

杭州电子科技大学

硕士学位论文

题目：基于机器学习的金属带锯条剩余寿命预测研究

研究生：俞 锋

专 业：电子与通信工程

指导教师：陈科明 副教授

完成日期 2021 年 6 月

杭州电子科技大学硕士学位论文

基于机器学习的金属带锯条剩余寿命预测研究

研 究 生： 俞 锋

指导教师： 陈科明 副教授

2021 年 6 月

**Dissertation Submitted to Hangzhou Dianzi University
for the Degree of Master**

Remaining Life Prediction of Metal Band Saw Blade Based on Machine Learning

Candidate: Yu Feng

Supervisor: Associate Prof. Chen Keming

June, 2021

杭州电子科技大学

学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

论文作者签名：俞锋 日期：2021 年 5 月 28 日

学位论文使用授权说明

本人完全了解杭州电子科技大学关于保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属杭州电子科技大学。本人保证毕业离校后，发表论文或使用论文工作成果时署单位名称仍然为杭州电子科技大学。学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公布论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存论文。（保密论文在解密后遵守此规定）

论文作者签名：俞锋 日期：2021 年 5 月 28 日

指导教师签名：陈明 日期：2021 年 5 月 28 日

摘 要

在“中国制造 2025”以创新驱动为方针的大背景下，传统制造业开始通过工业智能化解决方案实现升级。目前金属带锯床的智能化主要体现在锯切自动化和故障诊断方面，对带锯条剩余寿命的预测处于探索阶段。过早更换带锯条会带来带锯条的浪费，而过迟更换带锯条会带来锯切效率和质量的下降，准确预知带锯条寿命非常重要。目前对带锯条剩余寿命的判断方法主要有失效物理和数据融合，其在工业实践中存在效率低下、干扰过大等问题。

为了收集研究所需带锯条锯切数据，本文设计了一套完整的金属带锯床远程监控系统。通过智能网关与带锯床 PLC 对接，收集带锯床锯切信息上传到云平台进行分析，并设计 APP、Web 客户端实现系统多端监测控制。该系统完成了带锯床在线监测和数据收集功能，为带锯条剩余寿命预测提供了数据支持。

金属带锯条的剩余寿命需结合失效前的切削总面积分析，因此本文提出一套剩余寿命计算方案。通过对收集的锯条每一刀数据进行面积计算和优化计算方案，得到准确的切削面积，通过分析记录的已报废金属带锯条全寿命数据，寻找其合理失效点。从而得到失效前切削总面积，并计算每一刀后金属带锯条对应的归一化剩余寿命，作为机器学习中有监督学习的输出。为得到与剩余寿命相关的输入特征，本文提出一套基于数据驱动的剩余寿命预测特征选择方案。通过对远程监控系统收集到的其他数据进行整理和预处理，按照 4: 1 划分数据集，并结合与剩余寿命的相关性等实验测试完成输入特征的选择。

最后通过分析 K 最邻近算法、随机森林算法、LightGBM 算法三种机器学习算法原理，并基于网格搜索算法和五折交叉验证对训练集进行训练，并对测试集进行性能验证。验证结果表明 LightGBM 算法比 K 最近邻算法拥有更高精度，比随机森林算法运算快近 10 倍，其适合于工业实践。本文提出了基于 LightGBM 算法的带锯条预测模型，在 LightGBM 算法基础上进一步通过贝叶斯优化算法改进其调参过程，提高其预测精度，最终其预测结果的均方根误差为 0.0276，决定系数为 0.9908，满足带锯条剩余寿命预测的需求。

关键词：金属带锯条、远程监控系统、剩余寿命预测、机器学习

ABSTRACT

In the context of the "Made in China 2025" policy driven by innovation, the traditional manufacturing industry has begun to upgrade through industrial intelligent solutions. At present, the intelligence of metal band saws is mainly reflected in sawing automation and fault diagnosis, and the prediction of the remaining life of the band saw blade is in the exploratory stage. Replacing the band saw blade too early will result in waste of the band saw blade, while replacing the band saw blade too late will bring about a decline in sawing efficiency and quality. It is very important to accurately predict the life of the band saw blade. At present, the main methods for judging the remaining life of band saw blades are failure physics and data fusion, which have problems such as low efficiency and excessive interference in industrial practice.

In order to collect the band saw blade cutting data needed for the research, this paper designs a complete set of remote monitoring system for metal band sawing machine. Through the intelligent gateway and the band sawing machine PLC docking, the band sawing machine cutting information is collected and uploaded to the cloud platform for analysis, and the APP and Web client are designed to realize the multi-end monitoring and control of the system. The system completes the online monitoring and data collection functions of the band saw, and provides data support for the prediction of the remaining life of the band saw blade.

The remaining life of the metal band saw blade needs to be combined with the analysis of the total cutting area before failure, so this paper proposes a set of remaining life calculation schemes. By calculating the area of each blade data collected and optimizing the calculation plan, the accurate cutting area is obtained, and the reasonable failure point is found by analyzing and recording the whole life data of scrap metal band saw blades. In this way, the total cutting area before failure is obtained, and the normalized remaining life of the metal band saw blade after each cut is calculated as the output of supervised learning in machine learning. In order to obtain the input features related to the remaining life, this paper proposes a data-driven remaining life prediction feature selection scheme. Through the sorting and preprocessing of other data collected by the remote monitoring system, the data set is divided according to 4:1, and the selection of input features is completed by combining experimental tests such as the correlation with the remaining life.

Finally, through the analysis of the three machine learning algorithm principles of K nearest neighbor algorithm, random forest algorithm and LightGBM algorithm, the training set is trained

based on the grid search algorithm and the five-fold cross-validation, and the performance of the test set is verified. The verification results show that the LightGBM algorithm has higher accuracy than the K nearest neighbor algorithm, and is nearly 10 times faster than the random forest algorithm, which is suitable for industrial practice. This paper proposes a band saw blade prediction model based on the LightGBM algorithm. On the basis of the LightGBM algorithm, the Bayesian optimization algorithm is further used to improve its parameter tuning process and improve its prediction accuracy. The root mean square error of the final prediction result is 0.0276, and the coefficient of determination is 0.9908, to meet the demand for prediction of the remaining life of the band saw blade.

Keywords: Metal band saw blade、 Remote monitoring system、 Remaining life prediction、 machine learning

目 录

摘 要.....	I
ABSTRACT.....	II
第 1 章 绪论.....	1
1.1 研究背景和意义.....	1
1.2 国内外研究现状.....	2
1.2.1 带锯床远程监控系统研究现状.....	2
1.2.2 金属带锯条剩余寿命研究现状.....	2
1.2.3 机器学习技术研究现状.....	3
1.3 目前存在的问题和解决方法.....	3
1.3.1 目前存在的问题.....	3
1.3.2 解决方法.....	4
1.4 本文主要工作及章节安排.....	4
第 2 章 金属带锯床远程监控系统设计.....	5
2.1 系统总体设计.....	5
2.1.1 需求分析.....	5
2.1.2 总体框架.....	5
2.2 网关设计.....	6
2.2.1 网关硬件框架.....	6
2.2.2 网关软件设计.....	7
2.2.3 设备硬件展示.....	8
2.2.4 PLC 对接测试.....	9
2.3 云平台设计.....	10
2.3.1 设备云.....	10
2.3.2 业务云.....	10
2.3.3 性能测试.....	11
2.4 Web 设计.....	12
2.4.1 框架设计.....	12
2.4.2 功能设计.....	12
2.4.3 Web 客户端展示.....	13
2.4.4 功能测试.....	14
2.5 APP 设计.....	16
2.5.1 框架设计.....	16
2.5.2 功能设计.....	16
2.5.3 APP 客户端展示.....	18
2.5.4 稳定性测试.....	18
2.6 本章小结.....	19
第 3 章 基于数据驱动的剩余寿命预测特征选择.....	20

3.1 前期数据采集.....	20
3.2 剩余寿命计算方案.....	22
3.2.1 切削面积计算方案.....	22
3.2.2 失效点判断方案.....	26
3.2.3 剩余寿命定量计算.....	27
3.3 数据预处理.....	29
3.3.1 缺值补充.....	30
3.3.2 异值处理.....	30
3.3.3 归一化处理.....	30
3.4 实验测试.....	31
3.4.1 相关度热力图.....	32
3.4.2 线性关系可视化测试.....	33
3.4.3 数据分布平衡测试.....	34
3.5 本章小结.....	35
第 4 章 基于机器学习的带锯条剩余寿命预测模型研究.....	36
4.1 总体方案.....	36
4.2 基于 LightGBM 算法的预测模型.....	36
4.2.1 分类和回归树.....	37
4.2.2 梯度提升模型.....	38
4.2.3 LightGBM 模型.....	39
4.3 基于改进 LightGBM 算法的预测模型.....	41
4.3.1 贝叶斯优化算法.....	41
4.3.2 模型超参数设置.....	42
4.3.3 基于贝叶斯优化的改进模型.....	42
4.4 对比其他预测模型.....	44
4.4.1 K 最近邻法.....	44
4.4.2 随机森林模型.....	45
4.5 实验测试.....	47
4.5.1 模型评估指标.....	47
4.5.2 预测模型结果.....	48
4.6 本章小节.....	51
第 5 章 总结与展望.....	52
5.1 本文工作总结.....	52
5.2 展望.....	53
参考文献.....	55
附 录.....	58

第 1 章 绪论

1.1 研究背景和意义

我国是世界第一工业大国，论工业总产值我国远超美国、日本等发达国家。但是虽然我国工业总产值最高，却不是世界第一工业强国。原因是我国虽在人口红利下带来了劳动力充足，提高了工业生产的速度和规模，但在其背后导致了对环境、能耗等问题了解的不深入。2015 年 5 月，《中国制造 2025》正式签批，建立健全工业创新服务平台为工业产品研发、工艺技术改造、数据采集等需求提供技术服务支持的任务变得刻不容缓^[1]。

而锯切行业是工业中至关重要的一环，无论是基建、房地产、航空、汽车、零部件都需要用金属锯切功能。金属带锯床在锯切行业拥有很普遍的应用，但是市面上存在的大多是低端金属带锯床，如何使金属带锯床更智能、更自动、更安全、更节省资源，是金属带锯床研究的重点。如今市面上出现一些金属带锯床远程系统，大多只是记录金属带锯床的使用情况，记录位置和一些使用数据。伴随通信技术和故障监测技术的发展，智能金属带锯床也渐渐实现了故障监测和故障预警功能。由于市面上的工业互联网水平参差不齐，完成一个更标准的工业互联网平台建设准则是未来一段时间的发展方向。通过对金属带锯床远程监控系统的钻研，可以为建立统一规范的工业互联网平台提供参考^[2]。

据相关的工业统计资料显示，突发性的数控机床停机阻碍了加工行业的发展，数控机床一年中有 1/5~1/3 的时间是有刀具的失效引起的。在实际加工过程中，刀具一旦出现崩刃或者大于使用寿命还在继续加工的情况，容易出现工件质量明显下降甚至报废的情况^[3]。而早在 1999 年，徐子文等人在亚太期刊金属加工装备专栏表明了一个拥有 10 台带锯床的机械加工企业，若机器每天工作 24 小时，如果带锯条寿命提高 30%，其每年可节约 38.4 万美元^[4]。如今带锯条寿命通过带锯条科学使用、锯切参数的优化得到延长，但由于带锯条更换不及时发生加工停止等现象仍经常发生，很多企业的对此的处理方式是安排许多工人在锯切时实时查看带锯条的状态，根据他们的主观判断进行带锯条的更换。但是这样也会带来更换过早导致成本浪费，更换过晚产生次品而造成资源浪费的问题。因为带锯条的磨损与诸多要素相关，难以准确预测其磨损程度便于及时更换^[5]，因此大多数的企业都采取提前更换磨损程度较低的带锯的保守策略，牺牲一定的成本保证来总体的质量。

目前全国的金属带锯床监测平台正在建立，目的是为了带锯床锯切数据进行分析，如浙江省的缙云锯切行业作为中国锯切行业制造的最高水平的代表，其生产的锯切装备以及其特色机械装备产业在浙江省内乃至国内已经享誉盛名，然而国内的金属带锯床还存在许多不足，主要表现为：技术水平低、数控化率低、产品总体实力低和缺少理论研究。并且在带锯床数据的挖掘利用还不够深入和充分。一方面是因为金属带锯床数据本身不统一，不同的金属带

锯床使用不同的 PLC(可编程逻辑控制器),造成了数据不通用^[6]。另一方面由于金属带锯床需要挖掘的信息较多,他们往往使用不同的技术和算法来实现,如机器本身的故障监测和预警,带锯条的自动化控制,以及带锯条剩余寿命的预测。

1.2 国内外研究现状

1.2.1 带锯床远程监控系统研究现状

自 1957 年,中国制造了第一台锯床,金属带锯床的历经 63 年的积累,其加工的技术因素已经提高了数倍甚至数十倍,而数控化程度也提高了实现无人化作业,功能更加复合化^[7]。近几年来,设备远程监控系统的应用逐渐成熟,早在 2012 年武汉理工大学刘佳通过上位机、TC35 无线通信模块、数据采集模块、下位机以及 3G 视频模块进行了金属带锯床远程监控系统设计^[8]。而在 2020 年 6 月,河北大学刘腾发在 TC35 无线通信模块的基础上,集合短信发送,以及 BP 神经网络进行了金属带锯床在线监测与故障诊断系统的研究^[9]。

而国外的金属带锯床产业发展虽然较早,在金属带锯床的数控能力、加工因素上远超国内。但在智能带锯床在线监控系统上与国内发展相近,如波兰格但斯克科技大学的研究人员在 2020 年针对金属带锯床锯切的材料研究^[10],进行了金属带锯床远程监测系统的研究。

文献 8 中对金属带锯床监控的设计为 3G 通信,随着如今视频分辨率的提高已经无法满足市场的需求,并且在故障报警部分还存在缺失。文献 9 中通过神经网络进行了金属带锯床故障诊断系统,仍采用 TC35 模块用于发送短信通知,在人机交互客户端方面需要更好的优化提高,并且在视频监控部分存在缺失。上述系统中都没有做出对带锯条剩余寿命的预测功能。

1.2.2 金属带锯条剩余寿命研究现状

进行金属带锯条的剩余寿命预测前,先要了解通过什么信息来判断剩余寿命。在国外,Scheffer 等人在通过隔离刀具工作时存在的外在干扰,配合显微镜监测的方法,在微观的层面判断刀具的磨损程度,通过对刀具背面的磨损照片通过深度学习的方法,建立卷积神经网络进行寿命精准预测^[11],其通过刀面磨损判断剩余寿命;在了解金属带锯条剩余寿命如何定量,以及哪些因素影响金属带锯床寿命之后,下一步工作是如何进行数据处理和建立模型:Cho 等人通过集成各种传感器如噪音传感器、电流传感器以及振动传感器等,并且利用模糊逻辑的方法将众多数据融合进行了刀具状态的判断^[12]。

在国内,Jay Lee^[13]团队提出了一种基于振动信号的钻具剩余寿命在线评估与性能预测的混合方法。提取振动、噪音信号的奇异熵、能量熵等信息,并采用机器学习的回归算法进行模型来判断其寿命。Xu 等通过建立人工神经网络模型,通过对刀具工作时产生的复杂合力进行小波包变换,得到其动态信息量并且和其他加工时产生的数据组成训练集作为模型的输入进行训练,完成其切削各种工业金属时的刀具寿命预测的研究^[14]。

文献 11 中对金属带锯条等刀具的剩余寿命预测通过显微镜等工具识别刀具背面的磨损情况,需要拥有很高的成本,并且无法做到实时识别。文献 12 中使用多影响因素模糊逻辑开

发进行剩余寿命预测,该模型通过分析噪音、振动等因素来判断剩余寿命,一方面很多锯床没有配备噪音传感器和振动传感器,另一方面机器工作时环境因素影响过大,影响了预测准确性。而国内的研究虽然开展较晚,但是比较符合国内的现状,通过对已经可收集的数据进行大数据分析,进行数据预处理,再通过模型训练,得到剩余寿命预测值准确度较高。

1.2.3 机器学习技术研究现状

机器学习正在飞速发展,从而逐渐改变我们的生活。从 2016 年开始,阿尔法狗算法通过和与众多围棋手对弈,包括世界冠军都成了它的手下败将,之后仅仅过了一年,新的阿尔法狗 Zero 又创下新的成绩,它与知名围棋手对弈连胜了 100 盘。机器学习从本质上来讲可划为 3 种,包括监督学习算法、非监督学习算法还有强化学习算法^[15]。监督学习和其他不同的地方在于,其已经对特征进行了分类,来进行新的输入分类,而非监督学习指无分类的数据进行聚类的判断,而强化学习是通过选择评价高的行为改进学习策略^[16]。

在机器学习领域中,美国、中国、日本为主力军,占据专利总数的 80%,其中美国占据了 53%,而中国仅仅占据 23%。这是 2019 年的数据,我国从 2000 年初开始研究机器学习,从 2011 年才开始重视和发展,特别是近几年开始有了爆发性的上升^[17]。

在国内,叶志祥等人在 2018 年进行了经典预测方法的研究,并通过神经网络配合机器学习对比的方法进行了预测研究,提出 BP 神经网络和 KNN(K 最近邻)结合的模型^[18]。2020 年,姜鹏等人通过机器学习对煤炭燃烧产生温度的数据进行了处理,并通过随机森林、PSO-BP 神经网络、支持向量机等方法对比进行了煤自燃温度预测模型研究,通过根均方误差等值评价模型的性能^[19]。而王奕森进行了对随机森林和深度神经网络的深入学习和探索,完成了随机森林决策树的优化、特征子空间的选取以及深度神经网络在复杂噪声环境下的研究^[20]。在国外,今年的研究有澳大利亚墨尔本大学的 Babak Abbasi 等人通过机器学习中 ANN、KNN 等模型进行一种血液供应链的管理方案^[21]。

综上所述,在进行机器学习之前,先要根据具体问题确定使用哪大类的机器学习方法以及该类机器学习中的哪些算法。之后通过一系列预处理工作包括数据缺失补全、异常数据排除等,为之后的模型训练提供良好的输入量。最后通过根均方误差等因素来对模型进行性能评估,选择性能最好的机器学习模型进行预测。

1.3 目前存在的问题和解决方法

1.3.1 目前存在的问题

(1) 在金属带锯床远程监控系统方面,通信模块的更新升级是个问题,如今通信模块已经开始往 5G 发展,系统因据此进行升级。随着数据的积累,云平台的计算压力越来越大,如何进行算法的优化,负载均衡是需要考虑的问题。而系统设计需要拥有更好的人机交互,如何更直观更及时方便地机型设备维护和监测是共同难题。

(2) 在金属带锯条剩余寿命方面,影响带锯条的因素是复杂的,找到与寿命息息相关的因素是关键。而金属带锯条剩余寿命无法直接得到,而通过肉眼、显微镜等失效物理方法存

在没有效率、成本过高的问题，因此对金属带锯条剩余寿命进行定量分析是难题。

(3) 机器学习方面，如何获取到数据以及数据规模大小，数据噪音、缺失问题等问题的处理，以及选择监督学习还是非监督学习进行模型是关键问题。

1.3.2 解决方法

本文针对目前存在的问题，设计金属带锯床远程监控系统，使用拥有 4G、WIFI 等通信的网关传输数据，通过 Web、APP 客户端提高人机交互体验和多端监控能力，及时方便企业查看维护锯床和收到故障报警信息。并通过结合数据驱动方法对带锯条历史数据库进行分析找到合理更换点计算剩余寿命，并对金属带锯床的数据进行预处理。最后通过监督学习中的三个回归模型进行实验，用根均方误差、 R^2 得分、Z 统计值、运行时间进行模型性能评价对比得到最合适的模型，并对该模型进行改进。

1.4 本文主要工作及章节安排

本文主要研究的是基于机器学习的金属带锯床远程监控系统，实现数据展示、故障预警以及带锯条剩余寿命预测功能，通过自主研发的智能网关将金属带锯床 PLC 与设备云互联进行数据传输，设计 Web 端和客户端，实现设备基础信息、GPS 定位、智能锯切模式、视频监控、故障预警等功能，使工业制造更智能、更精密。通过对切削率的分析结合收集到的锯条全寿命信息对锯条剩余寿命进行定量计算供训练使用，对数据进行相关性测试，保障了机器学习模型的输入特征与剩余寿命的高相关性。最后通过机器学习的回归模型对带锯条的剩余寿命进行预测，实现带锯条的及时更换、高效使用，减少了人力资源的浪费。

文章的内容安排如下：

第一章：绪论，详细介绍基于机器学习的金属带锯床远程监控系统的研究背景和意义，对金属带锯床远程监控系统的现状和机器学习的现状进行分析，确定研究的方向。

第二章：金属带锯床远程监控系统设计，依照业务分析设计完成总体的系统框架，其次针对框架每个部分进行设计和测试，完成一套健全的系统为寿命预测提供数据。

第三章：基于数据驱动的剩余寿命预测特征选择，通过分析数据来驱动特征的选择。通过查阅文献了解带锯条的合理使用和寿命影响因素结合可收集数据实情，完成数据前期收集。通过面积计算方案和失效点判断方案完成金属带锯条寿命计算，得到有监督学习的输出。对收集的其他数据进行数据预处理，4: 1 分割训练集和测试集，与剩余寿命 RUL 之间进行相关性测试等可视化实验，完成机器学习有监督学习的输入数据特征选择。

第四章：基于机器学习的带锯条剩余寿命预测模型研究。再深入了解机器学习基础的上，通过 K 最邻近算法、随机森林算法、LightGBM 算法进行带锯条剩余寿命预测模型建立并实验。最后通过与目标值的均方根差等评价因素来评估模型，以得到最优的预测模型并改进。

第五章：对全文的系统设计和关键技术研究加以总结，并对更大数据、更多变量的金属带锯床远程监控系统提出建议和展望。

第 2 章 金属带锯床远程监控系统设计

针对目前金属带锯床维护存在的缺陷,本文设计了一种基于机器学习技的远程监控系统。系统包括金属带锯床、PLC 设备,通过智能网关通过接口将设备与云平台对接进行数据传输和 PLC 程序更新,并通过 WEB 端、APP 端来提高人机交互能力,通过数据中心的机器学习算法预测带锯条的剩余寿命,完成金属带锯床的远程监控以及带锯条的精准更换,本章主要介绍远程监控系统的平台设计与测试。

2.1 系统总体设计

2.1.1 需求分析

根据构建金属带锯床加工成本和质量的核​​心因素以及目前智能金属带锯床系统的不足,对于需要设计的金属带锯床远程监控系统,主要需求有如下几部分:

(1)改变传统设备维护策略,如:手持故障诊断仪、离线分析系统,实现设备状态在线监控、数据可视化。降低设备的人工维修成本,提高工作效率以及降低设备的故障率。

(2)系统功能实现关键数据采集和状态检测报警,关键数据如:锯条线速度、切削率、工作时间、进给力、工件材质、锯切规格等。状态监测报警包括:切削油温监测、锯床带速监测、故障报警等。

(3)实现各个设备通过网关接入物联网,开发友好的管理平台如:WEB、APP 客户端。让用户可以轻松便捷地进行金属带锯床远程在线监控,实现与设备的智能交互。

(4)通过数据库采集保存切削的数据,并通过人工智能技术,机器学习算法预测带锯床的关键因素:带锯的剩余寿命。

2.1.2 总体框架

金属带锯床远程监控系统由设备硬件、数据传输通道和人机交互三部分组成。

其中设备硬件部分由锯床本身、PLC、触摸屏组成。数据传输通道由 Bbox 智能网关、OneNet 设备云组成。人机交互模块由 WEB 端、APP 端组成通过业务云与设备云以及网关对接。如图 2.1 所示:

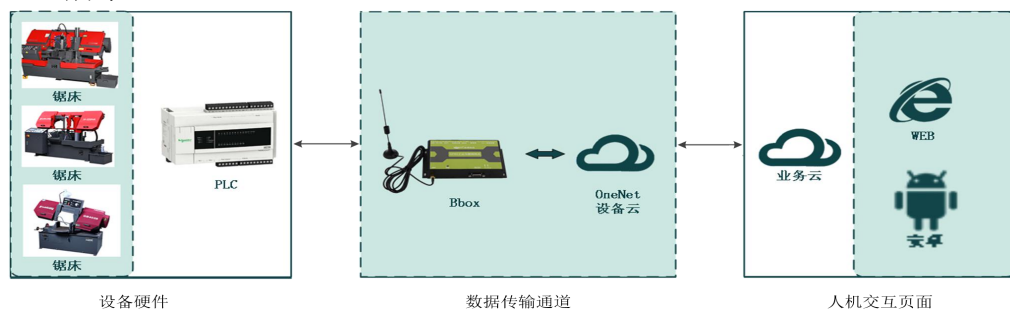


图 2.1 智能系统整体架构

系统构造主要为了完成前期数据采集工作，在采集数据的同时也横向地替提供企业金属带锯床进行故障预测和健康管理服务，并且将各个企业的金属带锯床不同 PLC 数据进行处理，统一保管数据。

网关是数据传输的基础。主要负责多 PLC 的对接和更新工作，通过更新 PLC 改变其锯切模式，保证数据质量良好符合机器学习标准。其自动连接 OneNet 设备上传数据，再通过业务云和人机交互客户端进行数据展示。在设计网关时需要考虑网关的多 PLC 接入功能，并且考虑到设备接入以及之后需要边缘计算功能，设计拥有高内存等条件。

人机交互客户端用于向企业、工厂提供一个良好的多端管理环境。通过列表、图表、地图以及摄像监控记录设备的信息、运行情况，方便统计使用情况和及时的维护。Web 客户端拥有更好的 UI 设计和功能设计，而手机 APP 端拥有更方便的交互环境，比如故障预警，在手机端可以更快地收到推送。

数据是整个系统的关键。由于需要接入不同企业不同 PLC 的数据，对数据如何选择，如何进行处理是一个问题。系统通过结合故障监测技术，进行有用数据的整理，将数据一部分用于故障预警功能，一部分用于带锯条剩余寿命预测功能。并且由于数据在复杂环境中容易受到干扰，造成数据缺失，数据异常，需对数据进行修复和排查，并且为了进一步进行机器学习模型训练，将数据进行可视化分析排除相关性低的数据，找到与带锯条剩余寿命相关性高以及符合条件的模型输入数据。

2.2 网关设计

网关是实现锯床 PLC 与设备云以及客户端数据传输的基础。考虑到网关需要自动连接上设备云，以及需要进行视频传输和环境不友好的问题，需要拥有较好的无线联网功能和有线联网功能，并且网关主要用于与不同协议的金属带锯床的 PLC 对接更新，并考虑需要预备较强的处理计算能力减轻服务器云端的计算压力。

2.2.1 网关硬件框架

网关硬件框架见图 2.2 网关进行数据传输主要使用的是 RS485 通信模块，其可进行一对多连接，且传输距离较长，适合用于对接多个金属带锯床 PLC。网关的关键模块为 MCU 模块和 JS9391 处理模块，由于金属带锯床数据量较大，需要较大存储空间，所以 MCU 模块型号选为 STM32F103RCT6。而金属带锯床带有监控摄像头，所以使用了 JS9391 处理模块。JS9391 处理模块自带 WIFI 模块，通过串口与 MCU 相连。LED 模块包括红色的工作状态指示灯、绿色的系统状态指示灯、蓝色的 4G 状态指示灯。而如果设备需要恢复出场设置，可长按复位按键 2s 至 4s。本智能网关具有存储内存、运行内存大的特点，当数据量过大时可进行边缘计算再上传服务器。

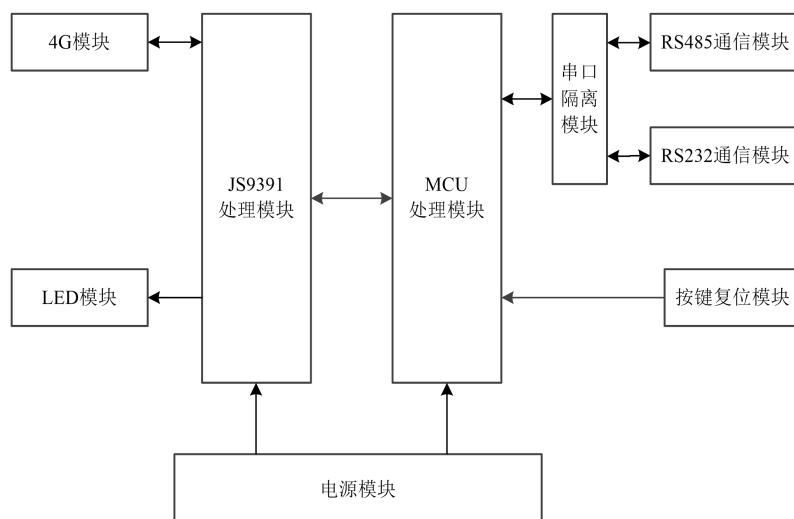


图 2.2 网关硬件框架图

2.2.2 网关软件设计

网关软件设计主要进行与云平台通信、传输协议设计处理工作，其流程为图 2.3。

网关与云平台连接设计见图，网关开机后需要和设备云 OneNet 保持连接状态，并且装有 4G 模块的网关需要在连接上设备云后进行注册，并且上报信息到业务云，若注册失败或者上传业务员失败，网关需要进行重启重新连接注册，保证数据传输无误。并且网关自身设计掉线监测和重连功能，以及按键复位恢复出场设置等功能，保证网关软件的健壮性。

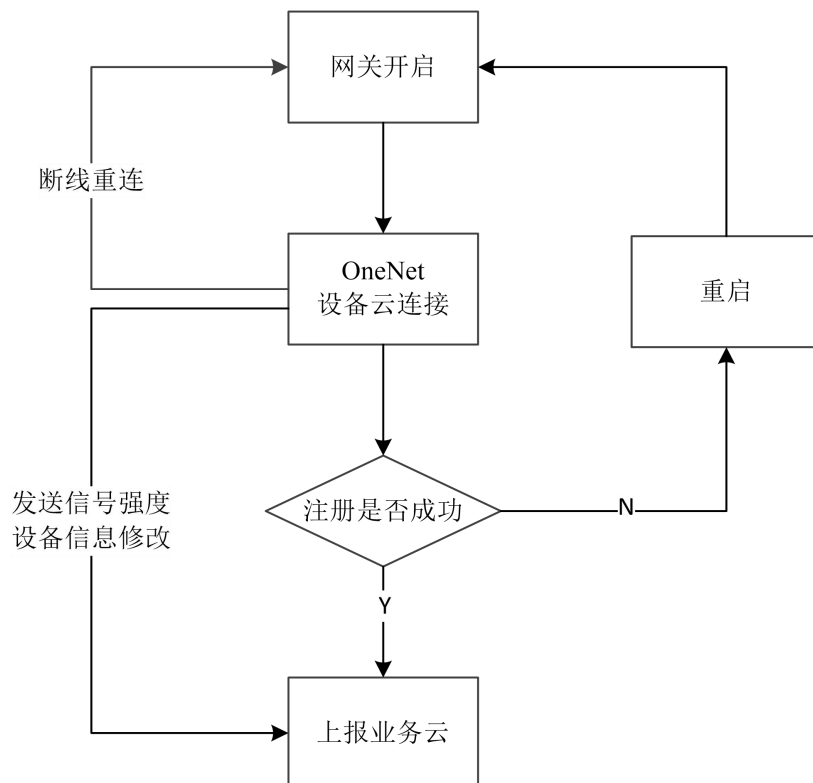


图 2.3 网关与云平台连接流程图

在完成与云平台的连接之后，进行传输协议的制定。网关通过 Modbus RTU 协议以及 Modbus ASCII 协议进行接口数据发送接受处理，通过 Modbus TCP 协议进行以太网接口数据处理。针对不同的对象制定不同的 C 结构体，如 RS485 通信结构体见下表。

表 2.1 RS485 通信结构体

结构对象	对象说明
uint8_t channel	通道号（0~10 表示哪个 485 接口，11~15 保留）
uint32_t baudrate	波特率
uint8_t wordlength	数据位（0: 8bit, 1: 9bit）
uint8_t stopbits	停止位（0 表示“1”，1 表示“0.5”，2 表示“2”，3 表示“1.5”）
uint8_t parity	校验位（0, 1, 2 分别指定无校验、奇偶校验）

数据在接受处理成功后以 Json 格式进行数据存储，保存为 txt 文件。数据传输时通过识别设备唯一标识码获取结构体头指针，再通过结构体头指针，进行数据增删改查，并最终都转换为 Json 格式数据保存在 txt 文件中。

2.2.3 设备硬件展示

智能金属带锯床实物如图 2.4 所示。



图 2.4 智能金属带锯床

网关电路板如图 2.5 所示，网关外形图如图 2.6 所示。



图 2.5 网关电路板



图 2.6 网关外形图

2.2.4 PLC 对接测试

网关的主要功能除了接受传送数据之外，还需要对 PLC 进行更新升级。由于本系统需要实现带锯条剩余寿命的预测，需要对加工程序进行严格的控制，因此需要不断优化 PLC 的锯切模式控制。从而保证剩余寿命预测的数据不是杂乱无章，而是可进行分析的优良数据。

图 2.7 是网关与 PLC 的上位机对接过程图，通过选择网关分组，填入设备 IP，加入网络组进行 PLC 对接。



图 2.7 上位机对接过程图

进行对接操作大约 30 秒后，上位机自动获得网关的 IP，通过上位机显示网关和设备在线情况以及网关与 PLC 之间通信延迟，当网关和设备都在线时可进行 PLC 的远程更新图 2.8，完成图见。通过对西门子 300、1200、1500、S7-200 的对接测试，其对接全部成功，延迟保持在 30-45ms 之间，具有优秀的通信性能。



图 2.8 上位机对接完成图

2.3 云平台设计

2.3.1 设备云

设备云使用中国移动的 OneNET 平台，负责与设备互联，设备数据存储和管理。网关通过 WIFI 建立 TCP 连接后，应用层使用 MQTT 协议将设备数据上传到 OneNET 设备云平台，同时对业务云提供 API 调用接口。平台的管理页面如图 2.9 所示。开发者可以根据具体需求进行数据流、触发器设置，监控管理接入设备。

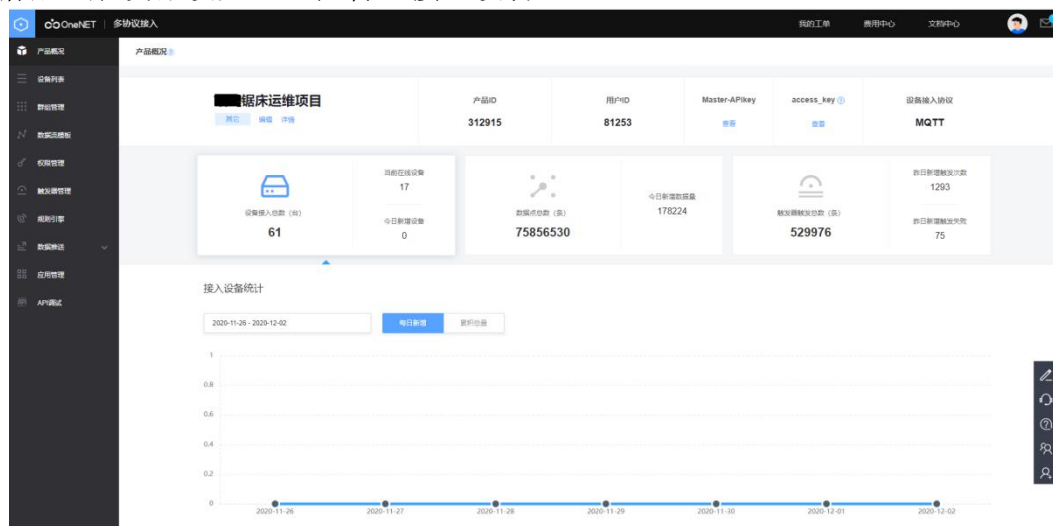


图 2.9 OneNET 锯床设备云界面

2.3.2 业务云

业务云架构如图 2.10 所示，功能是对外提供 API 调用，并且获取设备云平台的报警推送和调用设备云的 API。结构包括分发调用请求的负载均衡 SLB，运行后端程序的 tomcat 集群，持久化存储数据的数据库。负载均衡将外部的 API 请求转发到集群中的某一个 tomcat 上，tomcat 上运行着后端程序，程序根据不同的 API 请求有选择的与 OneNet，数据库交互，将业务处理结果返回给 API 调用者。业务云也负责接收处理 OneNet 平台推送的触发报警数据。

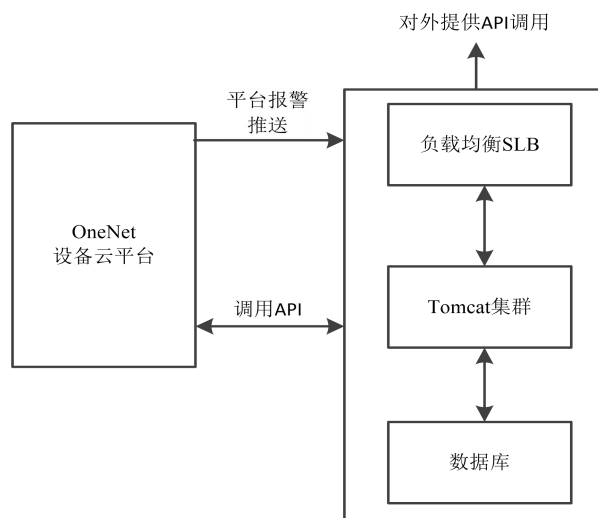


图 2.10 业务云架构

2.3.3 性能测试

测试观察设备查询接口和监控点查询接口,在加缓存前后以及使用负载均衡的性能表现,并分别求出最大并发数。设置响应事件为 1 秒、5 秒、10 秒,并分别在环境 A(不带负载均衡)的情况和环境 B(带负载均衡)的情况下进行测试见图 2.11 和图 2.12。



图 2.11 环境 A(不带负载均衡)



图 2.12 环境 B(带负载均衡)

测试结果见表格,在满足响应条件和成功率的前提下,并发数在 400-900 的范围,带负载均衡的性能优于不带缓存,拥有更大的吞吐量和更低的延时。

表 2.2 环境 A 测试结果表

线程数	请求数	1 秒	5 秒	10s	异常率	吞吐量
400	24000	18ms	3022ms	4030ms	0.00%	601.9/sec
600	36000	23ms	3215ms	9013ms	0.00%	507.0/sec
800	48000	25ms	3309ms	9220ms	0.00%	537.2/sec
900	54000	42ms	3708ms	10013ms	0.00%	515.8/sec

表 2.3 环境 B 测试结果表

线程数	请求数	1s	5s	10s	异常率	吞吐量
400	24000	17ms	3007ms	3997ms	0.00%	654.6/sec
600	36000	22ms	3208ms	9002ms	0.00%	600.0/sec
800	48000	24ms	3300ms	9104ms	0.00%	606.4/sec
900	54000	40ms	3612ms	9995ms	0.00%	552.3/sec

2.4 Web 设计

金属带锯床远程监控系统需要集成监控摄像头、并且根据业务需求，该系统一般用于企业或者工厂管理，很少用于个人。所以相对于 APP 端，Web 端是人机交互功能中的关键，设计得更加复杂和完善。

2.4.1 框架设计

Web 客户端采用的是 Vue 框架^[22]，它是一个轻巧、高性能的 MVVM 库。MVVM 框架在 Model、View、Controller 之间的处理更加高效，由于 Web 端设计相对较为复杂，需要将两个网站关联在一起，其中的数据处理频繁，故采用该框架。Vue 还有两个显著优点：响应式编程、组件化。响应式编程是一种代码风格，其面向输入和输出流，使用它可以更方便的处理数据流。而组件化方面 Web 端集成更多优秀的组件，具有更高的可扩展性。

Web 客户端的框架如下图 2.13 所示，框架以 Vue 为主体，通过其组件化特点集成第三方组件如 Bmap 来实现金属带锯床地理位置，Echart 来对金属带锯床详细数据进行表格分析。通过异步请求进行 Http 请求，完成对高并发的需求。

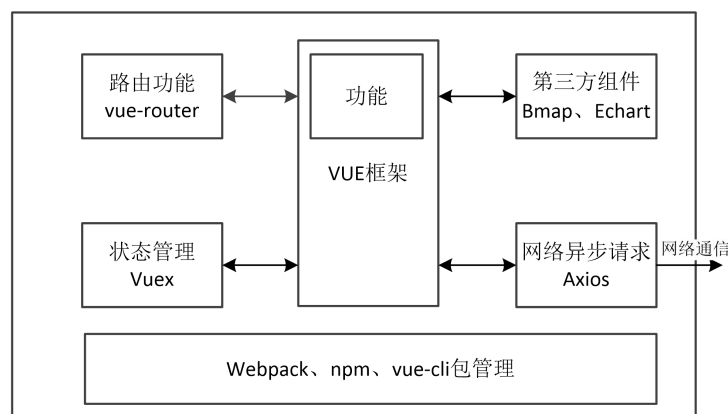


图 2.13 Web 客户端架构图

2.4.2 功能设计

在第一章中讲到，如今金属带锯床在线监控系统有许多，但是没有做统一的管理平台。所以本系统的 Web 端不光建立了针对金属带锯床数据采集管理的锯床远程运维网站，还建立了针对为不同企业提供统一管理的锯床企业管理网站。

(1) 设备数据采集端

设备数据采集端使用人员为不同的企业，要完成的功能包括设备模块、PLC 模块、监控点模块、报警模块和其他模块。设备模块首先需要在设备云进行注册，注册成功后方可进行添加设备、删除设备、设备转台推送、分享设备、移交设备等功能。PLC 模块需要进行 PLC 的添加、删除、修改以及状态推送功能，是在用户、网关和数据库之间进行通信。监控点模块进行监控点添加、删除、修改功能。报警模块进行报警推动和历史报警记录。其他模块包括图片上传功能、用户信息功能等。

(2)企业数据管理端

企业数据管理端要做的是管理上述设备数据采集端的信息，首先需要解决如何进行企业的区分，需要设计不同企业的账号编码，填入到文件模板进行导入数据。当有企业在设备数据采集端创建企业账号时，会调用设备运维的创建账号接口，在设备远程运维数据库插入一条账号记录，从而使用相同的账号密码也可以在设备远程运维平台登录。在设备远程运维系统开放设备数据的 API，企业数据采集系统调用即可获得。考虑到后期设备远程运维平台的企业会迁移，可以在数据库配置服务器地址，按企业配置。因为设备数据的 API 的格式是固定的，企业数据采集系统就可以通过调用不同服务器地址获取对应企业的设备数据。

Web 客户端相对与 APP 客户端在界面的设计上更有特点，APP 页面比较简单，目的是为了进行数据点展示和更及时的故障报警。而 Web 客户端在拥有 APP 客户端拥有的功能之外，拥有更友好的人机交互感受，比如针对不同企业可以进行企业背景设置，并且集成了视频监控功能在线监控机器的运行状态，以及地图信息更加详细，路由功能（跳转）更加复杂，详细成果在系统展示部分展示。

2.4.3 Web 客户端展示

Web 共设计两个平台，图 2.14 为企业数据管业平台，图 2.15 为远程运维系统。

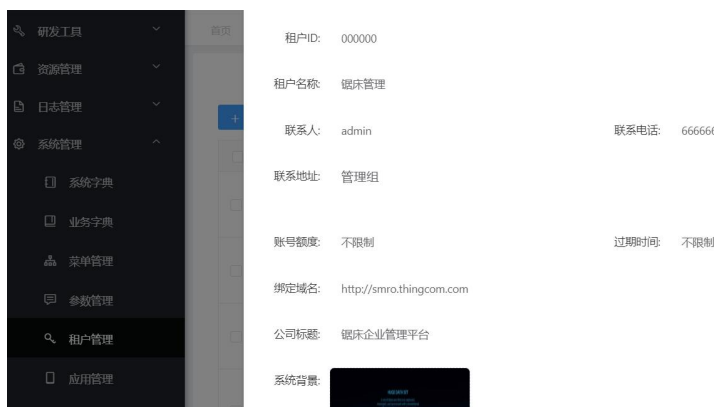


图 2.14 企业管理



图 2.15 远程运维

2.4.4 功能测试

(1)故障报警功能：将金属带锯床的故障分为三大类：锯切异常、伺服电机异常和其他问题。锯切异常分为断带故障、锯偏停机故障、锯轮故障，伺服电机故障分为进给伺服电机故障、送料伺服电机故障、升降伺服电机掉线、送料伺服电机掉线。其他问题包括水泵故障、油泵故障、水位过低报警，故障分类见图 2.16。

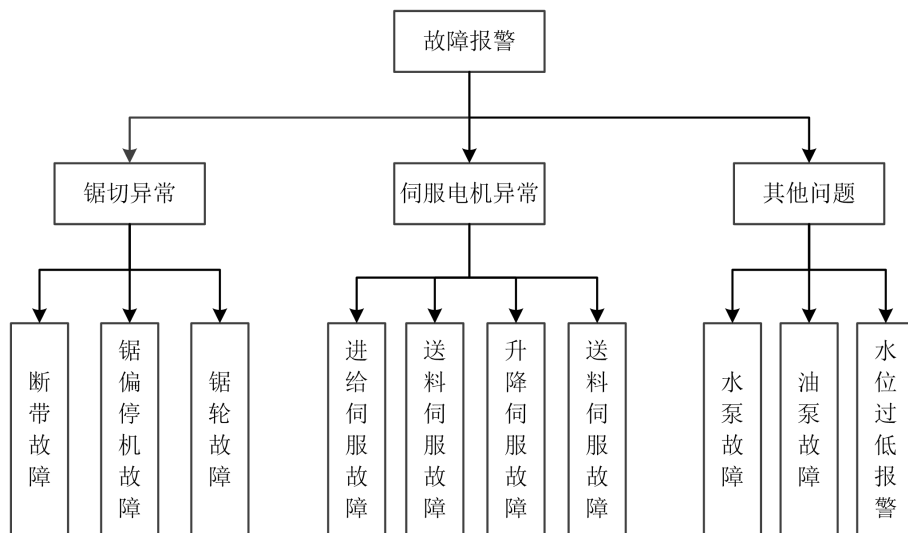


图 2.16 故障分类图

当 PLC 接收到数据点，通过网关 RS485 发送到云平台。普通数据只有在 Web 或者 APP 端请求时才获取，而故障报警数据通过推送的方式，一旦数据超过了安全界，云平台往 Web 端和 APP 端发送故障信息。下图 2.17 为 Web 故障报警接收功能界面，其提供了报警时间、处理方式、是否已处理以及处理操作。

数据点名称	报警内容	当前报警时间	数据值	状态	处理时间	应对方法	操作
水位过低报警	水位过低报警	2020-08-29 15:2 5:17	73	已处理	2020-08-29 15:2 8:22	点击查看	Ⓢ 点击查看历史报警
高压停机报警	高压停机报警	2020-09-06 10:4 7:50	79	已处理	2020-09-06 10:4 8:56	点击查看	Ⓢ 点击查看历史报警
Z锯偏停机报警	Z锯偏停机报警	2020-09-09 11:3 2:19	69	已处理	2020-09-09 11:4 2:32	点击查看	Ⓢ 点击查看历史报警
高压提示报警	高压提示报警	2020-09-15 18:0 4:39	78	已处理	2020-09-15 18:1 5:41	点击查看	Ⓢ 点击查看历史报警
低压停机报警	低压停机报警	2020-11-12 10:4 9:55	76	已处理	2020-11-12 10:5 1:00	点击查看	Ⓢ 点击查看历史报警

图 2.17 故障报警图

(2)视频监控功能:如果在金属带锯床操作期间发生上述故障情况，故障数据可以为使用者提供发现问题的依据。但是数据比较抽象，并且可能存在环境因素的影响，只有通过视频监控查看机器有无其他问题，如是否断电、正在维修等问题。通过视频监控功能，可以使问题展现更直观，采取方案更准确。因此本系统加入视频监控功能见图 2.18，提供 4G 接入和 WIFI 接入保证了高分表率的视频展示，提供更高清的影像供问题分析解决。



图 2.18 视频监控图

(3)地图功能:通过故障报警和视频监控找到问题所在之后,如果需要进行设备维修,那么地图显示功能能够帮助企业更直观地了解到金属带锯床发生故障时的地点,以便于后期开展维修等工作。并且地图功能还可以了解金属带锯床的销售分布、在线情况等。下图 2.19 为 Web 端设备地图定位功能:



图 2.19 地图总览图

(4)设备监测功能:除了进行对设备的故障进行报警,金属带锯床远程监控系统还需要时刻监测设备的运行数据并且对锯切工艺参数进行设置,进行锯切形状选择、锯切模式选择、锯切材料选择、材料高度、材料宽度、锯切面积、锯切线速度以及用户信息等展示。通过网关上传数据到云平台并且通过 Web 端和 APP 端进行显示,通过测试,Web 端工作正常可以完成设备锯切运行数据的实时展示见图 2.20。



图 2.20 设备监测

2.5 APP 设计

2.5.1 框架设计

安卓系统 APP 的设计框架发展多年，主要流行的有 MVC、MVP、MVVM。在进行程序设计之前，先要根据业务的需求分析选择合适的设计框架，才能事半功倍。如果一个 App 只有简单的页面和层次，甚至不需要使用到框架，因为使用框架反而增加了工作量。经过考察业务需求和业务体量，选用 MVC 设计框架作为金属带锯床远程监控 APP 的框架见图 2.21。

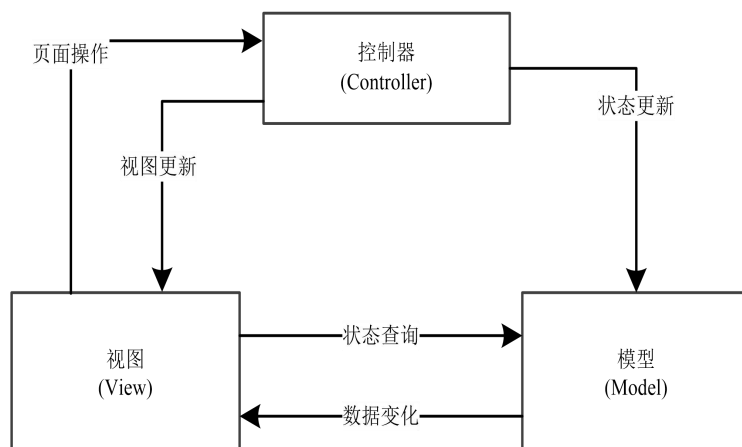


图 2.21 APP 框架设计图

在介绍 MVC 各个模块之前，先进行 Android 的简单分析，Android 的四大组件为 activity、service、content provider、broadcast receiver^[23]。其中 activity 本身就有 controller 的功能，因为 activity 页面不能进行超过 5 秒的耗时操作会造成 ANR，所以 activity 作为 controller 会通过监听 View 层的控制逻辑和变化来更新 Model 层的数据更新。而 Model 层主要进行一些耗时的操作，比如数据库的操作、网络操作以及一些耗时的算法。View 层主要是一些 xml 文件，通过在 xml 文件中布局 and 命名以及集成一些控件来完成优美的界面展示。

2.5.2 功能设计

在确定好设计框架后，开始进行业务分析。下图 2.22 是功能的总体框架，首先一个 APP 需要用户注册账号、登陆账号，之后进行主页面，主页面分为运行监控、工作台、个人信息三大块进行展示。

(1)运行监控模块：运行监控模块分为地图位置展示和设备健康管理模块，金属带锯床监控系统需要记录金属带锯床的地址，方便发生故障时维修人员及时修理，所以监控模块其中一部分是地图展示，通过对接百度地图 SDK，再放入设备的经纬坐标，可很方便得实现该功能。另一部分是设备监控，主要通过对接萤石云 SDK，在线监视机器的运行状态，以及显示故障详细信息等。

(2)工作台模块：工作台模块分为设备添加删除模块和设备详细详情展示模块，设备添加功能主要为了设备添加工程中，可以设置存放地址、设备型号、网关信息、拥有者信息等。而设备详情展示模块第一层用于显示设备列表以及设备 SN；当点击设备后显示第二层页面

设备详情列表，用于显示锯切速度。切削率、锯偏值、加工材质等信息；当点击每个详细信息的历史数据后进入可视化图表页面，该页面将历史数据分为表格显示历史数据和图表显示数据走势。

(3)个人信息模块：个人信息模块设计比较简单，主要进行名字修改、头像更换、账号退出、设备个数等信息，用一个列表展示。

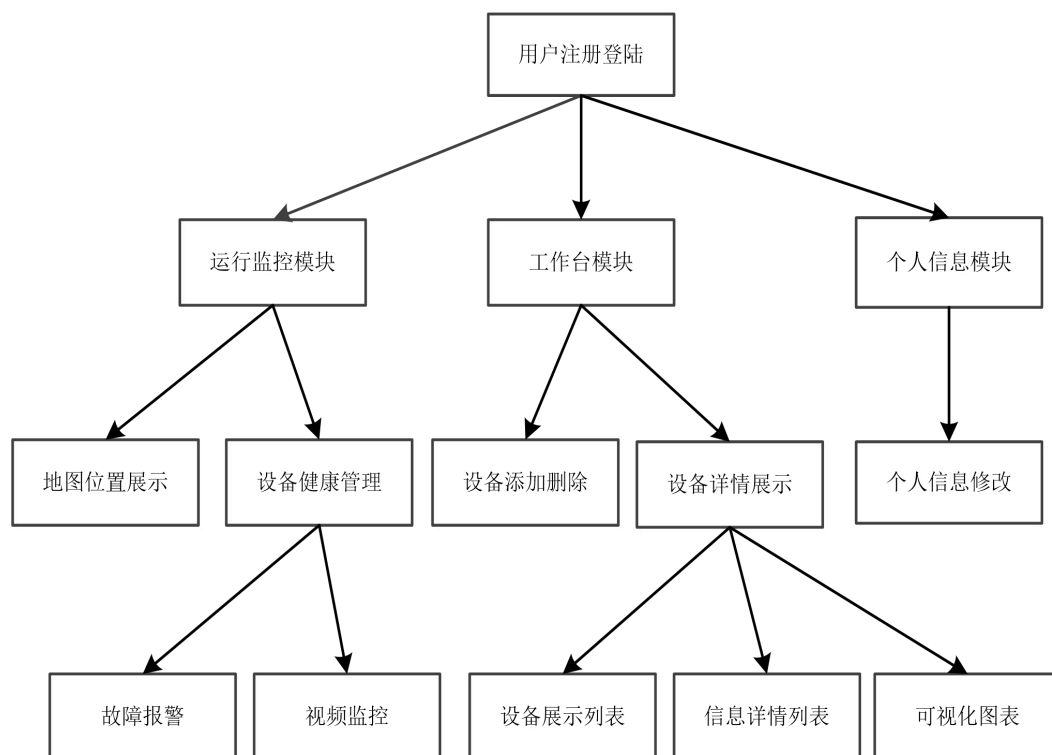


图 2.22 APP 功能设计图

通过上述业务分析了解 APP 软件需要的功能后，进行实现库选择开展工作。

整体界面：首先由于界面不多，且界面布局相对简单。我们使用到开源库 **Fragment** 第三方框架 **FragmentationX**，整个主页面为一个 **Activity**，运行监控模块、工作台、个人信息都为 **Fragment**。三个模块通过底部 **TabLayout** 和 **ViewPager** 完成页面路由功能，可通过点击 **TabLayout** 或者滑动 **ViewPager** 来实现三个 **Fragment** 的切换。

列表展示功能：在设计中会用到很多的列表，安卓开发自带的 **ListView** 使用不大方便，功能不够强大。虽然软件功能不复杂，但是为了提高代码质量和工作效率，可以使用 **RecyclerView** 配合第三方库 **BRVAH** 等完成列表的滑动删除、点击特效、装饰头等功能。

可视化图表展示功能：本软件使用到的是 **MPChartLib**，这是一款全球使用最多的软件图表显示第三方库，通过解析 **Json** 数据，将解析的数据放入 **Entry** 中，再通过对线条线性数据进行设计完成图表曲线的颜色、线形状等参数设置。

网络传输通信功能：本软件使用 **Okhttp+retrofit** 的框架，**Okhttp** 通过定义客户端的属性，对接收到的数据进行拦截。通过增删改查完成设备添加、删除、修改和查询功能，并且查询后台服务器的锯床详细信息发送到 **View** 中的列表显示。而 **retrofit** 主要使用到 **Java** 底层的注

解功能，从底层优化了 Okhttp 的使用，通过动态代理的方式减少了 Okhttp 的代码量，提高了解耦能力。

2.5.3 APP 客户端展示

APP 客户端是 Web 端的延伸，提高用户查看便捷性。其地图总览功能、设备和数据列表页面、以及数据详情页面见下图 2.23、图 2.24、图 2.25、图 2.26。

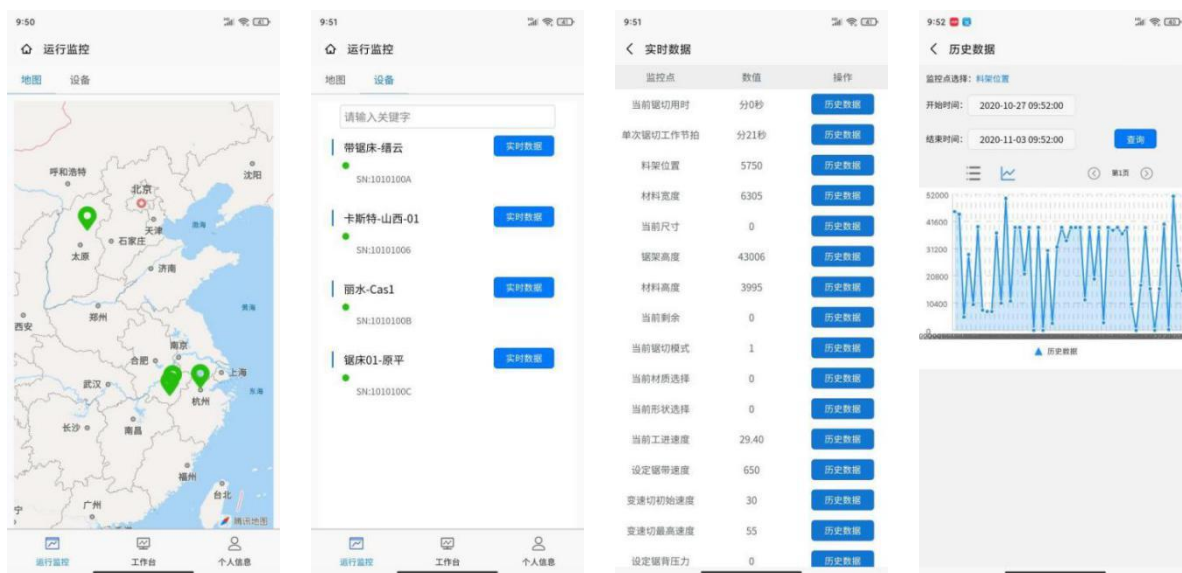


图 2.23 地图总览界面 图 2.24 设备列表界面 图 2.25 数据列表界面 图 2.26 数据详情界面

2.5.4 稳定性测试

通过上述分析设计后完成内测版 APP，经过测试后方可上传到 Google 商城以及小米、华为等各大商城。APP 的稳定性（耗电、CPU、崩溃率）是检验其好坏的因素。

(1) Android monitor 测试

Android monitor 是 Android Studio 自带的测试功能，可以用来监测 APP 运行中 CPU、各部件使用情况、耗电量等。下图是 APP 刚登陆时，查询数据库时的使用情况，这是 APP 内存使用最大的时刻。图 2.27 中底部柱状为 CPU 电量情况，其为 Light，耗电量较小。

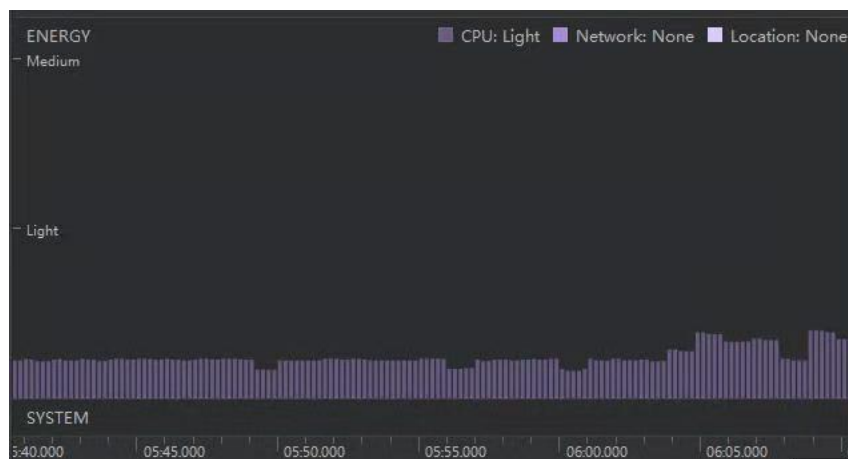


图 2.27 APP 耗电情况图

当 APP 处于前台状态时，查看 APP 的内存使用情况，通过一段时间的测试，APP 未出现卡顿现象，通过图 2.28 得到，从下往上 Java 占 26.3M、Native 底层占 63.6M、界面绘制 Graphics 为 40.7M、Code 为 16.3M、Other 为 8M，其总内存只用了 155.1M。

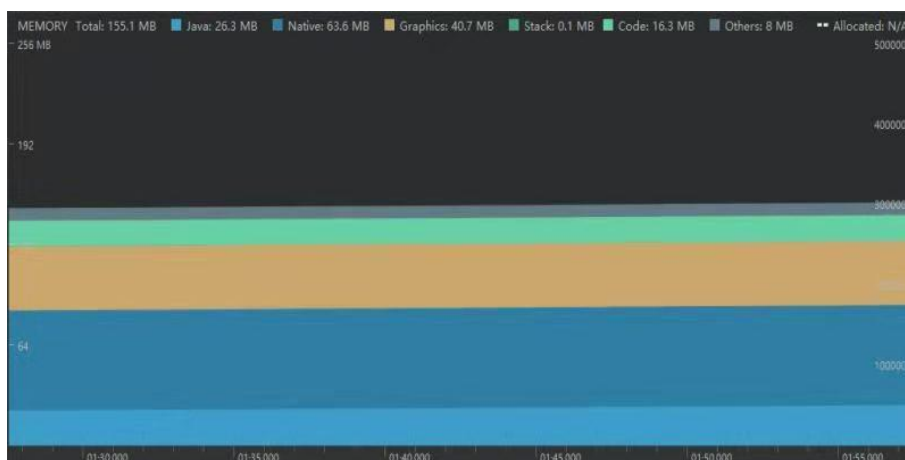


图 2.28 APP 内存使用图

测试结果表明，APP 在进行耗内存、耗电操作时，内存占用、电量消耗不高。

(2) Android Monkey 测试

“猴子测试”是 Android 系统自带的命令行工具^[24]，通过命令行在设备进行奔溃测试，在本文 APP 中测试主要进行了点击抗压，分别设置乱点乱按的测试事件为 1000、2000、5000、10000 进行测试。测试结果如表所示。

表 2.4 Monkey 测试结果表

设置点击次数	异常次数	异常概率
1000	0	0.0000%
2000	1	0.0005%
5000	3	0.0006%
10000	7	0.0007%

测试结果表明 APP 非常稳定，没有发现卡顿、死机、闪退、崩溃等事件，仅出现几次异常 exception，而没有报错 error。异常率并没有随着点击次数的增加有上升趋势，表明 APP 运行良好。

2.6 本章小结

本章主要介绍了金属带锯床远程监控系统的整体设计，从需求分析了解到需要设计一种低时延、高带宽、均负载海量业务的边缘网关来实现对接金属带锯床 PLC 与设备云，以及为将来边缘计算打下基础。之后通过框架选择、功能设计进行云平台设计、Web 客户端设计、APP 设计并且完成各部分的测试工作。通过远程监控系统的建立，保证了前期的数据采集工作，为带锯条剩余寿命预测提供了数据平台。

第3章 基于数据驱动的剩余寿命预测特征选择

设备部件剩余寿命预测一般有失效物理、数据融合和数据驱动三种研究方向^[25]。

失效物理方法：失效物理方法通常适用于已有明确性能退化数据，可以通过观察其材料的形变、氧化腐蚀、微观改变或者不可逆的失效等判断。该方法运用于带锯条剩余寿命预测中，可通过观察其掉齿情况，磨损、氧化情况甚至于显微镜观察微观结构来实现。但是带锯条在使用过程中不方便观察，因此失效物理方法并不适合。

数据融合方法：数据融合方法目前还处于发展阶段，其主要通过集成多种传感器如振动、声音、温度、压力等，并对这些特征进行融合。该方法主要考虑的是环境的多变以及昂贵设备的故障诊断，适用于作战管理等军事领域以及卫星等精密设备。

数据驱动方法：使用数据驱动方法的前提可获取大量的研究数据，随着信息传感技术的发展，该方法已得到了广泛的使用。当数据驱动方法配合机器学习进行寿命预测时，可将大量与寿命相关的因素作为模型的输入，通过机器学习自动判断这些因素与结果的关系，通过与预测寿命的损失函数不断训练得到高精度的模型。该方法适合与样本数量多的设备，而不适用于如卫星这类小样本设备。

在成功搭建工业锯床平台后，通过收集一段时间锯切数据已经具备了数据驱动的条件，根据上述的理论，带锯条不适合使用失效物理（取下观察磨损和掉齿）和数据融合（昂贵设备集合多种传感器）方法进行特征选择，而是通过数据驱动的方法根据数据本身特性来选择模型的特征输入，并进行机器学习来预测其寿命。

3.1 前期数据采集

一台典型的金属带锯床系统结构就如图 3.1 所示，夹紧液压缸进行加工件固定，进给液压缸和进给导向柱^[26]进行锯条下压，主从动轮负责控制锯条速度。

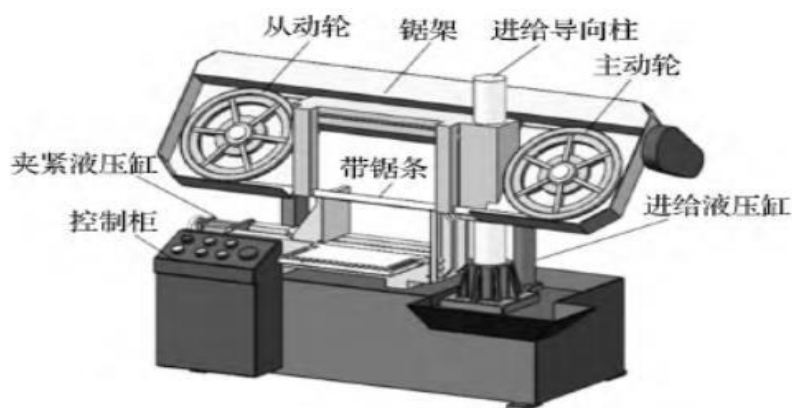


图 3.1 金属带锯床结构图

带锯床的功能是为了锯切工件，而带锯条是锯切的要素，哪些因素影响带锯条的寿命，

哪些因素造成带锯床的故障。该部分工作帮助我们在金属带锯床系统中发挥更好的故障预警以及带锯条剩余寿命预测性能。

通过上章系统的搭建，以及前人的研究，如钱晓健等人通过多年操作带锯床的经验，总结出很有价值的带锯条工作应收集的数据。通过阅读其文献，了解到影响锯切质量和影响带锯条寿命的因素有带锯条的选择、加工材料的选择、切削三要素的选择、带锯条的初期磨合、和其他如切削液、导向臂间距等因素^[27]。

在上章 WEB 端，划分了其故障判断数据，比如油泵故障、水位过低报警等，在进行寿命预测因素采集时有些数据也应该进入故障预警数据库，比如是否进行初期磨合、导向臂间距。同样是文献 27 中所讲述，若不进行初期磨合，锯条寿命将降低 40%左右。并且规定初期磨合时速度应该将为标准的 60%，通过切削屑的形状来判断初期磨合是否完成。通过带锯条的科学使用研究，把一些变量归入异常使用数据库，这些工作严重影响锯条寿命，若不符合既为故障因素。带锯条科学使用流程图见图 3.2，通过该流程其可保证锯条寿命不再受到切削液供给、是否初期磨合等巨大干扰因素的影响，减少预测误差。这样一方面影响带锯条寿命的因素将大大减少，另一方面此流程还在一定程度上延长了锯条寿命。

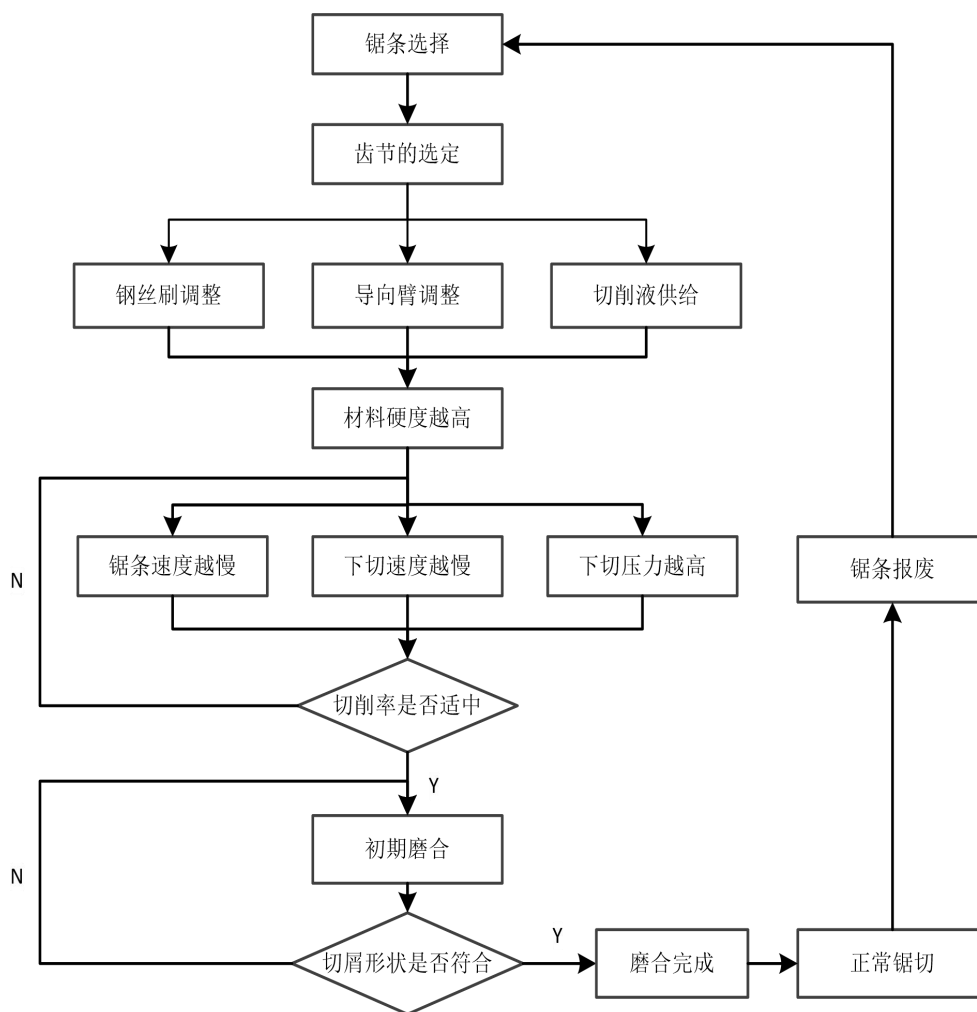


图 3.2 带锯条科学使用流程图

文献 27 所讲带锯条本身型号也为影响因素，由于不同型号材质的带锯条寿命本身的不同，不同带锯条需要进行分类讨论。本系统中大部分金属带锯床 PLC 并没有收集带锯条的材质信息，需要手动输入带锯条的型号。由于系统收集的数据为同一智能带锯床的锯条数据，其使用的是同一种双金属锯条 M42,故本文不将带锯条型号作为影响因素。因此本文主要研究某一型号的金属带锯条的剩余寿命预测，对影响因素进行了一定的限制。

去除了干扰因素后，平台收集如下 PLC 上传的数据，数据点：锯条速度、背压力、切削率、锯切面积、单次工作时间、材料高度、材料宽度、材料材质。最终我们在 Web 端展示了这些数据，见图 3.3。

刀数	材料名称	材料宽度	材料高度	锯切模式	设定锯带速度	变速切初始速度	变速切最高速度	设定锯背压力	锯切用时	锯切面积	切削效率	上报时间
第320刀	45#碳钢	380.2mm	396.2mm	手脉调速	65.0m/min	20.00m/min	55.00m/min	0.00kg	870s	1135cm²	78.33cm²/min	2020-08-29 15:52:49
第319刀	45#碳钢	380.2mm	396.2mm	手脉调速	65.0m/min	20.00m/min	55.00m/min	0.00kg	872s	1135cm²	78.15cm²/min	2020-08-29 15:36:59
第318刀	45#碳钢	380.2mm	396.2mm	手脉调速	65.0m/min	20.00m/min	55.00m/min	0.00kg	876s	1135cm²	77.79cm²/min	2020-08-29 15:20:59
第317刀	45#碳钢	379.7mm	404.2mm	手脉调速	65.0m/min	20.00m/min	55.00m/min	0.00kg	903s	1132cm²	75.27cm²/min	2020-08-29 11:57:10
第316刀	45#碳钢	379.7mm	404.2mm	手脉调速	65.0m/min	20.00m/min	55.00m/min	0.00kg	901s	1132cm²	75.44cm²/min	2020-08-29 11:41:10

图 3.3 Web 端收集的数据

通过上图可知，我们将工件材料、工件形状、锯切模式（速度、时间）、锯切用时、锯切面积、切削率等作为寿命影响因素。但是通过上述研究整理，我们只是记录了每个锯条切削的刀数和每刀的切削面积，我们找到了可作为剩余寿命模型的输入，可是对于有监督学习中的输出剩余寿命 RUL 无法直接得到。

3.2 剩余寿命计算方案

通过 3.1 节的研究，我们确定了影响寿命的因素，找到一些与带锯条剩余寿命相关的因素，而作为输出的带锯条剩余寿命 RUL 的定量变得尤为重要，由于带锯条的剩余寿命无法直接采集，因此我们通过 MTBD 理论找到其合理的失效点。并对失效点前的面积进行累加，通过计算得到其每一刀后的剩余寿命，得到锯条每刀的加工特征和对应的剩余寿命供模型训练。

3.2.1 切削面积计算方案

图 3.2 带锯条使用的流程图中，在识别工件材料后，需要进行切削要素的设置。其中包括锯带速度、背压力、切削率的设置。其中切削率对应材料的设置见表 3.1，该表结合近些年锯条的进步并参考了程廉贵等人多年操作的验证结果^[28]，只有设置合适的切削率才可以进行

安全的切削任务，保证锯条不会因为切削过快而降低寿命。

表 3.1 加工材料与切削率的关系

材料	型号	锯带速度 m/min	切削率 cm^2/min
低碳钢	15	55~80	80~95
中碳钢	45	55~75	75~90
合金结构钢	40CrNi	35~45	35~45
高速钢	W18Cr4V	30~40	25~35
轴承钢	GCr15	40~50	35~45
不锈钢	OCr18Ni9	40~50	25~35

而本系统在最初进行进行切削率分析时，发现其切削率始终无法保持在正常范围，如#45碳钢设置其切削率为 $80\text{cm}^2/\text{min}$ ，可实际上传的某些数据出现了上下波动现象。其切削率分析图见图 3.4。



图 3.4 部分异常切削率图

造成该现象的原因主要有两种:1.切削过程中，进行材料更换。而参数没有更正，导致切削完成时间的改变而导致切削率本身错误；2.切削率本身在范围之内，但切削率计算方式错误，主要因素是面积计算错误。而这两点都与加工材料的形状有关，在收集时比较容易得到的数据为加工材料高和宽，而加工材料的形状信息往往由于工人疏忽，忘记输入。

针对面积计算错误的问题，通过设计一种机制保证其加工材料形状识别的正确和加工面积计算的准确至关重要。通过研究后，该系统主要通过三种方法多重保证，来保证该功能的完成。其流程图见下图 3.5 而第二种情况通过摄像头识别工件形状见图 3.6，每当锯床进行更换材料时，工人需手动设置其形状；若有摄像功能，需用图像识别进行二次确认，并尝试根据压力的变化来三次判断工件形状，其中压力变化对于复杂形状较难判断。

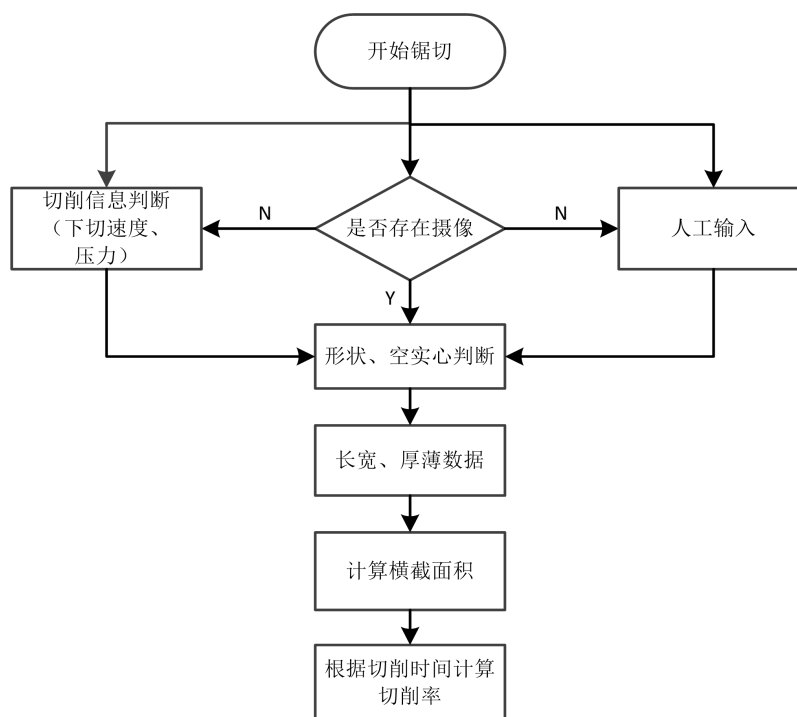


图 3.5 加工材料形状识别流程图

该流程中需要使用摄像图像识别，本系统主要通过 paddlepaddle 平台配合 opencv^{[29][30]} 进行图形形状判断，由于锯床摄像清晰，加工材料形状单一。通过判断其外围轮廓以及收集到的宽、高，结合形状数据库中对中空材料的厚薄数据的选择，进行加工材料面积的计算。



图 3.6 加工件形状监控图

通过以上工作，平台的锯切模式可以随着其自动化切削模式的升级细分成更多，通过不同形状设置不同的锯切模式，保证锯切过程中的切削率在科学范围内，而不是目前只通过面积除以时间，仅保证其一刀之后的切削率保证在科学范围。通过上述机制，平台的切削率已经足够对已报废带锯条历史切削率进行分析以找到合理更换点的判断。

该切削率识别更换点还存在一点问题，如果工人频繁更换加工材料，而其不只形状改变而是材料本身材质改变，则该带锯条的历史寿命将失去研究意义。通过数据和实地考察，该

情况几乎不发生，主要是工厂加工最多的为 45#碳钢。且一根带锯条大约加工 10m^2 碳钢。而加工材料部分如下图 3.7 所示，每个材料都需要切割上百刀，且不同材料往往在不同生产线进行加工，不存在频繁更换现象。



图 3.7 加工材料切削图

通过图像识别其为圆形、T 字形、中空等形状，结合长宽数据以及数据库形状切面厚度数据，得到其正确的切削面积以及切削率，保证其切削率在表 3.1 所示的范围内，减少上下波动跳跃现象的发生。通过对受到的切削率分析判断，使用智能网关往 PLC 更新其锯切模式代码，定期改善其加工切削参数，切削率得到明显改善见下图 3.8。由图可见其切削率趋于稳定，切削率计算得到改善，供失效点判断。



图 3.8 面积改善后的切削率图

经过图像识别加工件形状，并通过宽高得到大小后，使其对应材质的加工件切削率始终保持在科学范围内，可延长带锯条寿命。并且保证切削率稳定，保证当锯条磨损时，可通过切削率来判断其状态。

3.2.2 失效点判断方案

(1) MTBF 理论分析

MTBF 指产品的平均无故障时间，用来表示一个产品的可靠时间分布^[31]。一个产品从产生到永久报废，其过程中会出现图中所示的失效过程。首先是产品的早期失效期，产品由于设计、制造工艺等问题，产品在早期出现失效情况。如今很多电子产品在制造成功后，往往会进行较长时间的测试，通过对加工环境的改善、过程的修改来减少产品发布后早期失效的概率。MTBF 指的是正常使用的时间除以故障次数，下图 3.9 中正常使用时间分别为 T1、T2、T3，故障次数为 3，其 $MTBF = (T1 + T2 + T3) / 3$ 。

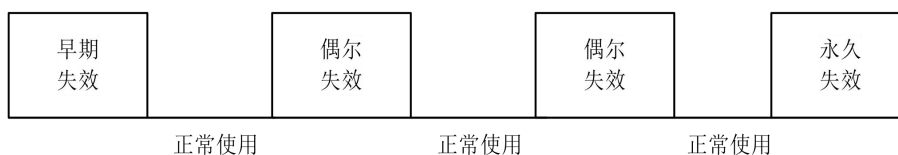


图 3.9 产品失效周期图

MTBF 不是作用在一个物品上，其通过对同一产品大数量产品的平均无故障时间进行统计计算。同一产品的无故障时间通常为正态分布，通过概率密度方程(3.1)表示 MTBF 为：

$$MTBF = \int_0^{\infty} tf(t)dt \quad (3.1)$$

MTBF 可以计算出产品的平均无故障时间，但是无故障时间符合正态分布的特点，而 MTBF 为正态分布中点。那么有近 50%的产品的故障时间小于 MTBF，如果根据 MTBF 去维修产品是不科学的。MTBF 通常用来计算产品的失效率，随着时间的增长，产品的失效率会增加，在 MTBF 时间未达到之前通过对失效率进行控制可提高维修效率。MTBF 可以用于产品的维修，但不适合用于消耗型产品的寿命分析。

(2) MTBD 理论分析

MTBD 是一个新概念，指的是器件衰退发生前的平均时间。其公式和 MTBF 相似，只是 MTBD 针对的是器件从使用到报废之间的过程，相比较 MTBF 理论其不存在进行维修的情况，也不考虑期间存在故障等异常现象。该理论讲述了器件存在稳定工作的区间，当器件过了稳定工作区间，器件将进入衰退期。该理论应用到带锯条，并通过收集到的数据分析，将带锯条分为“稳态”，“暂态”和“报废”三种健康状态^[32]。由于暂态维持时间短，期间锯条工作质量低，断裂概率高，应在暂态之前进行更换。

结合 Web 平台改善后采集到的带锯条切削率变化图见图 3.10，如图可知：其切削率在平稳期间保持不变或一个稳定下降的趋势。但是进入报废期之前的暂态，在同样的工作条件下，其切削率会急速下降。在这个期间带锯条的加工效率下降严重，导致工时的提高，加工材料切割质量降低，并可能使用切面不平滑，锯偏值变大甚至锯条断裂。

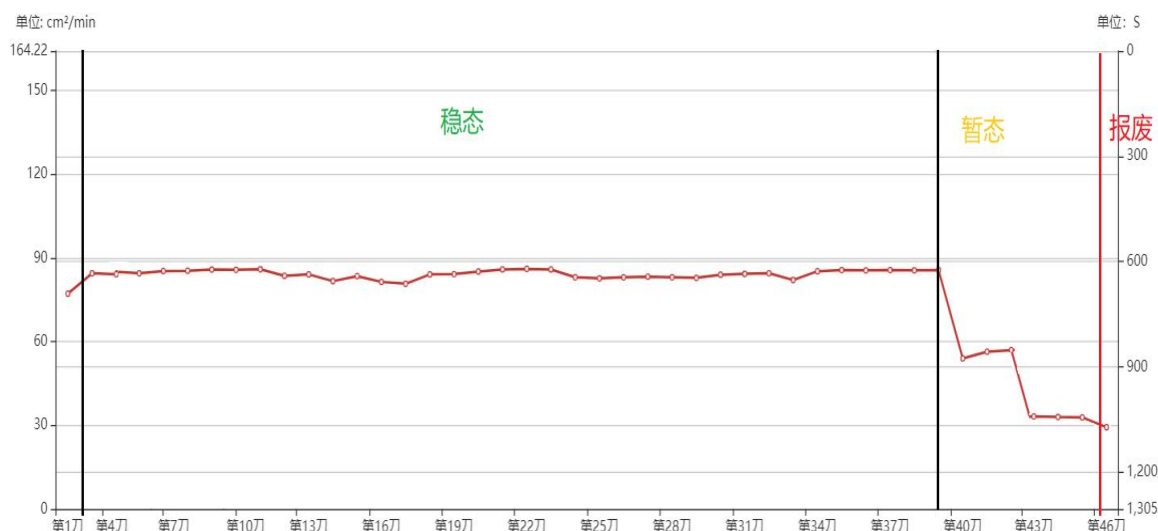


图 3.10 带锯条切削率变化图

通过上述分析，我们需要将 RUL 的终点设置在稳态的终点。一旦带锯条进入暂态，其效率与安全性将急剧下降，因此在未进入暂态的稳态末端进行带锯条的更换可以保证其工作效率、工件加工质量和安全性。

3.2.3 剩余寿命定量计算

根据上文所述锯条衰退失效前平均时间，在这段时间内，锯条的寿命是呈比较缓慢的线性下降，但是到了失效点后，锯条的性能会快速下降，甚至造成断裂等危险的事故。这种处理方式符合健康管理原则，在设备还没发生故障前，就提醒用户维护和更换，而不是等到事情已经严重，已经出现故障造成不可逆的损失后才提醒用户，而剩余寿命预测为也是同理。

而剩余寿命预测方法通常有相似度量方法和性能因素外推方法两种，而基于这两种方法，又存在数据驱动、物理驱动、集成/混合驱动等方法。本文因为建立了金属带锯床远程监控系统，并通过了近半年的运行，采集到的带锯条全寿命信息已经足够开展基于数据驱动的定量分析。之所以不进行相似系数度量方法是因为带锯条的寿命受加工材料不同等因素的影响较大，不易进行相似对比。

本文基于上述理论，又结合实际平台获取的数据，做出了以下分析工作：1.由于带锯条的剩余寿命在未报废前无法测量，所以我们对已报废带锯条的数据进行分析；2.以切削剩余时间还是切削剩余面积作为其剩余寿命的评判标准是一个问题，通过查阅文献如贾寓真等人2019 年对锯切参数与剩余寿命的研究^[33]，和市场调查后，了解到带锯条的价值在于加工了多少面积的材料，市场上以带锯条可加工约 10 m²碳钢来描述其寿命。

确定以切削累计面积作为带锯条寿命后，结合图 3.3 中平台收集的数据，从图中我们还可以了解到一个信息：刀数和面积。带锯条每一次下刀的面积都通过形状以及长宽计算得到并记录，直到其报废更换。这样通过面积加权得到其每个时间段的切削总面积 S_i 。因此根据上文所讲的 MTBD 理论结合经过改善后的切削率变化找到带锯条合理更换点。如果经过判断

锯条为提前更换，则不进行计算，选择换一组数据分析。

通过分析平台收集的数据，对该带锯条的锯切面积进行累加获得总的切削面积 S ，这样该锯条可以通过当前时间点累计切削的面积除以总面积处理得到每一时间点的剩余寿命归一化的 $RUL(i)$ ，其公式如下。

$$RUL(i) = \frac{S_i}{\sum_{i=1}^n S_i} \quad (3.2)$$

在进行剩余寿命定量计算过程中除了图 3.10 中切削率来判断带锯条合理更换，由于此工作致为重要、还需结合故障报警中的锯偏值等数据来提高判断准确性。经过探索，建立计算流程为：当切削率连续三刀低于正常值或者切削率低于正常值且锯偏值提高时判断其进入暂态，找到带锯条合理更换点并将其标为剩余寿命 0 点。其定量计算流程可见下图 3.11，选择锯条信息时主要判断是否其是否提前更换等异常，若是正常报废更换则进行下一步是否超过稳态判断，再通过上述方法找到带锯条合理更换点。最后通过公式（3.2）进行计算。本文通过对系统收集带锯条的全寿命信息进行定量计算，在去除异常锯条数据组后，共得到 200 组带锯条剩余寿命信息。

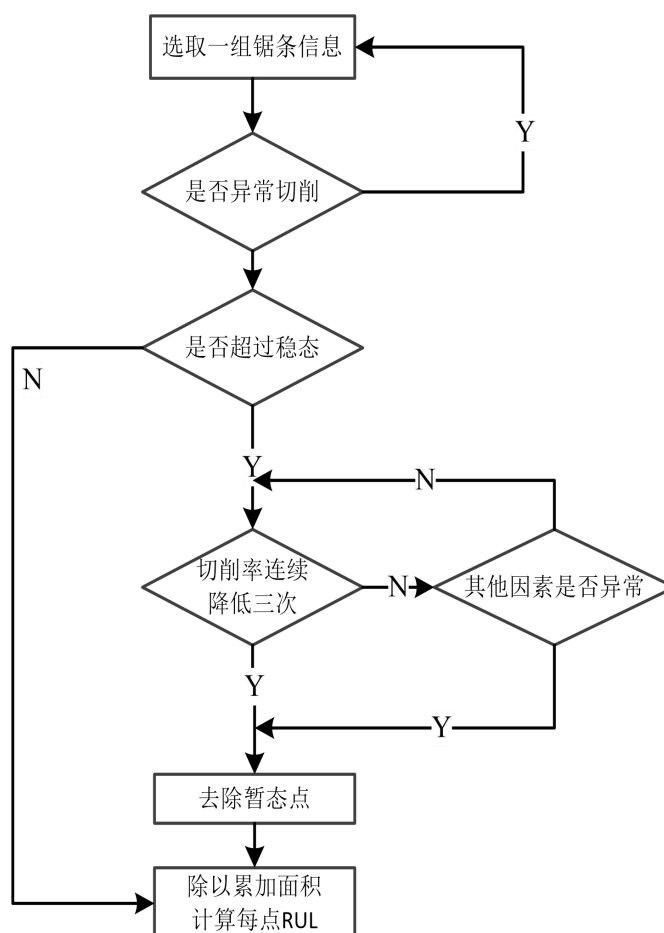


图 3.11 带锯条剩余寿命计算流程图

3.3 数据预处理

预处理工作决定了模型性能的上限，而不同模型的选择只是去逼近这个上限^[34]，通常预处理工作有缺值补充、数据异常去除、数据归一化等。本研究通过获取近 3 个月已报废的共 200 根金属带锯条，每个锯条切削刀数为 150-200 之间，共 34274 个数据。对每条锯条进行单独的数据处理，最终整合并抽取共 2000 个数据供模型训练。图 3.12 和图 3.13 展示了两锯条的部分原始数据。

材料名称	材料宽度	材料高度	锯切模式	锯带速度n	初始速度n	最高速度n	背压力kg	锯切用时	锯切面积	切削效率c	上报时间
A3碳钢	370.8	400.0	手脉调速	70.0	20.00	55.00	0.00	839	1080	77.24	2020-09-09 20:19:54
A3碳钢	370.8	400.0	手脉调速	70.0	20.00	55.00	0.00	829	1079	78.16	2020-09-09 20:04:34
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	845	1139	80.89	2020-09-09 18:37:44
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	840	1139	81.37	2020-09-09 18:19:24
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	842	1139	81.18	2020-09-09 18:04:34
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	857	1139	79.76	2020-09-09 17:49:44
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	831	1139	82.26	2020-09-09 17:34:34
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	842	1139	81.18	2020-09-09 17:19:54
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	845	1139	80.89	2020-09-09 17:05:04
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	864	1139	79.11	2020-09-09 16:50:04
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	849	1139	80.51	2020-09-09 16:33:24
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	840	1139	81.37	2020-09-09 16:18:24
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	841	1139	81.28	2020-09-09 16:03:34
A3碳钢	380.8	407.4	手脉调速	70.0	20.00	55.00	0.00	845	1139	80.89	2020-09-09 15:48:44
A3碳钢	379.1	407.0	手脉调速	70.0	20.00	55.00	0.00	844	1128	80.25	2020-09-09 14:39:58
A3碳钢	374.5	400.3	手脉调速	70.0	20.00	55.00	0.00	828	1101	79.82	2020-09-09 11:32:58
A3碳钢	374.5	400.3	手脉调速	70.0	20.00	55.00	0.00	829	1101	79.72	2020-09-09 11:17:48
A3碳钢	374.5	400.3	手脉调速	70.0	20.00	55.00	0.00	836	1101	79.05	2020-09-09 11:02:38
A3碳钢	374.5	400.3	手脉调速	70.0	20.00	55.00	0.00	842	1101	78.49	2020-09-09 10:47:18

图 3.12 带锯条 A3 碳钢部分数据图

材料名称	材料宽度	材料高度	锯切模式	锯带速度n	初始速度n	最高速度n	锯背压力k	锯切用时	锯切面积	切削效率c	上报时间
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	892	1140	76.71	2020-09-16 07:37:37
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	891	1140	76.8	2020-09-16 07:21:27
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	903	1140	75.78	2020-09-16 07:05:47
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	908	1140	75.36	2020-09-16 06:49:47
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	916	1140	74.7	2020-09-16 06:33:47
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	927	1140	73.82	2020-09-16 06:17:37
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	922	1140	74.22	2020-09-16 06:01:17
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	921	1140	74.3	2020-09-16 05:44:57
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	918	1140	74.54	2020-09-16 05:28:37
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	919	1140	74.46	2020-09-16 05:12:27
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	920	1140	74.38	2020-09-16 04:56:17
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	890	1144	77.15	2020-09-16 04:21:07
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	889	1144	77.24	2020-09-16 04:05:27
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	903	1144	76.04	2020-09-16 03:49:47
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	913	1144	75.21	2020-09-16 03:33:47
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	920	1144	74.64	2020-09-16 03:17:37
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	915	1144	75.04	2020-09-16 03:01:27
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	921	1144	74.55	2020-09-16 02:45:17
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	922	1144	74.47	2020-09-16 02:28:37
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	918	1144	74.8	2020-09-16 02:12:17
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	922	1144	74.47	2020-09-16 01:56:07
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	924	1144	74.31	2020-09-16 01:39:47
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	921	1144	74.55	2020-09-16 01:23:37

图 3.13 带锯条 45 碳钢部分数据图

由于智能锯床处于测试阶段，通过分析收集到的数据，显示其切削材料主要为碳钢材料，形状主要为圆形等实心形状。在进行数据处理过程中，为了方便机器学习 python 程序的编码的方便以及图标显示对英文的要求，各数据点用 A0、A1、A2.....标注。剩余寿命 RUL 无法直接获取，需通过上节方案进行计算。训练集中带锯条的剩余寿命 RUL 用于训练模型，而测试集中带锯条的剩余寿命 RUL 作为验证结果来检验模型的准确性。

3.3.1 缺值补充

预处理中缺值补充的方法通常为对缺失值数据的直接删除，或是进行数据补充。由于本研究单一数据对整体数据具有较大关联。因此若发现有缺值，采用数据补充的方法，不进行整条数据的删除。

并且由于剩余寿命 RUL 非直接得到，缺值补充工作主要体现在对 RUL 进行数据补充。通过 3.2.2 节中的通过对带锯条合理更换点的判断，以及寿命定量计算完成该工作。并且将其余寿命相关因素标上 A0、A1.....A10，我们得到每个带锯条完整的数据集，数据集部分数据如下图 3.14。

A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	rul
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	918	1140	74.54	0.7706
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	919	1140	74.46	0.7646
45#碳钢	381.0	416.0	手脉调速	65.0	20.00	55.00	0.00	920	1140	74.38	0.7586
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	890	1144	77.15	0.7526
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	889	1144	77.24	0.7466
45#碳钢	381.7	416.4	手脉调速	65.0	20.00	55.00	0.00	903	1144	76.04	0.7405

图 3.14 剩余寿命 RUL 数据补充

3.3.2 异值处理

预处理中异值处理通常采用直接删除、均值处理等，本研究的数据主要处理 PLC 数据传输中数据包受干扰产生的计算错误数据，数据异常体现在有些数据受到噪音干扰，数据偏移正常值数倍；以及剩余寿命计算中，失效点以后的数据。通过对单个锯条各数据进行可视化分析，找到其干扰数据进行均值处理，见下图 3.15 所示；找到其失效点后数据进行直接删除，见下图 3.16 所示。

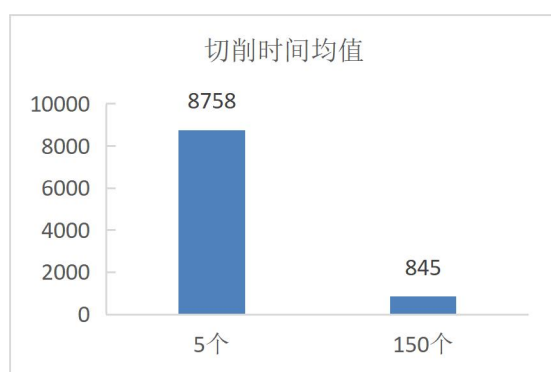


图 3.15 切削时间异值检测

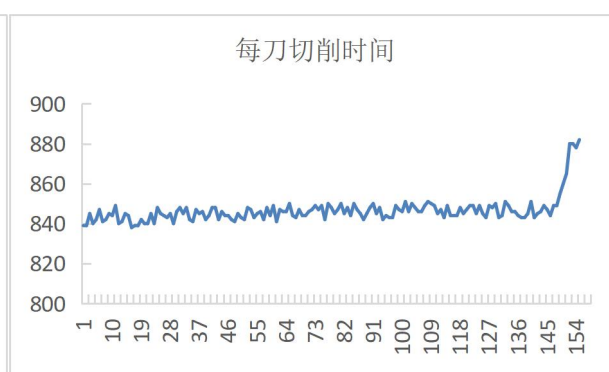


图 3.16 失效点后数据处理

材料的宽度和高度，其可由切削面积代表，进行变量去除。当其经过直方图分析后，背压力也处于异常状态，显示为 0，进行异值去除。

3.3.3 归一化处理

完成对剩余寿命 RUL 进行缺值补充以及其他异值处理后，对数据进行归一化处理。RUL 本身已经为归一化状态不进行处理，其余数据归一化处理与常见归一化公式（当前值/（最大值-最小值））不同，主要考虑到实际工作中新锯条数据加入后的最大最小值改变问题。

本文归一化方案为：材料进行 0.1-1.0 标记，间隔为 0.1；锯切模式按照其共有几种模式进行计算；锯带速度，变速切初速度都不大于 100，由整除 100 得到；切削率除以 100 得到；当前切削时间，由对每刀时间累加，并除以 200000 得到。该方案理论上对相关性的精度有一定的影响，但相关性精度对模型训练不造成影响，方便了工业实践。

通过以上归一化计算和上小节异值处理后，对剩余数据进行重排序。下图 3.17 是通过预处理后部分数据：

A0	A1	A2	A3	A4	A5	A6	rul
0.2000	0.5000	0.6500	0.2000	0.5500	0.3000	0.7454	0.7706
0.2000	0.5000	0.6500	0.2000	0.5500	0.3046	0.7446	0.7646
0.2000	0.5000	0.6500	0.2000	0.5500	0.3091	0.7438	0.7586
0.2000	0.5000	0.6500	0.2000	0.5500	0.3136	0.7715	0.7526
0.2000	0.5000	0.6500	0.2000	0.5500	0.3181	0.7724	0.7466
0.2000	0.5000	0.6500	0.2000	0.5500	0.3227	0.7604	0.7406

图 3.17 数据归一化处理结果

其中 A0 到 A6 分别对应了材料名称、锯切模式、锯带速度、变速切初速度、最高速度、切削总时间、切削效率，以及剩余寿命 RUL。

3.4 实验测试

通过完成对收集数据的预处理，200 条金属带锯条共计获取 34274 个数据，考虑到工程实践中无需对每一刀进行预测，以及训练单一锯条过多数据容易导致过拟合问题。故选取 160 锯条整合并随机抽取 1600 个数据，作为训练集。选取 40 条锯条整合并随机抽取 400 个数据，作为测试集，共 2000 条数据，对剩余寿命[0,1]之间进行模型训练。本章进行实验的总体流程图如下图 3.18 所示。

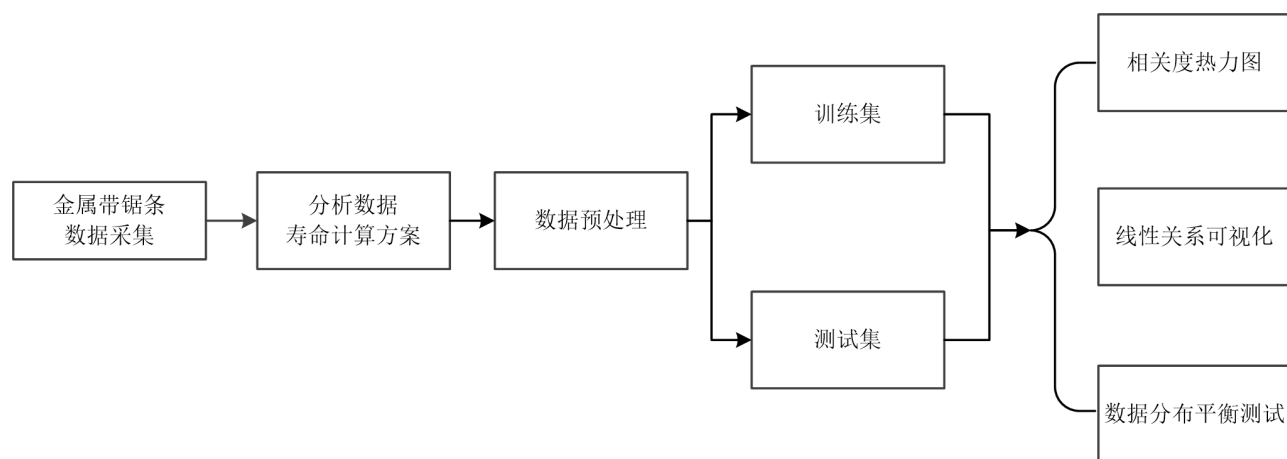


图 3.18 特征选择总体流程图

相对于抽象的数字，可视化分析将这些数字映射为图表，图表分析更加容易理解和直观。下面我们对训练集和测试集的数据进行相关测试，首先进行训练集数据之间相关性测试^[35]，通过 Python3.8 结合 numpy 库以及 matplotlib 库^[36]来进行图表分析。

3.4.1 相关度热力图

下图 3.19 是表示数据点间相关度的热力图

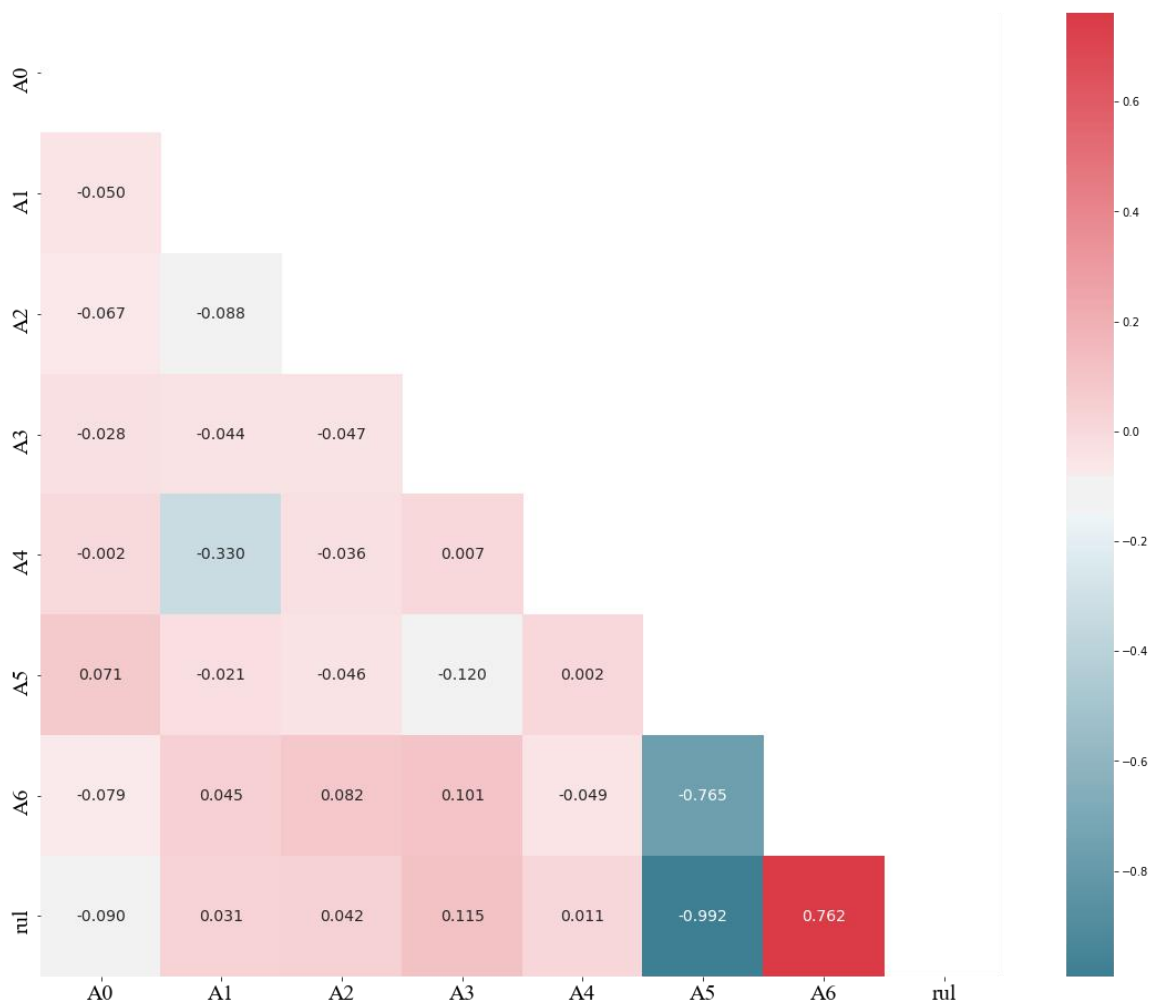


图 3.19 数据相关度热力图

其中关键的和 RUL 相关的部分数据如表 3.2 RUL 相关度热力表所示，由图表可见，与带锯条剩余寿命相关度较高的为 A5(切削总时间)、A6(切削率)。相关度低的为 A0(工件材料)、A1(锯切模式)、A2(锯带速度)、A3(变速切初速度)、A4(最大速度)。设置去除阈值为 0.1，去除相关性低于 0.1 的 A0、A1、A2、A4，对 A3(变速切初速度)进行下一步分析。

表 3.2 RUL 相关度热力表

数据点	工件材料	锯切模式	锯带速度	变速切初速度	最大速度	切削总时间	切削率
RUL	-0.090	0.031	0.042	0.115	0.011	-0.992	0.762
切削率	-0.079	0.045	0.082	0.101	-0.049	-0.765	
切削总时间	0.071	-0.021	-0.046	0.120	0.002		
最大速度	-0.002	-0.330	-0.036	0.007			
...	...	...					

3.4.2 线性关系可视化测试

通过相关性热力图，我们排除了一些数据，但是热力图只是单纯的数字，还不够直观的表达出各数据和 RUL 之间的关系，所以继续进行可视化分析。右图中粗线为相关度均数，左图点、右图柱为数据集数据，右图纵坐标为数量，但由于数量差异过大，刻度只为显示比例。

(1) 锯切初速度

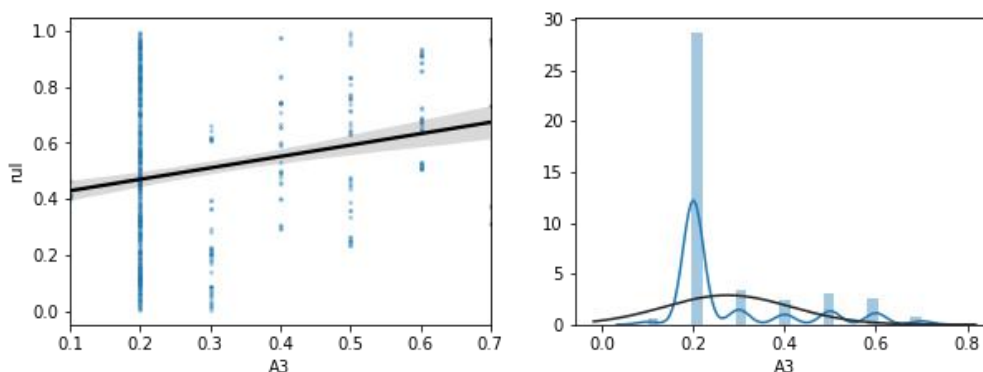


图 3.20 锯切初速度与带锯条剩余寿命的关系

锯切初速度较高代表整个锯切过程中整体速度较高。如图 3.20 所示，锯切初速度也出现了集中在 0.2 附近，代表锯切 20m/min。锯切初速度与带锯条剩余寿命之间的相关度比较低为 0.115 左右，并且根据图表进行猜想：带锯条初速度较高，能使锯切面积增加，也就相当于提高了带锯条的剩余寿命。但是初速度过快，容易导致切削液供给不充分从而提高断裂，缺齿等风险。因此锯切初速度是否作为影响带锯条剩余寿命的因素还需要进一步分析。

(2) 锯切累计时间

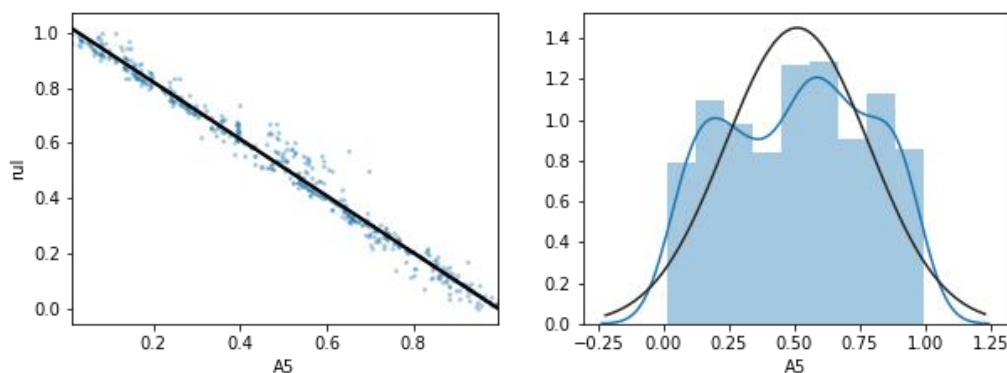


图 3.21 锯切累计时间与带锯条剩余寿命的关系

从某种意义上讲，带锯条剩余寿命可以用锯切时间来表示，但是带锯条的剩余寿命与加工的材料，切削效率等都有一定关系，所以本文用可锯切的面积来衡量一个带锯条的剩余寿命。因此从图 3.21 可见，锯切累计时间和带锯条剩余寿命之间的相关度较强，甚至热力图中出现了绝对值为 0.992 的相关性，且锯切累计时间分布比较均匀。但从左图中也可以看到，某些数据和 RUL 之间并不是非常的拟合，而是分散在线周围结合前文研究可以推测，锯切累计时间和剩余寿命之间还受到锯切效率、锯条质量不同等因素影响，并不能由锯切累计时间直接推断 RUL。综上锯切累计时间与带锯条剩余寿命的相关性较强，可作为模型训练的数据。

(3) 切削率

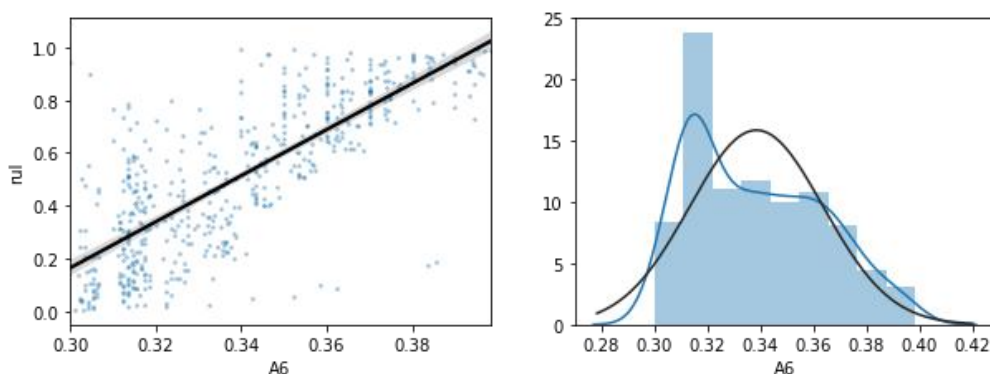


图 3.22 切削率与带锯条剩余寿命的关系

在 3.1.2 中讲到，切削率可以作为判断带锯条剩余寿命的一个条件，从图 3.22 中也可以看到，切削率与带锯条剩余寿命呈正相关性，热力图中显示两者相关性绝对值为 0.762。再从图表右图中看到切削率在 0.32 时分布最多，在 0.36~0.40 呈下降趋势，而 0.30 也比较少。猜测此现象与加工材料有关，不同材料的切削率不同，但随带锯条剩余寿命下降，切削率下降。切削率与带锯条剩余寿命的相关性较高，可作为模型训练的数据之一。

3.4.3 数据分布平衡测试

完成数据间的关系试验后，进行训练集和测试集的数据样本分布平衡测试分析，若训练集和测试集分布状态相差很大，会有一些干扰结果的数据点导致训练模型后得到一个较差的结果，因为通过模型训练的是训练集的数据，若测试集与训练集相差较大，会导致结果误差偏大。因此需要训练集和测试集的数据样本分布相对平衡，如果有数据点分布不平衡，也可以推断出该数据点分布太随机，没有一定的逻辑可能为干扰数据，需要进行进一步分析。本文基于 Python 语言以及 Spyder 平台和机器学习库对其进行训练集与测试集的分布平衡测试，运行平台环境见下图 3.23。

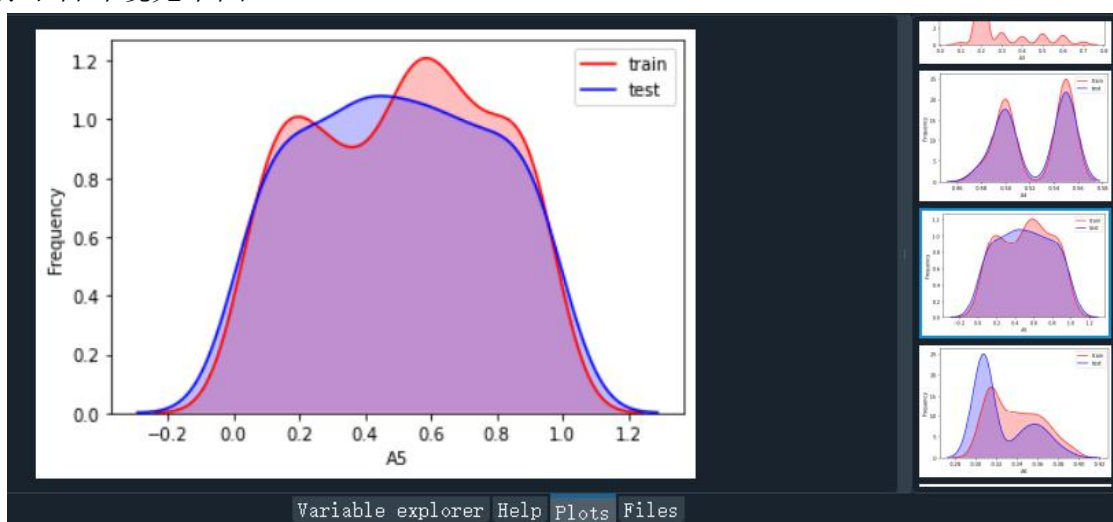


图 3.23 运行环境图

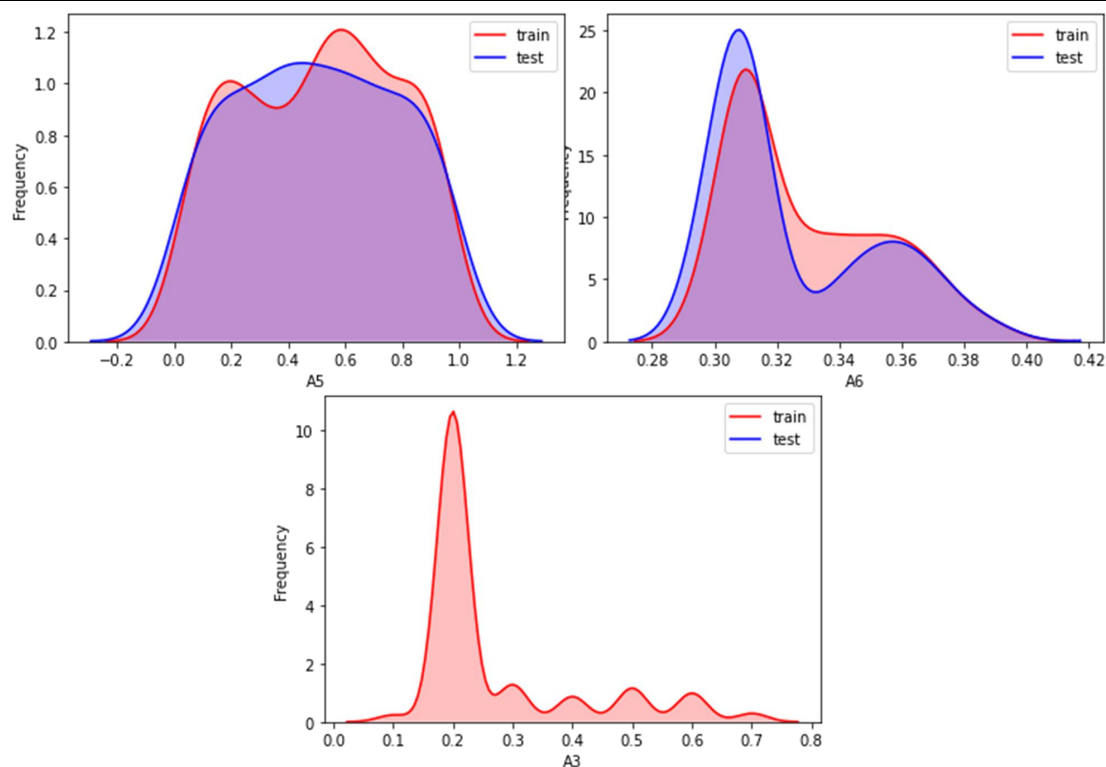


图 3.24 训练集与测试集数据分布图

通过相关性去除一些数据后，训练集和测试集中 A3(变速切初速度)、A5(切削总时间)、A6(切削率)数据分布情况如上图 3.24 所示，除了变速切初速度训练集和测试集相差较多，其他两个相关性高的数据点的分布状况都比较平衡，对预测结果没有过大干扰。变速切初速度测试集数据在图中没有显示，分析原因是因为数据比较集中单一，导致密度过大造成图像绘制失败，也说明了其数据样本分布比较不均。并且其与剩余寿命 RUL 之间的相关性仅为 0.115，为弱相关性，若选择其作为预测模型的输入数据，会造成误差增大，训练模型欠拟合等问题，因此舍去锯切初速度。

3.5 本章小结

本章主要完成数据采集、金属带锯条剩余寿命计算、数据预处理以及模型输入特征的选择。首先通过调研前人的研究结合 PLC 数据可采集实情，对影响带锯条寿命的数据进行采集。然后对采集的数据分析，通过图像识别、手动输入、切削数据的分析判断加工材料的形状，计算得到正确的切削面积。通过分析改善后的数据如切削率的变化以及锯偏值等找到带锯条合理失效点，从而对剩余寿命进行定量计算。接着在此基础上对数据集进行缺值补充、异值处理、归一化等预处理。最后对总体数据进行采样，得到 2000 条数据以 4: 1 划分训练集和测试集，完成相关度等实验测试，确定以当前锯切累计时间和切削率作为输入特征。

第 4 章 基于机器学习的带锯条剩余寿命预测模型研究

第三章介绍了带锯条剩余寿命定量计算，数据的预处理以及特征的选择。本章根据数据的特点进行模型训练，在进行模型构建之前，根据其输出值的特点进行机器学习算法的选择。

选择思路如下：

(1)若给定的数据没有输出值。可使用到机器学习中的无监督学习算法。无监督学习算法一般使用聚类以及相关分析，将关系相近的数据进行聚类得到想要的答案。

(2)若给定的数据有输出值，输出标记是离散的。进行离散数据的预测，可用到机器学习中监督学习的分类算法，常见如：支持向量机、决策树、朴素贝叶斯^[37]。

(3)若给定的数据有输出值，输出标记是连续的。可用到监督学习中的回归算法，如 K 最近邻、线性回归、Gradient Boosting 等。

参考上述思路，选择有监督学习的回归算法。选择模型后，需要对各模型参数进行设置，并且使用模型性能评价因素对模型进行评价，从而得到适合带锯条剩余寿命预测最优的训练模型，并且尝试对该训练模型进行算法优化。

4.1 总体方案

预测模型建立方案首先通过各种有监督学习的回归方法对数据集进行训练，数据集分为训练集和测试集，大小为 4: 1。各模型在对训练集训练时，采用网格搜索算法和五折交叉验证来得到各模型最优的参数。然后使用训练得到的最优参数进行模型训练，并对测试集进行预测。计算预测结果与实际值的 R^2 、RMSE，通过这些评估指标来选择合适的模型，并对其进行进一步优化研究。总体方案流程如图 4.1 所示。

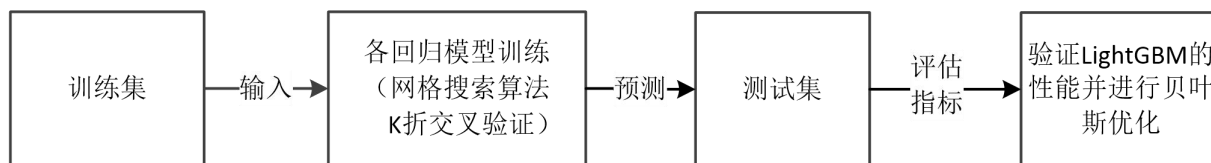


图 4.1 总体方案流程图

如上所述，本章预测模型研究包括各模型建立训练对比、各模型预测实验、以及合适模型的改进。在各回归模型训练过程中，线性回归模型如线性回归、岭回归、Lasso 回归等得分较低且种类较多，本文不加与讨论。本文主要对训练效果较好的 KNN 算法、随机森林算法、LightGBM 算法进行讨论研究。

4.2 基于 LightGBM 算法的预测模型

LightGBM 算法是一种专门用于工业实践的算法模型，他在 GBDT(梯度提升决策树)的基础上进行了算法优化，拥有更快的训练速度、更低的内存消耗和更好的准确度。GBDT 和随

机森林在做回归模型时都是基于 CART(分类和回归树) 弱学习器训练, GBDT 使用提升法, 对损失函数的负梯度进行训练, 随机森林使用袋装法进行训练, 其结构图如图 4.2 所示。

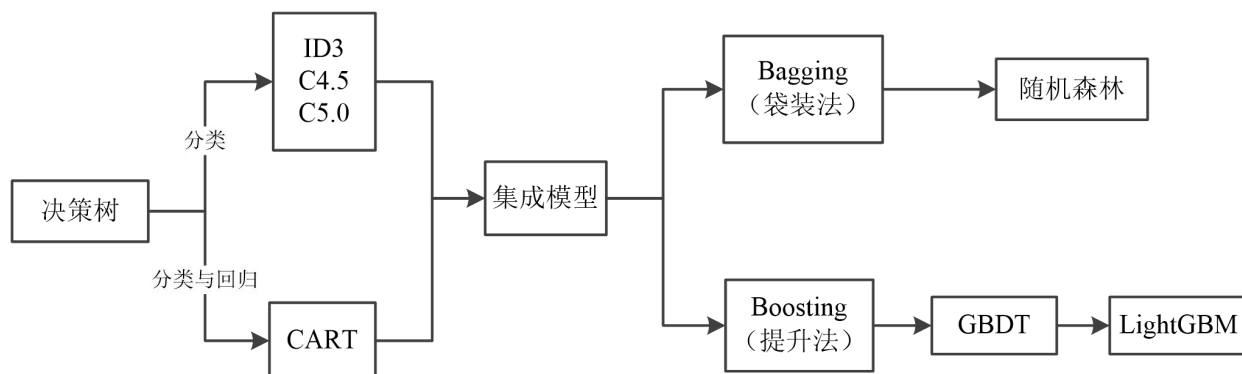


图 4.2 决策树算法结构

在建立 LightGBM 模型前, 先通过了解其算法原理: CART、GBDT 以及 LightGBM 的改进来帮助模型建立时完成参数的选择等工作。

4.2.1 分类和回归树

LightGBM 的底层弱学习器为 CART, CART 是决策树算法的一种。

决策树算法产生于 1966 年, 由 Hunt 等人提出, 其基础为 ID3 算法, C4.5 算法和 C5.0 算法都是对 ID3 算法的改进, 其适用于分类模型。而 1984 年 Leo Breiman 等人提出 CART(分类与回归树) 是通过决策树实现回归的算法^{[38][39]}。决策树分为三类节点: 根节点、叶子节点和内部节点^[40], 他从树的根节点出发不断建树一步步地将整个特征空间进行划分来逐步构造新的节点, 直到其节点数据无法再划分。决策树算法的关键是如何进行树节点的分裂, ID3 等分类树可以有多个叶节点, 而 CART 每个根节点只能有两个叶节点, 其算法模型见下图 4.4。

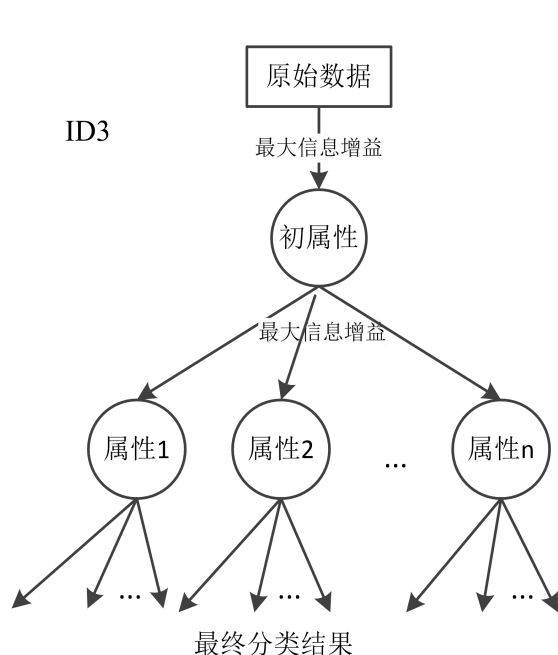


图 4.3 ID3 算法模型

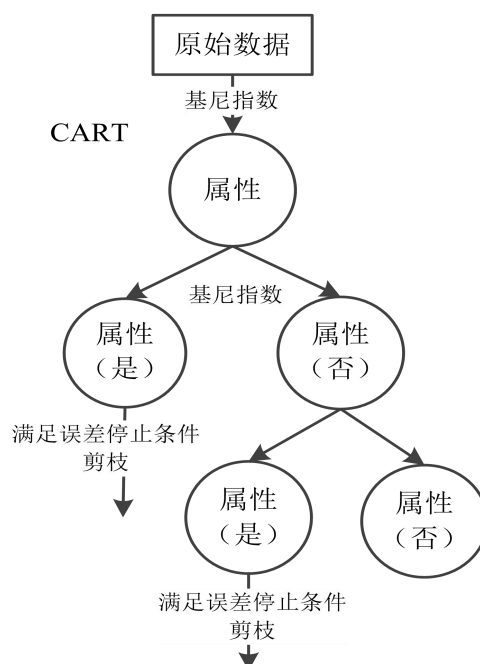


图 4.4 CART 算法模型

决策树算法中，最关键的节点划分算法公式如下：

(1)ID3 算法：熵在信息论和概率统计中是用来表示不确定性的度量，在 ID3 中，决策树基于香农熵理论结合贪心算法计算信息量来划分建树。香农熵的公式为：

$$H(X) = -\sum_{i=1}^n p(x_i) \ln p(x_i) \quad (4.1)$$

X 表示变量， $p(x)$ 表示每个变量对应的概率。

其中变量的不确定越大，熵就越大。当 X 均匀分布时，每个变量概率相等，香农熵最大。算法通过寻找最大的信息增益量来划分建树，信息增益公式为：

$$Gain(A) = H(X) - \sum_{i=1}^v \left(\frac{|x_i|}{|X|} \times H(x_i) \right) \quad (4.2)$$

$H(X)$ 表示数据集整体的信息量， v 表示特征 A 的数量。

其存在的缺点：当 A 特征的数量 v 最多时，其信息增益量一般越大。因此如 100 个值中有 99 个满足某个属性，ID3 因为其算法特性，往往会直接选择分布数量较多的属性，而这个属性很可能不是最优的。

(2)C4.5 算法：C4.5 算法解决了 ID3 的这个问题，他使用的是信息增益率作为分支的标准，信息增益率的公式为：

$$GainRatio(A) = \frac{Gain(A)}{-\sum_{i=1}^v \left(\frac{|x_i|}{|X|} \times \log \frac{|x_i|}{|X|} \right)} \quad (4.3)$$

由公式可知，当 v 较大时，分母增大导致整体减小，有效解决了 ID3 选择分布数量较多的属性进行分支的问题。

(3)CART 算法：上述两种算法更多用于分类问题解决，特别是 ID3，而 CART 算法为分类回归树算法，其更适合于回归问题的解决。

CART 算法在分支处理中分支属性的度量标准是基尼系数，设 X 为数量为 n 的训练集，第 i 个类别的概率其基尼系数公式为：

$$Gini(X) = 1 - \sum_{i=1}^n p(x_i)^2 \quad (4.4)$$

CART 算法构建的树为二叉树，其分支类于是和否的选择，其分支前，对所有的属性进行基尼系数计算，然后选择基尼系数最小的属性进行是该属性的左分支和不是该属性的右分支进行建树，直到达到树最大深度或者其他终止条件。

4.2.2 梯度提升模型

GBDT 全称梯度提升决策树，它在处理回归问题时是通过集成处理集成学习算法中的提升法（Boosting）将 CART 弱学习器进行串行训练，通过每次训练得到的损失函数往负梯度方向进行迭代，最终得到足够的学习率终止训练。GBDT 是一个加法模型，在进行回归任务

时,通过累加所有弱学习器的预测值来得到最终预测值^[41]。之所以用到负梯度计算,是为了防止过拟化,残差不一定到达 0 停止,而是通过负梯度不断靠近目标值,其模型流程见下图 4.5。

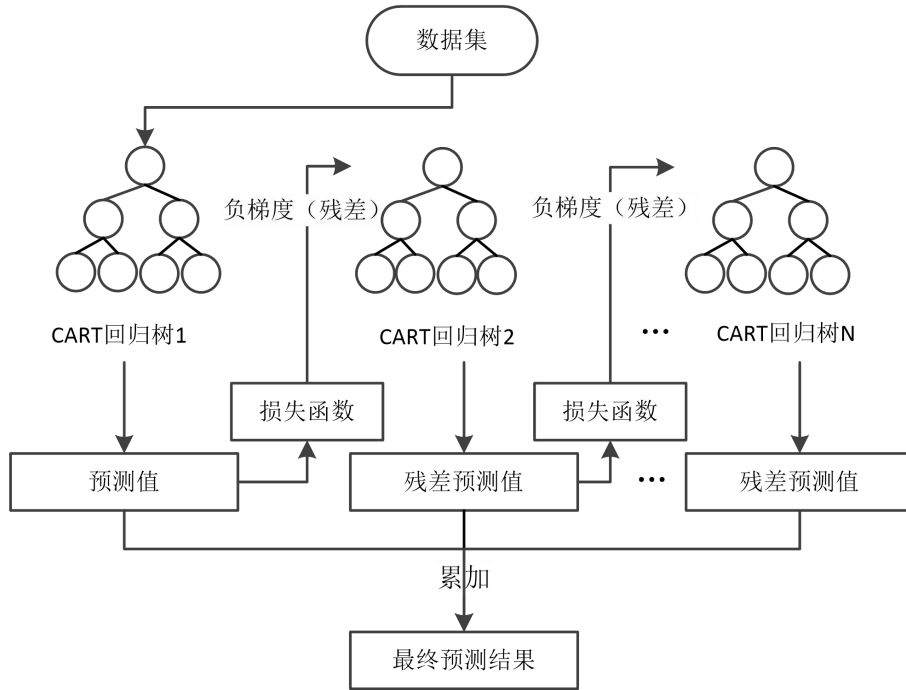


图 4.5 GBDT 模型流程图

GBDT 做回归任务时,损失函数使用均方误差,公式如下:

$$\mathcal{L}(y, \hat{y}) = \frac{1}{2} (y - \hat{y})^2 \quad (4.5)$$

它的负梯度为:

$$-\frac{\partial \mathcal{L}(y, \hat{y})}{\partial \hat{y}} = y - \hat{y} \quad (4.6)$$

其中 y 为实际值, $\hat{y}$ 为当前弱学习器的预测值,因此 GBDT 做回归任务时,损失函数的负梯度为预测值和实际值的残差。

4.2.3 LightGBM 模型

LightGBM 算法是对 GBDT 算法的改进,其之所以适合工业实践,是因为它对运算速度进行了很大的提升。主要有以下三点:

(1)直方图算法

GBDT 进行回归任务时,由于每个 CART 树的层数不高,保证其不容易出现过拟合现象,并通过提升法拥有了很高精度。但其训练每次几乎遍历全部数据,其时间复杂度和空间复杂度受到数据大小影响很大,假如数据大小为 N ,特征数量为 K ,那么 GBDT 进行一层树运算的

时间复杂度为 $O(N*K)$ 。LightGBM 将数据进行直方图将相同数值的数据累加到直方图上，如果数据大小为 N 的训练集，只有 n 个不同的数值，那么进行直方图算法后时间复杂度变为了 $O(n*K)$ ， $n \leq N$ 。在降低时间复杂度的同时，其内存也降为原来的 n/N ，见下图 4.6。

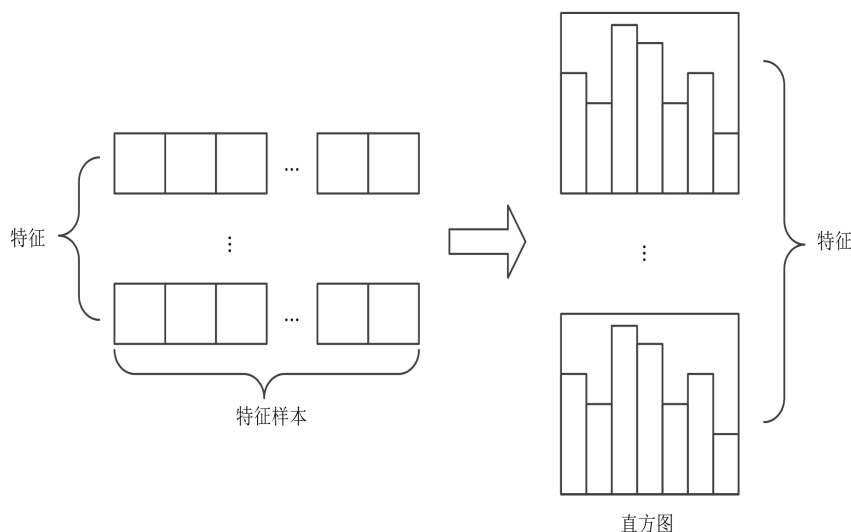


图 4.6 直方图算法

(2)带深度限制的 Leaf-wise 算法

GBDT 算法使用 Level-wise 的方式进行分裂，如果树的深度为 10，那么其叶子数量为 2 的 9 次方，有 512 个。而 LightGBM 算法使用 Leaf-wise 算法，他通过判断当前叶子的增益大小，只在分裂增益大的叶子进行分裂。这样在同样次数的分裂中，LightGBM 具有更高的精度，但也要注意其树深度可能过大，导致过拟合，因此在之后的模型建立中需要选择合适的树深进行实验，该算法示意如下图 4.7 所示。

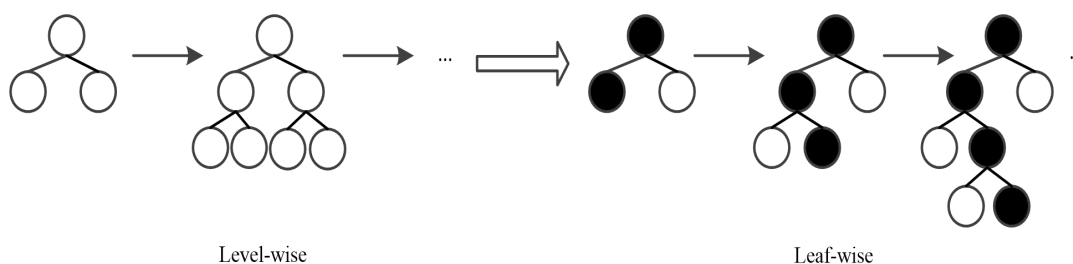


图 4.7 Leaf-wise 算法

(3)单边梯度采样算法

单边梯度采样算法简称 GOSS，GBDT 算法中，每一次学习后会对残差值进行新一轮的学习。而单边梯度采样算法的原理就是当数据的残差值已经足够小的时候，停止对其进行梯度计算，但是在计算中也并不是直接去掉了这些小残差值的数据，而是对数据进行绝对值大小的排序，抽取一定数量的大值和小值保留。之后对于残差样本仍然进行直方图算法优化，由于直方图每个桶中情况不一：一部分桶存在小值并全去除、另一部分桶仍有小值保留。需对桶中保留的最小值进行权重相乘，再加上保留的最大值，用于保证数据分布。通过去除一些残差较小的值可提高运算速度，并且其抽取和权重的策略，保证了运算的精度。

4.3 基于改进 LightGBM 算法的预测模型

4.3.1 贝叶斯优化算法

模型进行训练时，为了提高精度、减少运行时间以及防止出现过拟合和欠拟合，需要及性能参数的调优。目前调参通常有两种方法，一种是通过积累的经验 and 知识对超参数进行人工调参。另一种为通过网格搜索算法、随机搜索算法等进行自动调参。对于 KNN 这种训练简单的算法，可使用人工调参。而对于随机森林以及 LightGBM 这些拥有复杂超参数的算法，往往使用网格搜索算法和随机搜索算法进行调参。

网格搜索算法是一种性能较优的调参算法，它对设置的参数进行枚举的方法，对参数排列所有组合，逐一进行训练、验证。相对于随机搜索算法，它拥有更高的精度，但随着参数范围的增加，它容易出现参数爆炸的现象。如具有 5 个参数，其分别有 2、4、8 个离散值，其训练次数分别为 32、1024、32768 次。并且对于一些连续的参数，它无法做到精确的调参。

贝叶斯优化算法的核心过程为先验函数和采集函数，其基于高斯过程。通过先验函数初始训练，得到初始训练点。之后进行后验概率、均值、方差计算，采集函数根据计算得到的均值和方差决定是 exploration(探索未知空间) 还是 exploitation(现有最大值附近搜索)。

贝叶斯先验函数对 3 组超参数进行训练，每组参数对应的得分如图 4.8 红点所示。蓝色阴影区域为得分预测的 95%置信区间，采集函数进行了 exploitation(现有最大值附近搜索)。

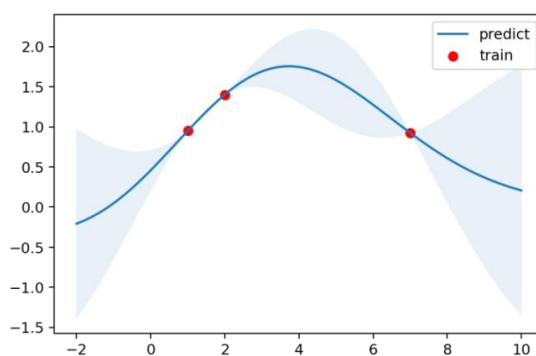


图 4.8 贝叶斯优化先验结果

对超参数继续迭代，进行下一组计算，结果如下图 4.9 所示，显示该点结果并不理想，进行下一组迭代，采集函数判断是 exploration 还是 exploitation，不断探索，以寻求最优解。

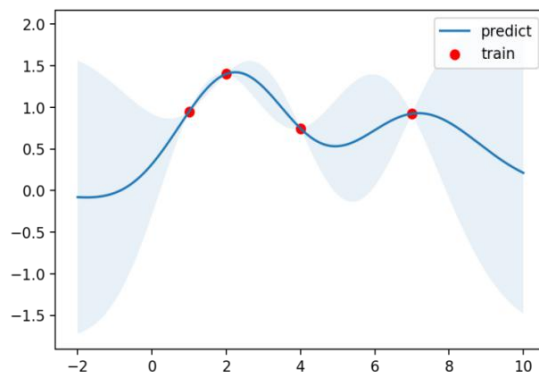


图 4.9 高斯过程 exploitation 结果

4.3.2 模型超参数设置

在阐述说明贝叶斯优化算法后，对 LightGBM 算法众多参数进行设置，由于贝叶斯优化算法的特性，可以训练一些连续的参数，如 `min_split_gain`（叶节点分支残差阈值）、`learning_rate`（学习率）、`bagging_fraction`（样本采用比例）、`feature_fraction`（特征采用比例）、`lambda_l1`（L1 正则权重值）、`lambda_l2`（L2 正则权重值）以及其他一些整数的参数，参数设置见下表 4.1。LightGBM 模型预测结果的精度以及训练速度与这些参数有很大关系，如树的深度和正则权重正确设置可以防止过拟合；学习率、迭代次数、叶节点分支残差阈值与结果精度相关；样本和特征采用比例、直方图分桶与训练速度有很多关系。

表 4.1 LightGBM 参数表

常用参数	参数含义	参数设置
<code>num_leaves</code>	叶子数量	[7, 31, 63]
<code>max_depth</code>	树的最大深度	[3, 5, 6, 7, 9, 12]
<code>min_data_in_leaf</code>	叶子最小数据数量	[30, 50, 70]
<code>min_split_gain</code>	叶节点分支残差阈值	(0.1, 0.8)
<code>learning_rate</code>	学习率	(0.01, 1)
<code>bagging_fraction</code>	样本采用比例	(0, 1)
<code>feature_fraction</code>	特征采用比例	(0, 1)
<code>max_bin</code>	直方图分桶	[2-20]
<code>n_estimators</code>	迭代次数	100
<code>lambda_l1</code>	L1 正则权重值	(0, 1)
<code>lambda_l2</code>	L2 正则权重值	(0, 1)

综上所述，LightGBM 模型需要设置的参数较多，各参数之间有一定联系，不适合人工反复调参来训练对比。并且拥有较多连续的参数，如 `bagging_fraction` 这类连续的参数，可能它的最优解为 0.84274，这点是网格搜索算法通过枚举无法做到的，因为连续值可以无限细分。因此使用贝叶斯优化算法进行对超参数进行迭代寻找最优解，可以提高 LightGBM 模型的预测精度和训练速度。

4.3.3 基于贝叶斯优化的改进模型

在了解贝叶斯优化算法和模型超参数设置以及 LightGBM 模型本身的特点后，进行基于贝叶斯优化算法改进的 LightGBM 预测模型，其流程见下：

(1)数据采集和预处理：智能网关收集锯床 PLC 数据，上传云平台。对收集的数据进行切削面积计算以及失效点判断，保证关键数据的准确。之后通过缺值补充、异值处理和归一化处理得到初始数据。

(2)特征选择：将初始数据整合并抽取 2000 条数据，按照 4: 1 划分训练集和测试集，通过数据间本身的相关性等可视化测试，驱动完成输入特征的选择。

(3)模型训练：将输入特征进行直方图算法提高运算速度和降低内存消耗，通过带深度限制的 Leaf-wise 和单边梯度采样策略构造 CART 回归树，进行单次预测，并对每次预测的损失函数负梯度（残差）进行迭代预测。最终达到终止条件后对每次的 CART 回归树的预测结果进行累加，得到一组超参数下的预测结果。

(4)贝叶斯优化：通过贝叶斯优化算法对模型的超参数设置进行改进，对于连续值的参数，该优化算法可以得到更精确的值。并且根据高斯回归过程的特点，其运算时间只和训练迭代次数相关，只需考虑各参数的设置范围和迭代次数即可，不用和网格搜索算法那样需要进行粗搜索和细搜索等分布操作。通过贝叶斯优化重复模型训练过程，最终得到最优的超参数设置。

(5)预测测试数据：根据改进的 LightGBM 模型，输入金属带锯条的特征测试集，完成预测结果，通过均方根和决定系数等评价指标进行评价。

基于贝叶斯优化算法对 LightGBM 的预测改进模型，建模流程如图 4.10 所示，并通过 Python 语言基于 Keras 平台以及 Sklearn 库完成模型预测。

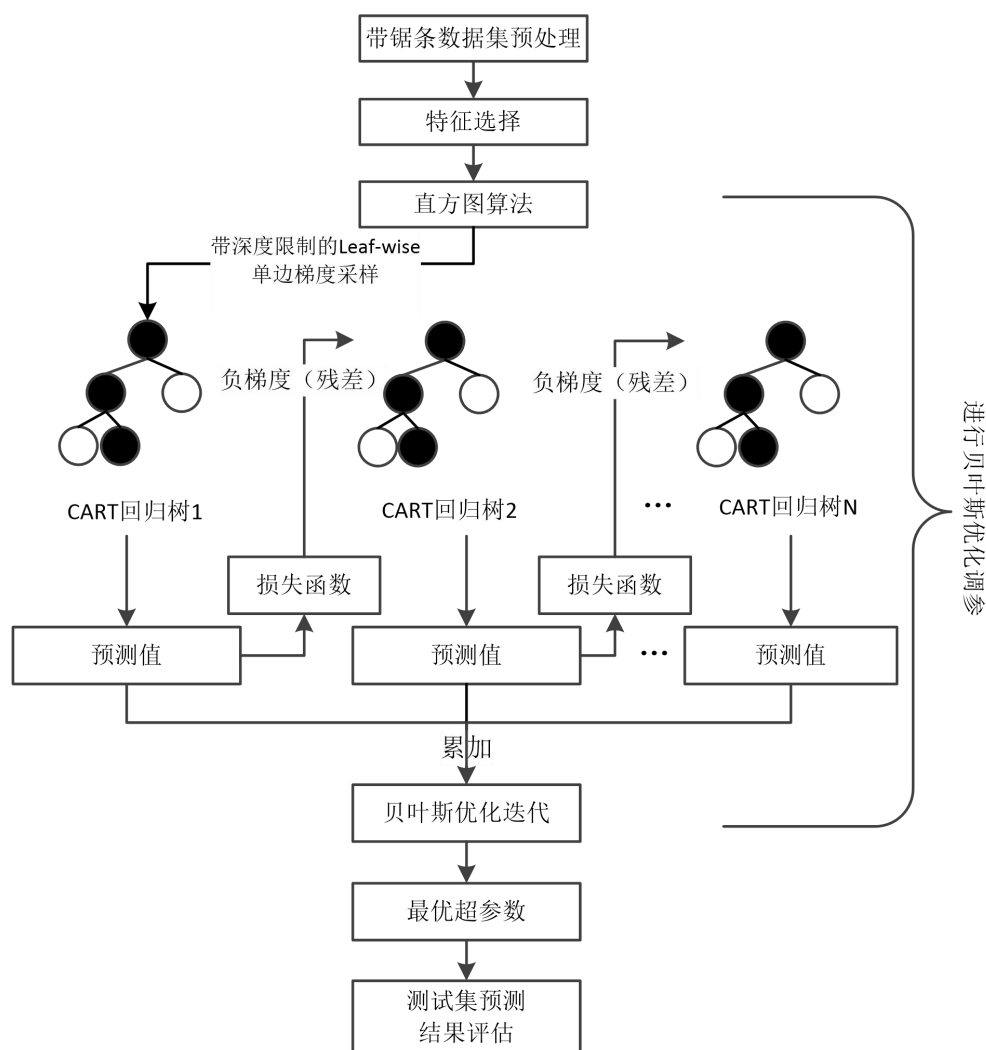


图 4.10 改进 LightGBM 预测模型的建模流程

4.4 对比其他预测模型

4.4.1 K 最近邻法

K 近邻法在 1967 年由 Cover T 和 Hart P 提出^[42], kNN 算法主要用于解决分类问题, 但是其在回归问题的解决上也有奇效。它有很多优缺点, 其中一个缺点是模型结果不具有可解释性, 而很多时候我们是希望得到的模型是能够作出合理解释的, 以便指导业务。其简单来讲就是通过计算不同数据信息之间的距离进行判断, 若进行分类学习, 则进行距离择近, 若用于回归学习则求其距离平均值。由于 KNN 是众多基于距离回归算法的典范, 故将 KNN 用于构建金属带锯条剩余寿命预测判别模型具有较强的代表性。

K 最近邻算法用于分类时, 其原理很好理解。从算法名就可以了解到其原理: 在数据因素选择 k 个相距最近的数据, 通过这些数据的类型标签, 通过多数投票的方法进行类别判定, 完成分类, 回归也是同理, 只是回归是进行均值计算。K 最近邻算法重点在于 K 的取值和距离计算方法上。

K 比较小时, 在进行分类时, 其可能与异常数据较近, 因此对数据质量有更高的要求。并且 K 最小, 其分类越多, 模型更为复杂。反之, 同理。

K 最近邻距离算法有许多, 如欧式距离、曼哈顿距离等^[43], 前者为两点间直线距离, 后者为两点间投影到 x, y 轴的距离之和。平面中两点的欧式距离公式为:

$$\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (4.7)$$

到多维空间中欧式距离公式为:

$$d(x, y) := \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4.8)$$

K 最近邻法参数最重要的为 K 值, K 值一般不超过 20, 所以 KNN 模型可使用人工调参也可以选用网格搜索算法进行自动调参, 本文选取 K 值进行 3-38 之间, 1 为间隔进行网格搜索算法进行调参训练。如图 4.11 通过测试当 K 为 18 时, 误差最小, 得到最优 K 值。

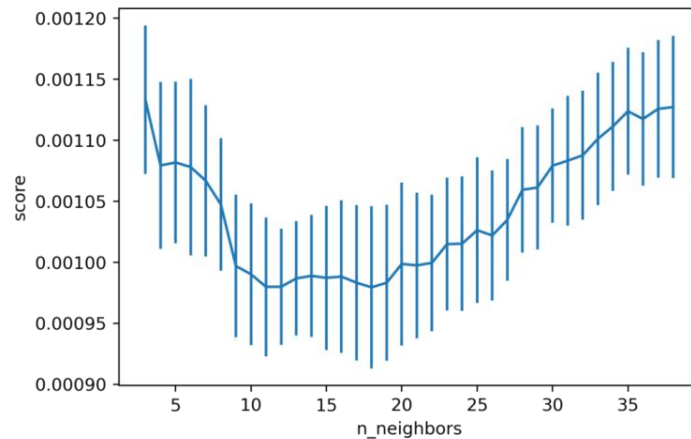


图 4.11 K 值训练图

4.4.2 随机森林模型

随机森林算法产生于 1995 年，他将多个决策树结合起来，形成了新的算法，这种方法叫做集成学习^[44]。集成学习模型建立多个弱学习器进行学习，再通过不同的算法将弱学习器联系在一起形成强学习器得到弱学习器不具备的能力。在随机森林算法中，决策树本身属于弱学习器，而将多个决策树集成在一起时就会具有很强的预测效果。而且随机森林算法是一种非线性的回归算法，随机森林不同于 XGBoost 以及 LightGBM 之处在于他是基于袋装法 (Bagging) 的集成方法，通过并行训练的方式使不同弱学习器之间彼此独立，其具有抗优秀的过拟合特性。综上所述，将随机森林用于构建金属带锯条剩余寿命预测判别模型具有较强的适用性。

随机森林算法主要从决策树算法和集成学习中的装袋法两方面进行分析^{[45][46]}。

(1) 装袋法

通过 4.2 节，我们得知随机森林是基于决策树算法加上集成学习思想，并且通过袋装法完成的一个算法。什么是装袋法，袋装法可以很形象地把自己概述成：分为几个袋子并且有放回的装袋，直到装满袋子并进行机器学习。但是其本质自助抽样集成方法 (Bootstrap AGGREGatING) 的缩写^{[47][48]}。

自助抽样集成方法的自助法体现在其将样本分成多个新样本是通过从自身抽取类似的新样本，因此叫做自助。举个例子，加入我们有数量为 M 的训练集，我们想有 N 个数量为 M 的子训练集来进行学习。那么我们从 M 个父训练集中随机抽取 1 个数据，将其放回再抽取 1 个数据，重复 M 次直至子训练集数量也到达 M ，这样就得到了一个数量为 M 的子训练集，总共实行 NM 次就可获得 N 个数量为 M 的子训练集。之后在每个子训练集上建立需要学习的模型，最终结果进行均值整合。

(2) 决策树

上文已经对决策树进行了简述，我们了解到他是机器学习中一个最重要的基础算法。而袋装法中，每个子训练集用上决策树算法模型，那么得到的就是随机森林算法模型^[49]。上文讲到决策树分为 ID3、C4.5、CART，其中 ID3、C4.5 是基于香农熵分裂的，CADT 是基于基尼系数分裂的。

(3) 多数投票

随机森林通过装袋法将原有的训练集重新装袋成 n 个同样数量的子训练集。之后通过 CART 算法弱学习器进行模型训练，得到结果。最终通过多数投票机制得到最终的结果，其在回归模型中往往使用均值的方法得到结果。

在阐述说明随机森林算法的基本原理后，本文利用 Python 语言基于 Keras 平台搭建基于随机森林算法的金属带锯条剩余寿命预测模型。

并且结合下图 4.12，完成模型构建的步骤和参数设置：

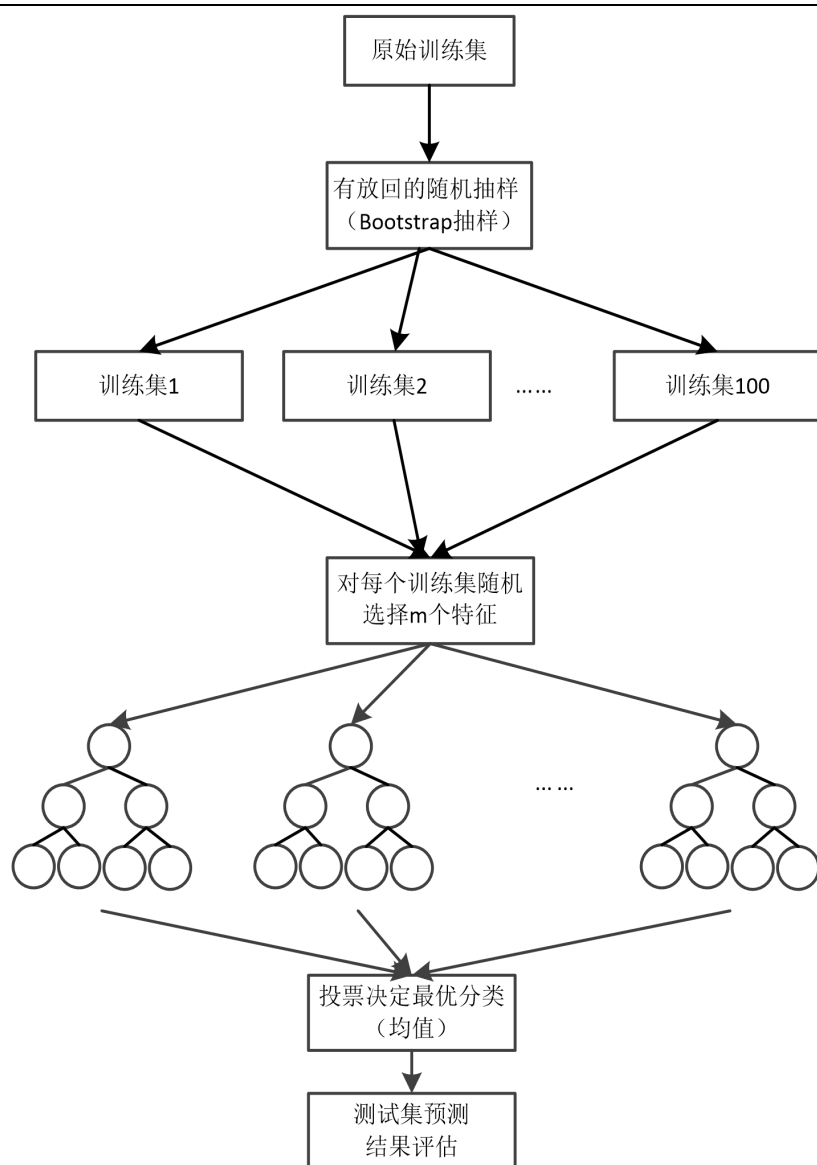


图 4.12 随机森林模型流程图

(1)本文采用的训练集的数据数量为 1600，通过装袋法从原始训练集中有放回的进行抽样，组成 100 个新的训练集。训练参数 $n_estimators$ 设置为 100、150、200，进行三种大小的训练，取最优的情况。

(2)使用决策树算法对每个训练集进行学习，使用最小的基尼系数进行分支。在每一层分支时需要设置最大特征数 $max_features$ ，设置为 1、2、3、4，进行 4 种特征数的训练。而每个节点最小样本数设置为 1、2，设置叶子节点数为空，不进行限制。设置决策树的最大深度为 12。节点最小样本数设置为 1、2 是因为训练集数量不大，不需要设置太多。而 1600 个数据树的层数不会超过 12 层，所以树的最大深度设置为 12，进行网格搜索算法调参。

(3)设置评价因素为 RMSE、 R^2 以及 Z 估计值，当所有分类模型完成后，进行测试集测试，通过投票决定最优分类。

4.5 实验测试

4.5.1 模型评估指标

机器学习在模型构造，模型参数确定后进行结果预测，而预测结果的好坏有机器学习评价因素来判断，我们在进行金属带锯条剩余寿命预测后，需要对结果使用评价因素进行评估，通过它得到性能较好的金属带锯条剩余寿命预测模型。

一般机器学习的评价因素分为回归和分类评价因素，因为本文模型使用的是回归模型，所以下面讲一下回归模型有哪些评价因素，主要介绍平均绝对误差 MAE、均方误差 RMSE、判断系数 R^2 以及在模型训练中使用的其他特征如 z 统计值。

(1) 平均绝对误差 MAE

在介绍平均绝对误差之前先要了解什么是（实际值与预测值）的均值。

$$\text{residual}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (4.9)$$

其中 y_i 实际值， $\hat{y}_i$ 为预测值。

残差均值表示实际值和预测值之间差的和，但是用残差如果遇到实际值减去预测值正负相抵，那么就没有意义了。所以给残差均值添加一个绝对值就成了平均绝对误差 MAE:

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.10)$$

平均绝对误差解决了残差均值的问题，但它仍存在问题：曲线平滑降低，无法处处求导。

(2) 均方根误差 RMSE

在介绍均方根误差 RMSE 之前，仍要先了解另一个评价因素：均方误差。

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.11)$$

同上其中 y_i 实际值， $\hat{y}_i$ 为预测值。

均方误差 MSE 解决了平均绝对误差 MAE 无法求导的问题，但是它也存在一个问题，那就是实际值减去预测值后再开平方，那么原函数的单位也进行了平方，导致单位不一。于是在此基础上，均方根误差 RMSE 通过对均方误差 MSE 开根号的方法，保证了单位一致问题，他在机器学习库中通常被表示为 std。

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.12)$$

(3) 决定系数 R^2 -score

决定系数 R^2 在机器学习中是一个关键的评价因素， R^2 在机器学习中通常被写成 score，其大小在 0 到 1 之间。 R^2 越大越接近 1，说明预测值和实际值越相近，说明训练结果越好。

R^2 越接近 0，说明预测值和实际值越远离，说明训练结果越差^[50]。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.13)$$

其中分子为预测数据和原始数据的误差平方和 SSE，也就是均方误差乘以了数量，本身没有什么意思^[51]。分母中 $\bar{y}$ 是实际值的算数平均值，在机器学习库中通常以 `mean` 表示，通过分子和分母相除可以减少实际值本身比较分散的现象，从而用来表征和的好坏。

(4) z 统计量

以上 3 点是机器学习评判回归模型性能的评价因素，而本文除了常用的机器学习评价因素外，还用统计学中的 z 统计量进行预测值和预测均值之间的差异判断，Z 统计值可以统计不同准确度的预测值。下面是他的公式：

$$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \quad (4.14)$$

其中 $\bar{X}$ 为检验样本的， μ_0 是已知总体的平均数， S 是样本的标准差， n 是样本容量。但是在机器学习中，Z 统计量通常被写成这样：

$$Z = \frac{y - \hat{y} - \overline{y - \hat{y}}}{\text{RMSE}(y_i, \hat{y}_i)} \quad (4.15)$$

其 $\overline{y - \hat{y}}$ 中为预测误差的均值。

4.5.2 预测模型结果

(1) 剩余寿命预测结果

通过不同模型对训练集训练，并对测试集进行预测。其预测值以及测试集的 400 个数据预测结果如下图 4.13 所示，由于四种模型预测精度相近，其总体结果不一一展示。

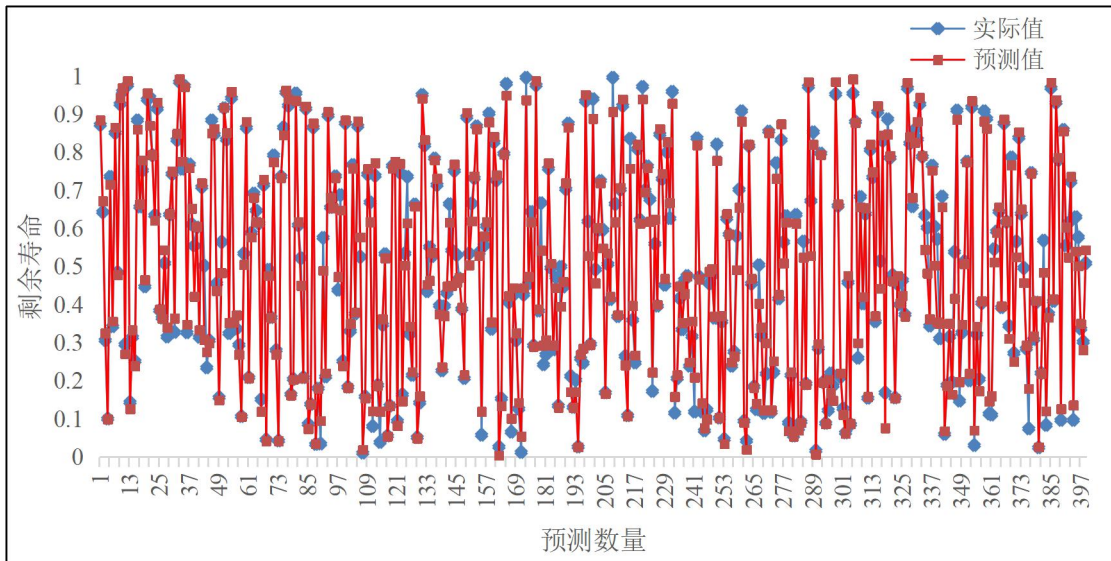


图 4.13 剩余寿命预测结果图

(2) 模型训练结果对比

预测结果折线图结果很相近，若直接观察很难看到各模型的性能优良。所以结合上小节的性能评估指标，对 KNN（K 最近邻）、随机森林、LightGBM 模型进行评估。各模型的决定系数、均方根误差以及 Z 统计值见下图 4.14，图 4.15，图 4.16。

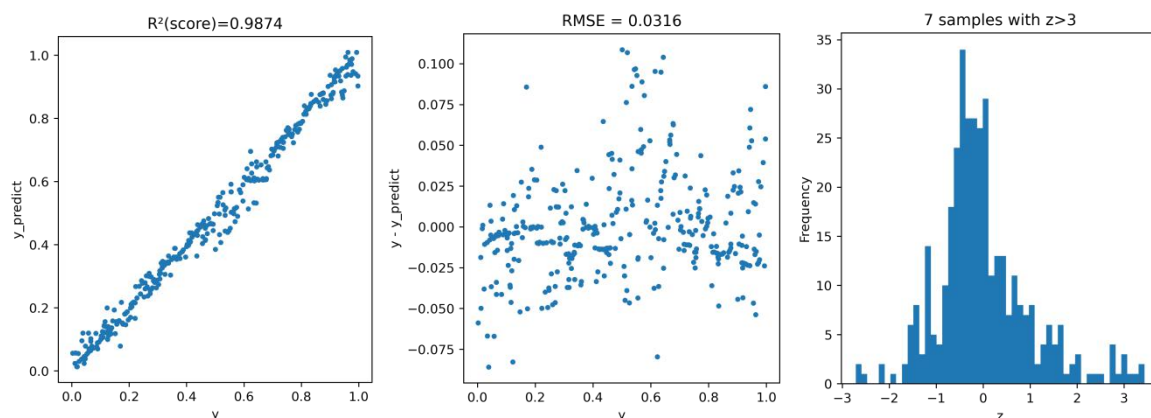


图 4.14 KNN 预测值评估指标图

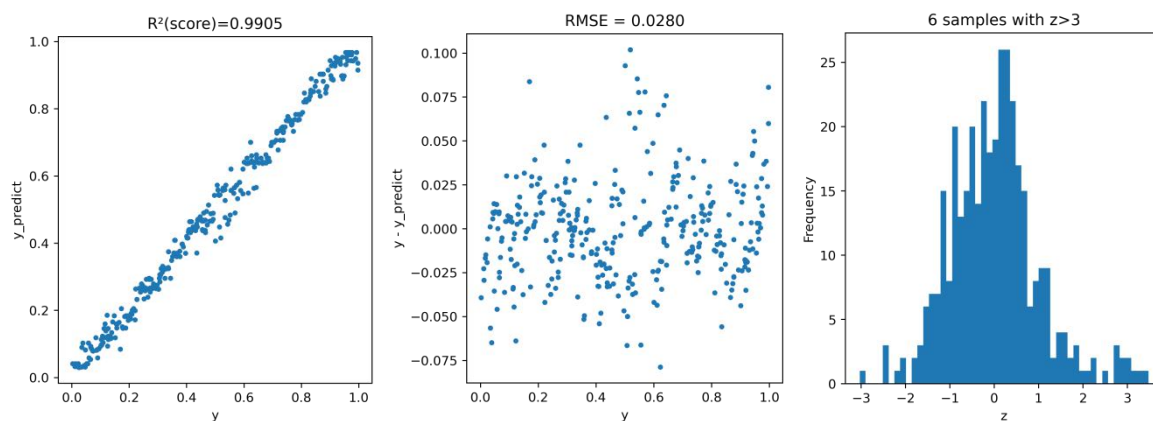


图 4.15 随机森林预测值评估指标图

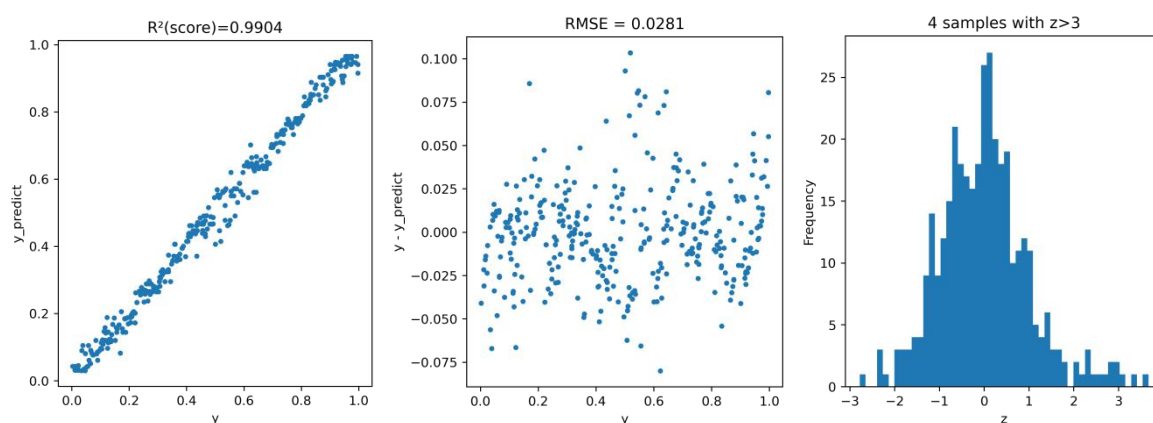


图 4.16 LightGBM 预测值评估指标图

根据上图，我们可以看到对于基于决策树算法以及集成算法的 LightGBM 和随机森林算法，KNN 算法的评估较差，三种指标都是相对最差的。但是随机森林的均方根误差为 0.0280，决定系数为 0.9905，而 LightGBM 的均方根误差为 0.0281，决定系数为 0.9904。随

机森林的预测结果和 LightGBM 很相近。最后仍然选择的是 LightGBM 模型进行优化的原因：LightGBM 的在训练速度方面的性能远远高于随机森林，三种模型训练时间对比图见下图 4.17。为了对比训练时间，三种模型都使用了网格搜索的调参方法，分别枚举 1、2、4...2n 次，再有五折交叉验证（乘五）。由图可知：1.随着参数的复杂程度上升（每次翻倍），模型的训练时间将成倍上升。2.LightGBM 模型的训练速度远远低于随机森林模型，相差近 10 倍，随着工业实践中数据样本的逐渐增加，LightGBM 将发挥更好的作用。

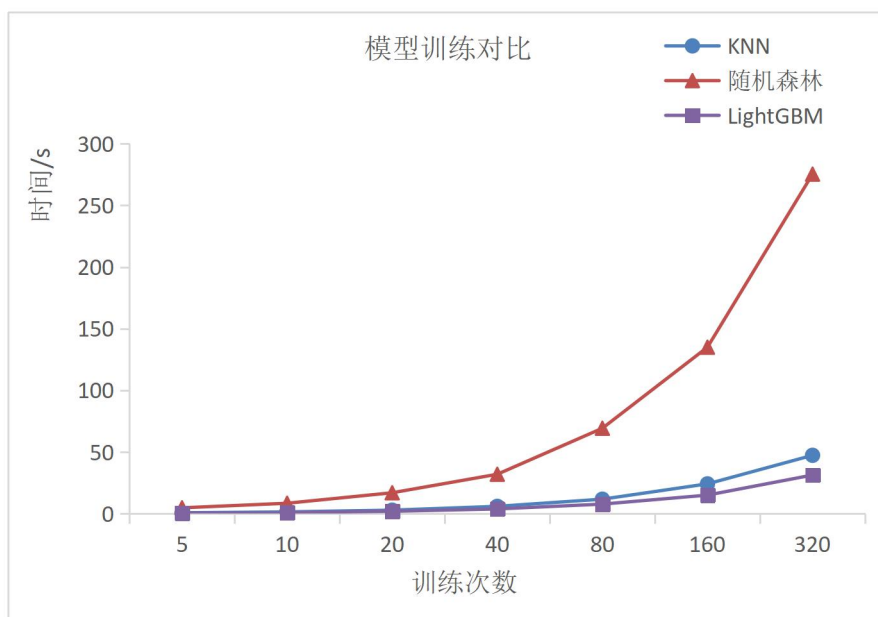


图 4.17 三种模型训练时间对比

(3)改进的 LightGBM

通过贝叶斯优化算法对 LightGBM 模型进行处理，得到更优和更精确的超参数组合后，其预测结果如下图 4.18 所示，相对于原来的 LightGBM 模型，其均方根误差和决定系数都得到提高。最重要的是：相对于网格搜索算法对参数一一进行枚举，贝叶斯优化算法训练实践主要与本身迭代时间相关，提高了训练效率。并且对于一些连续的参数，贝叶斯优化算法拥有更高的精准度，这是网格搜索算法无法做到的。

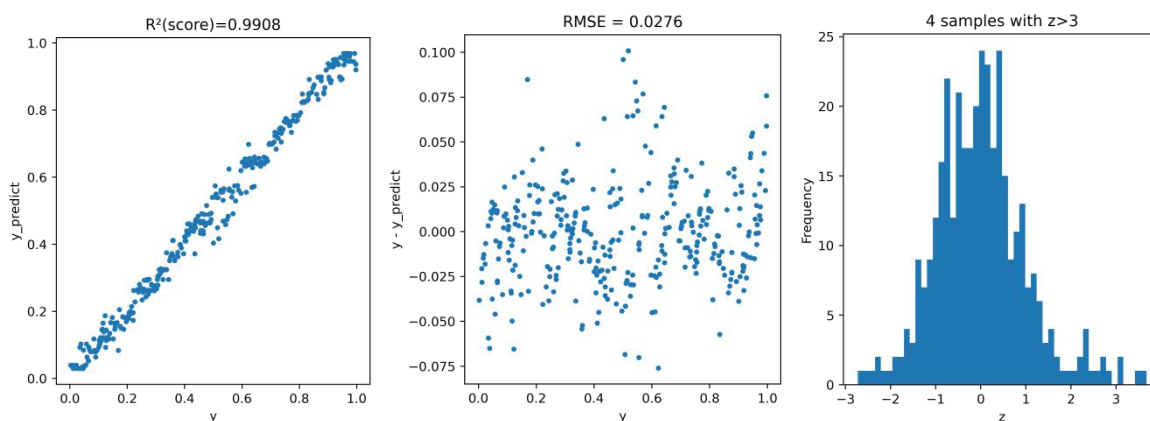


图 4.18 改进的 LightGBM 预测值评估指标图

对三种模型以及改进模型评价指标进行汇总，评价指标包括均方根误差（RMSE）、决定系数（ R^2 -score）和 $Z>3$ 统计值，其汇总结果见表 4.2。

表 4.2 模型性能评价指标对比

模型	评价指标		
	RMSE	R^2 -score	$Z>3$ 统计值
K 最近邻	0.0316	0.9874	7
随机森林	0.0280	0.9905	6
LightGBM	0.0281	0.9904	4
改进的 LightGBM	0.0276	0.9908	4

从表可以看到，在模型完成对训练集的训练之后，其在当前测试集下的验证结果为基于改进的 LightGBM 算法的金属带锯条剩余寿命的性能更为优秀，其次为随机森林算法、LightGBM 算法以及 K 最近邻算法。表中没有表述的为训练时间方面，随机森林算法需要花费大量的时间，不适用于有大量数据样本的工业实践中。而 LightGBM 算法针对工业分析的改进，其计算速度远远优于其他两种算法。改进的 LightGBM 模型对测试集的 RMSE、 R^2 -score 最优，表示其对训练集的预测值与实际值更接近， $Z>3$ 统计值最小，表示其误差大于误差均值三倍的情况较少，表示预测结果更稳定。

4.6 本章小节

本章根据 K 最近邻算法，随机森林算法，LightGBM 算法，分析其各自的原理，根据算法特性选择调参方式：人工调参或网格搜索调参。建立其各自的金属带锯条剩余寿命预测模型，通过模型评估指标的对比验证 LightGBM 算法是种适合工业实践的算法。并且对 LightGBM 进行贝叶斯优化算法进行调参优化，最终得到更优的模型。

实验结果为：基于 K 最近邻的金属带锯条剩余寿命预测模型的 RMSE、 R^2 、Z 统计值分别为 0.0316、0.9874、7。基于随机森林的预测模型的 RMSE、 R^2 、Z 统计值分别为 0.0280、0.9905、6。基于 LightGBM 的预测模型的 RMSE、 R^2 、Z 统计值分别为 0.0281、0.9904、4。基于改进的 LightGBM 的预测模型的 RMSE、 R^2 、Z 统计值分别为 0.0276、0.9908、4。并且通过训练时间证明了 LightGBM 在运算速度上的优势，实验结果验证了 LightGBM 的各性能指标都很优秀可作为带锯条剩余寿命预测模型，验证了贝叶斯优化算法对模型精度的改进。

第 5 章 总结与展望

5.1 本文工作总结

本文针对目前金属带锯床远程监控系统研究中存在的问题：1.网关通信能力差，无法在不友好环境下支持摄像头流畅工作；2.使用上位机来监控显示金属带锯床的数据，页面比较粗糙，人机交互不方便；3.仅有金属带锯床故障预警功能，没有对锯条剩余寿命预测的功能。针对以上问题，基于机器学习技术设计了一套远程监控系统。通过需求分析设计总体框架，之后完成网关设计、云平台设计、Web 设计、App 设计，完成一套功能健全、通信稳定、人机交互舒适的金属带锯床远程监控系统，并且通过该系统完成前期数据的采集，最终根据采集的数据对带锯条剩余寿命进行初探究。本文主要的工作和结论如下：

(1)对目前金属带锯床远程监控系统的文献专利进行分析,对目前金属带锯条剩余寿命预测的文献专利进行分析以及对近几年机器学习的发展进行分析，完成前期的理论基础，为带锯条剩余寿命预测的开展提供理论基础。

(2)对金属带锯床远程监控系统各部分进行设计。通过设计智能网关完成 PLC 的数据接收处理和对其远程更新功能，完成其 4G 和 WIFI 功能保证数据传输稳定。设计业务云进行数据储存管理，测试其带负载均衡时的性能。设计 Web 客户端展示设备数据、视频监控、故障预警等功能，设计企业管理 Web，为锯床行业提供管理平台，完成 Web 功能展示测试。通过设计 APP 客户端提高用户查看数据的便捷性，完成 APP 软件的稳定性测试。实现一套功能健全、通信稳定、人机交互舒适的金属带锯床远程监控系统。

(3)针对带锯条剩余寿命影响因素较多，影响关系复杂的问题。通过数据驱动的方式，直接通过数据间的关联来选择输入数据特征。在选择之间，通过面积计算方案降低异常率；通过已经更换的带锯条的历史信息结合 MTBD 理论对切削率的衰退趋势分析，找到锯条合理更换点，定量计算其每个历史信息中的剩余寿命。对剩余寿命以外的其他数据进行预处理，最后以训练集：测试集为 4：1 的比例分割，并进行数据相关性测试，数据分布平衡测试等实验的输入数据特征的选择。

(4)通过分析本文数据样本的特点，选择机器学习种有监督学习模型：KNN、随机森林、LightGBM 算法。分析各算法的原理，如决策树算法、集成学习以及贝叶斯优化算法等。结合 Python 语言以及支持的机器学习工具库，完成工作分析环境搭建。通过对机器学习 K 最近邻算法、随机森林算法、LightGBM 算法模型基于网格搜索算法调参，完成模型建立并进行评价指标对比。最终验证 LightGBM 模型的优良性能并使用贝叶斯优化算法进一步对其改进，其根均方误差仅为 0.0276， R^2 得分高达 0.9908，可实现金属带锯条剩余寿命预测功能且性能优秀。

5.2 展望

本文建立的金属带锯床远程监控系统在功能完整性、应用需求上达到了最初的设计目标，帮助企业在《中国制造 2025》的背景下完成智能化转型，通过良好的人机交互提高了效率，降低故障发生，通过金属带锯条寿命预测功能尝试解决带锯条浪费断裂问题。但是由于时间和客观条件的限制，虽然系统已经基本完成，但是研究内容还需进一步完善，主要体现在以下几个方面：

(1)远程监控系统中，当前每个网关接入设备较少，随着接入设备的增加以及计算预测速度的提高，需要进行边缘网关的设计，提高网关自身的计算能力，降低云平台的计算压力，如今接入设备还不多，数据还没达到海量的状态，暂时未充分使用对分布式数据库进行分布式的云计算处理。随着设备数据增加，需要将数据进行分布式管理分析，完善用于大数据分析的平台。

(2)在剩余寿命影响因素方面，本文在进行分析后只使用到锯切累计时间、切削率两个参数，还不能在更多因素影响下进行寿命预测。带锯条本身材质型号、带锯条齿数、下切压力等因素从理论上讲也是重要的影响因素，但是其或是缺失或是异常导致无法进行训练。另一方面机器振动、声音模块的集成还没普及，其与带锯条剩余寿命之间的关系有待研究。

(3)在机器学习方面，算法在更新换代，若要实现带锯条的剩余寿命预测，只有在大量数据的基础上通过不断尝试才能判断最优的模型。并且由于数据样本限制了带锯条的品牌以及切削环境比较友好，在此数据集下预测性能较优，改进 LightGBM 模型的均方根误差仅为 0.0276。但当其接受到更多剩余寿命预测特征以及更多数据样本的情况下，其均方根误差和得分系数可能上升。

致谢

光阴似箭，不知不觉硕士生涯已经走到尾声。在此论文完成之际，谨向三年来给予我关心、支持和帮助我的良师益友、亲人致以最真挚的谢意！

本论文是在导师陈科明副教授的悉心指导下完成的。论文选题、方案设计、实验的实施以及实验数据的处理分析，直至论文的撰写，每一个环节都倾注了导师的心血。我现在能够取得的成绩和进步都离不开陈老师的谆谆教诲。在此毕业之际，谨向陈科明副教授致以最崇高的敬意和最衷心的感谢，感谢他对我学业上的悉心指导和生活上的热情帮助。

感谢乔冠师兄以及其他师兄师姐，在研究生期间对我在学习和生活上给予的无私帮助，是你们让我能够很快的适应新环境，顺利完成研究生阶段的学习任务。

感谢我的同门王彪、檀宇、安建伟、王弈怡、吴新虎、郑杭明、司端杨、杜坤。感谢我们一起进步一起团结合作的时光，感谢你们在我失意时对我的鼓励，在论文上给我的建议。

感谢论文评阅和答辩委员会的老师，感谢你们在百忙之中对本文的审阅，并恳请提出宝贵的意见和批评。

最后，特别感谢我的家人和朋友在我求学生涯中给予我的鼓励和无微不至的关心与支持！衷心祝愿你们身体健康，永远幸福！

参考文献

- [1]孙燕妮,白晓军.《中国制造 2025》——中国特色的强国战略[J].智能制造, 2020 (10): 43-45.
- [2]郭震,郑松. 工业互联网平台参考架构与发展现状[A]. 中国自动化学会、杭州市人民政府、浙江省科学技术协会.2019 中国自动化大会(CAC2019)论文集[C].中国自动化学会、杭州市人民政府、浙江省科学技术协会:中国自动化学会,2019:6.
- [3]胡媛媛. 基于多源异构工业大数据的刀具健康监测方法及应用研究[D].哈尔滨工业大学,2019.
- [4]徐子文.怎样延长带锯条的寿命[J].世界制造技术与装备市场,1999(01):3-5.
- [5]任泽宇.刀具磨损监测技术的研究现状及发展趋势[J].中国设备工程, 2019 (02): 76-77.
- [6]胡福丽.PLC 技术在电气工程及其自动化控制系统中的运用[J].当代化工研究, 2019 (17): 55-56.
- [7]接鑫,高连兴.金属带锯床的技术现状及发展趋势分析[J].机床与液压, 2016, 44 (17): 188-190.
- [8]刘佳. 金属带锯床远程无线监控系统研究与设计[D].武汉理工大学,2012.
- [9]刘腾发. 金属带锯床在线监测与故障诊断系统研究[D].河北大学,2020.
- [10]Orlowski K A, Dobrzynski M, Gajowiec G, et al. A Critical Reanalysis of Uncontrollable Washboarding Phenomenon in Metal Band Sawing[J]. Materials, 2020, 13(20):4472.
- [11]Cho K , JH Müller, Scheffer C , et al. Application of an artificial neural network for the quantitative classification of trochlear dysplasia[J]. Journal of Mechanics in Medicine & Biology, 2013, 13(04):1350059-.
- [12]Cuka,Cho,Kim. Vision-based surface roughness evaluation system for end milling[J]. International Journal of Computer Integrated Manufacturing,2018,31(8).
- [13]Yan J H , Isobe N , Lee J . Fuzzy Logic Combined Logistic Regression Methodology for Gas Turbine First-Stage Nozzle Life Prediction[J]. Applied Mechanics & Materials, 2008, 10-12:583-587.
- [14]L.-J, Xing, J.-D, et al. Artificial neural network prediction on wear properties of high vanadium high speed steel (HVNSS) rolls.[J]. Materials Science & Technology, 2007, 23(3):315-319.
- [15]戴诗语. 基于结构支持向量机的目标跟踪算法研究[D].哈尔滨理工大学,2016.
- [16]娄棕棋.机器学习的理论发展及应用现状[J].中国新通信,2019,21(01):60-62.
- [17]黄鲁成,薛爽.机器学习技术发展现状与国际竞争分析[J].现代情报,2019,39(10):165-176.
- [18]叶志祥. 基于 BP 神经网络和 KNN 的店铺销售额预测研究[D].安徽理工大学,2018.

- [19]姜鹏. 基于机器学习的煤自燃温度预测模型研究[D].西安科技大学,2020.
- [20]王奕森. 随机森林和深度神经网络的若干关键技术研究[D].清华大学,2018.
- [21]Babak Abbasi,Toktam Babaei,Zahra Hosseinifard,Kate Smith-Miles,Maryam Dehghani. Predicting solutions of large-scale optimization problems via machine learning: A case study in blood supply chain management[J]. Computers and Operations Research,2020,119:104941.
- [22]周雪梅. 电动自行车智能充电系统设计实现及关键技术研究[D].杭州电子科技大学,2020.
- [23]孙腾飞. 基于 Android 的移动运维系统的设计与实现[D].东北大学,2015.
- [24]乔冠. 基于自适应温度控制的智能咖啡烘焙系统研究[D].杭州电子科技大学,2020.
- [25]刘琦. 卫星关键部件剩余寿命预测[D].西安理工大学,2020.
- [26]黄金发,陈贤生.金属带锯床人机工程设计研究[J].机械设计,2018,35(05):126-128.
- [27]钱晓健,冯加云.带锯床和带锯条的正确使用[J].机械制造,1998(04):34-36.
- [28]程廉贵,率云泉 ,代文科 ,何红霞 ,刘四忠 ,李红 ,王宏庆 ,王利民.带锯条使用寿命的因素分析[J].设备管理与维修,2005(04):13-15.
- [29]万尧. 基于深度学习的图像匹配算法研究[D].南昌大学,2019.
- [30]Hongbo Wang, QianTian, Zhihong Hu, Jingna Guo. Image Feature Detection Based on OpenCV[J]. Journal of Research in Science and Engineering,2020,2(7).
- [31]刘啸,柴会敏,罗俊波,安汝伟,崔岗卫,王元伦.数控机床可靠性试验研究[J].内燃机与配件,2019(11):93-94.
- [32]柴尔烜. 基于机器学习的电力系统故障诊断与暂态稳定评估研究[D].杭州电子科技大学,2020.
- [33]贾寓真,吴懿萍,刘刚,刘国跃.基于田口方法的双金属带锯条锯切参数优化[J].现代制造工程,2019(08):78-81+55.
- [34]廖伟琛. 基于聚类的数据挖掘技术在未来网络基站部署策略中的应用[D].北京邮电大学,2018.
- [35]李小英. 基于相似性网络的疾病 miRNAs 预测方法研究及应用[D].湖南大学,2018.
- [36]彭敦诚. 基于 Python 机器学习的大气环境下材料失效数据分析[D].机械科学研究总院,2020.
- [37]张亮亮,唐有,张翌维,王新安.针对密码芯片数据搬移能量曲线的机器学习攻击[J].电子科技大学学报,2019,48(03):415-419.
- [38]张晓贺. 决策树分类器的实现及在遥感影像分类中的应用[D].兰州交通大学,2013.
- [39]贺捷. 随机森林在文本分类中的应用[D].华南理工大学,2015.
- [40]张洁. 基于深度学习方法的胶质瘤图像分割的初步研究[D].西安电子科技大学,2017.
- [41]武亦文.基于决策树的分类算法研究[J].数字通信世界,2017(12).
- [42]王邦军,李凡长,张莉,于剑,何书萍.基于改进协方差特征的李-KNN 分类算法[J].模式识别与

人工智能,2014,27(02):173-178.

[43]何晴,郭捷.非重叠域行人再识别算法研究[J].信息技术,2018,42(07):34-38.

[44]梁书彤. 基于 GPS 数据的个体出行方式识别方法研究[D].北京建筑大学,2020.

[45]叶君炎. 异构环境下 MapReduce 任务推测执行算法研究[D].湖南大学,2016.

[46]刘欢.基于随机森林算法建模的糖尿病预警系统设计与实现[D].山东师范大学,2019

[47]Otok B W , Musa M , Purhadi, et al. Propensity score stratification using bootstrap aggregating classification trees analysis[J]. Heliyon, 2020, 6(7):e04288.

[48]郭东锋,刘新民,姚忠达,舒俊生,胡海洲.基于机器学习的焦油预测模型研究[J].安徽农业大学学报,2015,42(03):473-477.

[49]艾上杰. 基于机器学习的 COX-2 抑制剂虚拟筛选方法研究[D].海南大学,2018.

[50]张宁. 基于模糊模型的区域线损预测研究[D].燕山大学,2018.

[51]李培畅. 肌肉疲劳进程中的超声信号衰减特性研究[D].陕西师范大学,2015.

附 录

作者在读期间发表的学术论文及参加的科研项目

参加的科研项目：

- [1] *智能割草机 WiFi 联网控制系统研发，横向课题
- [2] *智能健康蒸汽房系统研发，横向课题

专利：

- [1] 一种咖啡豆升温率平滑可控及烘焙状态自动识别方法. 中国专利,授权公告号:
CN112056591A 申请公布日：2020.12.11