

基于Stacking模型融合的工程机械核心部件寿命预测研究

梁 超

(浙江理工大学信息学院, 浙江 杭州 310018)

摘 要: 预测性维护是工业互联网应用的重点, 实现预测性维护的关键是对设备系统或核心部件的寿命进行有效预测。随着近年来机器学习的发展, 机械设备海量数据已成为工业互联网分析核心部件剩余寿命的关键指标, 也成为设备健康管理决策性数据。基于工程机械设备大数据, 结合XGBoost、随机森林、LightGBM等多种机器学习模型, 多维度探究影响机械核心部件寿命的机器学习模型效果, 建立Stacking算法模型融合的部件寿命预测模型, 并在核心部件数据上验证模型预测有效性, 从而减少设备非计划停机时间, 推进智能制造和预测性维护的进步。

关键词: 工程机械; 寿命预测; 机器学习; Stacking

中图分类号: TP181 **文献标识码:** A

Life Prediction of Construction Machinery Core Components Based on Stacking Model Fusion

LIANG Chao

(School of Information Science and Technology, Zhejiang SCI-TECH University, Hangzhou 310018, China)

Abstract: Predictive maintenance is the focus of industrial Internet application. The key to predictive maintenance is to effectively predict the life of equipment system or core components. With the development of machine learning in recent years, the massive data of mechanical equipment has become the key index of the residual life of core components in industrial Internet analysis and the decision-making data of equipment health management as well. Based on big data of construction machinery, multi-dimensional search of sample characteristics affecting the life of mechanical core components is conducted by combining multiple machine learning models, such as XGBoost, Random Forest and LightGBM, where Stacking algorithm is used to construct the component life prediction model, and the validity of model prediction is verified on the core component data. This reduces unplanned downtime and advances the intelligent manufacturing and predictive maintenance.

Keywords: construction machinery; life prediction; machine learning; Stacking

1 引言(Introduction)

在工业4.0的环境下, 工程机械设备自动化发展高速, 设备部件在传统机械行业运行通信环境恶劣, 为工程机械设备监控管理带来了新的挑战。设备部件长期运行, 零部件寿命减少, 可靠性降低将大大影响工程设备使用, 甚至威胁人类生命财产安全。那么及时维护更换工程设备核心部件成为设备健康管理的关键。寿命预测是工程机械设备安全运行的重要基础^[1]。早期部件寿命研究指的是基于理论物理学和统计学, 计算部件寿命, 探索其电压、电流、转速、工作时长、温度等数据挖掘出指标之间的规律^[2]。统计学模型则选择合适的寿命分布模型, 建立统计学可靠性高的概率性公式研究零

部件特征分布, 如正态分布、指数分布等^[3]。物理学寿命预测模型则是根据零部件运行过程物理感应失效模型, 损伤力学, 能量等方法对零部件失效类型进行了定义分析, 对各类零部件退化机能对应物理模型定义, 均偏重理论研究^[4,5]。目前已经广泛运用在汽车零部件等制造业中, 但是在工程机械设备使用过程中, 存在着数据非线性、不等长、维度多等波动, 众多参数甚至是传感器无法及时传播的, 导致理论寿命计算出现较大误差。

本文运用大数据分析机器学习的方法评估零部件运行中晚期健康状态, 数据挖掘分析核心部件电动机在相同情况下历史数据, 探究影响机械核心部件寿命指标关联关系,

如电流电压、振动、噪声、转速、温度等，结合机器学习K-means聚类分析、线性回归算法、随机森林、LightGBM和XGBoost等方法，构建与寿命相关的高质量特征Stacking集成算法预测模型，可预先知晓设备零部件结果，及时检测更换，有利于在机械设备使用中期发现设备零部件异常状态，增强零部件健康质量把控。

2 工程机械核心部件寿命预测数据处理(Processing of life prediction data for core components of construction machinery)

中期寿命预测是工程设备监控管理的重要内容，主要针对对机械设备运行过程中出现的状态把控，防止运行过程中发生意外^[6]。很多设备零部件在未充分使用到设计寿命规定年限时，就已经损耗严重需要报废。如果不能及时发现，将会造成很大的影响，设备不合理使用也将会造成极大的浪费^[7]。通常在实际使用环境中预测存在较大偏差。因为现实采集到的数据数量庞大，复杂多变，基础数据质量差，能都对大量数据探索性分析研究，从中挖掘出与寿命相关隐藏的信息非常重要。

针对工程机械设备耗损性部件电动机，获取到数据集包含训练集和测试集两部分。训练集中涵盖电动机全寿命物联网采样数据，即从安装后一直到更换之间的对应数据，形式为多维时间序列^[8]。字段“部件工作时长”的最大值，即为该部件实例的实际寿命。测试集中包含部件一段时间内的电动机物联网采样数据，基于该段数据，预测电动机此后的剩余寿命。数据集中样本EDA后特征数据字段如表1所示。

表1 特征数据字段

Tab.1 Feature data field

字段名称	数据类型
部件工作时长	数值型
累积量参数1	数值型
累积量参数2	数值型
转速信号1	数值型
转速信号2	数值型
压力信号1	数值型
压力信号2	数值型
温度信号	数值型
流量信号	数值型
电流信号	数值型
开关1信号	0或1
开关2信号	0或1
告警信号1	0或1
设备类型	字符串型

数据标准化(归一化)处理，将消除指标之间的量纲影响，解决数据指标之间的可比性，使各指标处于同一数量级，适合进行综合对比评价^[9]。采用Min-Max

Normalization如公式(1)，对数据进行归一化构造量化[0, 1]区间，特征缩放后梯度下降过程会更加笔直，收敛可以得到更快的提升，得到数据归一化处理结果如表2所示。

$$Z = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (1)$$

表2 数据归一化处理结果

Tab.2 Min-Max normalization

字段名	Min	Max
部件工作时长_shape	10.0	580508.0
累积量参数1_max	2227.5	401764.0
累积量参数2_max	0	378345.5
转速信号1_argmax	0	429846.0
转速信号2_argmax	0	93459.0
压力信号1_argmax	4.0	502504.0
压力信号2_argmax	1.0	207233.0
温度信号_max	47.8	150.0
温度信号_argmax	0	311276.0
Life	246.75	22801.75

3 工程机械核心部件寿命预测模型构造(Construction of life prediction model for core components of construction machinery)

3.1 K-Means聚类

使用K-Means聚类无须进行模型的训练，无监督学习效率高，处理大数据集，算法保持可伸缩性和高效性当簇接近高斯分布时，效果较好^[10]。本文利用优秀的特征进行聚类，使用最小距离分类器MDC进行测试获得最佳K均值聚类k值为5，结果得分如表所示。

表3 K-Means聚类结果

Tab.3 K-Means result

k_class	0	1	2	3	4
部件工作时长max	0.446	0.044	0.108	0.54	0.263
部件工作时长std	0.435	0.043	0.106	0.525	0.258
累积量参数1std	0.464	0.046	0.115	0.525	0.269
累积量参数2max	0.481	0.054	0.126	0.56	0.285
累积量参数2std	0.483	0.055	0.126	0.558	0.287
转速信号1argmax	0.092	0.01	0.019	0.172	0.048
转速信号2argmax	0.41	0.041	0.077	0.465	0.177
压力信号1argmax	0.08	0.008	0.019	0.159	0.039
压力信号2argmax	0.137	0.02	0.03	0.186	0.069
温度信号_max	0.521	0.227	0.509	0.576	0.413
流量信号argmax	0.714	0.047	0.09	0.052	0.184
得分score	0.028	0.007	0.029	0.023	0.026

3.2 Stacking模型融合

Stacking堆叠(元组合)是指组合多个预测模型信息生成新模型的一种模型组合技术。堆叠模型(也称为二级模型)，可以

组合每个基本模型优秀的能力，忽略其他模型缺点，组合起来将优于单个模型^[11]。因此，当基本模型模拟效果不同各有优缺点时，选择模型堆叠是最有效地。本研究选择了八个单模型进行预测，有Lasso回归、ElasticNet弹性网络、核岭回归Kernel Ridge Regression、决策树DecisionTree、随机森林RandomForest、GBDT、XGBoost、LightGBM。

各个模型预测效果精度得分如表4和图1所示，可以观察到决策树、XGBoost和LightGBM模型效果较好平均精度得分可以达到0.03左右，三种模型入选第二层学习器，为了防止过拟合现象发生，将随机森林模型也加入第二层学习器中。

表4 八种基础模型预测效果得分

Tab.4 Scores of predictive effect of eight basic models

模型名称	预测得分
XGBRegressor	0.031725
lightgbm	0.031193
DecisionTreeRegressor	0.022978
GradientBoostingRegressor	0.029988
RandomForestRegressor	0.013027
Lasso	0.005298
ElasticNet	0.005319
KernelRidge	0.017499

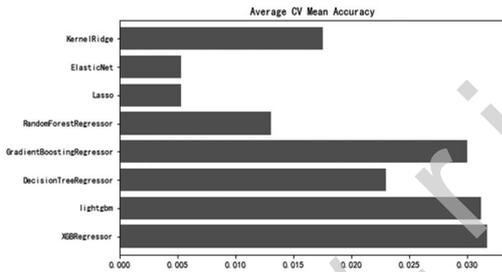


图1 八种基础模型预测效果对比

Fig.1 Comparison of prediction effects of 8 basic models

使用Stacking(堆叠)模型对多个单一模型进行组合。以初级训练集八种模型训练出初级训练器，生成一个新数据集训练刺激第二层学习器，结合四个LR顶层模型进行次级训练^[12]，多模型融合过程如图2所示。

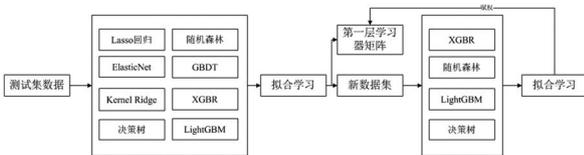


图2 Stacking多模型融合过程

Fig.2 Stacking multi-model fusion process

$n=8$ 个初级学习器，对初级学习器 M_n ，利用测试集数据 D 进行训练，将训练完成的 M_n 预测训练集 D 和测试集 T 的标签列，得到结果为 P_n 和 T_n ^[13]。随后将 n 个学习模型结果合并，得到次级学习器 M' 的训练集 $P' = (P_1, P_2, P_3, \dots, P_n)$ 和测试集 $T' = (T_1, T_2, T_3, \dots, T_n)$ 。利用 P' 训练次级学习器 M' ，并预测 T' 得到最终的预测结果。

3.3 模型融合交叉验证

交叉验证指的是通过估计模型的泛化误差，选择模型的方法。优点在于不做任何假定前提，具有应用的普遍性，操作简便，是一种行之有效地模型选择方法^[14]。通过将数据集均分成 K 个子集，并依次将其中的 $K-1$ 个子集作为训练集，剩下的1个子集用作测试集。在 K 折交叉验证的过程中，每个子集均会被验证一次。本研究使用四折交叉验证方法构造，稳健性强，交叉验证过程如图3所示，对每一折进行预测，采用两层循环，第一层循环控制基模型的数目，第二层循环控制交叉验证四次，则对每一个基模型会训练四次，最后求均方根拼接得到预测结果。

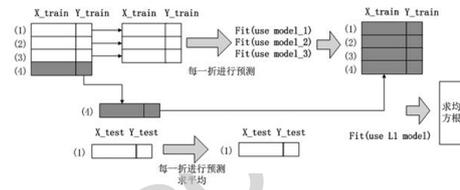


图3 四折交叉验证预测过程

Fig.3 Four-fold cross validation for prediction process

4 模型预测结果(Model prediction results)

使用Stacking模型融合方法对所有测试集进行预测，使用时间为四天以内的电动机生命周期在8000—12000天，预测寿命和实际寿命相对均方根误差在0.1以内。使用时间为30天以内的电动机生命周期在2000—6000天，预测寿命和实际寿命相对均方根误差在0.4以内。使用1600天以内的电动机生命周期在4000天以内，相对预测寿命和实际寿命相对均方根误差在0.75以内。由此可见，基于Stacking算法的工程机械核心部件寿命预测模型预测效果良好。

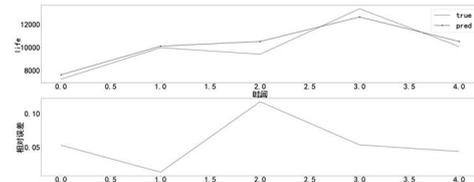


图4 四天寿命模型预测结果

Fig.4 Prediction results of four days life model

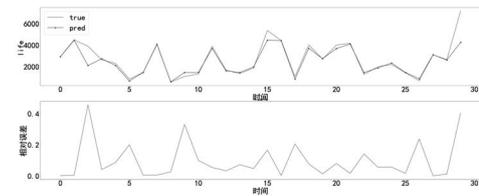


图5 30天寿命模型预测结果

Fig.5 Prediction results of 30 days life model

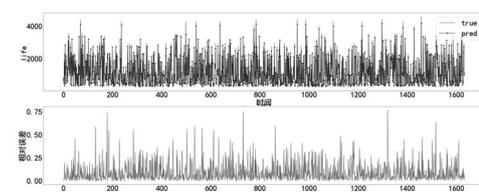


图6 1600天寿命模型预测结果

Fig.6 Prediction results of 1600 days life model

5 结论(Conclusion)

本文将大数据分析机器学习的方法运用到工程机械自动化领域,提出了一种基于Stacking模型融合的工程机械核心部件寿命预测算法。从原始电动机传感数据入手,经过数据准备和预处理,K-means聚类生成训练样本,选择Lasso回归、ElasticNet弹性网络、核岭回归Kernel Ridge Regression、决策树DecisionTree、随机森林RandomForest、GBDT、XGBoost、LightGBM。八种子模型作为Stacking模型融合的基分类器,随机森林、GBDT、XGBoost和LightGBM作为Stacking模型融合的次级分类器进行识别,最终证明本算法具有良好的寿命预测效果,均方根误差小,精度高。有助于改变机械自动化领域传统被动的维修模式,转向预测维修自动化方面,从而降低核心部件使用风险,提高设备使用效能。

参考文献(References)

- [1] Sbarufatti, M. Corbetta, A. Manes, et al. Monte-Carlo sampling based on a committee of artificial neural networks for posterior state estimation and residual lifetime prediction[J]. International Journal of Fatigue, 2016:35-38.
- [2] Xiang Wan, Dong Wang, Peter W. Tse, et al. A critical study of different dimensionality reduction methods for gear crack degradation assessment under different operating conditions[J]. Measurement, 2016(02):22-24.
- [3] 杜党波, 司小胜, 胡昌华, 等. 基于随机退化建模的共载系统寿命预测方法[J]. 仪器仪表学报, 2018(08):61-62.
- [4] 牛乾. 机械旋转部件的性能退化及其寿命预测方法研究[D]. 浙江大学, 2018:6-8.
- [5] 封杨, 黄筱调, 陈捷, 等. 基于小样本试验的大型回转支承剩余寿命预测[J]. 中南大学学报(自然科学版), 2015(09):12-13.
- [6] 尤明懿. 基于状态监测数据的产品寿命预测与预测维护规划方法研究[D]. 上海交通大学, 2012:43-46.
- [7] ChunLin Zhang, Bing Li, BinQiang Chen, et al. Weak fault signature extraction of rotating machinery using flexible analytic wavelet transform[J]. Mechanical Systems and Signal Processing, 2015(06):89-90.
- [8] 周志华. 机器学习[M]. 北京:清华大学出版社, 2016:67-69.
- [9] Audi Ahmad, Pierrot-Deseilligny Marc, Meynard Christophe, et al. Implementation of an IMU Aided Image Stacking Algorithm in a Digital Camera for Unmanned Aerial Vehicles. [J]. Sensors (Basel, Switzerland), 2017(7):11-13.
- [10] Truong Giang Nguyen, Muller Jean-Michaël, Rupnik Ewelina, et al. Second Iteration of Photogrammetric Processing to Refine Image Orientation with Improved Tie-Points[J]. Sensors (Basel, Switzerland), 2018(7):21-22.
- [11] YiJun Chen, Man-Leung Wong, Haibing Li. Applying Ant Colony Optimization to configuring stacking ensembles for data mining[J]. Expert Systems With Applications, 2014(6):46-47.
- [12] 沈长青. 旋转机械设备关键部件故障诊断与预测方法研究[D]. 中国科学技术大学, 2014:67-68.
- [13] 肖婷, 汤宝平, 秦毅, 等. 基于流形学习和最小二乘支持向量机的滚动轴承退化趋势预测[J]. 振动与冲击, 2015(09):11-13.
- [14] 孙闯, 何正嘉, 张周锁, 等. 基于状态信息的航空发动机运行可靠性评估[J]. 机械工程学报, 2013(06):90-92.

作者简介:

梁超(1993-), 女, 硕士生. 研究领域: 工业互联网, 软件应用, 机器学习.

(上接第40页)

管理者和学校的管理决策者都能实时地了解实验室资源的使用情况, 并且通过大数据分析发现一些长期闲置不用的“僵尸”资产, 从而使得校产绩效考核管理趋于精细化。通过这些数据来决策是否应该添置某种类型的设备, 是否应该扩展楼宇等, 进一步提高决策的准确性, 提升国有资产的利用效能。

6 结论(Conclusion)

校园效能监管平台的构建一方面解决了目前高校所存在的由于信息不对称、管理效率不对称, 造成的管理者无法科学决策, 以及基础设施严重浪费等现象。解决了长期困扰我们的如何有效提高学校管理效率, 降低管理成本, 提高为师生服务的水平的问题, 进一步为学校的发展添砖加瓦。

参考文献(References)

- [1] Barrera JF, Mira A, Torroba R. Optical encryption and QR codes: Secure and noise-free information retrieval[J]. OPTICS EXPRESS, 2013, 21(5):5373-5378.
- [2] Sun YC, Song HB, Jara AJ, et al. Internet of Things and Big Data Analytics for Smart and Connected Communities[J]. IEEE ACCESS, 2016(4):766-773.
- [3] Kim AA, Sunitiyoso Y, Medal LA. Understanding facility management decision making for energy efficiency efforts for buildings at a higher education institution[J]. ENERGY AND BUILDINGS, 2019, 199:197-215.
- [4] 徐雯, 刘永贵. 美国高校: 云服务超越基础设施成为战略走向[J]. 中国教育网络, 2017(02):67-69.
- [5] 先晓兵, 陈凤, 王加年, 等. 基于“云大物移”的高校资产管理的实践与创新[C]. 高等学校智慧校园应用案例集, 2017:100-108.
- [6] 黄叶超. 物联网技术在高校智慧实验室构建中的应用[J]. 佳木斯职业学院学报, 2018(12):402-403.
- [7] 陈少涌, 李哲夫, 徐晓旋, 等. 基于微信小程序的签到系统设计与实现[J]. 中国教育信息化, 2018(11):87-92.
- [8] 朱士中, 先晓兵, 张尔喜, 等. 高校资产效能管理决策研究与实践[J]. 办公自动化, 2017(08):53-56.

作者简介:

朱峰(1985-), 男, 助理工程师. 研究领域: 计算机科学与应用.

先晓兵(1971-), 男, 硕士, 副教授. 研究领域: 数据挖掘, 数据分析.