learning_rate = 0.06$ 、 $max_depth = 5$ 和 $n_estimators = 200$ 。这表明在不同的情况下最佳超参数的取值差异很大,充分验证了参数调参的必要性和有效性。

表 7 三折交叉验证实验结果

Table 7 Results of three-fold cross-validation experiment

超参数	Dim	4D			14D			24D		
	γ	0.1	0.01	0.001	0.1	0.01	0.001	0.1	0.01	0.001
C	32	0.039	0.357	0.376	0.000	0.069	0.394	0.000	0.016	0.389
	64	0.039	0.372	0.385	0.000	0.069	0.378	0.000	0.016	0.387
	128	0.039	0.366	0.402	0.000	0.069	0.381	0.000	0.016	0.387
	256	0.039	0.363	0.406	0.000	0.069	0.380	0.000	0.016	0.387
	512	0.039	0.364	0.413	0.000	0.069	0.380	0.000	0.016	0.387

2.7 面向保内保外的故障预测结果分析

本节对面向保内保外的故障预测结果进行分析与比较。

2.7.1 模型评估指标

本文基于混淆矩阵,使用准确率(accuracy)、精确率(precision)、召回率(recall)、F1分数(F1-score)和AUC(Area Under Curve)这5个指标来全面衡量故障预测模型的性能。

如混淆矩阵作为一种可视化工具可用来评估二分类模型的效果。根据模型的预测结果与真实分类结果将所有样本划分为4类。本文将“设备故障”定义为正类,“设备正常”定义为负类。真正例(True Positive, TP)表示将真实发生故障的设备标签预测为故障;真负例(True Negative, TN)表示将真实未发生故障的设备标签预测为正常;假正例(False Positive, FP)表示将真实未发生故障的设备标签预测为故障;假负例(False Negative, FN)表示将真实发生故障的设备标签预测为正常。

准确率是预测正确的样例占整体的比例,正确包括两种情况,即将故障设备预测为故障,真实是正常的设备预测为正常,计算方式如公式1。

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%。 (1)$$

精确率是预测为故障设备中真实是故障设备的比例,召回率是真实是故障的设备中预测为故障的比例,计算方式如公式2。

$$Precision = \frac{TP}{TP + FP} \times 100\%。 (2)$$

召回率是真实是故障的设备中预测为故障的比例,计算方式如公式3。

$$Recall(TPR) = \frac{TP}{TP + FN} \times 100\%。 (3)$$

F1分数是对准确率和召回率做一个权衡,是精确率和召回率的调和平均数,计算方式如公式4。

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\%。 (4)$$

ROC(Receiver Operating Characteristic)曲线全称是“受试者工作特征”曲线,以FPR为横坐标,TPR为纵坐标,可准确反映分类器真正例率和假正例率的关系。TPR为真正类率,等同于召回率,反映了分类器的灵敏度,FPR为假正类率,计算的是真实是正常的设备中分类器错认为是故障的比例,计算方式如公式5。

$$FPR = \frac{FP}{FP + TN} \times 100\%。 (5)$$

AUC为ROC曲线下的面积,表示预测的正例排在负例前面的概率,其数值越大代表模型的性能越优。

2.7.2 模型预测结果分析与比较

深度学习模型MLP可用于学习从输入到输出的映射,其结构简单、模型训练难度低、具有较好的识别率且分类速度快,适用于本研究中的表格数据集,故选择此深度学习模型进行对比实验。本文配置的MLP模型包含一个隐藏层,隐藏层中的神经元数量为100,激活函数为relu,此时模型的损失函数在训练集上收敛在测试集上发散,如图5(a)所示(其中MLP model loss是MLP模型损失函数,epoch是迭代轮次)。不断调节units最终确定神经元数量为5时损失函数收敛,此时训练集和测试集的损失函数曲线如图5(b)所示。对SVM模型、XGBoost模型与深度学习模型MLP在不同预测周期下进行实验,使用20D-SVM模型、20D-XGBoost模型与深度学习模型MLP预测一天后(pred_1d)、一周后(pred_1w)和一个月后(pred_1m)设备故障情况,对比分析模型的预测性能。3个模型的实验结果见表8。

从同一模型不同预测周期角度来看,20D-SVM模型预测一天的准确率和召回率最高,其次

为预测一个月,最后是预测一周;20D-XGBoost 模型预测一个月的准确率和召回率最高,其次为预测一天,最后是预测一周。

使用 MLP 模型对构建的 20 维数据进行故障预测,从结果来看,MLP 模型的预测准确率不足 80%,召回率很低,不到 30%,与 SVM 和 XGBoost 模型相比预测效果较差,证实了 20D-SVM 模型与 20D-XGBoost 模型的优越性。

综上所述,对于保内期的设备,由于数据完整,将采取基于 20D-SVM 模型进行 pred_1d 的每日故障预测。由于保外期的设备数据难以追踪,故不宜进行每日故障预测。对于保外期的高风险设备,其发生故障的频率较高,将采取基于 20D-SVM 模型 pred_1w 的短周期故障预测;对于保外期的低风险设备,其发生故障的频率较低,将采取基于 20D-XGBoost 模型 pred_1m 的长周期故障预测。

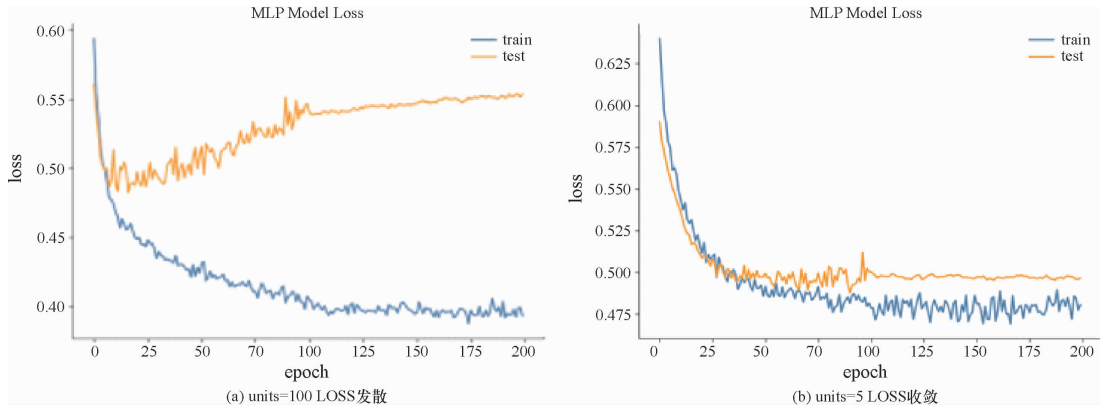


图 5 不同 units 下 MLP 的 LOSS

Fig. 5 Loss of MLP under different units

表 8 SVM、XGBoost 和 MLP 3 种模型的预测结果
Table 8 Prediction results of SVM, XGBoost and MLP models

Time period	Evaluation index	20D-SVM	20D-XGBoost	MLP
Pred_1d	Accuracy/%	87.04	85.22	78.45
	Precision/%	66.79	64.56	72.59
	Recall/%	100.00	96.01	27.92
	F1	0.800 9	0.772 1	0.403 3
Pred_1w	Accuracy/%	80.83	77.70	74.94
	Precision/%	57.49	53.76	92.31
	Recall/%	100.00	100.00	3.49
	F1	0.730 1	0.699 2	0.067 3
Pred_1m	Accuracy/%	84.67	95.19	74.08
	Precision/%	62.69	84.26	50.00
	Recall/%	100.00	100.00	2.18
	F1	0.770 7	0.914 6	0.041 8

2.7.3 特征空间构建效果分析

为检验特征构建在故障预测方面的效果,本文将不同特征空间下模型的预测性能进行对比分析。如图 6 所示,在预测周期为一天、一周和一个月 3 种场景下,分析 3 种特征空间效果,3 种特征空间分别是原始 4 维特征(4D)、以设备运行数据为基础构建的 14 维特征(14D),以及以设备售后

历史数据为基础构建并提取的 20 维特征(20D)。本文使用 SVM 与 XGBoost 两种模型,通过 ROC 曲线检验特征空间的构建对模型预测效果的影响。

如图 6(a)、图 6(b)、图 6(c)分别是 SVM 模型在 3 种预测周期下不同特征空间的 ROC 曲线。越靠近左上角的 ROC 曲线所代表模型的准确性越高,可以看出在任意预测场景中,随着特征工程的构建,SVM 模型预测准确性不断提高,14D 优于原始 4D,20D 优于 14D,在最终特征构建后达到最优的预测效果。从 AUC 值来看,3 种预测场景下,原始 4D 的 AUC 值在[0.65,0.81]区间,预测模型的效果较差;14D 的 AUC 值在[0.85,0.91]区间,预测效果一般,但较 4D 有所提升;20D 的 AUC 值在[0.93,0.97]区间,较 14D 进一步提升,预测效果很好。图 6(d)、图 6(e)、图 6(f)分别是 XGBoost 模型在 3 种预测周期下不同特征空间的 ROC 曲线。对于 XGBoost 模型,在 pred_1d 和 pred_1w 预测场景中,构建的 14D 和 20D 特征没有原始 4D 预测效果好,但 20D 预测效果优于 14D;在 pred_1m 预测场景中 20D 预测准确性优于 4D 和 14D,说明 20D 特征构建在 XGBoost 模型长周期预测上效果更好。

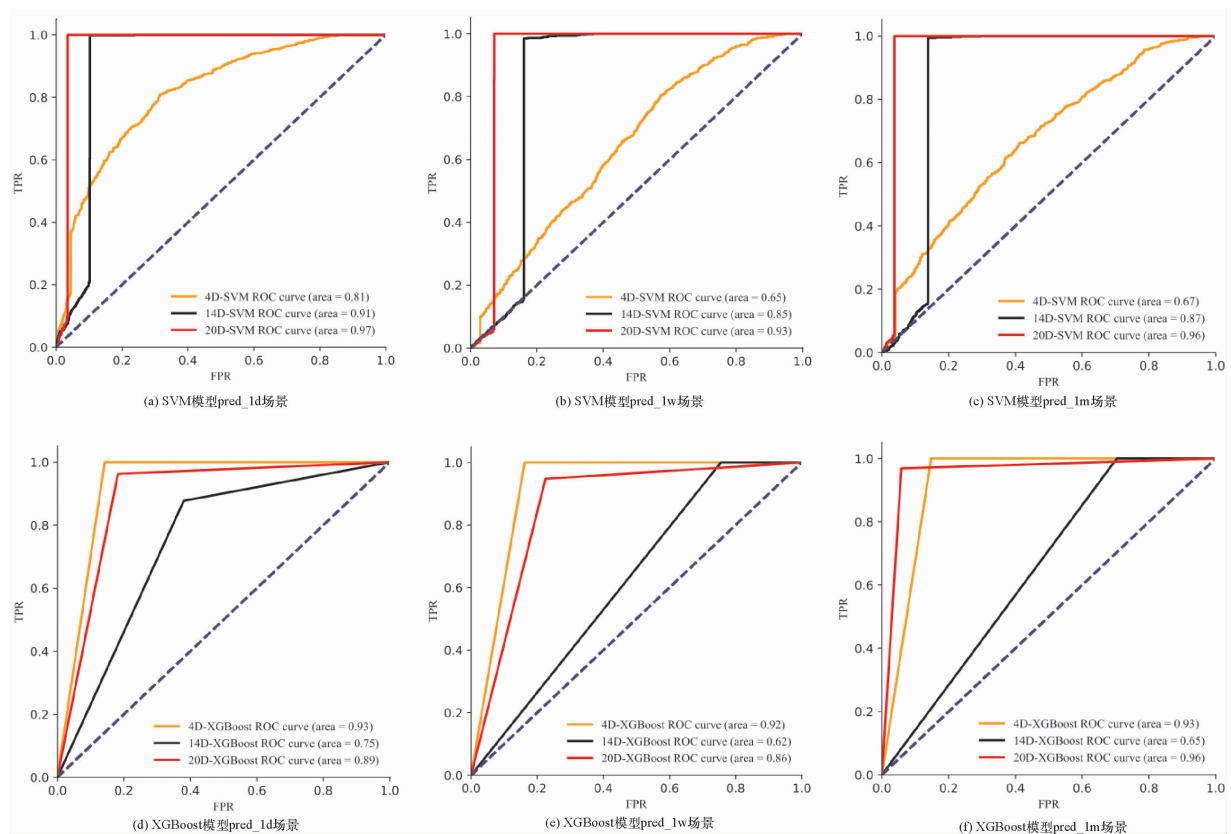


图 6 3 种预测场景,3 种特征空间下 SVM 与 XGBoost 模型的 ROC 曲线

Fig. 6 ROC curves of SVM and XGBoost models in three prediction scenarios and three feature spaces

由此可见,无论是 SVM 模型还是 XGBoost 模型,20D 的 AUC 值在 $[0.86, 0.97]$, 原始 4D 的 AUC 值在 $[0.65, 0.93]$, 说明特征工程的构建在整体上具有较好的预测效果。

3 总结与展望

本文基于保内保外场景中对设备进行故障预测研究,设计出一种基于模型集成的保内保外设备故障预测方法,该方法具有较高准确率的故障预测能力。该方法中,保内保外故障预测协同模型的设计包括 3 个部分,分别是面向保内保外设备属性特征构建、基于灰色关联度分析的主特征提取,以及结合 SVM 和 XGBoost 的故障预测集成

模型构建。其中,面向保内保外的 4D-24D 特征构建包含以设备运行数据为基础的 4D-14D 特征构建和以设备售后历史数据为核心的 14D-24D 特征构建。为减少模型参数,通过灰色关联度分析从 24D 特征空间中提取 20D 主要特征。基于 20D-SVM 和 20D-XGBoost 的故障预测协同模型在构建中结合了保内保外实际业务场景和设备的故障风险等级两个因素,并结合多种预测机制,经过实验测试,该方法具有良好的预测能力。

本文以预测性维护数据集为基础,提出了基于模型集成的保内保外设备故障预测方法,通过实验验证了该方法预测的有效性,但在真实工业生产环境中的效果未知,还需要通过实际运维数据进一步验证模型的通用性。

参考文献:

- [1] Zio E. Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice [J]. Reliability Engineering & System Safety, 2022, 218: 108119.
- [2] 沈保明,陈保家,赵春华,等. 深度学习在机械设备故障预测与健康研究中的研究综述[J]. 机床与液压,2021,49(19):162-171.
- [3] 黎巧能,徐勇,刘凤伟,等. 铁路货车制动系统和轴承的故障诊断及预测[J]. 铁道车辆,2022,60(6): 139-145.

- [4] Qian H, Sun B, Guo Y, et al. A parallel deep learning algorithm with applications in process monitoring and fault prediction[J]. Computers and Electrical Engineering, 2022, 99:1-13.
- [5] Hu Y, Li J, Hong M, et al. Industrial artificial intelligence based energy management system: Integrated framework for electricity load forecasting and fault prediction[J]. Energy, 2022, 244: 123195.
- [6] 林东明, 医疗设备的维修与保养工作研究[J]. 中国设备工程, 2022(22): 17-19.
- [7] 纪红波, 李瑜, 龙张亮. 地铁车辆架大修维修模式研究[J]. 城市轨道交通研究, 2022, 25(12): 216-220.
- [8] 王希, 叶耀祖, 巩美娟. 飞行试验机务维修保障信息化发展思考[J]. 航空维修与工程, 2022(12): 17-19.
- [9] Uppal M, Gupta D, Juneja S, et al. Cloud-based fault prediction for real-time monitoring of sensor data in hospital environment using machine learning[J]. Sustainability, 2022, 14(18): 11667.
- [10] 金晓航, 王宇. 工业大数据驱动的故障预测与健康管理[J]. 计算机集成制造系统, 2022, 28(5): 1314.
- [11] PHM 行业: 亿渡数据-2022 年中国 PHM 行业研究报告-230111 [EB/OL]. (2022-12-30) [2023-02-15]. <https://www.waitang.com/report/417400.html>.
- [12] Katona T. Decentralized finance: The possibilities of a blockchain“money lego”system[J]. Financial and Economic Review, 2021, 20(1): 74-102.
- [13] Xu T A, Xu J. A short survey on business models of decentralized finance (DeFi) protocols[EB/OL]. (2022-02-15) [2022-02-21]. <http://arxiv preprint arXiv: 2202. 07742, 2022>.
- [14] 程中华, 董恩志, 王荣财, 等. 延伸保修策略下故障相关多部件系统可用度优化模型[J]. 陆军工程大学学报, 2022, 1(5): 31-37
- [15] Corbet S, Goodell J W, Gunay S, et al. Are DeFi tokens a separate asset class from conventional cryptocurrencies? [EB/OL]. (2021-04-20) [2022-01-13]. Available at SSRN 3810599, 2021.
- [16] Popescu A D. Transitions and concepts within decentralized finance (DeFi) space[J]. Social Sciences Research, 2020(9):40-61.
- [17] 柳尧云, 林润玮, 阎虎勤. 福建省 2023 年碳达峰前二氧化碳排放趋势研究——基于 GM(1,1), GM(2,1) 与 GM(1, 1) 邓聚龙灰色预测模型[J]. 应用数学进展, 2021, 10(6): 2013-2020.
- [18] Sairamesh L, Sabena S, Selvakumar K. Energy efficient service selection from IoT based on QoS using HMM with KNN and XGBoost[J]. Wireless Personal Communications, 2022, 124(4): 3591-3602.
- [19] 帅云峰, 张友卫, 孙栓柱, 等. 基于多尺度主元分析的给水泵故障检测方法: CN201810113302.8[P]. 2022-02-08.
- [20] 张运铎. 基于 GA-BP 神经网络技术的计算机网络故障诊断模型研究[J]. 无线互联科技, 2023, 20(1): 149-152.
- [21] Behnke R. Explained: The superfluid hack[EB/OL]. (2022-02-01) [2022-02-06]. <https://halborm.com/explained-the-superfluid-hack-february-2022/>.
- [22] Rahimian R, Eskandari S, Clark J. 2019 IEEE European Symposium on Security and Privacy Workshops (EuroS & PW) [C]. Piscataway: IEEE, 2019. doi:1019/Euro SPW.2019.00042.
- [23] Rekt. Explained: The furucombo evil contract hack[EB/OL]. (2021-02-27) [2021-12-06]. <https://halborm.com/explained-the-furucombo-evil-contract-hack-feb-2021/>.
- [24] Rahman D, Mohammad D, Hamid M, et al. A survey of fault prediction and location methods in electricat energy distribution networks[J]. Measurement, 2021, 184: 109947.
- [25] Wang A, Wang H, Jiang B, et al. 2020 7th International Conference on Dependable Systems and their Applications (DSA) [C]. Piscataway: IEEE, 2020. doi:10.1109/DSA51864.2020.00031.
- [26] 周才杰, 汪玉洁, 李凯铨, 等. 基于灰色关联度分析-长短期记忆神经网络的锂离子电池健康状态估计[J]. 电工技术学报, 2022, 37(23): 6065-6073.
- [27] Lei H, Govindaraju V, Beal M J. Speeding up multi-class SVM evaluation via principle component analysis and recursive feature elimination[J]. Computer Science, 2015(4): 1-23.
- [28] Wu Y, Zhang Q, Hu Y, et al. Novel binary logistic regression model based on feature transformation of XGBoost for type 2 Diabetes Mellitus prediction in healthcare systems[J]. Future Generation Computer Systems, 2022, 129: 1-12.
- [29] Amjad M, Ahmad L, Ahmad M. Prediction of file bearing capacity using XGBoost/algorithm: Modeling and performance evaluation[J]. Applied Sciences, 2022, 12(4): 2126.