

杭州电子科技大学

硕士学位论文

题 目： 带锯床在线监测系统的设计与实现

研 究 生 蔡坤

专 业 电子与通信工程

指导教师 陈科明 副教授

企业导师 曾佳

完成日期 2022 年 5 月

杭州电子科技大学硕士学位论文

带锯床在线监测系统的设计与实现

研 究 生： 蔡坤

指导教师： 陈科明 副教授

2022 年 5 月

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

Design and implementation of on-line monitoring system for band sawing machine

Candidate: Cai Kun

Supervisor: Associate Prof. Chen Keming

May, 2022

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：

蔡坤

日期：2022 年 5 月 20 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：

蔡坤

日期：2022 年 5 月 20 日

指导教师签名：

陈利明

日期：2022 年 5 月 20 日

摘 要

在智能制造的背景下，锯床行业寻求智能化解决方案实现产业升级。目前国内生产的锯床普遍存在以下问题：不能远程获取设备的实时运行状态、锯切参数；锯床在工作过程中，无法获知带锯条磨损状态，很难预测锯条的剩余寿命；实时监测数据量过大，可视化绘制实时性不高，并且数据处理开销大。

为了实现远程数据获取和实时监测，本文设计和实现了一套基于物联网技术的带锯床在线监测系统：通过自主研发的智能网关，向下对接锯床控制器，实时收集终端设备采集的参数信息，向上通过 Wi-Fi 上传到云平台进行处理分析。云平台将处理的数据分发给 Web 端和小程序，提供可视化展示、远程控制等功能。该系统实现了在线监测、远程运维、故障预警等实际功能需求。

为了进行带锯条剩余寿命预测和提高锯条磨损识别精度，本文提出了一套基于自适应噪声完备集合经验模式分解（Complete Ensemble Empirical Mode Decomposition with Adaptive Noise, CEEMDAN）和改进人工蜂群（Artificial Bee Colony, ABC）算法的支持向量机（Support Vector Machine, SVM）锯条磨损识别和预测方案。将采集到的切削力、振动信号进行 CEEMDAN 分解，选取峭度值作为特征向量，通过改进的时序混淆 ABC 算法优化 SVM 模型参数 c 和 g 。测试表明，改进算法可以将磨损识别准确率提高到 94.01%，大于标准算法的 87.88%，满足需求设计。

为了降低海量监测数据的处理开销、优化可视化绘制性能，本文提出了一套基于改进的最大三角形三段（Large Triangle Three Buckets, LTTB）算法的降采样方案。为了采样后尽可能保持原始数据曲线特征，通过基于距离和离散程度的自适应选取分段点的改进设计，减少 LTTB 因平分各段导致的“尖峰”丢失问题。实践表明，降采样十倍后，改进算法对尖峰点的覆盖率仍有 87.4%，比标准算法提高了 13.3%。SVM 前期数据处理速度提升了一个量级，且提取的峭度值相对误差在 6% 以下。结合 ECharts 的实例拆分优化设计，可视化绘图操作的耗时减小数十倍，内存占用减小超过 30%。该方案大幅降低了锯床监测数据的处理开销，优化了数据可视化的性能和内存占用。

关键词：带锯床系统、剩余寿命预测、支持向量机、人工蜂群、降采样、数据可视化

ABSTRACT

In the context of intelligent manufacturing, the sawing machine industry seeks intelligent solutions to achieve industrial upgrading. At present, domestic sawing machines generally have the following problems: the real-time operating status and sawing parameters of the equipment cannot be remotely obtained; during the working process of the sawing machine, the wear state of the band saw blade cannot be known, and the remaining life of the saw blade cannot be accurately predicted; the amount of real-time monitoring data is large, Visual rendering is not very real-time and has high processing overhead.

In order to achieve remote data acquisition and real-time monitoring, this thesis designs and implements a set of online monitoring system for band sawing machines based on the Internet of Things technology: Through the self-developed intelligent gateway, it connects downward to the sawing machine controller, real-time collects the parameter information collected by terminal equipment, upward through Wi-Fi uploads to the cloud platform for processing and analysis. The platform distributes the processed data to the web and mini programs, providing interactive functions such as visual display and remote control. The system realizes the actual needs of online monitoring, equipment remote operation and maintenance, fault warning and so on.

In order to predict the remaining life of the band saw blade and improve the recognition accuracy of saw blade wear, this thesis proposes a complete set of Support Vector Machine (SVM) saw blade wear identification and prediction scheme based on Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and improved Artificial Bee Colony (ABC) algorithm. The collected cutting force and vibration signals are decomposed by CEEMDAN, and the kurtosis value is selected as the eigenvector, and the SVM model parameters c and g are optimized by the improved time series obfuscation ABC algorithm. Tests show that the improved algorithm can increase the wear recognition accuracy to 94.01%, which is greater than 87.88% of ABC algorithm, which meets the needs of the remaining life prediction of the band saw blade.

In order to reduce the processing overhead of massive monitoring data and optimize the visualization rendering performance, this thesis proposes a set of downsampling methods based on the improved Large Triangle Three Buckets (LTTB) algorithm. In order to maintain the original data curve characteristics after sampling, the improved design of adaptively selecting segment points based on distance and discrete degree reduces the problem of "spike" loss caused by bisected segments of LTTB. Practice show that after downsampling ten times, the improved algorithm still

covers 87.4% of peak points, which is 13.3% higher than LTTB. The data processing speed in the early stage of SVM is improved by an order of magnitude, and the relative error of the extracted kurtosis value is less than 6%. Combined with the instance splitting optimization design of ECharts, the time-consuming of visual drawing operations is reduced by dozens of times, and the memory usage is reduced by over 30%. This solution greatly reduces the processing overhead of sawing machine monitoring data, and optimizes the performance and memory usage of data visualization.

Keywords: Band saw system, Remaining life prediction, Support Vector Machine, Artificial Bee Colony, Downsampling, Data visualization

目 录

摘 要.....	I
ABSTRACT.....	II
目 录.....	IV
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 锯床及锯床系统发展现状.....	2
1.2.2 剩余寿命预测研究现状.....	3
1.2.3 海量数据处理研究现状.....	4
1.3 本文研究内容和章节安排.....	5
第 2 章 带锯床在线监测系统设计.....	7
2.1 系统总体设计.....	7
2.1.1 系统需求分析.....	7
2.1.2 系统总体框架.....	8
2.2 网关设计.....	9
2.2.1 网关的通信方式和协议.....	9
2.2.2 网关硬件设计.....	10
2.2.3 网关软件设计.....	11
2.3 应用层设计.....	12
2.3.1 云平台设计.....	12
2.3.1.1 设备云平台.....	12
2.3.1.2 业务云平台.....	12
2.3.2 Web 端设计.....	13
2.3.2.1 架构设计.....	13
2.3.2.2 业务功能设计.....	14
2.3.2.3 兼容性适配.....	16
2.3.3 小程序设计.....	16
2.4 系统测试和效果.....	17
2.4.1 通信延迟效果.....	17
2.4.2 设备和网关实物展示.....	18

2.4.3 应用层效果展示.....	20
2.4.3.1 云平台部分效果展示.....	20
2.4.3.2 Web 端部分效果展示	20
2.4.3.3 小程序部分效果展示.....	23
2.5 本章小结.....	23
第 3 章 基于 CEEMDAN 和改进 ABC 的 SVM 锯条剩余寿命预测方案.....	24
3.1 锯条磨损分析和定量.....	24
3.1.1 锯切运动分析.....	24
3.1.2 磨损监测信号分析.....	25
3.1.3 数据集和磨损定量.....	26
3.2 SVM 模型预测方案分析	28
3.2.1 SVM 和 SVR.....	28
3.2.2 优化方向分析.....	30
3.2.2.1 自适应时频分析.....	30
3.2.2.2 寻参优化算法.....	32
3.3 基于 CEEMDAN 和改进 ABC 的 SVM 锯条剩余寿命预测设计与实现.....	32
3.3.1 总体方案设计.....	32
3.3.2 磨损监测信号的选择.....	33
3.3.3 基于 CEEMDAN 的信号处理和特征提取.....	35
3.3.3.1 CEEMDAN 分解算法.....	35
3.3.3.2 磨损信号的 CEEMDAN 分解实现.....	38
3.3.3.3 分量的特征提取和选择实现.....	39
3.3.4 基于改进 ABC 的 SVM 回归预测算法.....	42
3.3.4.1 SVM 的核参数优化.....	42
3.3.4.2 ABC 算法及其改进	42
3.3.4.3 基于改进 ABC 的 SVM 回归预测算法实现.....	45
3.4 实例测试和结果分析.....	45
3.4.1 模型评估指标.....	45
3.4.2 实例测试.....	46
3.4.3 结果对比分析.....	48
3.4.3.1 改进 ABC 算法寻参的对比分析	48
3.4.3.2 模型识别结果的对比分析.....	49
3.5 本章小结.....	51
第 4 章 基于改进 LTTB 的海量数据处理的优化方案.....	52

4.1 LTTB 算法的方案分析.....	52
4.1.1 LTTB 算法步骤.....	52
4.1.2 LTTB 算法的问题和改进方向.....	54
4.2 基于改进 LTTB 算法的海量数据处理的优化方案设计与实现.....	54
4.2.1 总体设计流程.....	54
4.2.2 改进的 LTTB 算法设计.....	55
4.2.3 ECharts 的数据可视化优化设计.....	57
4.2.3.1 内部渲染优化.....	57
4.2.3.2 实例拆分优化设计.....	59
4.3 实验测试和结果分析.....	60
4.3.1 改进的 LTTB 算法降采样测试.....	60
4.3.2 降采样前后的 ECharts 绘图测试.....	61
4.3.3 降采样前后的 SVM 模型数据处理测试.....	63
4.4 本章小结.....	65
第 5 章 总结与展望.....	66
5.1 工作总结.....	66
5.2 工作展望.....	66
参考文献.....	68
致 谢.....	72
附 录.....	73

第1章 绪论

1.1 研究背景和意义

从工业生产总值来看，我国虽然已经成为世界第一工业大国，但在很多场景下仍然存在着工业精密化、数字化、智能化程度不高的问题。机床行业也是如此，我国机床产值在 2010 年左右已经位列世界第一，但是中高档数控机床设备仍然依赖国外。

金属锯切工艺是零件加工的重要一环，锯床设备使用范围很广，比如航空、机械、汽车轮船等重工业领域^[1]。因此，发展国产锯床装备产业是我国机床发展过程中的至关重要的任务。

目前国内主要使用的是国产金属带锯床，我国在 1957 拥有了自主生产金属锯床的能力，此后技术水平不断发展，国产金属带锯床越加精密、高效^[2]。

但是国内锯床行业很多还是半自动化，技术创新能力较弱。而国际领先的企业如贝灵格等已基本实现数控化，正在进一步实现智能化。而国内的锯床行业，如浙江省缙云县锯床装备产业群，其生产的锯床在国内很有名气，但是仍有许多不足，比如信息化程度低、产业集群大而不强^[3]。

所以，目前阶段的任务就是要实现锯床设备自动化、数字化，并进一步走向智能化。其中的重要一步就是实现设备状态监测（Equipment Condition Monitoring, ECM）系统，并在此基础上有效挖掘监测数据进行，进行智能化改造。比如采用传感器实现关键数据的监测和上报，实现故障检测、预警、远程运维等预测性维护的功能。

近年来，国内外学者和企业也对此做了大量研究工作，然而目前锯床行业的带锯床监测系统仍然存在一些可以改进的地方：

- (1) 目前实现的带锯床监测系统功能并不十分完善，比如视频监控、大屏设计、丰富的用户侧交互产品等。另外由于锯床行业集约化程度不高，网关、Web 界面、数据云平台等没有形成统一的标准，对锯切数据没有很好地利用起来。
- (2) 锯床行业正在积极探索智能化改造，比如故障诊断等，但是对于带锯条的磨损和健康状态研究不多。带锯条是有损耗的关键部件，但其健康状态无法直接有效获得，难以准确识别其磨损状态，这不利于企业生产和设备维护。
- (3) 状态监测系统的数据量十分庞大和多变，造成云平台的数据处理效率不高、Web 系统界面的可视化展示性能也较低，因此需要对海量时序监测数据的处理进行优化。

因此，本论文根据上述的一些问题进行改进优化，实现了一套功能完善的带锯床在线监测系统，来满足工厂使用者和管理员对带锯床设备的数据监测、远程运维、数据分析等需求，并在此基础上，着重对急需解决的痛点进行研究：

- (1) 研究带锯条剩余寿命预测的方案,在完善状态检测系统的基础上,从锯床设备的历史运行状态数据中,提取特征信息,识别锯条的磨损状态,进而预测锯条剩余寿命。这对于提高锯条利用率、降低加工和维修成本、实现锯床设备智能化改造有着重要意义。
- (2) 研究海量数据处理的优化方法,在保持原始数据曲线的细节特征的前提下,提高处理海量数据的效率。这对于提高剩余寿命方案前期数据处理的效率、优化数据可视化绘制性能有着重要意义。

1.2 国内外研究现状

1.2.1 锯床及锯床系统发展现状

锯床设备发展至今,目前比较常用的锯床设备有三类,如图 1.1 所示,依次是:弓锯床、圆盘锯床、带锯床。



图 1.1 三类锯床

其中,弓锯床因为速度太慢导致效率过低,圆盘锯床因为锯片太厚导致浪费材料,缺点明显。而带锯床切削效率高、材料损耗小、适用的材料范围广,优点明显,是目前的锯床设备采取的主要形式。但它存在有锯条损耗、锯切精度受锯条振动约束等缺点,这需要在人工使用中加以注意。

带锯床有立式和卧式,其中卧式带锯床较多,目前市面上最常见的是图 1.1 所示的卧式双立柱金属带锯床(右 1)。它拥有速度快、效率高、液压系统灵活、双立柱结构操作稳定等优点,广受加工厂商欢迎,下文所述的带锯床均默认为此类锯床。

近年来针对带锯床设备的智能改造,主要集中于系统监测和故障诊断。在国内,孙荣霞等人使用 DS18B20 温度传感器、A3144 霍尔传感器等采集监测信号,基于 TC35 通讯模块实现对锯床的远程监测以及系统故障报警^[4]。刘佳采用移动通信网络实现视频通信,增强了带锯床在线监测系统的交互功能^[5]。刘腾发同样在 TC35 通信模块的基础上,采用多种传感器实现在线监测与故障诊断^[6]。

国外的带锯床产业发展较早,不局限于故障诊断和研究,开始关注于刀具的磨损研究。Potočník 等人通过多传感器振颤检测,基于对切削力、机床振动和发出声音信号的分析,检测切削状态中的颤振现象,以提高锯切面质量^[7]。Turchetta 等人通过研究圆锯在加工花岗岩过

程中切削力在三个空间方向上的分量变化，来研究锯片的磨损情况^[8]。

在线监测系统需要与时俱进，通过对多组采集的传感信号提取特征，借助机器学习、大数据分析等现代化工具，可以对锯床一些无法直接观察得到的状态进行归纳和总结，挖掘出重要的有用信息，比如故障状态、健康状态等，实现锯床的智能化改造，这也是本文的研究重点。

1.2.2 剩余寿命预测研究现状

工业设备进行预测性维护的流程图^[9]，如图 1.2 所示。在数据处理分析的基础上，进行异常检测、故障识别、寿命预测。其中一个核心步骤就是对设备的关键耗损部件进行寿命预测，在此基础上调整维修保养决策，从而实现健康、高效生产。带锯床设备的锯条部件也是如此，其中的关键是带锯条的剩余寿命预测（Remaining Useful Life, RUL）。

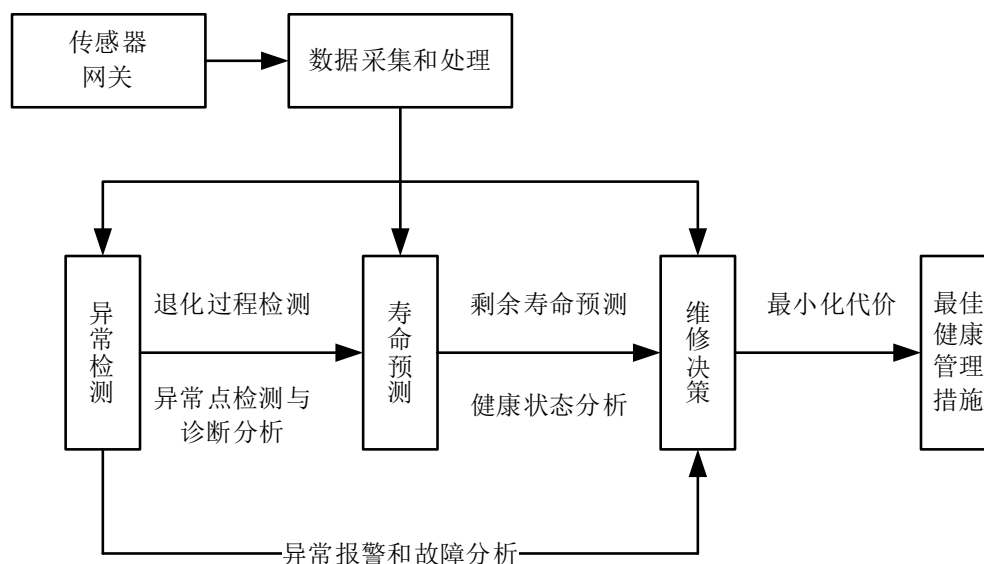


图 1.2 异常检测、寿命预测、维修决策

目前比较常用的设备部件剩余寿命预测方法很多，姑且可以分为“基于故障物理”和“基于数据驱动”的方法。

以刀具为例，基于故障物理的方法适用于部件有可见的性能退化现象，该方法以刀具的健康状态识别为目的。比如使用显微镜微观观测刀具的磨损量；使用深度学习和图像识别分析刀具表面磨损。但对于锯床设备来说，关键设备即带锯条的寿命情况通常难以被具体感知的，尤其在带锯条高速运行阶段，很难进行有效的观测。

基于数据驱动的方法，是指对产品退化过程中产生的退化数据，基于统计、退化建模或机器学习的方法对其进行分析和预测。带锯条的研究显然更适合采用这种方法，通过对一些关键锯切数据指标的统计和分析来进行寿命预测。而数据方法又可以大致分为三种。

基于传统统计的方法，比如统计出一千根锯条从开始到报废的过程，就容易统计出一根锯条的失效时间，用以推测接下来的损坏趋势。这时可以用总切削时间、总切削面积作为统

计量来计算失效点。但是,切削过程复杂、加工环境恶劣,锯条的状态变化非线性、随机性大,传统根据线性的失效时间、失效点来预测寿命的方法可能会效果不佳。

基于退化模型,是试图捕捉退化过程数据的潜在失效过程。这种方法不要求样本量且能够把实际现场的不确定性因素考虑在内。Le Son 等通过非均匀伽马过程对退化建模,用 Gibbs 采样对退化的随机性做近似,估计系统的剩余使用寿命分布^[10]。李烁使用逆高斯过程模型研究退化,用蒙特卡洛方法估计复杂模型参数,给出了进应力退化试验的寿命分布评估方法^[11]。但是,随机退化过程的分析假设零件的退化符合某种随机模型,比如上述的逆高斯过程、伽马过程等,模型的选择、何种退化参数符合模型都是不确定的,需要大量的试验来验证,不适合锯切过程。

最后是基于机器学习的数据分析方法。近年来机器学习和深度学习使用愈加广泛,在 RUL 预测上进行了一些研究应用,其中典型的是,人工神经网络(ANN)和支持向量机(SVM)。

神经网络使用隐藏层和输出层两层网络,以大量神经元进行连接交互。使用的损失函数大多是平方损失,可以很好地反映非线性特性,适用于切削过程。Kalidass 使用模拟退火算法预测工具磨损,建立并优化工艺参数的神经网络,将过程参数与工具磨损相关联,以确保优化^[12]。然而,ANN 等神经网络模型更适合大量训练样本的场景。根据锯切现场的工作情况看,虽然锯切数据庞大,但是一根带锯条大概只能锯切百刀,且实际现场提供的可观测刀具样本并不多。所以,虽然神经网络模型可以有效捕捉数据序列的长期依赖关系,但对于样本有限的切削过程 RUL 预测不完全适用。

SVM 算法由 V.N.Vapnik 在 1995 年提出^[13],回归形式是支持向量回归(Support Vector Regression, SVR)。在小样本情形下 SVM 模型也能表现良好,非常适合用来解决模式识别、故障诊断、预测分析等问题^[14-17]。因此也适用于识别切削过程的非线性信号特征,如电子科技大学的吴昊使用声发射信号提取时频参数和 EMD 分解的 IMF 分量作为特征,建立 LS-SVM 刀具磨损识别模型^[18]。Jingchao Guo 等人研究铣削信号的大小波动与刀具磨损的关系,采用多重分形趋势波动分析法计算特征向量的频谱参数作为输入,使用 SVM 进行刀具状态诊断研究^[19]。

不过, SVM 算法也存在问题:核函数的选取没有很好的选取准则,尤其 SVM 的精度表现对惩罚参数比较敏感,只能步步试错才能逐步调优。这是需要着重研究和解决的。

1.2.3 海量数据处理研究现状

在当今的信息化时代,信息量过大已成为我国各行各业所必须面对的问题。如何在庞大的信息中寻求对企业或者个人有用的信息来推动经济的发展,已成为我国学者所共同关注的问题。同样在本文的工业场景下,监测数据往往更新频率高,每次更新的数据量较大,产生的数据集非常庞大,这些数据在使用时存在效率和性能低的问题。

针对这个问题,国内外主要是研究海量数据下的大容量存储和访问优化问题。比如,Google 设计并实现了面向大规模数据处理的 MapReduce 编程模型,能够同时让内部员工在大

规模集群上开发各种数据密集型应用，目前 MapReduce 每天处理的数据量超过 20PB，以 MapReduce 为核心的 Hadoop 技术广泛应用于 Web 网页索引、文本处理和数据挖掘等领域。针对数据密集型应用中数据规模庞大、难以管理的问题，学术界提出了 MonetDB 等基于列存储的分析型数据库管理系统，和一定的缓存与索引技术相结合，MonetDB 可以在不影响其他查询效率的前提下，使得分析型查询的效率提高 10~100 倍。Singh Amritpal 等人设计了可调节布隆过滤器，优化海量数据的存储和查询性能^[20]。Mujahid Taha 等人的研究通过使用 GPU 加速数据处理算法^[21]。Kun Zheng 等人针对边缘计算中的实时海量矢量场数据的高速度、高体积特点，提出一种基于边缘计算的数据流化调度，以减少数据块量和输入-输出的延迟^[22]。

国内的胡素娟基于云计算技术实现分布式海量数据处理系统^[23]，杨晓云主要关注于通过分布式储存层进行海量数据存储的优化^[24]。卢民荣等人研究海量数据的压缩分析，根据极点的特殊性进行特征识别，按交叉区间、趋势明显和趋势不明显改进基于趋势自适应处理的极点提取算法^[25]。

海量数据处理的优化比较复杂，上述方法大多是改进数据结构，对海量数据的存储、查询进行性能优化，没有从根本解决数据量级大的问题，反而加大了对数据的前期处理压力。为此近年来降采样方法逐渐兴起，可以从根本上把数据数量级降低。采样方法类似于上文的压缩方法，主要也是通过极点的识别来确定采样点。不过需要注意的是，对时序数据进行降采样处理，同时必须尽可能保证原始数据的波动细节、趋势、形状，具体量化来说就是序列的统计特征（如均值、极值、方差）以及形态特征（如距离、角度、斜率）等，不然就失去了原始数据的物理意义。

1.3 本文研究内容和章节安排

本文针对带锯床在线监测系统的研究现状和问题，并结合企业实际生产情况，探索锯床监测系统的改进和优化。

对带锯床在线监测系统的功能进行设计和完善，增加数字大屏、视频监控、工业终端设备 PDA 等，实现对生产的整个流过程的全面监测，提供多端全方位的交互产品。同时，针对目前比较缺失的带锯条寿命预测内容加以研究，主要是研究一种基于 SVM 的锯条磨损识别模型，进而对带锯条进行寿命预测。另外，针对监测数据量级过大导致的性能问题，研究对海量数据处理的优化，重点是探索更加适应工业环境的降采样算法，从根本上解决问题，降低算法处理的开销，优化监测系统的可视化绘制能力。

全文共五章，内容安排如下：

第 1 章 绪论。引出研究背景，阐述研究带锯床在线监测系统及其关键技术意义；接着综述了带锯床在线监测系统、剩余寿命预测检测技术、海量数据处理优化方案的研究和分析，分析目前的可行方案；并结合实际锯切场景，确立论文的研究任务以及研究方法。

第 2 章 带锯床在线监测系统设计与实现。本章完成一套功能完善的锯床在线监测系统，实现系统的在线化、数据化。按照系统开发流程，先进行业务需求分析和系统总体设计，然

后对通信协议、网关、应用层等模块进行详细的设计和实现，最后进行系统各部分的效果展示。

第3章 基于 CEEMDAN 和改进 ABC 的 SVM 锯条剩余寿命预测方案。首先从锯切现场出发，分析锯条磨损的因素，并确定磨损状态的定量表达式。接着，针对磨损模型建立的各个环节的实际问题进行分析和改进，包括传感信号的选择、振动信号的 CEEMDAN 分解、改进的人工蜂群算法用于优化 SVM 参数 c 和 g 。最后对比未优化模型，测试和验证设计的优越性。

第4章 基于改进 LTTB 的海量数据处理的优化方案。对锯床实时监测过程中产生的海量时序数据处理进行优化，阐述传统 LTTB 算法的原理和不足，并提出基于距离和离散程度的改进设计。然后就海量数据可视化展示的难点加以研究和解决。最后加以测试，验证改进 LTTB 算法的效果，并应用于第三章 SVM 的前期数据处理、ECharts 可视化绘图的实践中。

第5章 总结与展望。对带锯床在线监测系统设计与实现及其创新成果进行总结，并分析设计的不足和作出展望。

第2章 带锯床在线监测系统设计

针对当前锯床产业数字化、智能化改造的实际需求,且不能及时掌握锯床工作运行状态、锯条健康状态不明确、数据量过大、数据可视化展示能力差等一系列问题,本文设计并实现了一套基于物联网技术的带锯床在线监测系统。通过定制开发的物联网网关与锯床控制器(PLC)对接,实时获取传感器的采集数据,并通过无线通信技术上报到物联网云平台。云平台进行数据处理分析,将锯切力等关键设备数据传输给 Web 端展示和交互,还可以通过小程序、短信通知等方式提供故障预警提醒,便于实现对锯切设备、锯切过程的在线监测和远程控制。

2.1 系统总体设计

2.1.1 系统需求分析

针对锯切过程和实际应用中金属带锯床系统存在的缺陷,本文设计的带锯床在线监测系统,需要满足以下几点需求:

- (1) 实时监测,锯床设备便捷组网入网,通过网关将采集的传感数据实时上报,实现数据的数字化、在线化。比如锯切力、锯偏值等数据。
- (2) 联网稳定,网关与 PLC、网关与云平台的连接方式要合理、稳定,尤其是与 PLC 的连接本就不太稳定,需要保证较小的通信延迟。
- (3) 远程控制,操作用户侧页面可以对监测设备下发指令,网关解析指令后下发给 PLC,由 PLC 对锯床设备进行控制,比如锯切模式、开关机等。
- (4) 系统运维,对设备故障进行预警设置,通过短信提醒管理人员,并提供设备的保养维护管理。
- (5) 数据可视化和分析,对锯切过程的关键数据,提供大屏展示、图表分析等功能,直观展示企业生产状况。并提供设备历史数据的查询展示和可视化图表分析。
- (6) 完善的管理平台,提供各种交互产品。如 Web、小程序,让用户可以便捷地远程操作设备和设备数据。

除上述需求分析之外,后文需要着重研究的有:

- (1) 带锯条健康状态预测是重难点问题,需要对数据库保存的切削数据,辅以机器学习和数据分析的技术,对历史数据进行分析,尤其针对锯条寿命问题,能够提出预测和维护建议;
- (2) 数据可视化能力加强,针对于实时监测的庞大数据进行性能优化,保证提供的可视化产品交互流畅、直观美观、信息明确。

2.1.2 系统总体框架

根据以上需求，系统的总体框架设计如图 2.1 所示。整个系统由设备硬件、网关、云平台 and 智能应用共同组成。物联网通常可以看作是三层结构，其中，设备硬件、传感器采集部分、网关合为感知层，网关与上、下的通信桥梁是网络层，云平台、智能应用则组成了应用层。

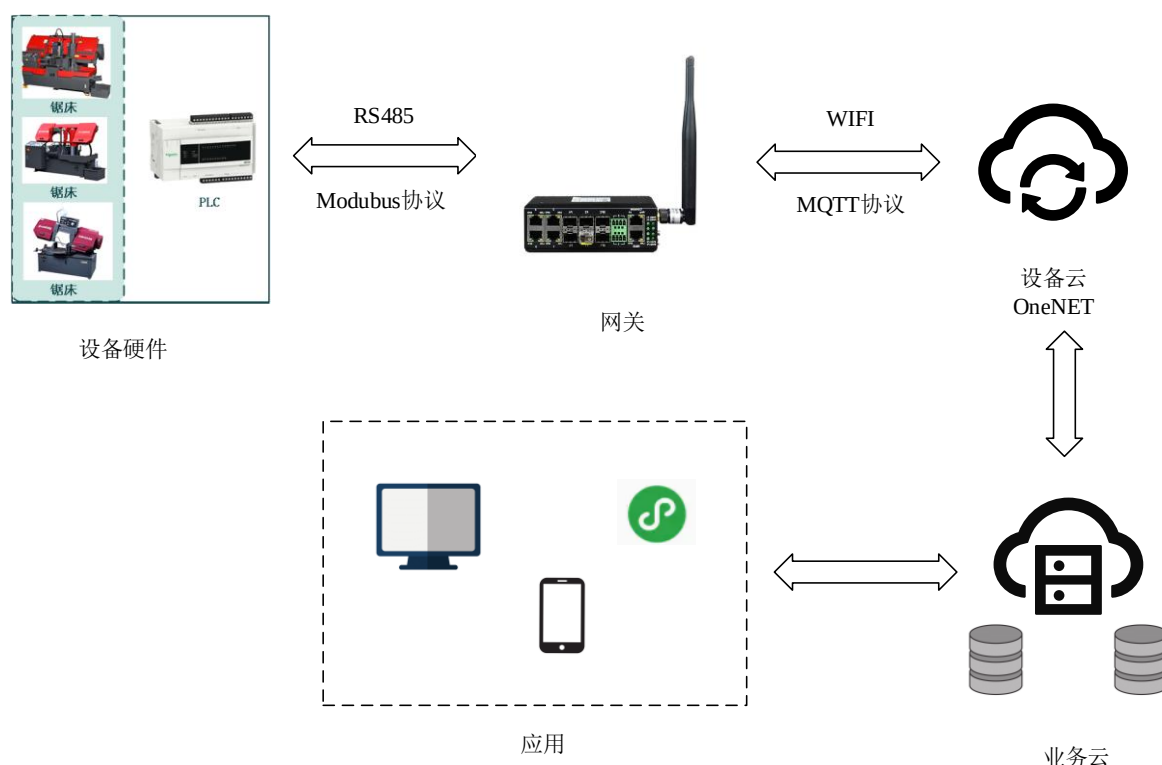


图 2.1 系统总体框架

设备硬件部分主要包括锯床、数据采集传感器、PLC 控制器、操作触摸屏等。其中 PLC 控制器最为重要，负责调控各传感器的采集数据并交付给网关，与网关通讯接收远程指令，解析后对设备进行控制。

网关是数据上传传输的桥梁。向下对接设备 PLC，采用 RS485 总线连接和 Modbus 通信协议，实时传输 PLC 采集数据。向上对接设备云平台，采用 Wi-Fi 方式和 MQTT 通信协议，把下挂设备的 PLC 数据传输到设备云平台。同时把云端传来的某些数据解析指令，下发给下挂设备 PLC，进行设备的远程控制。

云平台、智能应用是系统的应用层。设备云对接物联网网关，负责接收 PLC 数据，企业内建的业务云平台将这些数据进一步地处理，并提供给人机交互应用如 Web 端、APP、小程序使用，通过列表、图表、地图等形式将各类设备信息和工作数据展示给用户。用户在智能应用端的某些交互操作，再以数据指令的形式反馈到网关进行 PLC 控制。

由此完成对锯床的在线监测功能。其中，进行设计时需要着重考虑的有：

- (1) 网关的通信方式和能力。比如网关与 PLC 的连接方式和协议、是否有能力支持多 PLC 接入、与云平台的连接速度是否足够、以及一些新形式如视频监控的传输方式等。
- (2) 云平台的业务能力。设备云平台的选择、接入协议的选择、企业自建的业务云平台要符合业务需求。
- (3) 智能应用的交互性。比如 Web 系统、APP、小程序等，要尽可能为用户提供全方位的产品，以及注重用户体验，如设计感、可视化图表的交互性能等。

2.2 网关设计

带锯床在线监测系统的目的是为了实现锯床设备的在线化、数字化，物联网网关是整个系统中的关键，网关下连锯床 PLC、上接设备云平台，在二者之间起到数据获取、传输的桥梁作用，同时又是云端下发控制指令的返回通道。

2.2.1 网关的通信方式和协议

(1) 网关与下挂设备

RS485-Modbus 是常用的工业组网方式，网关和下挂设备之间，使用 RS485 有线连接，通信协议采用 Modbus。Modbus 协议由施耐德集团专门专为应用 PLC 所研发，代码开源，使用简单，应用广泛，已经是工业电子上标准的主-从式串行通信协议。

Modbus 区分主机和从机，本系统中，智能网关属于主机，下挂的锯床 PLC 属于从机。网关不断向从机发送读监控点的 Modbus 指令帧，帧数据采取 RTU 格式。RTU 格式的 Modbus 帧结构如图 2.2 所示。

地址 8位	功能码 8位	数据域 N×8位	校验位 16位
----------	-----------	-------------	------------

图 2.2 RTU 格式 Modbus 协议帧

上述的帧格式中，地址指的是下挂设备 PLC 的地址，大小 1 个字节。功能码是协商好的操作指令码，大小 1 个字节。数据域的值会根据功能码的不同有不同的格式，数据域值一般包含是否需要应答，是否有连续读取操作，连续读取数量等。校验方面，RTU 采用 CRC16 校验，高位在后，低位在前。

(2) 网关与设备云平台

根据本文锯床系统的实际情况，选用传输速率快的 Wi-Fi 方式进行连接，频段方面选择使用 2.4G 和 5G 双频，以增加接入设备的数量。通信协议采用 MQTT^[26]，MQTT 是一个发布-订阅式的轻量传输协议，其网络开销低，适用于嵌入式场景。

MQTT 的数据包格式由固定头、可变头、消息体构成，如图 2.3 所示。其中，固定头部标识数据包的类型和分组，消息体承载数据内容。

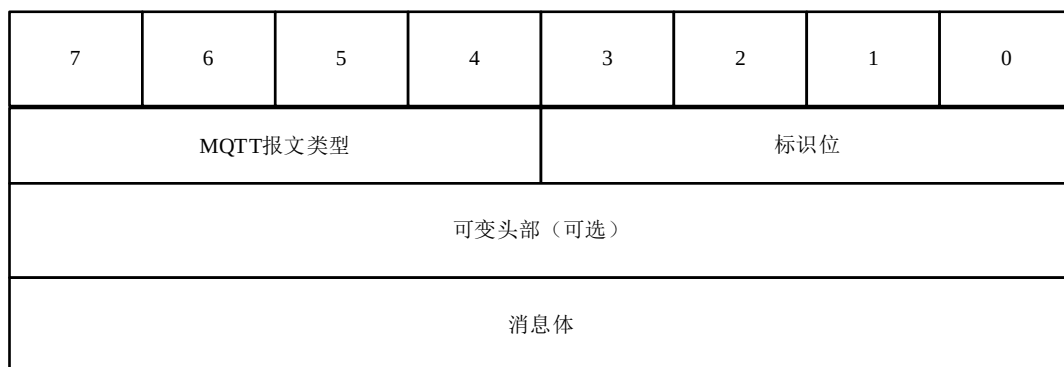


图 2.3 MQTT 数据包格式

可以看出，MQTT 数据包的头部相比于传统的 HTTP，开销很小，可以节省传输带宽，很适合物联网场景的终端设备。

2.2.2 网关硬件设计

网关硬件设计框图如图 2.4 所示。主要包括：MCU、Wi-Fi 模组、RS485 以及 LED、按键等外设。主控芯片以串口和 Wi-Fi 模组相连，配置 AT 命令来控制模组。再通过 RS485 和下挂设备进行异步传输通信，传感器检测模块的数据也通过串口传输给 MCU。MCU 还控制 LED 的亮灭，指示网关的联网状态和下挂设备的通信状态。按键模块可以对网关进行开关、重启复位操作和 AP 配网模式的选择。

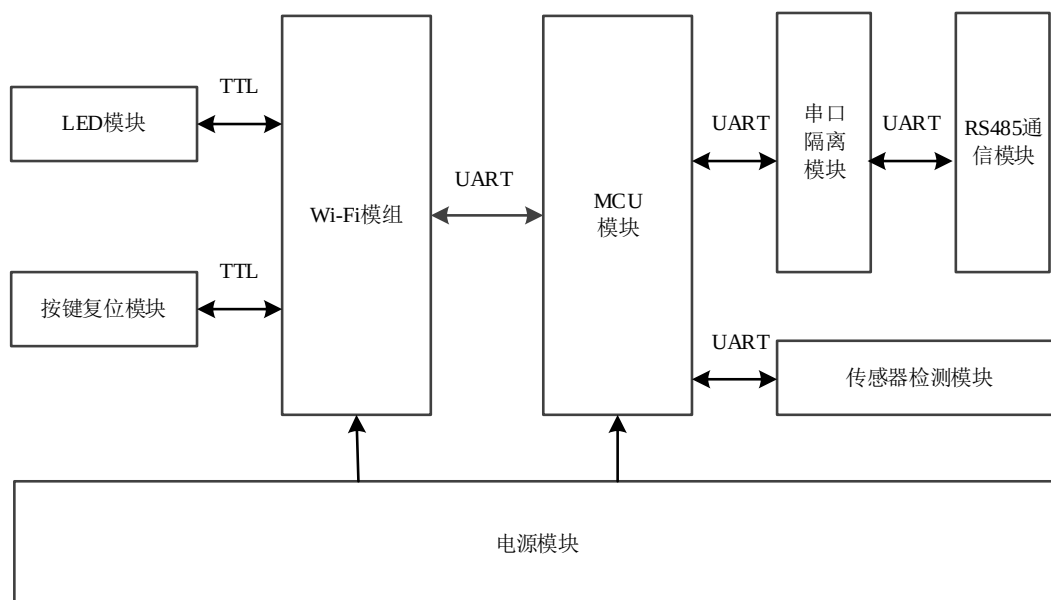


图 2.4 网关硬件设计框图

MCU：主控芯片选用 STM32F103RCT6，该芯片功能强、存储空间大、成本低，能满足本系统设计要求。

Wi-Fi 模组：通过串口与 MCU 相连，负责把数据通过 Wi-Fi 传输到设备云平台。选用 ESP8266 Wi-Fi 芯片，性能稳定、功耗较低。

RS485 模块：与下挂的锯床 PLC 设备相连，以接收锯床数据，可用于对接多个 PLC。

传感器检测模块：通过串口与 MCU 连接。用于采集锯切过程的重要数据，选用 ACC345 型号加速度计传感器采集锯条振动数据，选用 HYL0-010-2 称重传感器采集锯背压力数据来表征切削力。

电源模块：提供电源的管理和转换功能。

2.2.3 网关软件设计

网关运行过程中，需要处理多种不同优先级的任务，比如数据处理、上传、控制外设等。基于 OpenWrt 嵌入式系统，软件程序由 4 个线程实现，分别是：网络管理线程、下行的 RS485 数据收发线程、上行的云平台数据收发线程、数据处理线程。这些线程在操作系统的管理下，根据实时性和优先级大小进行调度，实现图中的网关功能。其中，网络优先级最高，上下行数据的接收和发送都需要设备处于联网状态。网络连接成功后，网关注册并连接到 OneNET 设备云服务器，而后进入业务数据处理，进行实时的数据收发。

网关软件的整体工作流程如图 2.5 所示。

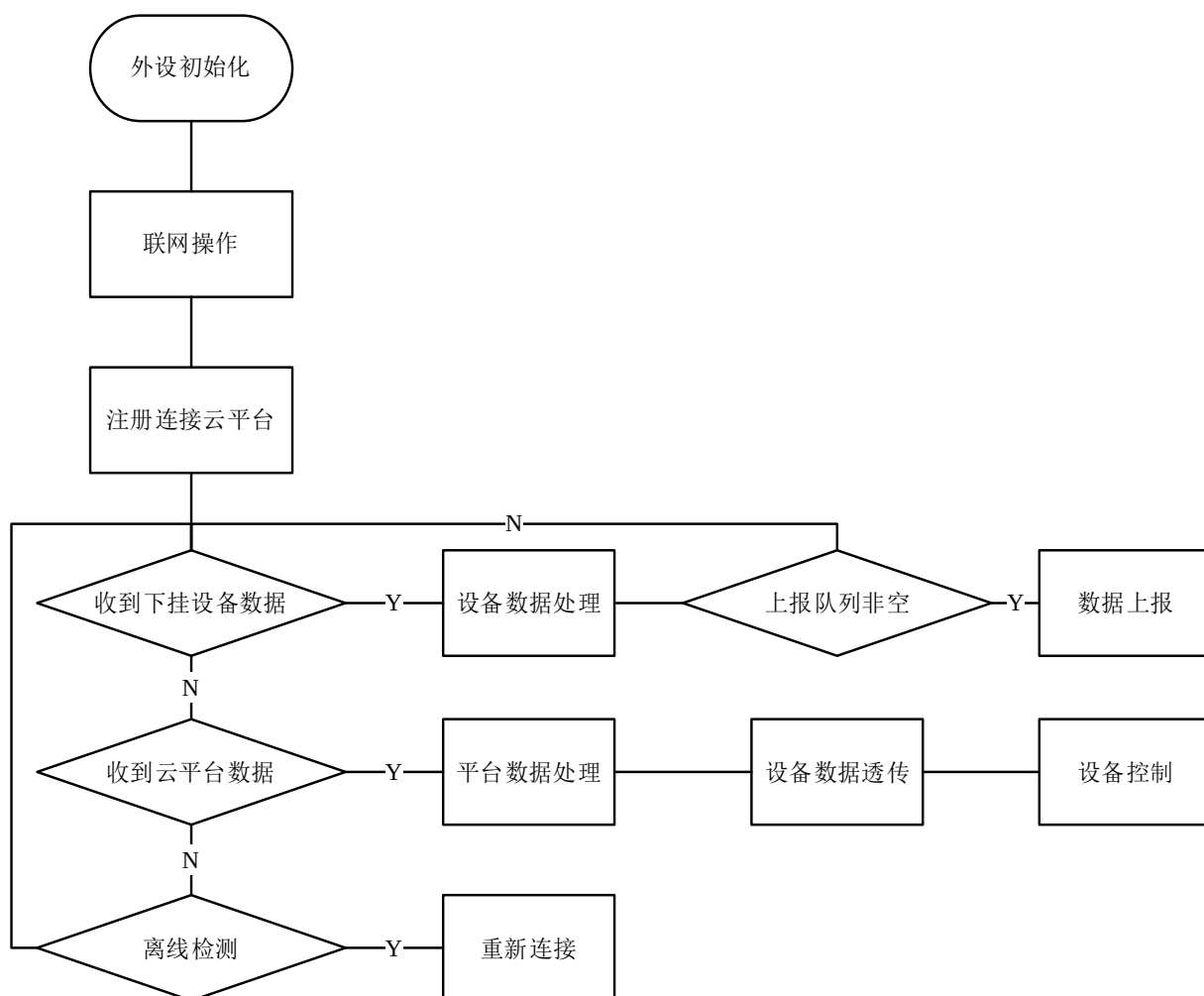


图 2.5 网关工作流程图

2.3 应用层设计

2.3.1 云平台设计

2.3.1.1 设备云平台

设备云平台是物联网网关经由基站直接相连的平台，本设计选择的是中国移动 OneNET 物联平台。

设备云的架构设计如图 2.6 所示。网关注册设备后，使用 Wi-Fi 方式与 OneNET 连接，并通过 MQTT 协议上传由下挂设备发来的锯切数据。业务云通过 HTTP 协议调用 OneNET 的存储数据进行处理。开发者还可以在 OneNET 平台进行在线调试、设置数据流的触发器，实现对网关设备的实时数据监控和报警设置。

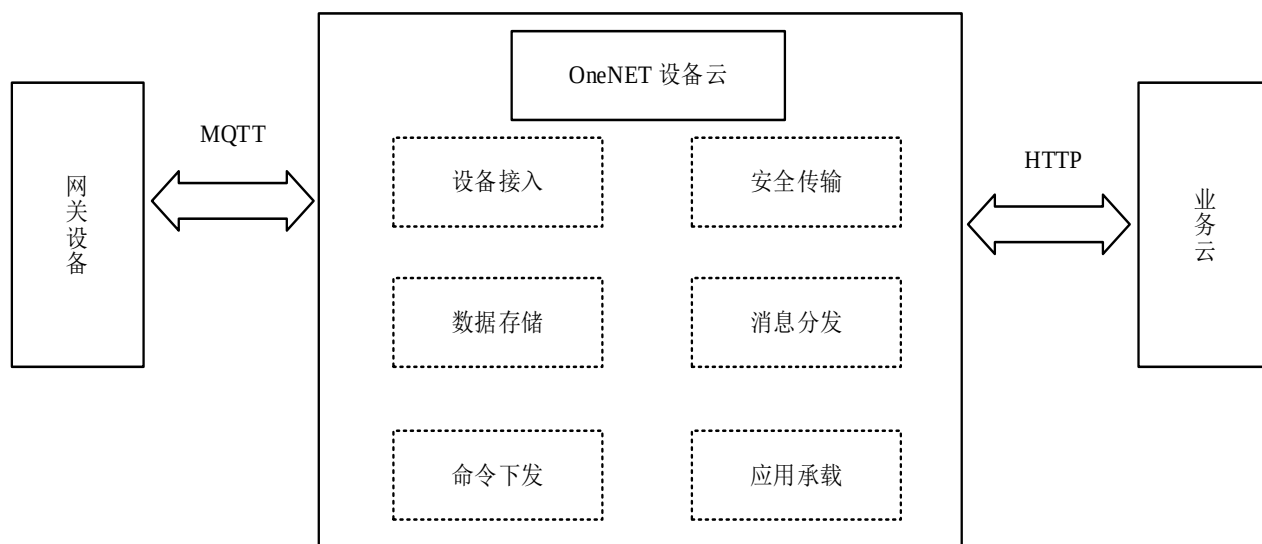


图 2.6 设备云架构

2.3.1.2 业务云平台

业务云的主要工作是调取设备云的锯床数据，根据企业内部的具体业务需求，对其进行处理、分析以及应用，业务云架构如图 2.7 所示。主要功能包括：

- (1) 设备云对外提供了接口调用，业务云可以调用数据接口获取锯切数据，使用 Redis 数据库存储；
- (2) 锯切数据的业务处理和分析功能，如锯切数据的统计、图表分析等；
- (3) 向开发者提供 RESTAPI 进行调用，使用 Nginx 进行负载均衡，将请求分发到 Tomcat 集群；
- (4) 调用设备云的 API，如报警推送，用于向用户发送短信提醒；
- (5) 使用 Docker 容器部署，降低计算开销，提高灵活性；
- (6) 为客户端系统的用户鉴权、人员权限等提供安全认证等。

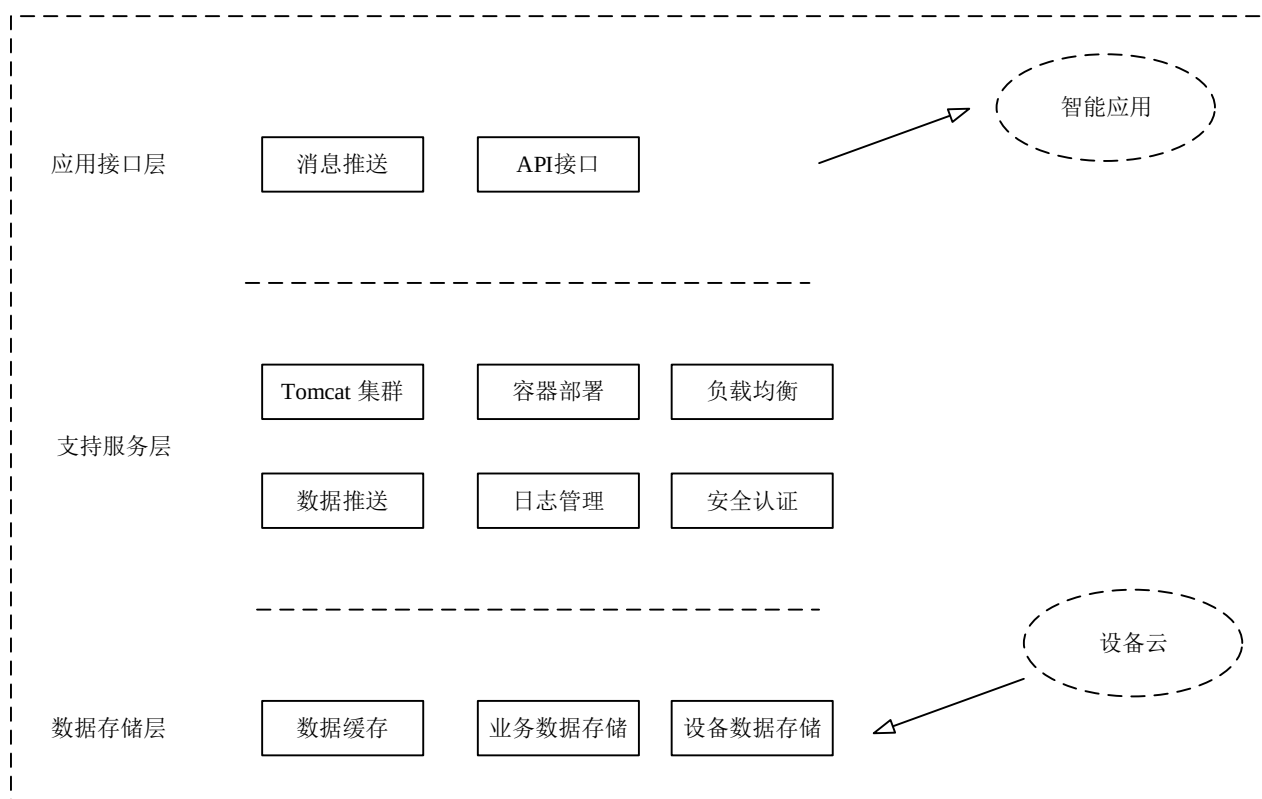


图 2.7 业务云架构

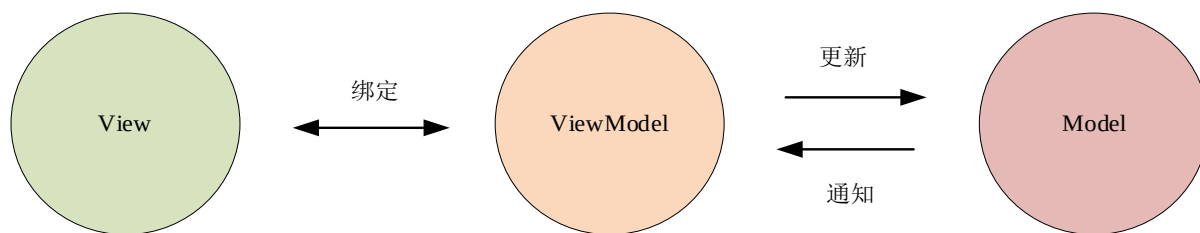
2.3.2 Web 端设计

带锯床在线监测系统是用户侧的最主要的产品，它的用户主要是锯床企业的人员、工厂设备的工作人员等。主要功能就是为锯床企业及用户提供生产过程的数字化管理，将云平台的重要数据以各种直观友好的形式提供给用户使用。

由于目前锯床行业趋于联合，所以本设计的 Web 端针对性地提供两个产品。一个是锯床在线监测和远程运维系统，供企业用户使用，是租户性质的前台应用；一个是锯床企业管理网站，由管理员来统一管理锯床企业，是租户的管理者。二者的设计和实现类似，只不过功能不同，前者使用的数据是锯床企业采集到的现场数据，后者只是提供对所有锯床企业的系统管理功能而已。所以本节主要介绍前者。

2.3.2.1 架构设计

Web 端的开发采用的是 Vue，它是一套用于快速开发界面应用的 JavaScript 框架，采用了 MVVM（Model-View-ViewModel）的架构设计，如图 2.8 所示。其中，Model 是数据层，它只关心数据本身，不关心数据如何操作和展示；View 是视图层，负责将数据转化为视图；ViewModel 为视图提供引擎，负责监听和处理数据以及数据到视图的映射。于是，View 和 Model 实现了解耦，也使得组件复用更加方便。



Vue 的核心库只聚焦 ViewModel 层的引擎功能，其余的生态功能需要使用第三方库进行集成。现代 Web 开发比较关注工程化能力，进行功能开发时，可以使用 NPM 对第三方的类库进行安装和管理，使用 Webpack 进行依赖包管理和配置管理，使用 VueCLI 脚手架快速部署开发环境。

最终，Web 端的框架设计如图 2.9 所示。以 Vue 为主体，集成多个三方组件库，包括：Vue-Router 管理路由、Pinia 管理数据状态、Element-UI 提供 UI 组件、Axios 用于收发 HTTP 请求、ECharts 开发可视化图表、Bmap 用于嵌入百度地图、Lodash 提供工具函数等，组成 Web 开发的架构。

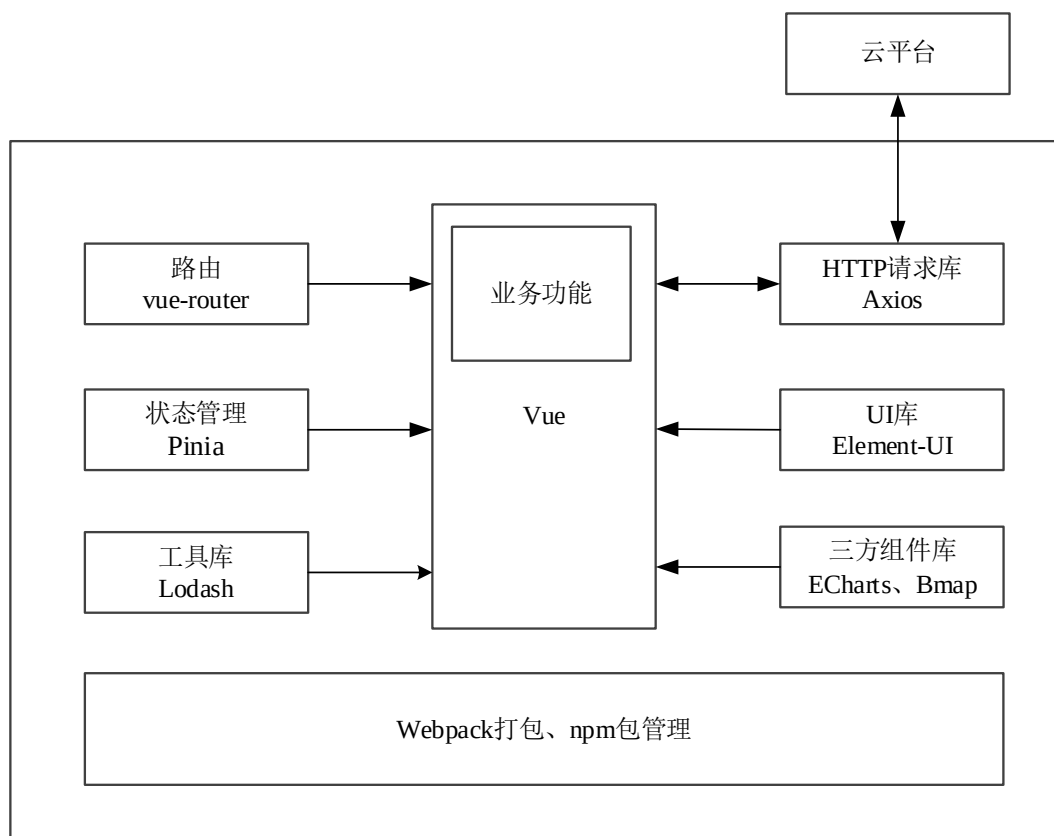


图 2.9 Web 端框架图

2.3.2.2 业务功能设计

在确定好设计框架后，根据系统的需求分析，带锯床在线监测和远程运维系统的主体功

能都是在 Web 端呈现给用户的。系统实现的功能模块如图 2.10 所示。

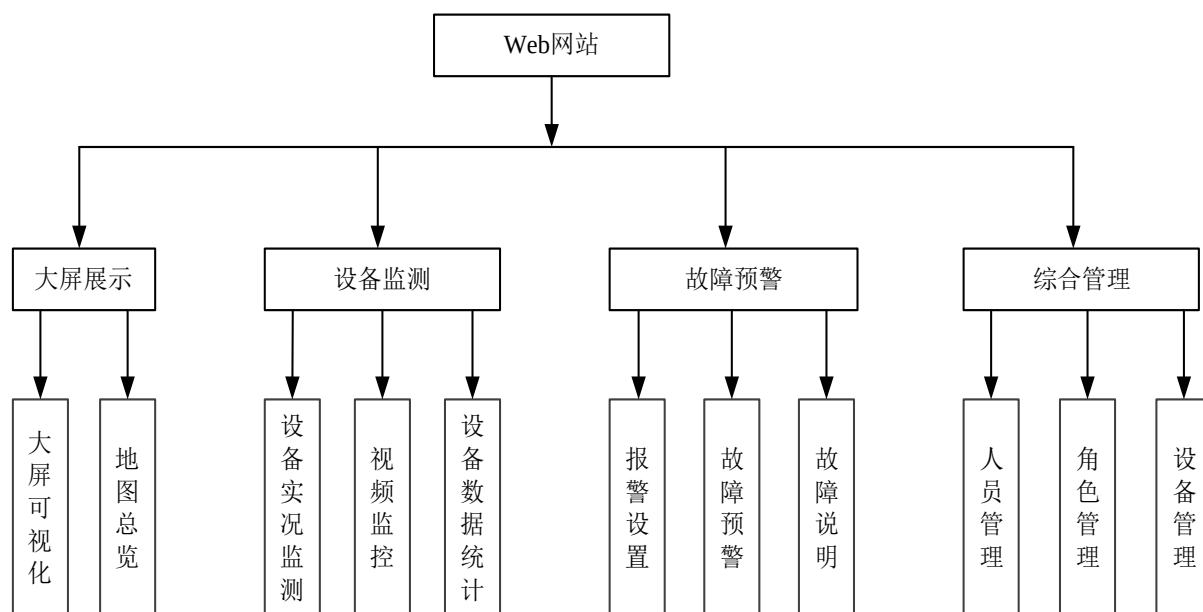


图 2.10 Web 功能模块

综合管理：提供账号的注册管理，对企业的人员、角色、设备的管理。

大屏展示：基于 ECharts 可视化组件库绘制数据大屏，基于自研的低代码平台搭建组态大屏，为锯床企业用户提供数据化的大屏展示。

地图总览：使用百度地图组件库，标注企业的锯床设备，便于全局的管理和查看，帮助企业了解设备的安装分布、运行情况。还提供地图定位，在收到故障报警后帮助维修人员定位导航到具体地点。

设备监测：实时监测设备的运行状态，得益于网关设备对锯切现场的实时数据采集和上报，Web 系统能够从云平台调用锯切数据，对锯切过程的各参数进行统计分析和直观的展示。在实时监测的同时，Web 端还提供远程操作的功能，交互界面可以调用接口向业务云平台发起请求，用于向网关下发指令，从而进行远程控制，比如对锯切工艺参数进行设置、开关机等。

视频监控：除了数据的实时监测功能，得益于当前互联网络的高速度，视频功能流行。在 Web 端嵌入萤石云 SDK，可以很方便地通过 Wi-Fi 方式，实时上传和展示锯床锯切过程的高清视频摄像，实现视频监控。

故障预警：金属带锯床的故障可以分为三种，锯切异常、伺服电机异常以及其他问题^[27]。在设备云平台可以对数据流设置触发器，监控数据点，当某个数据点异常时，设备云触发报警并推送给业务云平台。业务云可以直接调用短信推送接口进行推送，或者使报警接口与客户端应用进行长连接，客户端就可以实时监控到异常报警，弹出提醒。其中，故障分类图^[27]如图 2.11 所示。

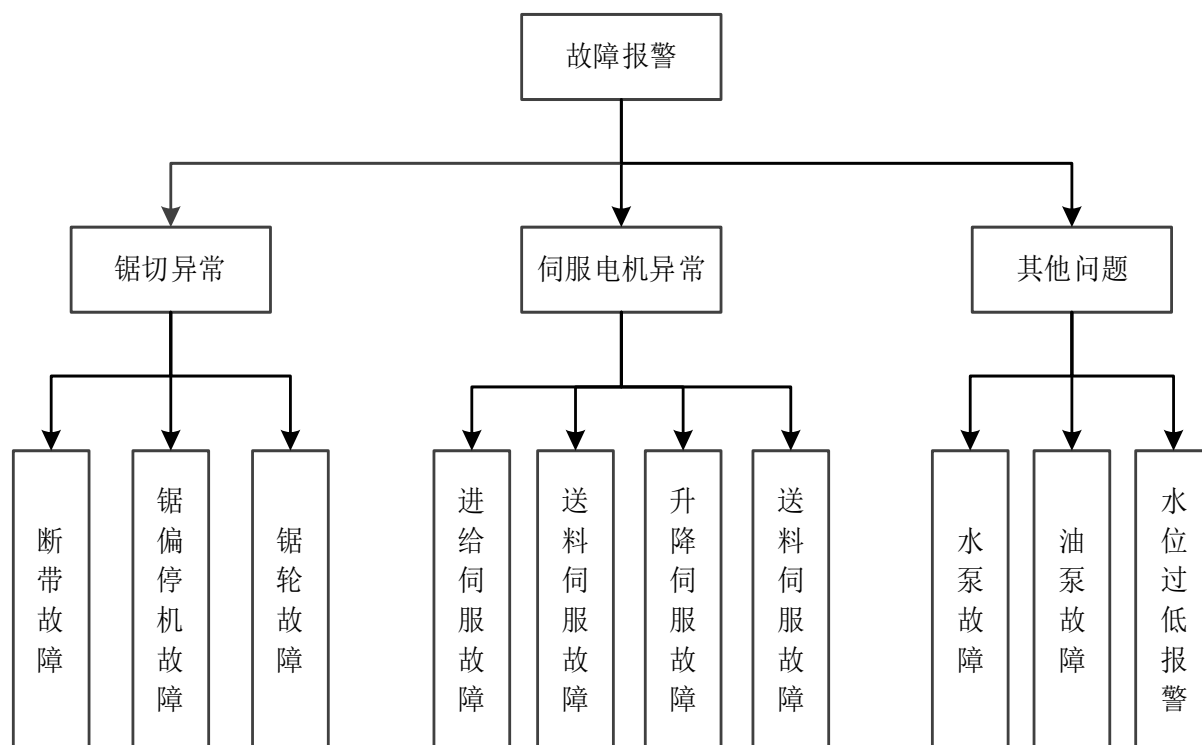


图 2.11 故障分类图

2.3.2.3 兼容性适配

Web 端还需要处理兼容性问题，本设计使用的 Vue 框架本身不需要直接操作浏览器原生的 DOM，比较多的问题是事件侦听器的差别，本设计采用适配器模式，先通过浏览器标识来判断当前的浏览器环境，把对事件侦听器的使用行为进行了统一封装，内部根据当前的浏览器环境去调用不同的 API，从而进行了简单地统一。

另外，样式的兼容问题是 Vue 无法避免的。所以在开发过程中，先是调用了 Normalize.css 来标准化各浏览器平台的初始样式，在后续的开发过程中尽量使用 Flex 这种兼容性优秀的布局方式，对于不太大众的特性，需要到 CanIUse 平台查询兼容性情况再决定能否使用。对于中后台的 Web 项目，本身并不十分重视页面效果，稳定性才是最优先考虑的，所以需要尽可能优先采用支持良好的特性来开发。

2.3.3 小程序设计

小程序是 Web 端的延伸，是 Web 端的功能简化版，提供一些关键数据点展示和及时的故障报警，以及设备的扫码添加等。也正是因为需求简单，所以本设计将其实现为跨端的应用，所谓跨端是指开发一套代码，可以一键编译到不同的平台，这为移动端开发带来了很大便利。因为移动端平台众多，尤其是国内的各家小程序平台。本设计主要提供微信小程序产品，如有业务需要也可以一键转换成 APP 和其他平台的小程序应用。因此选择基于类 Vue 语法的 uni-app 框架来开发，其应用架构如图 2.12 所示。

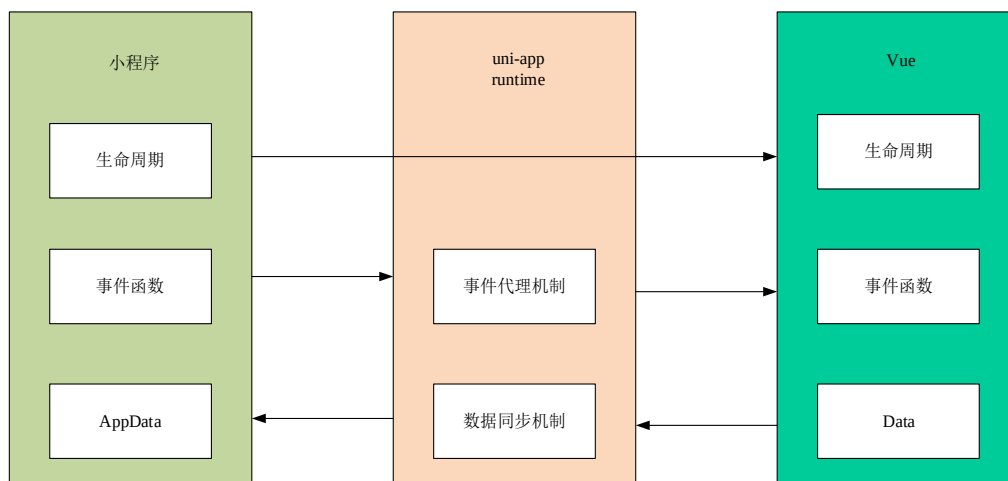


图 2.12 uni-app 架构

小程序和 Web 端功能基本相同，故不再赘述。其不同在于：提供扫码添加设备的功能，工作人员可以使用终端，扫描锯床设备 SN 码，设备进行设备添加以及简单的设备管理。

2.4 系统测试和效果

2.4.1 通信延迟效果

实时监测系统的通信能力至关重要，主要包括以下三方面：

- (1) 设备云和业务云的 HTTP 网络传输延迟。一般来说云端之间的延迟主要来自于 TCP 传输延迟，网络状态良好的情况下可以低至 10ms 级别。
- (2) 网关和设备云 OneNET 的 MQTT 通信延迟。得益于 OneNET 的 MQ 消息队列，其可以支持百万级 TPS 处理能力，ms 级的端到端时延，确保消息实时同步。
- (3) 网关和下挂 PLC 的 RS485 通信延迟，下面主要对此进行测试。

网关与 PLC 上位机的对接过程如图 2.13 所示，通过选择网关分组，填入设备 IP，加入网络组进行 PLC 对接。



图 2.13 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 的数据更新，如图 2.14 所示。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 ms 级别的通信延迟。

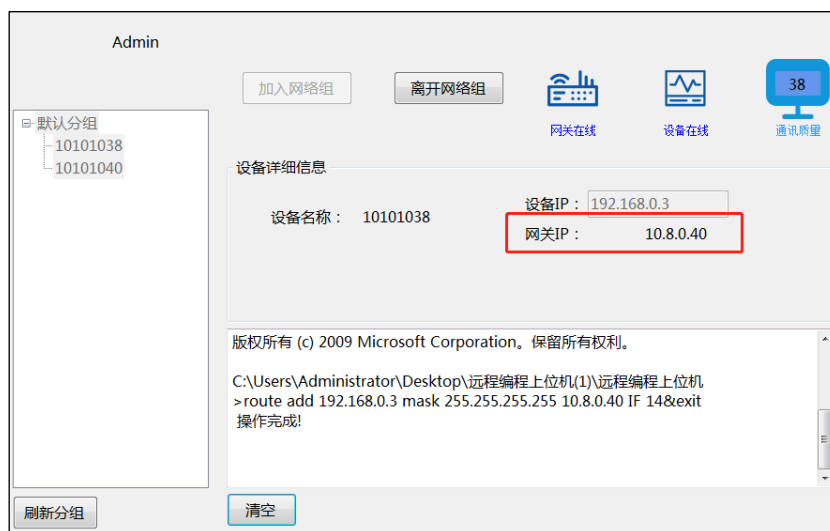


图 2.14 上位机对接完成图

2.4.2 设备和网关实物展示

金属带锯床设备实物包括锯床、PLC 控制器、操作触摸屏、传感器等。设备实物如图 2.15 所示，PLC 控制器如图 2.16 所示，触摸屏如图 2.17 所示，采集传感器如图 2.18 - 图 2.21 所示。网关电路板和外形图如图 2.22 所示。



图 2.15 带锯床设备



图 2.16 锯床 PLC 控制箱



图 2.17 操作触摸屏

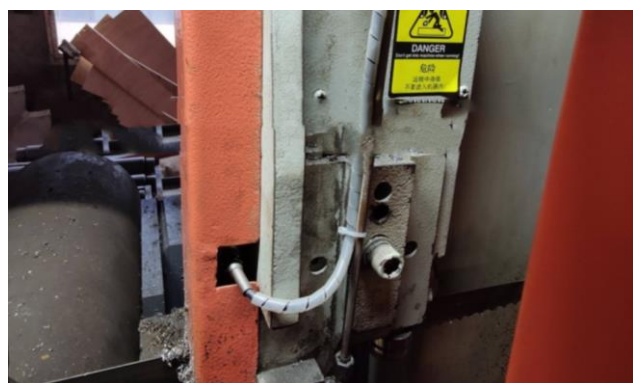


图 2.18 锯背压力检测



图 2.19 锯偏值检测



图 2.20 送料检测

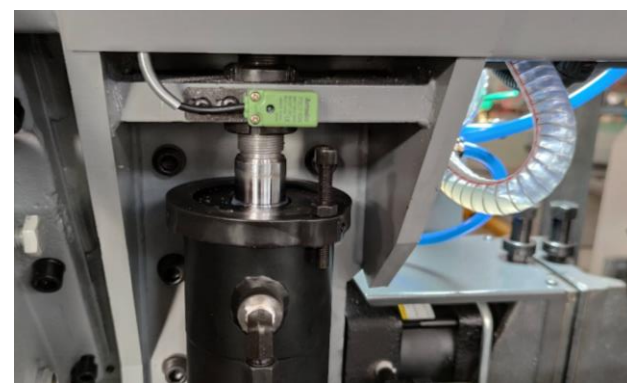


图 2.21 下限位切断检测



图 2.22 网关电路板和外形图

2.4.3 应用层效果展示

2.4.3.1 云平台部分效果展示

云平台的系统主要用于数据的处理和展示，这里截取展示设备云平台界面部分效果如图 2.23 -图 2.24 所示。

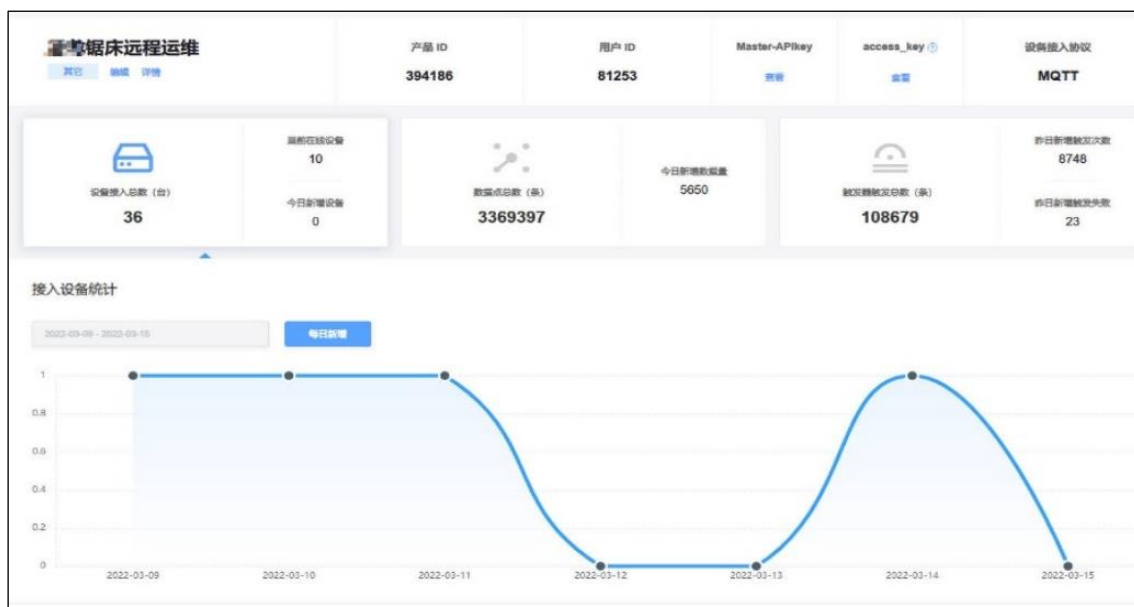


图 2.23 锯床云平台产品概况

设备数量 (个)	在线设备数 (个)	设备注册码	数据导出工具	批量添加	添加设备
36	10	A4p7UD1Q5ZxoCEOS			
在库状态(全部)	设备名称	设备状态	最后在线时间	操作	
在库	00ca2107bbaa	离线	2022-03-16 10:33:50	详情 数据流 更多操作	
在库	00ca2107bd32	离线	2022-03-16 11:55:00	详情 数据流 更多操作	
在库	00ca2107bbb2	离线	2022-03-09 14:03:53	详情 数据流 更多操作	
在库	859638498	离线	2021-11-30 14:29:51	详情 数据流 更多操作	
在库	859638193	离线	2021-11-30 14:28:39	详情 数据流 更多操作	
在库	859637741	离线	2021-11-30 14:27:12	详情 数据流 更多操作	
在库	859637350	离线	2021-11-30 14:25:21	详情 数据流 更多操作	
在库	859637053	离线	2021-11-30 14:23:27	详情 数据流 更多操作	
在库	859636528	离线	2021-11-30 14:21:05	详情 数据流 更多操作	
在库	859636204	在线	2022-03-15 16:15:10	详情 数据流 更多操作	
共36页					

图 2.24 锯床云平台设备列表

2.4.3.2 Web 端部分效果展示

Web 端界面部分效果如图 2.25 -图 2.31 所示。



图 2.25 数据大屏

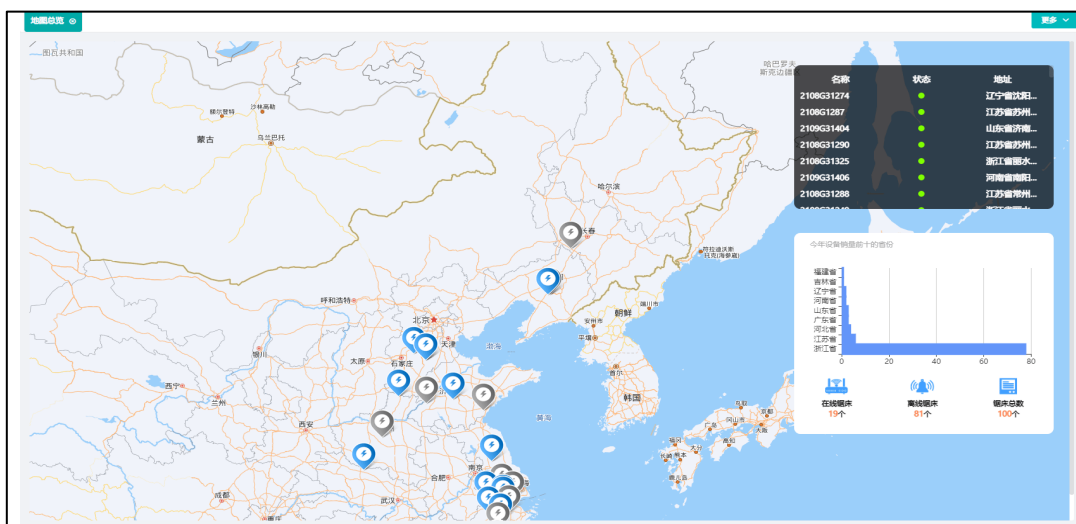


图 2.26 地图总览



图 2.27 锯切负载监测

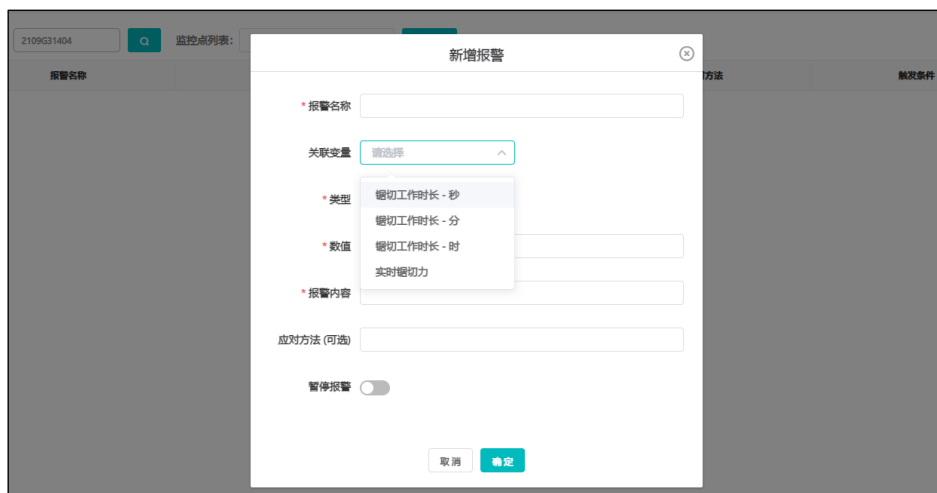


图 2.31 故障报警

2.4.3.3 小程序部分效果展示

小程序端界面部分效果如图 2.32 所示。

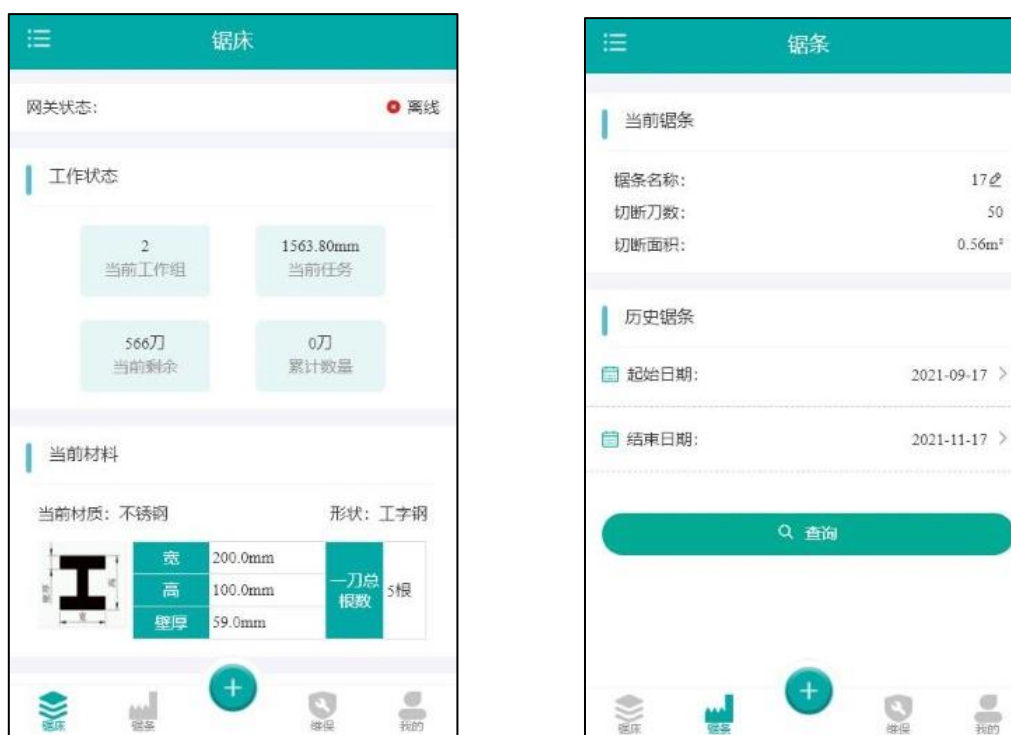


图 2.32 小程序部分界面

2.5 本章小结

本章设计和实现了带锯床在线监测系统。首先，对系统进行了完整的需求分析和总体设计；然后，就系统各部分进行详细设计，其中着重说明了：网络层的通信方式和通信协议，物联网网关软硬件程序的设计及其工作流程，应用层的云平台、Web 端、小程序应用的设计和实现。最后，分模块对系统的整体效果进行展示。本系统功能完整，切合实际需求，初步完成了设计目标，后文在此基础上就其他关键技术问题进行研究。

第3章 基于 CEEMDAN 和改进 ABC 的 SVM 锯条剩余寿命预测方案

目前锯床行业的智能化改造集中在故障诊断等方面，对带锯床锯条的寿命预测研究仍处于探索阶段。带锯条是锯床设备的关键部件，第二章的阐述表明了锯条存在磨损和退化的过程，退化到一定程度后，会引起锯带崩坏断裂，破坏正常的生产过程，严重的甚至会引发安全事故。但其健康状态无法直接有效获得，因此研究锯条的剩余寿命预测技术非常重要。

在第二章的锯床在线监测系统中，已经实现了在线监测的系列功能，同时在过去一年多的设备现场也采集到了大量锯床设备数据，具备了后续数据分析的基本条件。所以本章在此监测数据的基础上，开展基于退化过程的锯条剩余有效寿命预测的相关技术研究。在第一章节绪论的研究现状部分已经阐释了本设计选择基于 SVM 模型的剩余寿命预测方案的原因，下面就各个部分进行分析和设计。

3.1 锯条磨损分析和定量

寿命预测的关键是对锯条的磨损状态的正确识别，所以有必要分析锯切运动和锯条磨损影响因素，进而明确各锯切参数和采集信号中应该选取哪些作为磨损监测信号输入到磨损识别模型、使用什么参数来定量表示磨损量和剩余寿命。

3.1.1 锯切运动分析

仍以卧式双立柱金属带锯床为例，带锯床设备的锯切过程如下：夹紧模块的前钳夹具固定工件，送料时松开前钳，后钳夹紧工件送至指定位置，前钳继续夹紧固定住工件。双导向柱调整角度，由进给液压缸提供压力下压到初始位置。张紧液压缸给压用于张紧锯带，在电机驱动下主从锯轮转动，带动锯带高速转动。进给液压缸持续加压，下压锯条来切割工件。在切割的同时喷洒油基或水基冷却液，对锯条进行降温。一刀切割完后触发下限位信号表明切割完成，减速箱停止转动，快降杆辅助导向柱把锯背复位到初始位置，完成一刀。

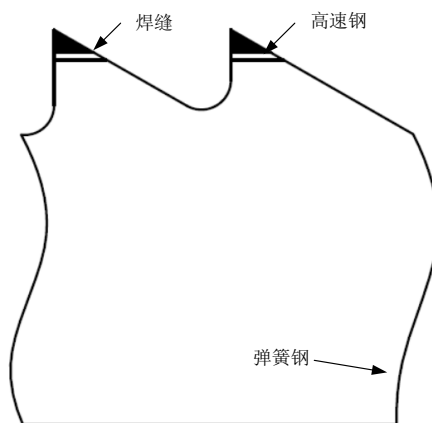


图 3.1 双金属带锯条锯齿示意图

锯条的锯齿示意图^[29]如图 3.1 所示。锯床的刀具是一边开齿的钢带锯条，目前常见的是双金属带锯条，它的齿部和背部分别由两种合金焊接而成^[28]。其中锯齿部分使用的材料一般为高强度合金钢。除材质外，锯条规格的区分主要是齿距的大小。

切割时，由电机提供动力，驱动锯轮带动锯条运动，同时锯架垂直下压产生压力，工件水平进给，来完成切割。锯切运动可以分解为两个运动：锯带方向上的锯切运动，用于切割工件，瞬时速度为锯切线速度 v_s ；垂直方向的进给运动，用于下压锯架使切割连续，由进给液压缸的下压力提供进给速度，记作 v_a 。最终，锯带运动、进给运动合成为完整的锯切运动。在运动过程中，锯齿会受损退化，导致切削力不足，锯条表面张力过大引起坏损。

3.1.2 磨损监测信号分析

研究锯条剩余寿命预测之前，要结合锯切运动过程考虑，应该选取什么磨损监测数据作为训练模型的特征输入。

现阶段的刀具磨损监测有直接法和间接法^[30]。直接法是直接观测获知，间接法是指切削运动中某些参数可能包含了刀具磨损的信息，对其进行数据分析来替代直接观测。优点是目前的传感技术采集数据相对准确，缺点是需要对磨损特征进行复杂的分析。本文采取数据驱动的研究方法，自然是选取后者。

首先考虑的是锯切参数和一些统计参数，如下表 3.1 所示。

表 3.1 锯切参数和统计参数

参数	量化单位	含义
锯切速度	m/min	锯带的线速度
进给速度	mm/min	单位时间的进给距离
工件材料	/	材质 形状 尺寸
导热条件	/	温度 导热系数 切削液
进给距离	mm/min	单位时间的进给深度
切削效率	cm ² /min	单位时间的切削面积
切削时间	min	累积的切割时间
切削面积	cm ²	累积的切削面积

其中，工件材料的特性、导热条件不易量化，在实验时尽量控制变量即可。文献[27]对上述其他参数进行了相关性分析，并与剩余寿命 RUL 进行图分析，最终认为切削效率、切削时间与 RUL 相关较大，其余参数相关性不强。

实际上，切削参数不能直接反映出磨损信息，直接得到的传感信号蕴含的磨损信息更多。王艺玮等人总结了常用的刀具磨损监测信号有：切削力、振动、声发射、电机的电流信号^[31]。

切削力信号用压力传感器测量锯背压力来表征。Mei Wang 等人研究表明, 切削力对刀具磨损敏感, 且在加工过程波动较稳定, 是比较好的选择^[32]。

振动信号用加速度计测量。其与刀具磨损相关, 但易受到机床抖动和噪声干扰, 后期处理有难度, 需要进行额外的时频分析来提取参数以降低干扰。比如 Lotfi 等人分析了超声波振动对刀具表面磨损的影响, 用于预测和改进刀具的磨损^[33]。

声发射信号是刀具加工工件时产生应力波的现象。其没有机械干扰, 监测过程可以不影响刀具和工件。比如电子科技大学的吴昊就是使用声发射信号中提取的时频参数和 EMD 分解的 IMF 分量作为特征参数, 效果不错^[18]。但声发射采集器相对昂贵, 且信号处理方式复杂, 效率较低。

电流信号比较好测量, 而电流、功率的变化主要是为了调节切削应力的大小, 可以在不适合采用切削力信号时作为替代品使用。比如武滢以主轴电流信号作为监测信号, 提取特征参数和能量熵进行特征融合, 来作为刀具磨损识别模型的输入特征^[34]。

多传感器共同使用理论上可以增强数据中的磨损信息量, 优势互补, 达到优于单一传感采集的效果。但如果造成冗余反而会影响决策, 在使用时必须进行相关性分析以减少冗余, 选择少数的关键传感特征参数。比如 Chuangwen Xu 等人认为工具磨损和铣削深度显著影响切削力和振动, 使用切削和振动信号来研究刀具侧面的磨损情况, 比单一的参数更加科学^[35]。

根据上述文献分析, 锯条的振动和切削力信号比较有价值, 本文采取多传感信号混合的方法。由于本系统实现的数据采集部分由于实际原因并未使用采集振动的加速度传感器, 所以使用 PHM2010 刀具磨损公开数据集来完成基本的设计和实现, 该数据集公开了三向的切削力、振动以及一个声发射信号的均方根值, 共计七个信号数据, 以及一个磨损量的物理测量数据。至于这三种信号哪些是关键参数以及具体采用哪些, 还需要后文进行相关度分析。

3.1.3 数据集和磨损定量

研究锯条剩余寿命预测之前, 还要明确数据集, 以及选取什么数据作为训练模型的输出, 即训练模型的样本标签。这也就是磨损状态和剩余寿命的定量表示问题。

1.2.2 节的研究现状分析也提到, 可以使用统计数据来表征寿命, 比如文献[27]选择切削面积的失效点来量化 RUL, 同时选择切削效率与切削时间作为特征输入。但是切削加工过程非线性、随机性大, 传统根据失效时间、失效点、失效面积是统计上的线性关系, 寿命计算可能会效果不佳。

所以磨损状态的定量问题最好还是通过实验测量, 限于本文写作过程中的实际生产条件无法进行实验, 本文采取使用 PHM2010 公开测量数据集。该数据集有 6 组刀具(记作 c1~c6)的测量样本, 其中 c1、c4、c6 为训练样本, c2、c3、c5 为测试样本。在切割过程中, 对三向的切削力、三向的振动以及声发射信号, 共计七个信号数据进行采集, 分别记作 X-Force (N)、Y-Force (N)、Z-Force (N)、X-Vibration (g)、Y-Vibration (g)、Z-Vibration (g)、AE-RMS (V)。汇总为约 300 个数据采集文件, 每个文件内有近 10 万条上述七个特征的采集数据。另外, 在每

个数据采集文件（每一刀）的结束，对刀具磨损量进行测量，形成一个测量磨损记录文件，记录这 300 次切割过程后的刀具磨损量测量值（取刀齿最大磨损处，单位 10^{-3}mm ）。

下面截取展示部分数据，取采集数据文件 c_1_001.csv，如图 3.2 所示。每一行数据代表一组采集信号数据，第一到第七列则代表上述其中信号。对应地，取磨损数据文件 c1_wears.csv，如图 3.3 所示。每一行数据代表一刀完成后的磨损数据，第一列代表刀数，后面三列分别代表当前刀下的三个刀具的物理磨损量。

```
c_1_001.csv
0.704,-0.387,-1.084,0.018,0.031,0.027,-0.004
0.772,-0.573,-1.153,-0.056,-0.057,-0.058,-0.004
0.828,-0.673,-1.242,0.037,0.019,0.031,-0.004
0.866,-0.601,-1.311,-0.028,-0.027,-0.022,-0.004
0.91,-0.377,-1.341,-0.05,-0.049,-0.058,-0.004
0.896,-0.095,-1.333,0.022,0.036,0.017,-0.004
0.91,0.239,-1.27,-0.062,-0.038,-0.047,-0.004
0.9,0.614,-1.209,-0.089,-0.067,-0.07,-0.004
0.879,0.974,-1.137,-0.028,-0.009,0.003,-0.004
0.906,1.291,-1.047,-0.093,-0.082,-0.072,-0.004
0.832,1.496,-1.039,-0.046,-0.048,-0.04,-0.004
0.785,1.567,-1.021,0.01,0.016,0.021,-0.004
0.66,1.552,-1.033,-0.088,-0.088,-0.106,-0.003
0.593,1.513,-1.027,-0.042,-0.04,-0.056,-0.004
0.461,1.513,-1.007,-0.024,-0.004,-0.012,-0.004
0.386,1.524,-0.965,-0.09,-0.072,-0.089,-0.004
0.305,1.509,-0.911,0.097,0.102,0.1,-0.004
0.262,1.522,-0.784,-0.056,-0.041,-0.038,-0.004
0.171,1.618,-0.653,-0.07,-0.061,-0.072,-0.004
0.102,1.771,-0.46,-0.034,-0.027,-0.03,-0.004
0.014,1.905,-0.264,-0.064,-0.045,-0.054,-0.004
-0.109,1.888,-0.096,-0.023,-0.021,-0.026,-0.004
-0.218,1.773,0.065,-0.057,-0.058,-0.037,-0.004
```

图 3.2 c_1_001 过程信号数据

```
281,156.0613707,144.0991016,141.8572923
282,156.6281987,144.5983129,142.1576774
283,157.1903603,145.1033072,142.4671341
284,157.7477835,145.6141647,142.7858874
285,158.3004046,146.1309628,143.1141568
286,158.8481684,146.6537756,143.4521558
287,159.3910281,147.1826743,143.8000917
288,159.9289454,147.717727,144.1581657
289,160.4618907,148.2589982,144.5265724
290,160.9898432,148.8065493,144.9054997
291,161.5127908,149.3604379,145.2951287
292,162.0307303,149.9207185,145.6956336
293,162.5436673,150.4874416,146.1071812
294,163.0516165,151.0606542,146.5299311
295,163.5546014,151.6403995,146.9640355
296,164.0526545,152.2267169,147.4096385
297,164.5458173,152.8196419,147.866877
298,165.0341403,153.419206,148.3358794
299,165.5176828,154.0254367,148.8167664
300,165.9965134,154.6383575,149.3096502
301,166.4707092,155.2579876,149.8146348
302,166.9403565,155.8843423,150.3318158
303,167.4055503,156.5174323,150.8612801
304,167.8663945,157.1572644,151.4031059
```

图 3.3 c_1 过程磨损量数据

上述数据集中对训练样本的磨损量进行了物理测量，能用来表征磨损状态。原始数据采用的是绝对磨损量（单位 mm），这里先进行归一化，取当前磨损值为：

$$y_i = \frac{\text{当前磨损量}}{\text{锯齿VB值}} \quad i = 1, 2, \dots, 300, \dots \quad (3.1)$$

其中， i 代表每一刀，锯齿 VB 值由固有的锯条规格决定，这里使用 200（单位 10^{-3}mm ）作为可供磨损的总量。磨损状态值 y_i 是区间 $[0,1]$ 上的无量纲值，使用它来表征刀具的归一化磨损状态。同时认为，磨损状态 $y_i \geq 0.9$ 健康度欠佳，需要提醒更换。也就是说，刀具的剩余有效寿命 RUL 可以表征为：

$$rul = 1 - y_i \quad (3.2)$$

由于 rul 与磨损值 y_i 线性相关，可以直接使用 y_i 作为识别目标和样本标签。

3.2 SVM 模型预测方案分析

经过上一小节锯条的磨损分析、磨损监测信号的分析 and 选择、磨损定量、以及数据集的构造，下面可以使用数据集中的特征输入和样本标签，进行 SVM 模型的应用研究。

3.2.1 SVM 和 SVR

在二维，两类点能被直线完全分开叫线性可分，推广到多维，将两类空间完全分开的平面称为一个超平面。最佳超平面，就是以最大间隔把两类样本分开的超平面：

- (1) 两类样本分割在超平面两侧；
- (2) 两侧在超平面附近的样本点，到超平面的距离最大。

就是说，SVM 就是希望各类样本点被超平面 $f(x, w) = w^T x + b$ 分割开，并且到平面距离最远。所以在 n 维空间中，SVM 优化的主问题是：

$$\min \frac{1}{2} \|w\|^2 \quad \text{s.t.} \quad y_i (w^T x_i + b) \geq 1 \quad (3.3)$$

其中 $w = \sqrt{w_1^2 + L + w_n^2}$ 。

但真正线性可分比较少见，这时就要采用软间隔，允许个别样本点错误出现在间隔带内：

$$y_i (w^T x_i + b) \leq 1 \quad (3.4)$$

为了度量这个间隔，为样本引入松弛变量 ξ_i ，令 $\varphi_i \geq 0$ 且 $1 - y_i (w^T x_i + b) - \xi_i \leq 0$ ，于是优化目标变成：

$$\min \frac{1}{2} w^2 + C \sum_{i=1}^n \xi_i \quad \text{s.t.} \quad y_i (w^T x_i + b) \geq 1 - \xi_i \quad (3.5)$$

其中正实数 C 为惩罚参数，代表对错误样本的惩罚程度。

上式(3.3)的优化目标是目标函数，亦称损失函数，是模型对误差的度量，用来量化预测值和真实标签间的差异。模型训练的目的就是求式(3.3)的最优解，使误差函数尽可能小。在

使用时，由于 SVM 模型本身已经选定了损失函数，比如最小二乘，直接使用即可。

但是，即使 SVM 增加了软间隔来尽可能实现线性可分，但是很多样本仍然无法做到。如图 3.4 所示，鸢尾花数据集无法仅根据花萼长度、花萼宽度两个特征来分类。

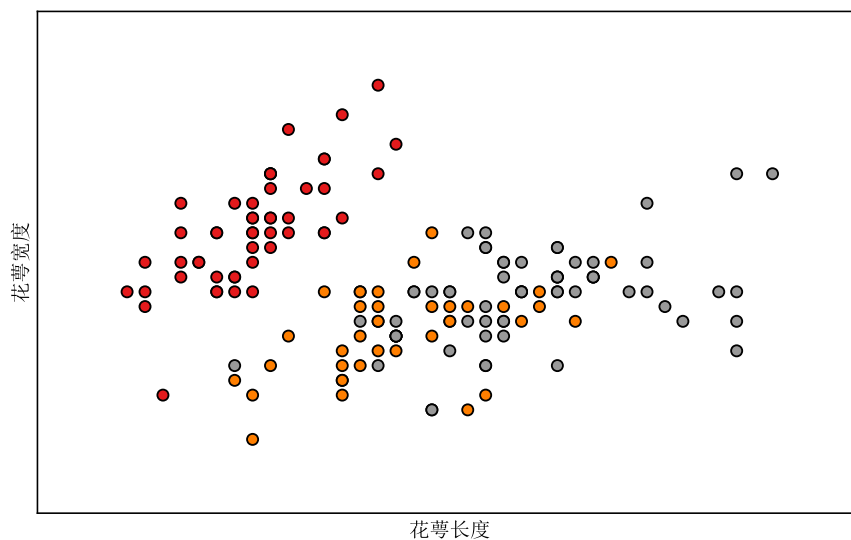


图 3.4 低维线性不可分

如图 3.5 所示，再增加一个特征比如花瓣长度或宽度，将二维的点 (x_i, x_j) 映射为高维中的 $(\phi(x_i), \phi(x_j))$ ，从而线性可分。

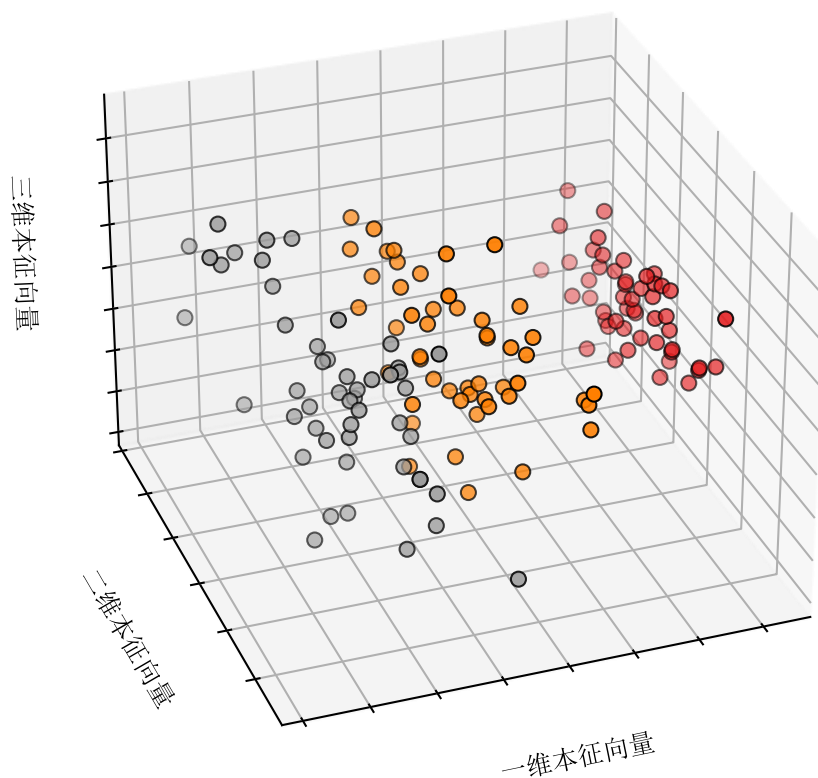


图 3.5 高维映射后线性可分

常用核函数如表 3.2 所示：

表 3.2 常用核函数

序号	核函数种类	公式
1	线性核	(x, x')
2	多项式核	$(\gamma(x, x') + r)^d$
3	径向基高斯核	$\exp(-\gamma \ x - x'\ ^2)$

其中，对于高斯核， $\gamma = \frac{1}{2\sigma^2}$, $\gamma > 0$ ， σ 为核的作用宽度，也就是高斯分布标准差。

对应地，使用 SVM 理论进行回归的方法叫支持向量回归（SVR）。其不同在于，SVM 寻找最大间隔超平面使样本点分割开，并处于隔离带之外；SVR 是寻找最小间隔超平面使样本点落在面上，并尽可能多地落在隔离带内，从而达到拟合预测的目的。也就是在求解过程中，式(3.3)中的约束条件等号取了反而已。

虽然 SVM 模型建立简单且能解决非线性数据问题，但是有时候会陷入局部最优解的情况，需要对参数选择方式进行优化。

3.2.2 优化方向分析

研究金属带锯条的 SVM 模型剩余寿命预测，需要考虑这样几个问题：

- (1) 如何选择磨损监测信号，这一点可以对 3.1 章节提到的诸多磨损监测信号做相关性分析来选择。
- (2) 要明确金属带锯条剩余寿命如何定量，或者说锯条的磨损量如何量化。这一点也已经在 3.1 章节分析明确，在构造数据集时加以计算即可。
- (3) 如何对使用的传感信号进行有效的特征提取与选择。尤其对于易受到噪声干扰的振动信号等，就需要使用到一些时频域的信号分析方法。
- (4) 磨损定量后，使用 SVM 回归预测模型进行磨损量的识别和回归预测。而且要解决 SVM 寻优过程对参数敏感的问题，提高识别和预测精度，这需要使用一些优化算法。需要着重研究进行优化的是第三点和第四点，下面分别进行分析。

3.2.2.1 自适应时频分析

对于干扰过多的振动传感信号，提取特征时需要排除干扰，来保证磨损量尽量识别精确，这需要进行信号分析。传统的信号分析方法主要是分析时域的信号谱、傅立叶变换后的频谱等。但是简单的时频参数不适合非线性随机信号的分析。对此做出改进的是小波变换和经验模式分解（Empirical Mode Decomposition, EMD）技术。董伟航等人通过离散小波变换提取主轴功率信号的近似系数、主轴转速、铣削深度作为特征，监测木工刀具磨损状态^[36]。但是

小波变换的性能过度依赖于小波基，自适应性不佳，使用不灵活。

EMD 自适应时频分析技术，由 Huang NE 在 2001 年提出^[37]。相比于小波变换，EMD 的基函数具有自适应性，无须事先预设，就可以直接分解信号。比如 Xinhua Shi 等人基于 EMD 将切削方向的声音信号与背景噪声分离^[38]。

EMD 认为，任何信号都可以分解成一系列本征模态函数（intrinsic mode function, IMF），得到的各阶分量包含了原始信号在不同时间尺度的局部特征^[39]。IMF 分量要满足两个条件^[40]：

- (1) 数据段内，极值点与零点数量相等或相差不大于一，即数据曲线要反复穿过 X 轴。这就要求 IMF 不可以缺少极大/小值点，即不可单调。
- (2) 局部极大值点、极小值点各自连接而成的上、下包络线相对于时间轴局部对称。实际就是要求，均值线的平均值趋近于 0。

EMD 算法的过程简单直观，就是要保证每一次分解的 IMF 符合约束条件。也就是说，分解的截止条件是，发现某一次的 IMF 是单调函数（不满足条件 1）。

具体步骤如下：

- (1) 求极值点，使用寻峰（Find Peaks）算法获取原始信号序列 $x(t)$ 的全部极值，分别画出上下包络线。
- (2) 以上、下包络的均值，画出均值包络线。
- (3) 将信号 $x(t)$ 减去包络均值 $m_1(t)$ ，得到中间信号 $h_1(t)$ 。

$$h_1(t) = x(t) - m_1(t) \quad (3.6)$$

- (4) 判断 $h_1(t)$ 是否满足 IMF 约束，若满足， $h_1(t)$ 就是 IMF 分量，记为 IMF1；若不满足，继续对该分量做 1-4 的分离，直到分量符合约束。
- (5) 将 IMF1 从原信号 $x(t)$ 分离出去，即 $x(t)$ 减 IMF1 得到 $r_1(t)$ 。对 $r_1(t)$ 进行步骤 1-4 的分离，得到 IMF2。以此类推，直到 $r_n(t)$ 不能继续分解为止，此时 $r_n(t)$ 是一个单调的余项，称为剩余信号 Res。由于它无法继续进行 IMF 分解，于是就完成了 EMD 分解。

最终原始信号 $x(t)$ 可以表示为：

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (3.7)$$

原信号经 EMD 分解的 IMF 分量，能够反映原始信号的局部特征。但是，在处理复杂和弱特征信号时，EMD 存在模式混叠的问题^[39]。模式混叠是指 EMD 分解在求取包络线中，信号中的异常破坏极值点的分布，求得的包络中就有异常包络。之后的分量就会包含异常、或包含相邻特征时间尺度的 IMF，造成混叠^[41]。

集合经验模态分解（EEMD）和自适应噪声完备集合经验模式分解^[42]（CEEMDAN）是从 EMD 的基础上加以改进的算法，可以有效解决 EMD 的问题。这也是本方案的改进方向之一。

3.2.2.2 寻参优化算法

近年来,人们针对复杂优化问题的求解,尤其是最优解问题提出了一系列的群智能算法,原理是模拟生物种群的协作行为,其具有灵活性、自组织性、稳健性、分布性等特点,非常适合用来进行机器学习的模型参数寻优。文献[43]综述了常用的群智能优化算法,其中,粒子群算法搜索速度较快、蚁群算法动态搜索能力较强、狼群算法深度搜索能力较强,但又各有缺点。本文选取的是简单易用且收敛效果较好的人工蜂群(ABC)算法^[44],并对其进行改进。

ABC 算法模拟蜂群觅食行为,人工蜂个体只需要对原有解和寻到的可能解做适应度的优劣比较,在进化中,不断对比局部优解和全局优解,从而在进化结束时表现出全局最优解。ABC 操作简单、控制参数少、局部收敛速度较快,适用于连续函数的全局优化问题,因此适用于 SVM 的超平面参数优化过程。

3.3 基于 CEEMDAN 和改进 ABC 的 SVM 锯条剩余寿命预测设计与实现

3.3.1 总体方案设计

所谓剩余寿命预测,需要实现的目标就是根据测量的传感数据识别出当前锯条的磨损状态量,进而预测其剩余寿命。本方案是基于 SVM 的锯条寿命预测实现,关键在于磨损量的识别,在此基础上进行回归预测。以每一刀切削完毕的时刻为节点,将此刀切割期间的运行状态数据作为输入,在训练集中,对输入数据做改进 EMD 时频分析和特征提取,调参训练得到基于改进的 ABC 算法的 SVM 磨损状态模型。识别并输出当前的磨损状态(0-1)后,接着根据磨损状态与剩余寿命的关系式,得到下一次的磨损预测量,直到得到一个低于健康阈值的量(假定为 0.9),即可得出该锯条的剩余寿命。整体方案设计流程如下图 3.6 所示。

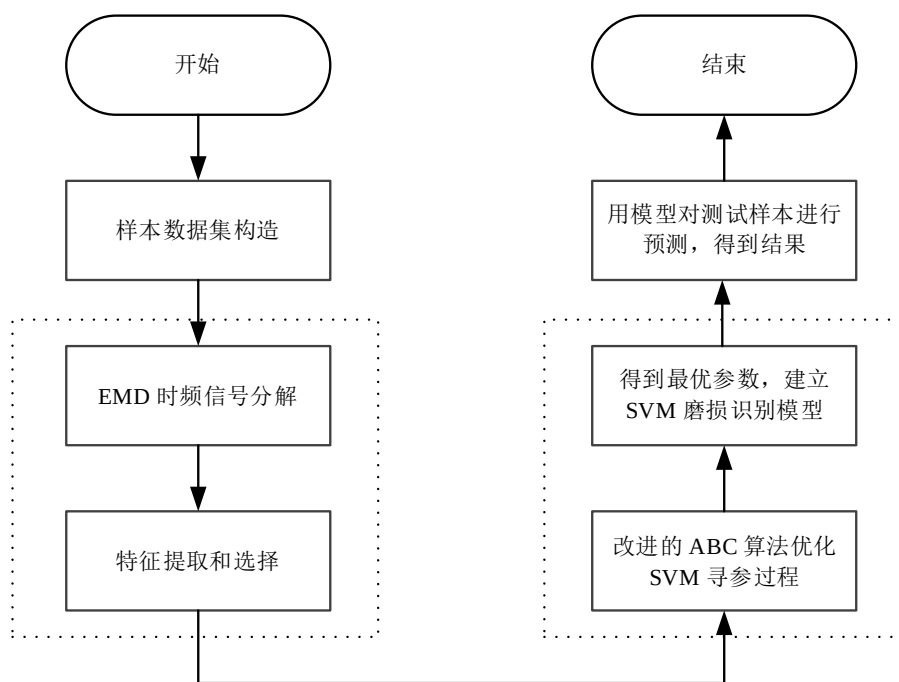


图 3.6 总体设计流程

其中，针对各环节的实际问题，本设计作出了一些改进：

为了更好地识别出磨损量，针对磨损趋势做了图表定性分析，从数据中分离出前中后三个磨损阶段的样本，以此来选择更好的特征。

针对刀具采集数据过程中的显著噪声数据以及非线性信号问题，先对信号做相关性分析做选择，再针对噪声过大的振动信号做分解，对分解得到的分量进行选择，舍去相关性过小（小于 0.01）的分量，并选出特征良好的 IMF 分量，计算峭度值作为 SVM 模型的输入向量。

寿命预测的关键是识别出磨损量，在模型的选择上，选择了适合切削过程这种小样本场景的 SVM，且针对 SVM 性能对核参数敏感的问题，对人工蜂群算法进行改进，使之用于寻找最优参数建立模型。

3.3.2 磨损监测信号的选择

使用 3.1.3 中所述的数据集，首先绘制采集到的信号波形做初步的判断。取前 10000 个信号数据，绘制如图 3.7 - 图 3.9 所示。简洁起见这里各类型只展示一个。

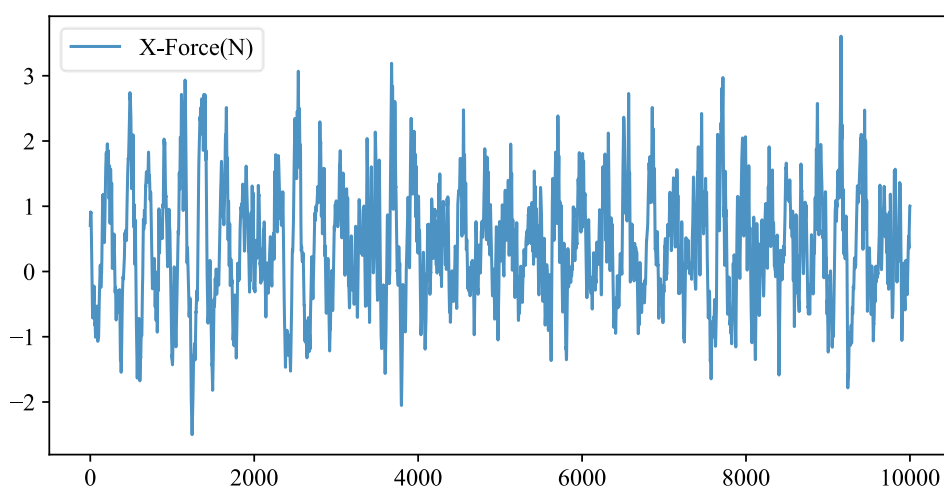


图 3.7 X 轴切削力

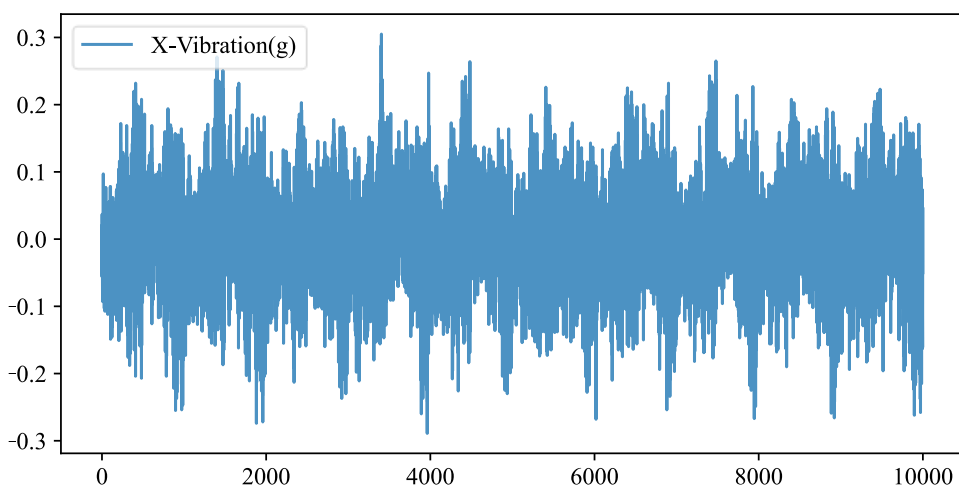


图 3.8 X 轴振动信号

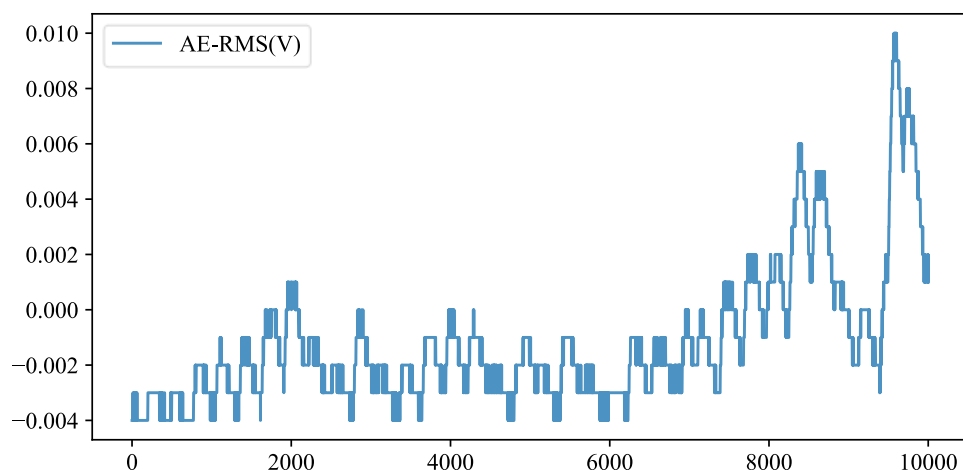
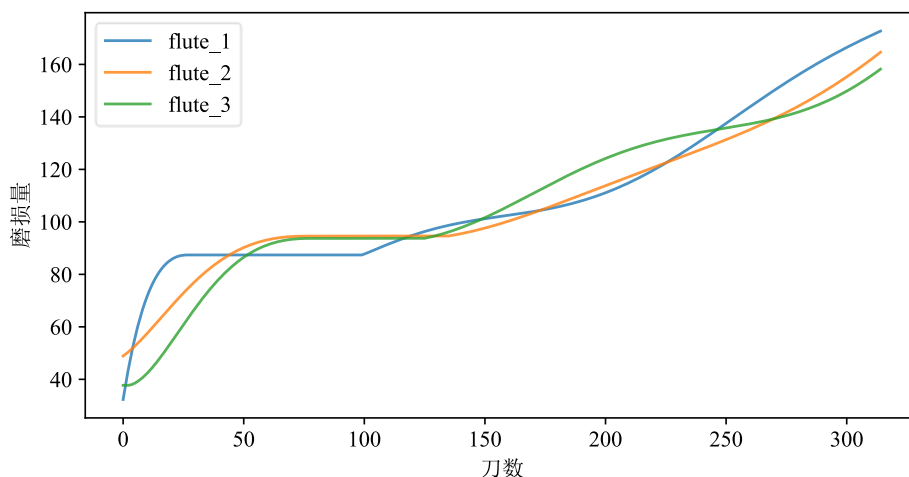


图 3.9 声发射信号

可以看出，采集数据尤其是振动数据，混杂难辨，特征不明显，且实际工业过程存在噪声，很难直接使用，后续需要进行信号的分解来得到比较稳定的分量信号。再以训练组 c_1 的三个刀具为例，取图 3.3 所示的 $c1_wear.csv$ 的磨损数据，绘制磨损量趋势，如图 3.10 所示。

图 3.10 c_1 组随刀数增加的磨损量趋势

从图中可以看出，磨损量数据曲线存在先快后缓再较快的总体趋势。在起始阶段，锯条初始试锯，在初期没有完全磨合时，磨损出现加剧现象是正常的。中间过程则相对比较平稳，后期锯条达到一定量的磨损后，又出现了变快的趋势。

所以，后续设计可以根据曲线的明显拐点，把数据样本划分成前期-中期-后期三个磨损阶段。这样根据不同的阶段样本来识别磨损状态，可以避免过大的误差。

下面对这些传感信号进行选择，将训练样本的磨损量值归一化为磨损状态值 y 作为样本标签，再取数据集的七个信号，分别表示为 A0 到 A6，汇总成一组数据。考察各特征和磨损量 y ，用 Pandas 库的 `corr` 函数计算相关系数，并用 Seaborn 库的 `heatmap` 函数绘制热力图，如图 3.11 所示。

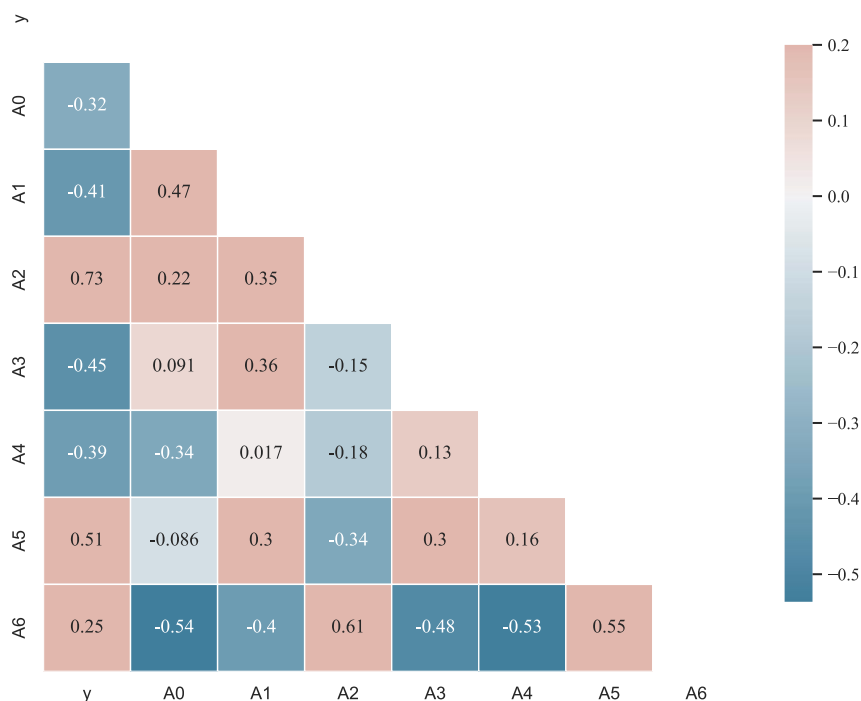


图 3.11 信号的相关度热力图

上图表明，A6 的声发射信号与其他信号的相关系数较高，这可能是因为声发射很大程度上是切削力、振动的双重作用，是因变量，所以舍去这一特征，避免冗余信息。仅采用切削力、振动信号进行后续特征的提取和选择。

3.3.3 基于 CEEMDAN 的信号处理和特征提取

锯条的振动信号非线性、非平稳，噪声较多，不太容易直接用来作为特征，在提取特征之前，先对振动信号进行时频分析。

3.3.3.1 CEEMDAN 分解算法

EEMD 利用白噪声均值为 0 的特性，多次叠加并平均以抵消噪声^[45]。而 CEEMDAN 为了得到更加精准的上下包络，参考了 EEMD 的特性。首先确定处理次数 M 、添加的白噪声幅值，其中幅值表征为“白噪声幅值的标准差与原始信号幅值标准差之比”，根据文献经验设定为 0.2。算法流程如下：

首先，第一次迭代过程，以 $x(t)$ 作为输入信号，向其中分别加入白噪声进行 EMD 分解，得到的各 $IMF1$ 分量成分求均值得到 $IMF1$ ，并更新残值 $r1$ 。流程如图 3.12 所示。

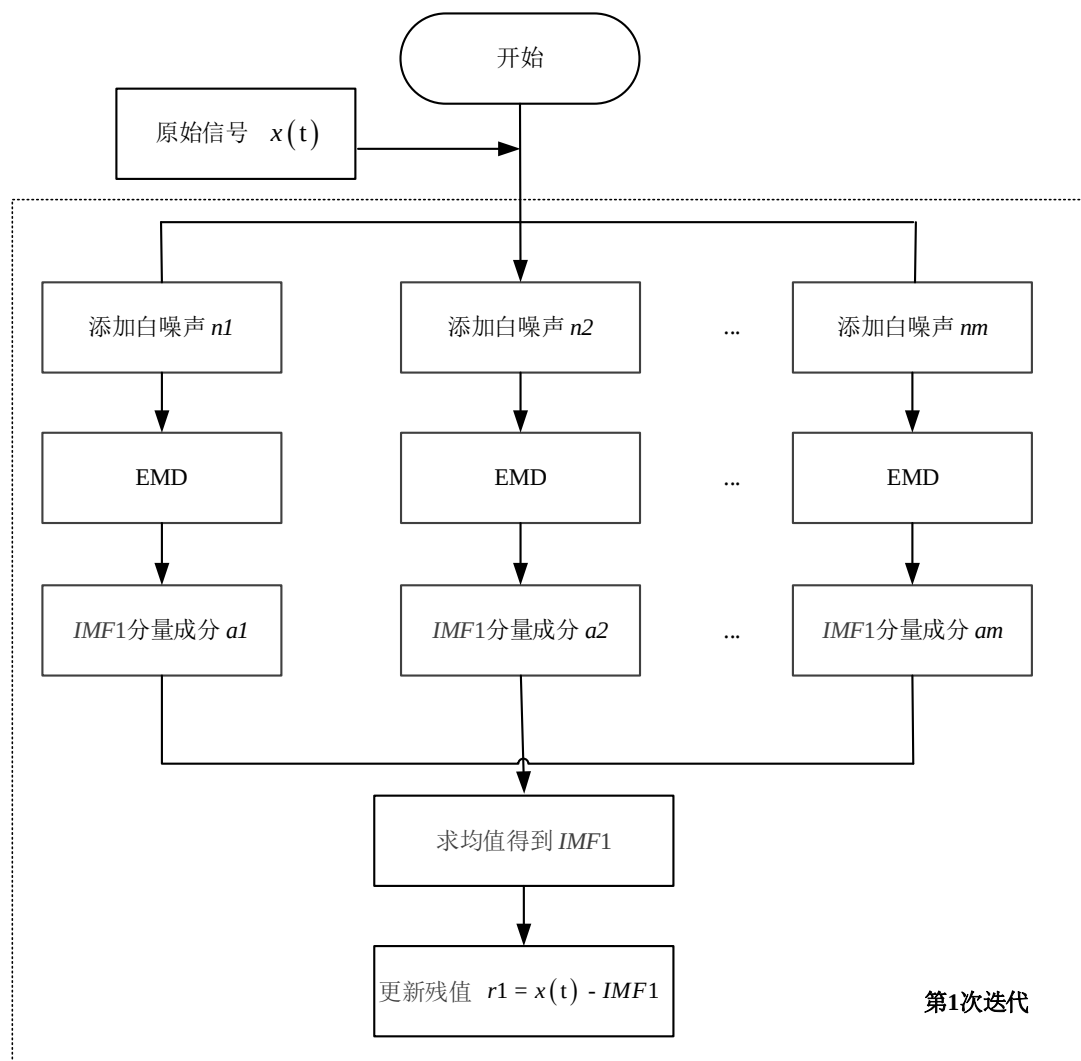


图 3.12 CEEMDAN 算法第一次迭代

以此类推，进行第 K 次迭代之后，得到分量 IMF_k 并更新残值 r_k ，最终得到分解完成的原始信号 $x(t)$ ，如图 3.13 所示。

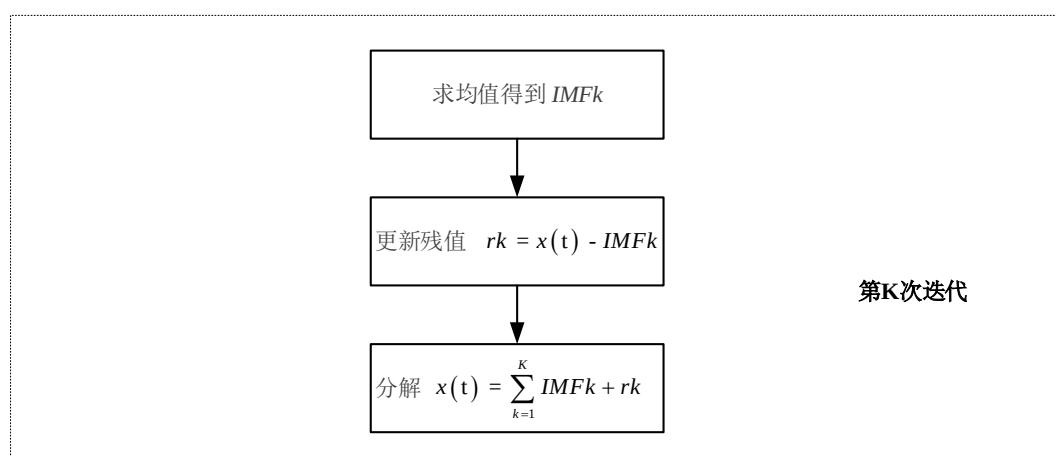


图 3.13 CEEMDAN 算法最后一次迭代

EEMD 是直接给 M 个信号添加白噪声，进行 EMD 分解得到 IMF 分量，并求取均值^[46]。而 CEEMD 是每求完一阶 IMF 分量，重新给残值（即余项）加入白噪声，并求此时的 IMF 分量均值，逐次迭代，最终完成分解。

相比于 EMD 和 EEMD 方法，CEEMDAN 的优点明显：

- (1) CEEMDAN 未对原始信号加噪声，所以在较小的平均次数下就可以有很好的完备性，同时因为不需要太多的平均次数，所以有更快的计算速度。
- (2) CEEMDAN 方法可以自动滤除掉对信号分析意义不大的低幅度分量，有效减少分量个数。

为了简单说明效果，随机生成一个频率混叠的正弦信号作为输入，分别经过 EEMD、CEEMDAN 分解，得到如图 3.14、图 3.15 所示。其中红色为原始信号，绿色为各分量。

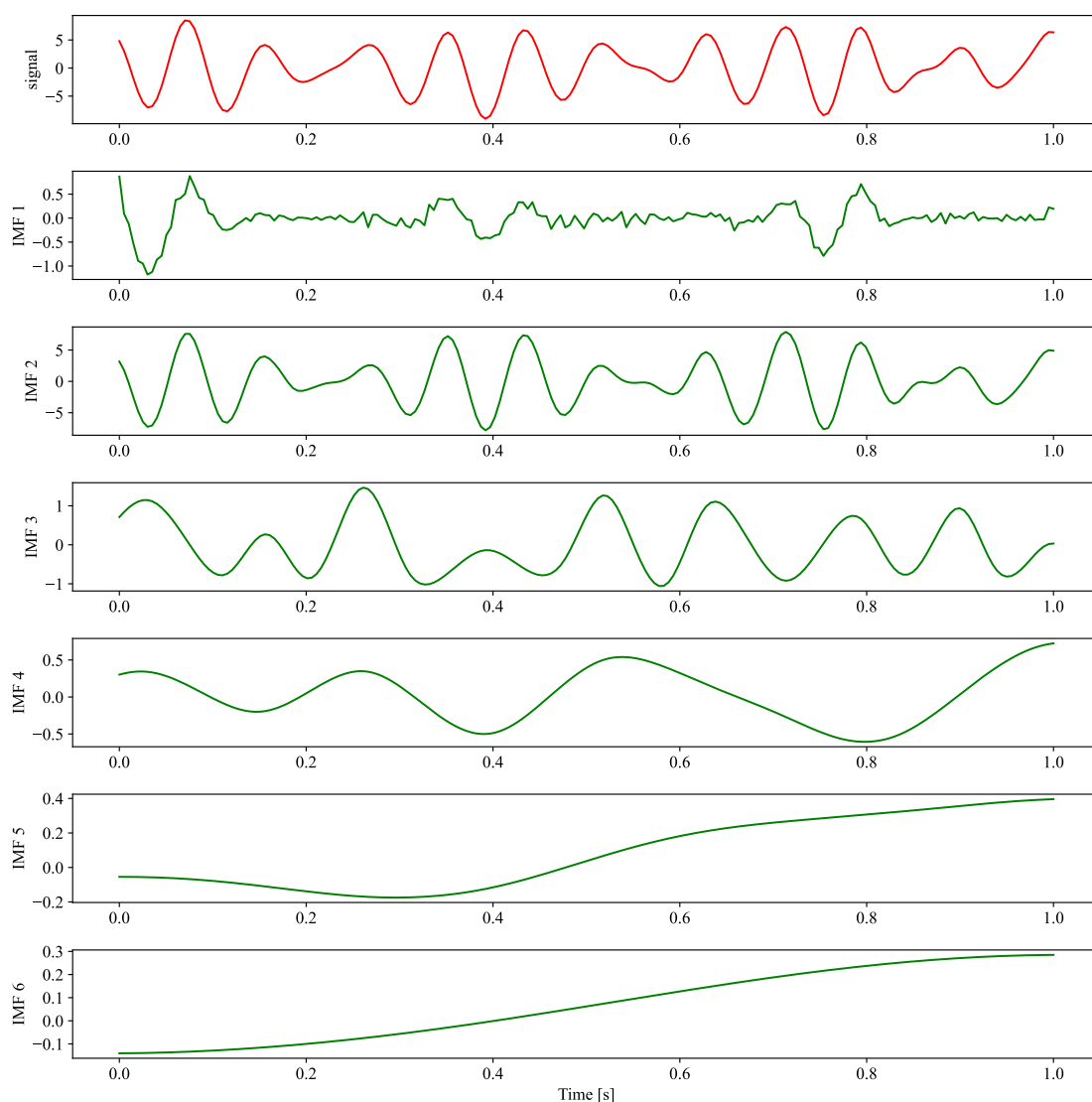


图 3.14 EEMD IMF 时域分量

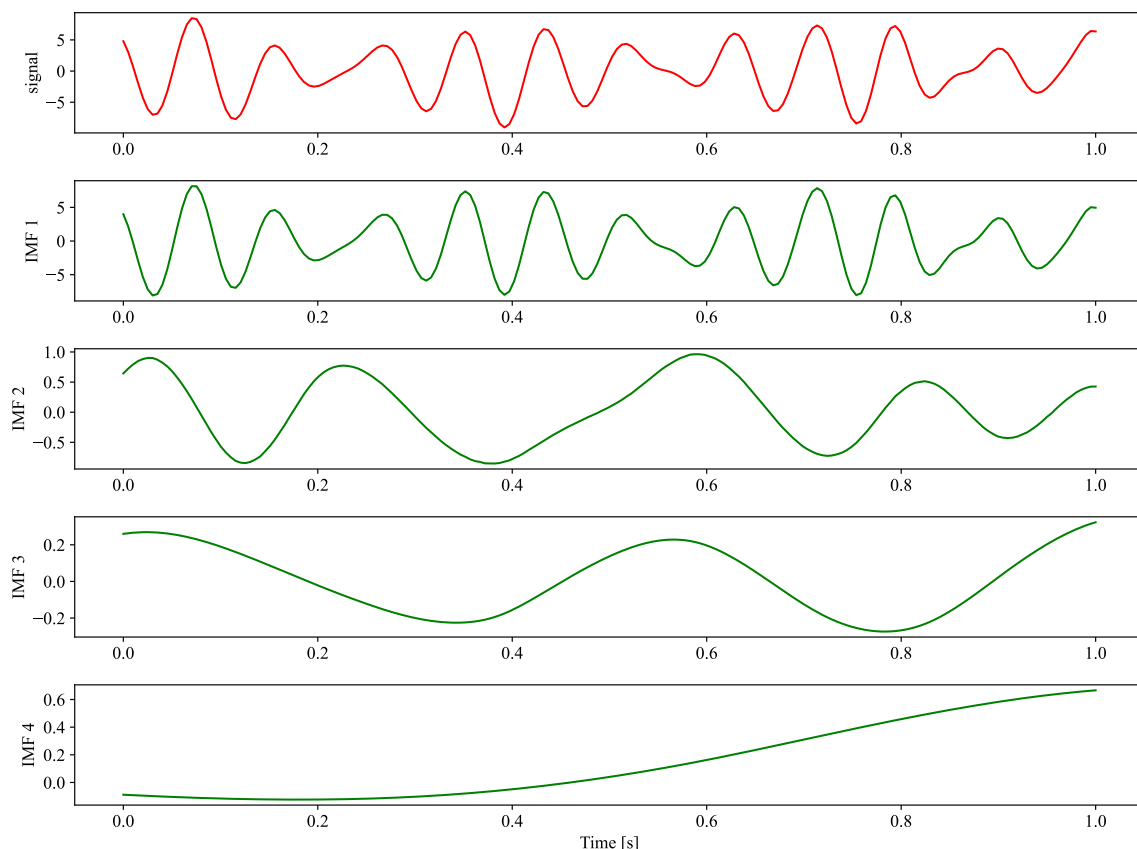


图 3.15 CEEMDAN IMF 时域分量

从上图可以看出，相比于 EEMD 方法，CEEMDAN 方法可以有效减少 IMF 分量个数，滤除了杂乱分量。因此它可以有更快的计算速度，更好的模态分解效果，这在后续的特征处理中也可以有效减少特征个数。

3.3.3.2 磨损信号的 CEEMDAN 分解实现

选取三个磨损阶段的样本，对样本内的传感信号数据进行 CEEMDAN 分解处理。CEEMD 的实现比较简单，可以使用 PyEMD 库封装好的模块实现，导入 CEEMDAN 函数后进行实例化，输入参数信号 *signal* 和时间 *t* 就可以做时域分解，分解的时域分量直接绘制波形图。

将时域波形进行希尔伯特变换得到频域图，调用 SciPy 库的 Hilbert 函数，并输入时域分量即可。频域转换也可以借用 PyEMD 库封装好的 `plot_instant_freq` 绘制函数，输入时间 *t* 和 IMF 时域分量。

比如，使用上述方法，对样本数据集的刀具 c1 中磨损量为 32.31711361 的对应 X 轴振动信号做 CEEMDAN 分解，得到 12 阶的 IMF 分量，其时域波形、频谱分别如图 3.16、图 3.17 所示。

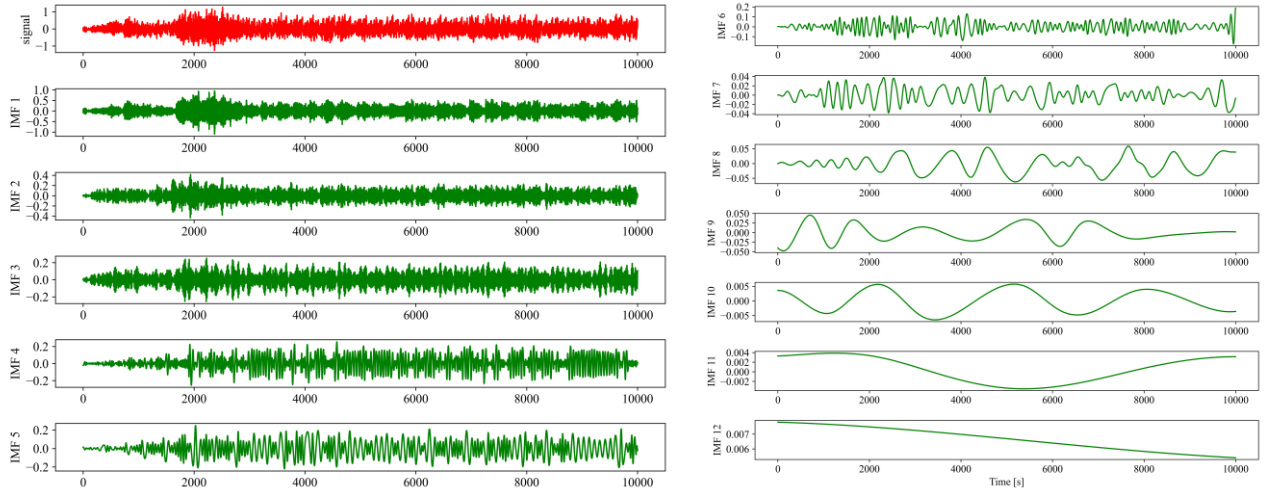


图 3.16 IMF 分量时域图

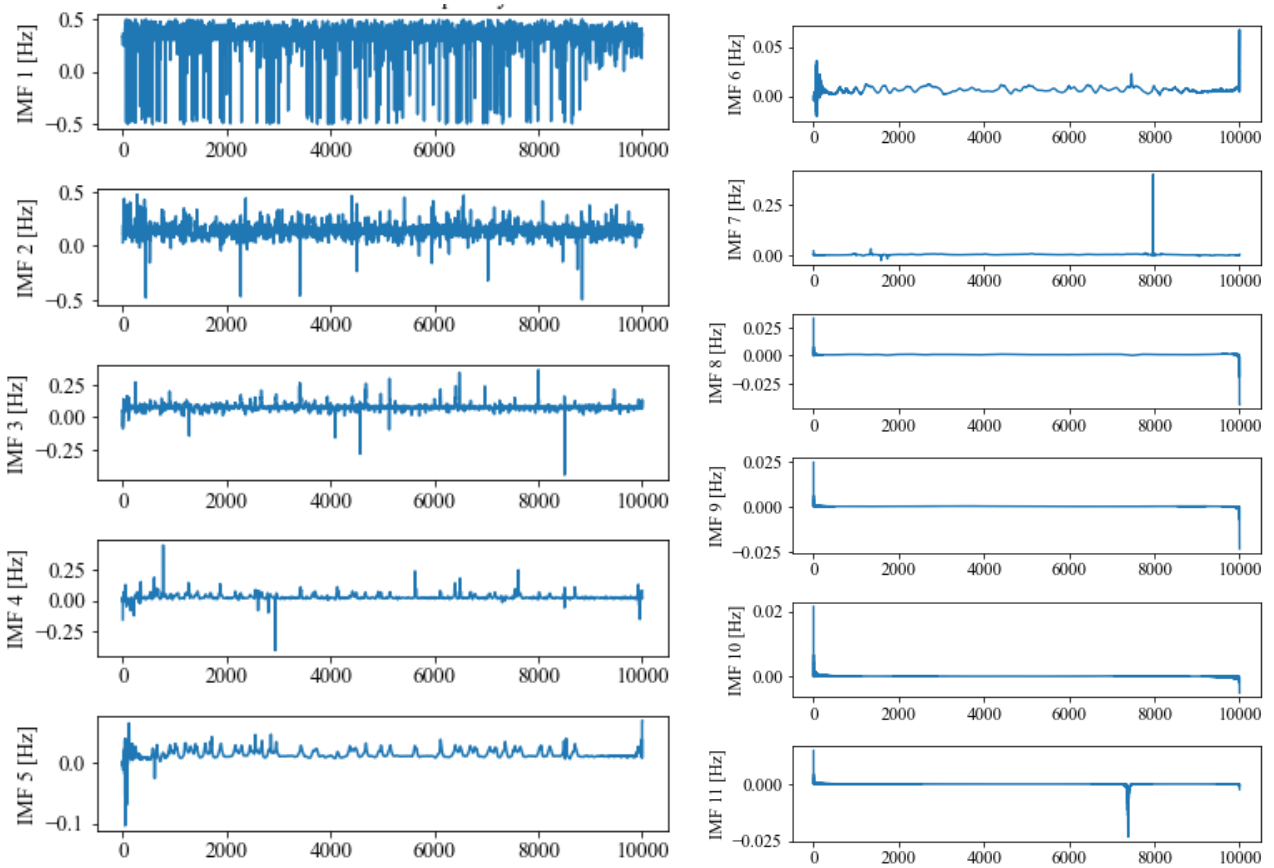


图 3.17 IMF 分量频域图

3.3.3.3 分量的特征提取和选择实现

改进的 CEEMDAN 把非平稳信号分解成较平稳的分量，并且可以有效滤除一些低幅度分量。除此之外，一般认为相关性小于 0.01 的分量就几乎没有作用，所以在后续实现过程中，选取分量时还可以去除一些相关性低的低频分量。计算相关性公式如下：

$$Q_{xy} = \frac{\sum_{k=0}^n x(k)y(k)}{\sqrt{\sum_{k=0}^n x^2(k) \sum_{k=0}^n y^2(k)}} \quad (3.8)$$

从公式可以看出，分解阶数越高，包含特征信息越少，与原始信号特征的相关性就越低。

还要对各 IMF 分量进行特征提取。选取峭度参数，峭度是反映随机变量分布特征的数值量，它是无量纲参数，对冲击信号敏感，适用于表面损伤类场景^[47]。

$$C_k = \frac{1}{n} \sum_{i=1}^n x_i^4 \quad (3.9)$$

从参数意义上讲，峭度衡量的是实值随机变量的概率分布相对于正态分布的偏差，均匀分布是最极端的情况。根据随机动力学，锯条无损伤工作时，信号的幅值分布接近正态，峭度值接近于 3。随着损伤累积，大幅值的概率密度增加，分布偏离正态，峭度值增加。峭度越大，说明锯条震动异常，可能出现了较大磨损^[48]。

实现过程如下：

(1) 剔除相关性小于 0.01 的 IMF 分量

根据式(3.8)的相关性计算公式，对上一节中分解的 12 个 IMF 分量计算相关性。为了便于展示，略过低阶的重要分量，最终如下表 3.3 所示。

表 3.3 IMF 各分量相关性

样本 磨损量	相关系数							
	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10	IMF11
0.45	0.5317	0.5293	0.3244	0.1243	0.1275	0.0112	0.0112	0.0043
0.55	0.3466	0.2312	0.1541	0.2292	0.1052	0.0031	0.0031	0.0011
0.63	0.4166	0.3452	0.2342	0.2193	0.3053	0.0532	0.0432	0.0112
0.75	0.0432	0.0376	0.2275	0.1087	0.0723	0.0008	0.0102	0.0033
0.85	0.2379	0.3746	0.0955	0.2523	0.1112	0.0012	0.0046	0.0065

可以看出，IMF9~IMF11 与原始信号的相关性都较小，特征信息量少，故舍去后三个分量，只保留前 8 个特征 IMF 分量。

(2) 求各 IMF 分量的标准峭度

根据式(3.9)的峭度计算公式，对选择的 8 个 IMF 分量计算峭度，计算比较简单，可以使用 Pandas 库的 kurt 函数一键求出。上述磨损量 32.31711361 的刀具 c1 数据样本的 IMF 分量计算峭度，得到散点图如图 3.18 所示。从图中可以看出，该样本此时的测量数据信号各 IMF 分量都位于正常范围内，没有出现较大的离群异常值。

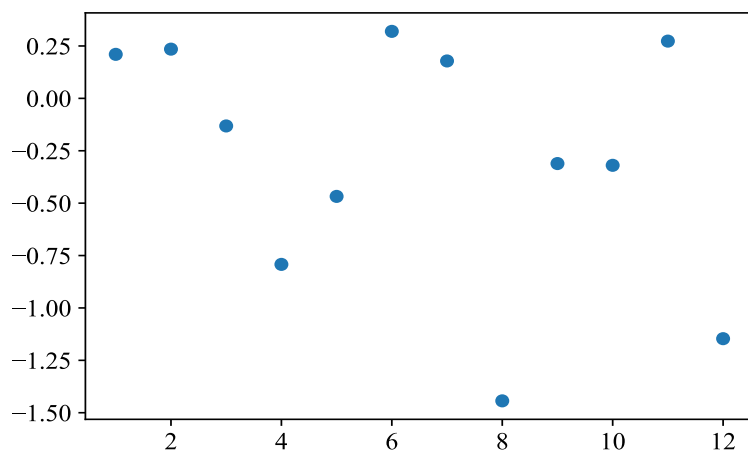


图 3.18 峭度散点图

(3) IMF 各分量的特征选择

对三个阶段的样本数据做信号分解，并求取峭度，绘制峭度曲线，观察哪些分量可以很好地区分出三个阶段的样本。仍然对 c1 刀具，随机选取前期样本、中期样本、后期样本，计算各样本的 8 个 IMF 分量的峭度绝对值，绘制三阶段峭度趋势折线图，如图 3.19 所示。



图 3.19 各样本的 IMF 分量峭度趋势图

从上图可以看出，随着刀具磨损从前期、中期到后期的阶段，整体上是各 IMF 函数的峭度值不断增大。也就是说，信号幅值的分布对正态分布的偏离程度，随着刀具磨损而增加。

并且从折线走势来看, IMF1~IMF4 分量下的三个阶段的曲线区分度较好。而后面四个分量, 样本点在中期磨损和后期磨损曲线重叠部分较多。因此选择 IMF1~IMF4 峭度值作为特征值。

(4) 归一化处理

支持向量机算法不能很好地容忍非标准化数据, 所以在使用之前需要先标准化数据。由于本文选取的峭度值已经是归一化的量, 不需要单独进行处理。而其他的信号特征在使用前可以使用 scikit-learn 库的 preprocessing 和 pipeline 模块先完成归一化, 这对于后续的调参过程很有好处。

3.3.4 基于改进 ABC 的 SVM 回归预测算法

3.3.4.1 SVM 的核参数优化

文献[49]指出, SVM 需要优化两个参数: 惩罚参数 C (记为 c) 和核参数 $gamma$ (记为 g)。

向量机共有的惩罚参数 c 是一种正则化系数, 用以调节两个指标“间隔大小、分类准确度或者说噪声容忍度”的权重。默认可以选取为 1, 当 c 趋于 0 时, 是放宽间隔, 代表不过度关注正确性, 可以容忍一些错误样本, 就容易产生欠拟合。当 c 趋于无穷大时, 就是减小间隔, 代表只允许正确, 不允许出现误差样本, 旨在希望正确识别所有训练样本, 但这就容易过拟合。

高斯核特有的 g 参数与核作用宽度 σ 反相关, 其定义了单个训练样本有多大的影响力。 g 很大时, 支持向量影响的宽度范围 σ 小, 但是附近的样本一定会被影响到, 训练准确率可以很高。反之, g 越小, σ 越大, 支持向量影响的范围变大, 但附近的样本被影响的可能性变低, 会向着欠拟合方向走, 但好处是对未知样本的分类效果好。

所以, 当数据中有很多噪声(错误样本)时, 就要尝试调小 c , 调小 g ; 如果发现训练数据上的准确率低, 可以试着调大 c , 调大 g 。

综上所述, SVM 的性能依赖于核函数以及核参数的优劣, 在使用时 c 、 g 参数要不断进行试错调参, 过程麻烦。所以实际在选取参数时, 可以使用网格搜索的方式, 简化手工调参的繁琐过程, 比如借助 scikit-learn 库的 GridSearchCV 函数对 c 和 g 以指数间隔来选取到一个合适的参数数值, 防止过拟合和欠拟合。但是这可能造成算法收敛速度缓慢、识别结果不准确的问题, 因为识别结果的准确性常常伴随一定程度的过拟合, 而网格搜索的过程摒弃了这些。所以本方案设计使用人工蜂群智能优化算法来优化参数, 目的在于提高 SVM 算法收敛速度和识别预测的准确性。

3.3.4.2 ABC 算法及其改进

ABC 算法中, 解的进化由三类蜂共同完成: 引领蜂、跟随蜂、侦察蜂, 问题的求解对应于蜂群寻找蜜量最大的蜜源目标^[50]。引领蜂利用现有的蜜源信息邻域搜索新蜜源并分享, 跟随蜂等待并依据引领蜂的分享信息寻找蜜源, 侦查蜂负责随机寻找一个新的有价值蜜源^[51]。

在搜索的过程中保留较好解，不断进化，最终得到全局最优解。

假设问题的解空间 D 维，ABC 的求解就是在 D 维空间中进行搜索。引领蜂、跟随蜂以及蜜源的个数都是 N ，一个引领蜂与一个蜜源相对应，每个蜜源代表一个可能解，蜜源的花蜜量对应于解的适应度值 fit_i ^[52]。其中适应度则依赖于评价函数值 f_i 。

ABC 算法原理和过程如下：

- (1) 首先初始化参数：蜜源数量 N ，蜜源搜索限制次数 N_l ，最大迭代次数 N_c 。
- (2) 第 i 个引领蜂根据初始解产生新解，并根据贪心策略将新的可能解与原来的解比较，保留较好解。

公式如下：

$$x_{id}' = x_{id} + \phi_{id}(x_{id} - x_{kd}) \quad (3.10)$$

其中， $\phi_{id} \in [-1, 1]$, $k \neq i$, $i = 1, 2, \dots, N$, $d = 1, 2, \dots, D$ 。

- (3) 当有蜜源被引领蜂选择，第 i 个跟随蜂依据概率去选择新的蜜源。标准算法采取随机的轮盘赌策略，概率公式为：

$$p_i = \frac{fit_i}{\sum_{j=1}^N fit_j} \quad (3.11)$$

其中 fit_i 是可能解 x_i 的适应度值，其计算公式为：

$$fit_i = \begin{cases} \frac{1}{1 + f_i}, & f_i > 0 \\ 1 + \|f_i\|, & f_i < 0 \end{cases} \quad (3.12)$$

其中， f_i 为第 i 个蜜源的评价值，即评价函数值，参数训练时可自行选择。

- (4) 当引领蜂、跟随蜂搜索完解空间，如果蜜源适应度值在限制次数 N_l 内没有被提高，则认为蜜源太差，丢弃。而与该蜜源对应的引领蜂变侦查蜂^[53]。
- (5) 当有侦查蜂产生，由它随机产生新的蜜源，公式如下：

$$x_{id} = x_d^{\min} + r \cdot (x_d^{\max} - x_d^{\min}) \quad (3.13)$$

其中， r 是区间 $[0, 1]$ 的随机数， $x_d^{\max}$ 、 $x_d^{\min}$ 是第 d 维的上、下界。

- (6) 重复 2~5，每次记录较优解。直至达到最大迭代次数 N_c ，输出得到种群的最优解。最终 ABC 算法流程如图 3.20 所示。

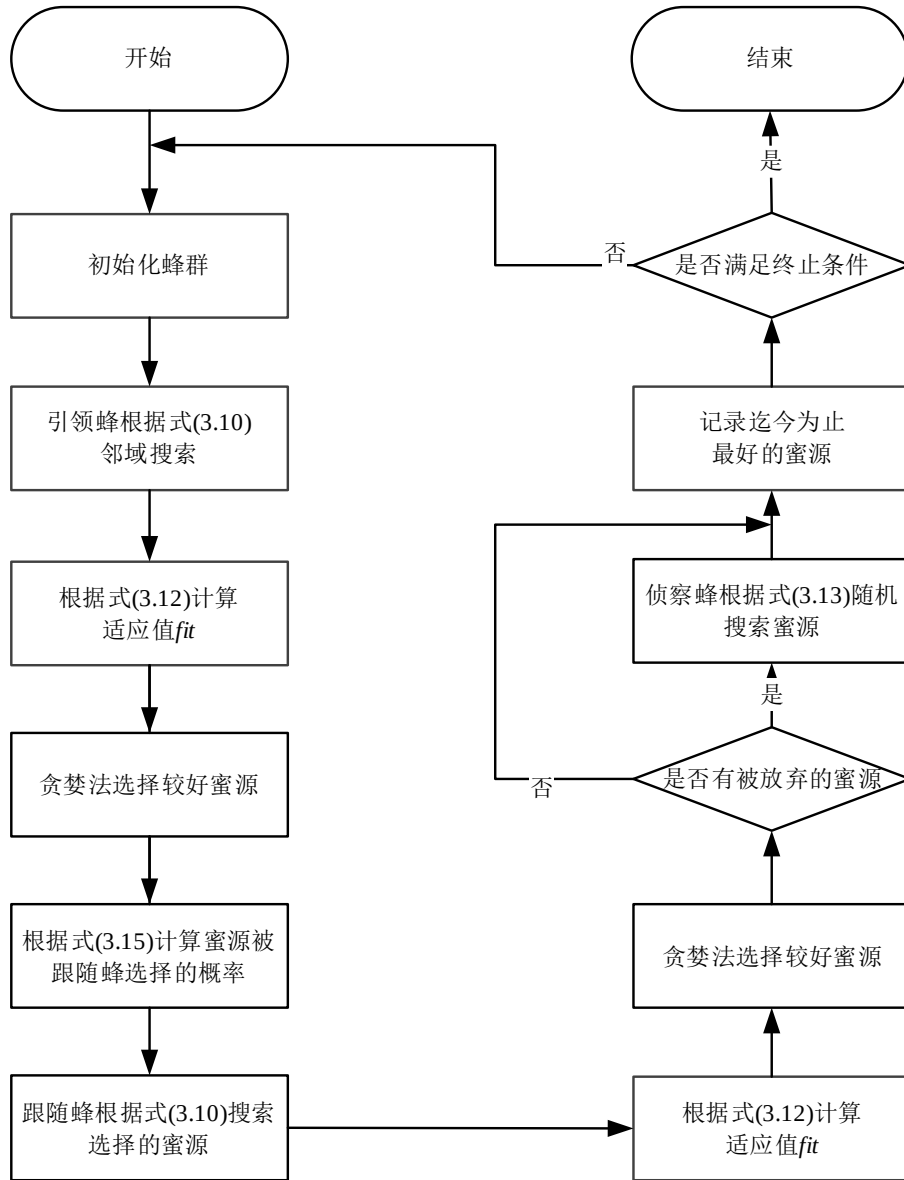


图 3.20 人工蜂群算法流程示意图

ABC 的泛化能力很大程度上由跟随蜂的概率选择能力来决定，而上述标准 ABC 算法的问题在于，跟随蜂是基于式(3.11)选择蜜源，对蜜源的选择概率完全依赖于计算的适应值 fit_i 。显然，较高适应值的最优蜜源更容易被选择，但是并不能排除较差蜜源仍包含有用信息的可能^[54]。所以在算法初期，就容易迭代出超常个体，控制整个概率选择过程。后期的较差蜜源就没有被选择的可能了，也就是会产生局部极值。而且，后期不同个体的适应值差异会越来越小，完全根据适应值来选择，基本等同于完全随机^[55]。这些问题都极大弱化了 ABC 算法的全局泛化能力。

本设计的改进之处在于，把蜜源的选择概率公式，用反序函数加以混淆。混淆因子如下：

$$a_i = \frac{1/i}{\sum_{i=1}^N 1/i} \quad (3.14)$$

结合式(3.11)和式(3.14)，混淆后的概率选择如下：

$$p_i' = p_i \cdot a_i \quad (3.15)$$

使用式(3.15)代替原有的式(3.11)即可。

上式中， a_i 随着 i 增大而增大，这实际上是对个体进行了一种排序。个体在后期的迭代过程中，混淆因子的值变大，因子对概率选择的影响也就越大，就使得不同个体不再完全依赖于适应值的差别，后期的个体相比之前会有更大的可能被选择，从而减少超常个体出现的可能，提高泛化能力。

3.3.4.3 基于改进 ABC 的 SVM 回归预测算法实现

使用上节的改进 ABC 算法，对 SVM 参数 c 、 g 搜索最优解，建立改进 ABC-SVM 刀具磨损状态识别模型的步骤如下：

- (1) 准备训练样本，预处理将数据归一化，尤其是 3.2.2 节确定的磨损的定量表示，这将作为后续的样本标签。
- (2) 初始化 SVM 参数：设置惩罚参数 c 、核参数 g 的搜索范围，并将 c 、 g 作为蜜源的二维坐标，训练刀具磨损状态模型。
- (3) 初始化 ABC 算法的参数，包括：蜜源数量 N ，蜜源最大循环次数 N_l ，迭代次数 N_c ^[56]。
- (4) 根据式(3.12)计算该蜜源解的适应度，其中评价函数如下：

$$f(x_i) = \frac{1}{T} \sum_{i=1}^T (y - y_i)^2 \quad (3.16)$$

式中 y 、 y_i 分别表示实际磨损值和预测磨损值。

- (5) 根据式(3.15)选择新蜜源。对每个蜜源解，根据式(3.12)计算自身最优值，将适应度与自身最优值进行比较，更新自身的最优适应度，并比较自身与全局最优值，更新种群最优适应度。
- (6) 在搜索解的过程中，检查蜜源丢弃条件，直到超过蜜源搜索限制次数 N_l ，就舍弃该蜜源，继续根据式(3.13)搜索可能解。
- (7) 检查终止条件。是否达到最大迭代次数 N_c ，如不满足，转到步骤 3；如满足，则模型参数寻优结束，返回当前种群的最优适应度以及最优核参数 c 、 g 。
- (8) 根据最优核参数建立改进 ABC-SVM 磨损识别模型。

3.4 实例测试和结果分析

3.4.1 模型评估指标

对模型的训练结果进行评估，就是找到样本标签的真实值与预测值的差异，对于本文的回归算法，是评判是否预测到正确的数值、是否拟合足够的信息。本文选取的回归模型的评估指标有：

(1) 均方误差 (MSE)

均方误差是预测值和真实值的误差的平方和的均值，值越小说明拟合效果越好。

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.17)$$

其中 y_i 实际值， $\hat{y}_i$ 为预测值。

但是由于 MSE 做了平方，为了保持量纲要进行开方，得到均方根误差 RMSE。由于本文样本标签进行了归一化，所以 RMSE 取值范围是 [0,1]。可以使用 scikit-learn 的 mean_squared_error 模块计算。

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.18)$$

(2) 决定系数 (R²)

方差是任意值和样本均值的差异，差异越大，值所带的信息越多。由此衡量数据上所拟合的信息量。

决定系数 R² 计算方差得分，表征预测值和真实值的拟合程度。取值范围是 [0,1]，越接近 1，说明拟合线越接近实际曲线，训练结果也就越好。使用 scikit-learn 库的 r2_score 模块计算。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.19)$$

(3) 算法性能

对比测试时同样要考虑运行性能，在其他条件不变的情况下，模型训练的时间可以直观表现模型的运行性能。

3.4.2 实例测试

根据 3.3.4 节的模型建立步骤进行实例预测，步骤如下：

(1) 准备训练样本

根据上面的信号分析，振动信号选择了 4 个 IMF 分量作为特征值，六组信号就是 3+3*4=15 个特征值。选择不同磨损阶段的数据作为训练样本，样本标签是求得的磨损状态值。

(2) 参数初始化设置

对改进 ABC-SVM 的参数进行初始化，如表 3.4 所示。

表 3.4 改进 ABC-SVM 初始化参数

蜜源数量 N	蜜源搜索限制次数 N_l	迭代次数 N_c	c 范围	g 范围
20	50	200	0.1~10000	0.1~1000

(3) 模型训练

使用 3.3.3 节经信号分解提取到的特征值，按照 3.3.4 节中 ABC-SVM 模型的具体步骤进行模型训练，经过 200 次迭代，得到最优参数， $c=203$ ， $g=43.8$ ，此时最优适应度值为 0.0146。

(4) 模型验证

以上述最优参数建立刀具磨损状态的识别模型，用测试样本进行模型验证，识别结果如图 3.21 所示。磨损识别效果越好，得到的预测寿命越精确。

为了说明改进 ABC-SVM 模型的有益效果，本文将其与标准的 ABC-SVM、未优化的 SVM 进行对比。其中，标准的 ABC-SVM 寻参过程使用标准 ABC 算法。未优化的 SVM 采取自动网格搜索方法进行寻参，对 c 和 g 以指数间隔来选取到一个合适的参数值。最终识别结果分别如图 3.22、图 3.23 所示。

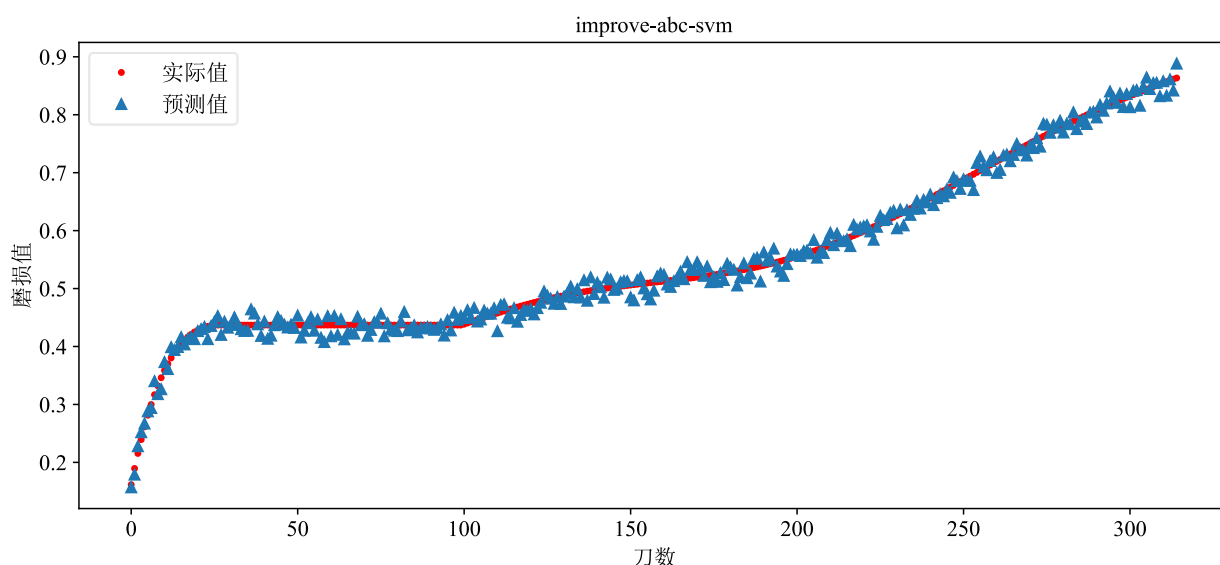


图 3.21 改进的 ABC-SVM 结果

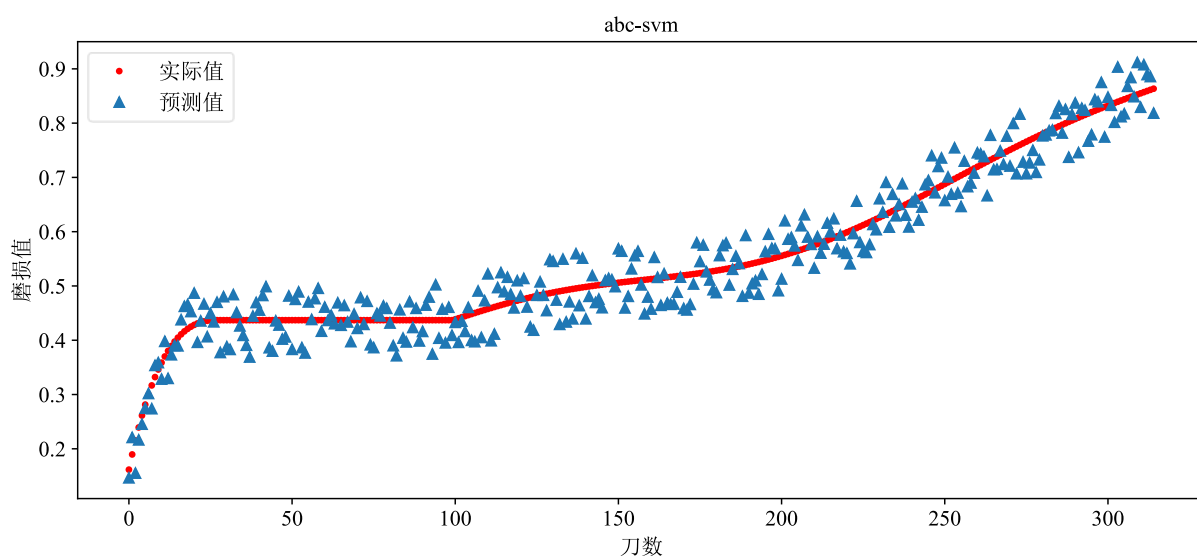


图 3.22 ABC-SVM 结果

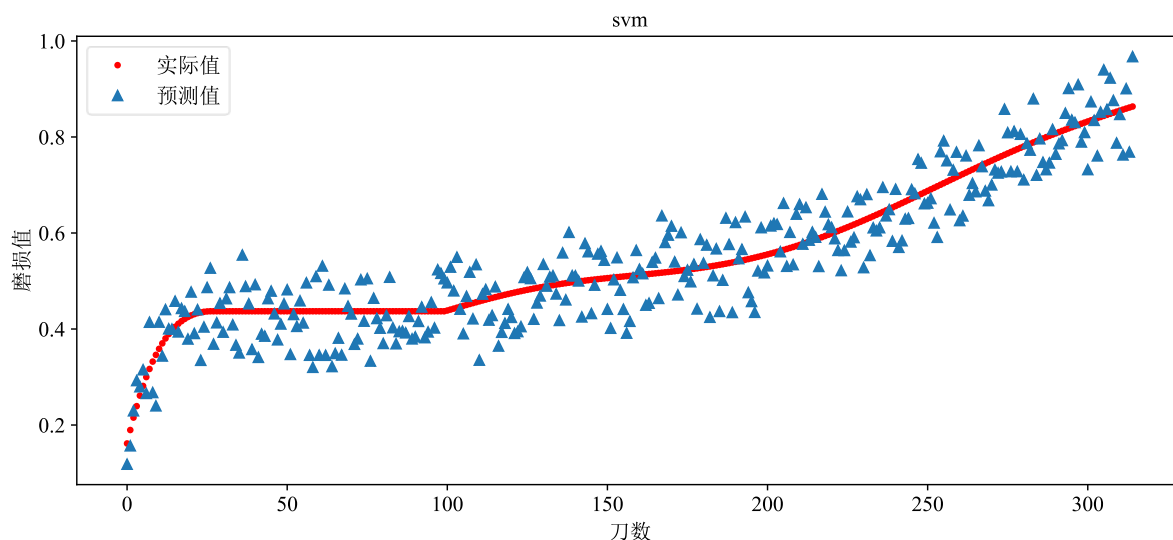


图 3.23 SVM 结果

从上述各图可以看出,改进 ABC-SVM 预测效果较好,而 ABC-SVM、SVM 相较于真实值偏离较大,效果不好。具体定量的比较分析见下节。

3.4.3 结果对比分析

3.4.3.1 改进 ABC 算法寻参的对比分析

为了说明改进的 ABC 算法对 SVM 的寻参优化的作用,对改进的 ABC、常规 ABC 进行对比测试,考察收敛性和识别精度。

多峰值函数求极值问题是测试连续优化算法的试金石,所以这里选用两个经典的多峰测试函数来测试算法的寻优性能,它们在各自的定义搜索空间内的最优解(最小值)都是 0。可以通过考察最终解与最优解之间的误差来考察其识别精度、通过考察收敛到最终解的迭代次数来考察其收敛性。显然,误差越小、迭代次数越少,算法越优秀。

测试函数如表 3.5 所示,它们的绘制图像如图 3.24 所示。

表 3.5 测试函数

函数	表达式	搜索空间	维度
Rastrigin	$f_1(x) = \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i) + 10)$	$[-5.12, 5.12]$	50
Ackley	$f_2(x) = -20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^n \cos(2\pi x_i)} + 20 + e$	$[-5, 5]$	50

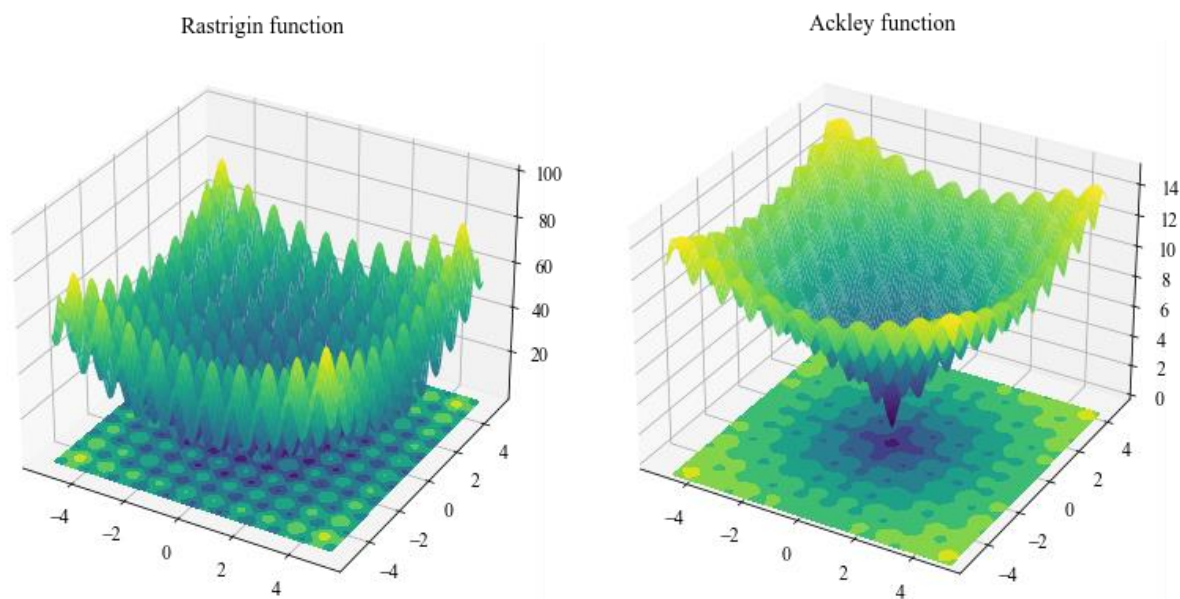


图 3.24 Rastrigin、Ackley 的函数图像

参照 3.4.2 节中的参数初始化步骤，本测试仍然设置改进 ABC、ABC 的初始参数为上表 3.4 所述。进行 10 次测试，取测试结果的数据均值。最终，两个测试函数的结果数据如表 3.6、表 3.7 所示。

表 3.6 Rastrigin 测试结果

算法	最优值	最差值	标准差	最小迭代次数	运行时间/ms
ABC	0.135	0.185	0.142	121	193
改进 ABC	3.126e-04	0.087	7.3e-04	127	204

表 3.7 Ackley 测试结果

算法	最优值	最差值	标准差	最小迭代次数	运行时间/ms
ABC	0.09	0.173	0.112	97	182
改进 ABC	1.546e-04	0.067	2.42e-04	103	197

上述两个表格反映改进 ABC 与 ABC 对于各测试函数的求解结果。可以看出，改进后的 ABC，最终解的标准差大幅小于标准算法，但迭代次数和运行时间略有增加，这是因为增加的混淆因子使得后期的个体有更大的可能被选择，所以需要多迭代几次完成收敛。但这样的好处是使得寻得的最优解不会因为超大个体而突然收敛，从而提高了识别精度。用牺牲的一点时间换取精度的提升是值得的。由此说明，本文对 ABC 算法的改进设计是有效的。

3.4.3.2 模型识别结果的对比分析

3.4.2 节的实例测试得到了三个对比模型的识别预测结果，结合 3.4.1 节的模型评估指标，对实例测试中使用的改进 ABC-SVM、标准 ABC-SVM、SVM 模型的识别结果进行评对比。

首先是各模型的训练时间对比，使用各模型训练多次，记录运行过程时间的均值，结果如图 3.25 所示，其中纵轴代表训练次数，横轴代表该次数下的训练时间。

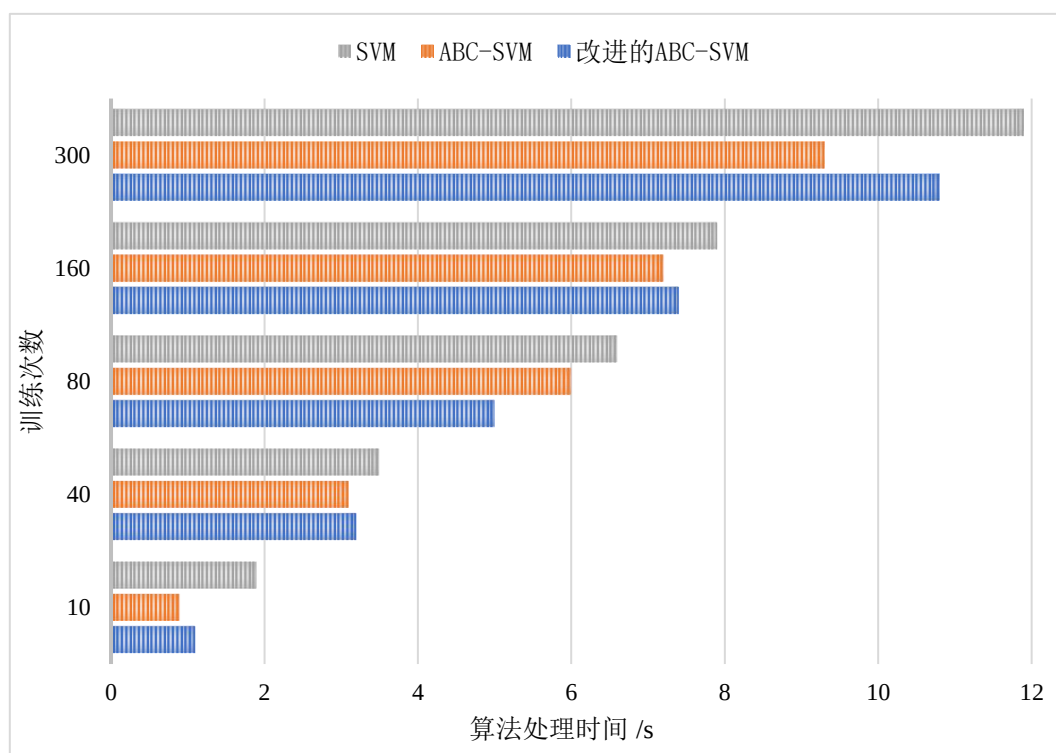


图 3.25 各模型训练时间对比

由于实验过程使用个人的台式电脑，计算能力没有云服务器强大，所以训练时间相对较长。从图中可以看出，改进的 ABC 算法应用于 SVM 寻参，虽然通过混淆因子使后面存在的个体被选中的概率加大，从而导致迭代次数增加，但最终的收敛时间增加得并不明显，对性能无碍。

然后考察模型的误差和拟合准确度。将各模型预测数据和真实磨损数据，使用 scikit-learn 库的 metric 模块，计算评估指标，最终汇总结果如表 3.8 所示。

表 3.8 模型评估结果对比

模型	评价指标		
	RMSE	R ²	全样本训练时间/ms
SVM	0.06078	0.83265	776
ABC-SVM	0.03871	0.87883	498
改进 ABC-SVM	0.01394	0.94041	514

从表中可以看出，当前测试集下，基于改进的 ABC 算法对 SVM 的寻参优化效果更好，模型预测结果更优秀。相比于未优化的 SVM、传统的 ABC-SVM 模型，RMSE 误差分别减小了

3.36 倍、1.78 倍， R^2 分数分别提高了 11.8%、6.2%。优化效果比较明显，且算法训练时间仅略有增加，所以综合来说最为优秀。

3.5 本章小结

本章设计和实现了带锯条的剩余寿命预测方案。首先，对带锯条的磨损过程作简要概述，进而分析磨损监测信号的选择过程、磨损和剩余寿命的定量计算。其次，简单介绍 SVM 方案原理，指出可行的改进方向。然后，着重进行方案的整体设计和实现：先进行信号数据的处理、相关度分析，确定跟磨损定量密切相关的信号；对于振动信号这样的含噪信号，先使用基于 CEEMDAN 的分解方式提取本征分量，对各分量进行相关分析，并通过曲线对比找出对于区分磨损更良好的分量用作特征值；提出改进的时序混淆 ABC 算法，优化 SVM 的参数搜索过程，得到 c 、 g 参数，建立磨损识别模型。

最后，经实验测试对比，本方案相较于未优化的 ABC-SVM、SVM，通过改进的时序混淆 ABC 算法优化模型参数，可以将磨损识别准确率提高到 94.01%，有较大的精度提升。

第4章 基于改进 LTTB 的海量数据处理的优化方案

第二章的监测系统中，现场频繁采集到的大量时序监测数据，直接使用会存在效率和性能低下的问题。

首先，在第三章中，寿命预测的数据处理过程容易出现耗时严重的问题。比如前期采集的数据信号多达十万条，在算法实现过程中明显存在运行时间较长的现象，所以需要预先对信号数据做优化处理，来降低后续算法的复杂度。

其次在系统交互产品的可视化绘制时存在一些问题：第一，数据趋势不明显且数据量级大、波动频繁，绘制得到的数据点折线连接后相互重叠，变得杂乱，无论是用于后续的数据挖掘还是直接进行绘制，趋势变化特征都不太显著，难以被人眼接受，不利于可视化效果的展示。第二，过多的数据点严重影响图形的绘制渲染性能。终端系统页面的可视化技术如 ECharts、HighCharts 等的绘制性能会随着绘制点的不断增多而快速下降，数据点过多时渲染会比较卡顿，而且系统内存占用高。第三，数据点是时序频繁变化的，新增数据时不仅会绘制新数据，还会连带着所有的旧数据重新渲染，加剧了第二个问题。比如在第二章的交互应用产品如 Web 端、APP、小程序中，需要进行可视化绘制，来直观地向人展现数据的趋势和变化。但是这些大量数据在可视化绘制时容易出现耗时严重、内存占用过大的性能问题。

在第一章的研究现状已经分析过，目前比较好的海量数据处理优化方案是进行降采样，降低数据数量级。所以本章提出一种基于改进最大三角形三段算法的海量数据优化方案，用以处理海量数据，提升数据处理算法的速度，并进行交互应用的绘图优化。

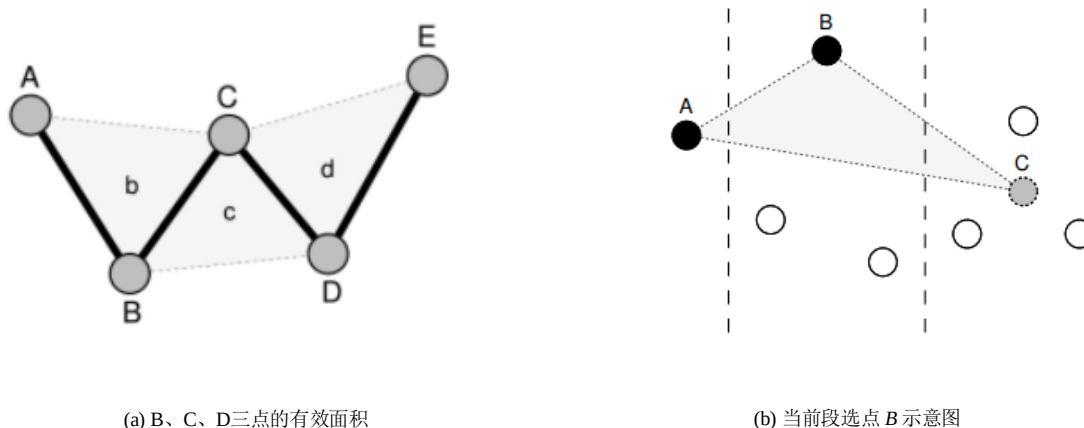
4.1 LTTB 算法的方案分析

使用降采样算法来对海量数据处理做优化，可以降低数据量级，是治本之策，但必须注意的是，采样后的数据要尽可能保留原始数据的细节特征，否则对于后面的数据分析和处理就是一种破坏，得不偿失。目前得到了广泛使用的是冰岛大学的 Steinarsson Sveinn 在 2013 年提出的最大三角形三段（Large Triangle Three Buckets, LTTB）算法^[57]。

4.1.1 LTTB 算法步骤

LTTB 算法通过计算最大有效面积来选择采样点，一个点的有效区域定义为与该点相邻两点所围成的三角形，示意图如图 4.1 所示。

通过选取当前段内使得有效区域最大的点，可以有效地保持住点与点间的距离、角度和轮廓特征。相比于均值采样等简单采样方法，虽然 LTTB 较为复杂，但非常适合处理海量频繁变化的时序数据。



(a) B、C、D三点的有效面积

(b) 当前段选点 B 示意图

图 4.1 有效区域示意图

LTTB 算法的步骤如下：

- (1) 确定分段大小。为了便于改变分段大小，把分段大小当作参数传递给算法，这样如果需要采样 100 倍，只需要传递参数 *threshold*（总数据大小/倍数）。把 *total* 个数据点平分到这所有的段内，共分成 *threshold* 段。另外为了确保数据分完后可以选中首尾，首和尾都单独占据一个段。
- (2) 选中第一个点（也就是第一个段）。
- (3) 从第二段开始，遍历段中所有点，计算每个点的有效区域面积，并选取面积最大的点作为该段的选中点（采样点）。其中，所述的有效三角区域正如图 4.1(b)所示，是指以前一段选中点 *A*、当前点 *B*、后一个段的平均点 *C* 三个点为顶点的三角形区域。
- (4) 遍历，直到选中最后一个点（也就是最后一个段），算法结束。

LTTB 算法的流程如图 4.2 所示。

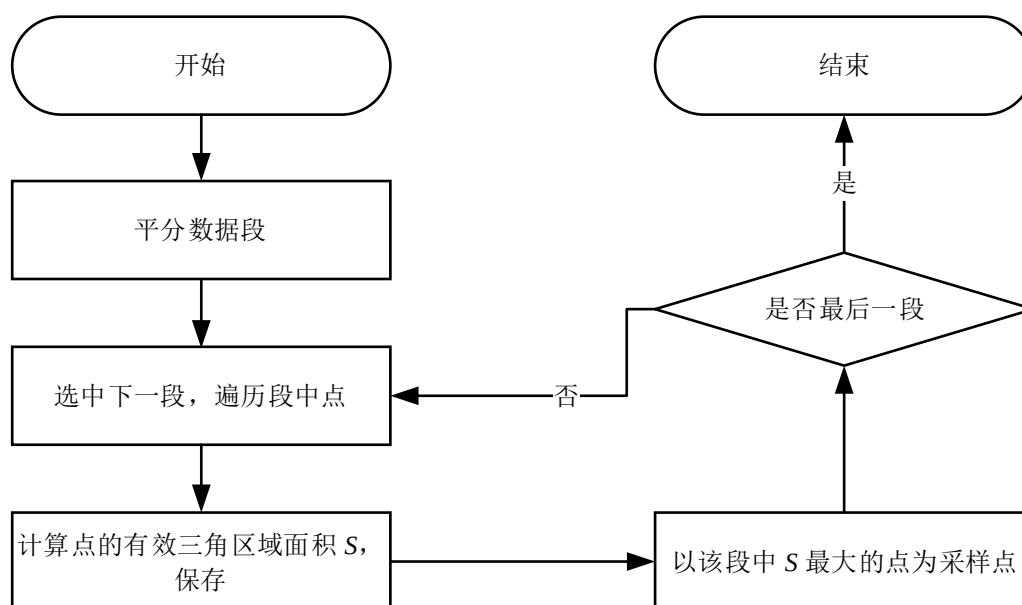


图 4.2 LTTB 算法流程示意图

4.1.2 LTTB 算法的问题和改进方向

在实际的工业场景中，标准的 LTTB 算法在数据陡峭处容易丢失“尖峰”，并不完美。广州虎牙公司的陈键冬等人采取在 LTTB 处理之前先进行一次均值采样的改进方法，如分段聚合近似（PAA），意图先降低尖峰后再进行 LTTB 采样^[58]。但是 PAA 容易丢失数据曲线更多的形状细节，得不偿失。王立峰等人先对时序数据以当前数据点与前一数据点的时间差值为权重值，选择某个权重和作为阈值，把数据切分成段，来分段进行 LTTB 采样^[59]。其分段的想法是正确的，但是以时间差值为权重忽视了尖峰的产生是由于数值的突变，而非时间差值过大的特点。而且阈值的选取并不严谨，没有合适的固定值，实际在处理数据进行设置时也不容易快速选取到合适的阈值。

基于此，本章提出了一种基于距离和离散程度的 LTTB 算法的改进设计，用以解决上述存在的缺陷。

4.2 基于改进 LTTB 算法的海量数据处理的优化方案设计与实现

4.2.1 总体设计流程

本章基于改进的 LTTB 算法的海量数据处理的优化方案，核心点在于对 LTTB 的改进设计，并在此基础上应用于数据处理和可视化绘制的实践：

第一点，基于改进的 LTTB 算法进行降采样，用于减少海量时序数据的量级，同时尽可能地保留数据曲线的细节特征，具体量化来说就是序列的统计特征（如均值、极值、方差）以及形态特征（如距离、角度、斜率）。这样可以为后续的数据处理算法减负，而且可以在尽量不损失原始数据曲线形状的前提下，减小绘图所用的数据点，把数据量级控制在一个合理的范围内，缓解渲染性能问题。

第二点，针对海量数据绘图问题，基于 ECharts 可视化技术进行改进：首先，提出实例拆分的优化设计。对降采样后的数据设定阈值，将整体数据进行进一步的切片，对每一个切片数据单独创建实例进行渲染绘制，最终将各个子图表定位叠合为一个完整的图表，从而完成绘制。阈值的确定需要由具体的绘图试验来选取，选择到一个合适的阈值数据点量级。这样不仅可以人为地把每个渲染实例的数据量级控制在一个可以接受的阈值范围内，缓解绘图性能问题，而且在数据变化时不需要完全重绘全部实例，只需要关注自己所在阈值段的子实例的数据变化即可，这种局部刷新渲染的方式，缓解了重绘时的性能占用。其次，选择 ECharts 技术是因为其内部进行了渲染优化，性能相对来说比较好。比如它做了脏矩形渲染优化，也是一种局部渲染手段，用于控制只重绘画面改动部分，而不是直接全屏刷新。带来了比较好的页面性能提升，交互动作也更加流畅。

第三点，将改进的 LTTB 降采样算法应用于 SVM 模型的前期数据处理实践，由于与第三章关联大，故只在最后的测试部分进行实验测试，不纳入方案之中。

最终该设计的整体流程如图 4.3 所示。

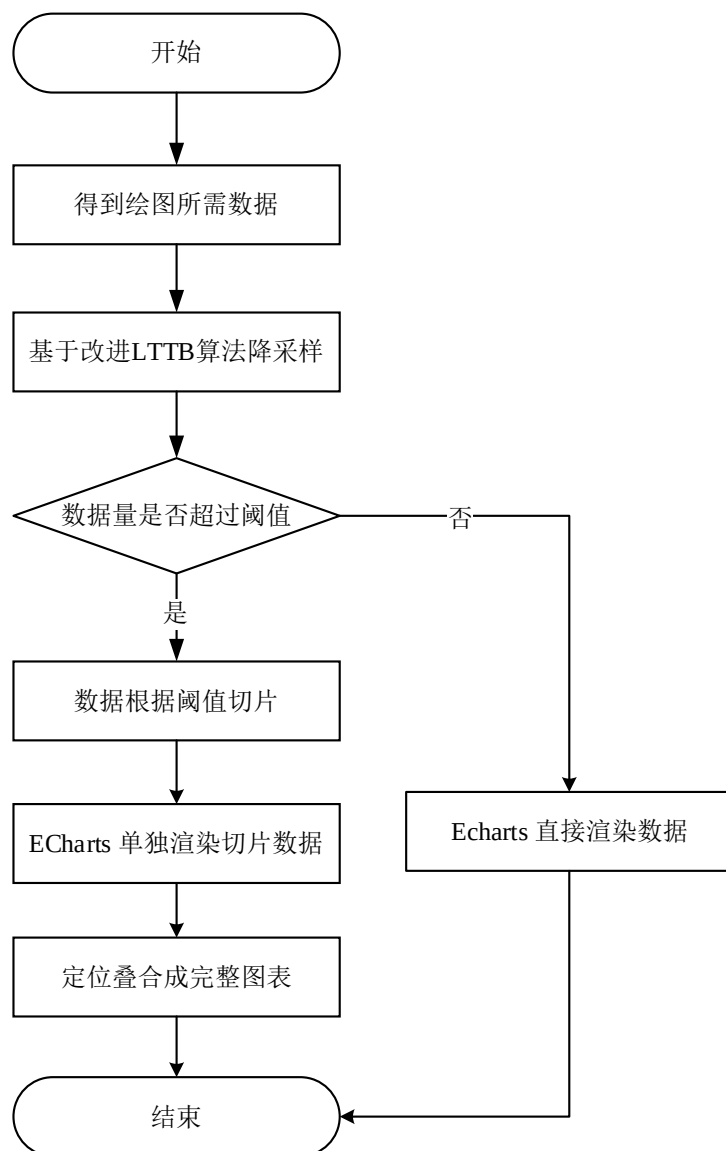


图 4.3 总体设计流程

4.2.2 改进的 LTTB 算法设计

LTTB 算法的问题在于平分所有段,这在一些数据曲线比较陡峭的情况下表现不好。因为在平缓处较少的点就能反映出细节,而在陡峭处则需要较多的点,而 LTTB 为了节省分段操作需要消耗的时间,直接简单地进行了平分,就会导致一些陡峭的尖峰、波谷被意外过滤掉。

本设计在此做了基于距离和离散程度的动态分段的改进,主要原理是先近似拟合一条渐近线,用数据点到渐近线的距离来表征其离散程度。将所有点到线的距离之和进行三等分,那么在均匀分段时,只要拟合线尽量准确,那么遇到的尖峰由于远离渐近线,就会有极大的概率单独落在一个等分段,从而被采用。而其他的很多点趋近于拟合线,不会或只会产生较小的距离。于是对 4.1.1 中算法步骤(1)可以拆分改进如下:

- 1) 还是把分段大小当作参数传递给算法, 先和 LTTB 一样平分各段, 首和尾都单独占据一个段。
- 2) 从第二段开始遍历段中所有点 (共计 $m = \text{total} / \text{threshold}$), 以前一段的选中点 A 、后一个段的平均点 C 作渐近线 L , 计算得到段中每一个点到直线 L 的垂直距离得到距离数组 SS 、所有垂直距离之和 SSE 。
- 3) 以 $1/3SSE$ (记作 target) 为目标值, 将数组 SS 进行值的三等分。对于寻找到的三等分索引, 也作为分段点进行分段, 由于垂直距离一定程度上表征了相对于渐近线的离散程度, 所以一个 SSE 的分段切分成了三个比较均匀的 target 的分段, 比较动态地应对了急剧变化的恶劣情况。其中, 如果不存在恰好等于 target 的三等分点, 就取左右比较近似的点。
- 4) 这样进行分段后, 再进行标准 LTTB 算法的步骤(2)即可。其中, 该算法拟合数据曲线的方法比较粗糙。如果为了保证拟合准确, 可以采用一些更复杂的数据曲线拟合算法。最终改进后的算法流程如图 4.4 所示。

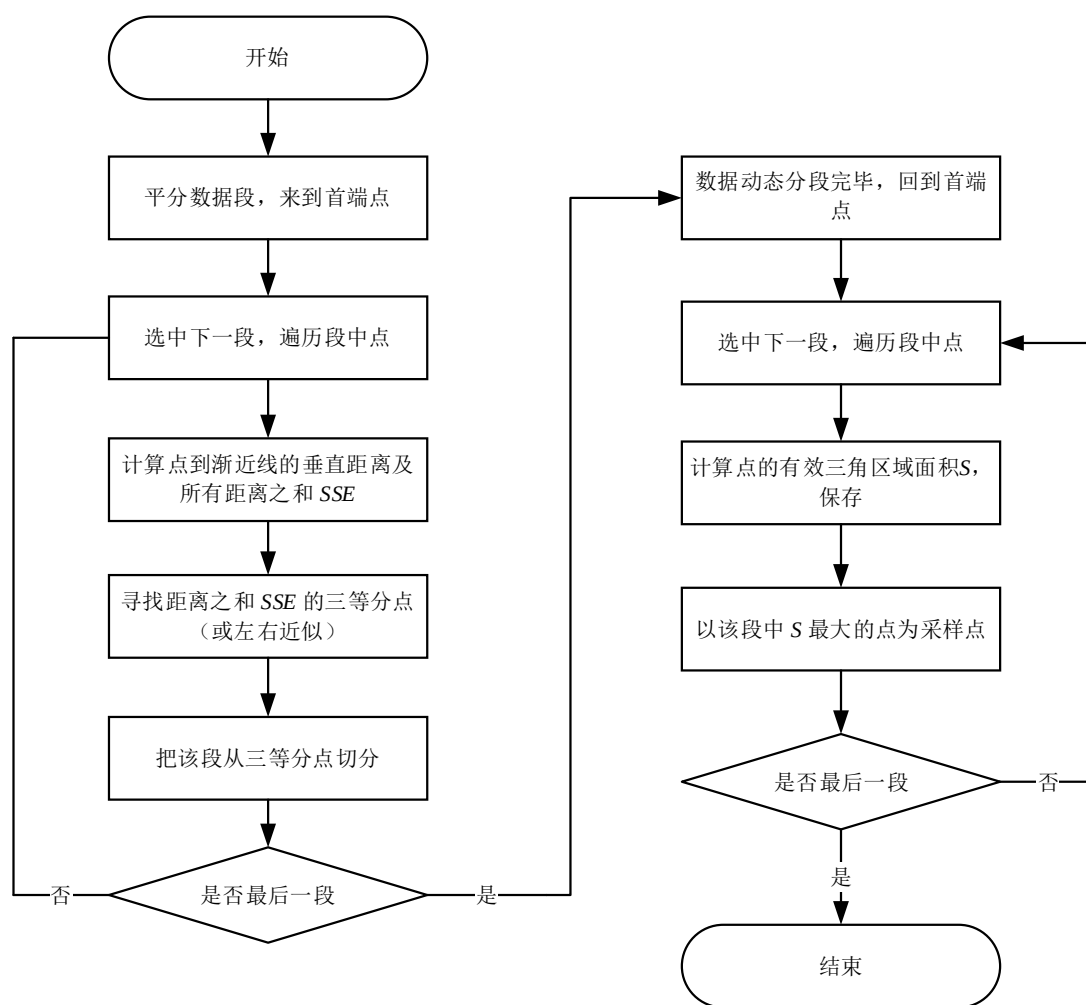


图 4.4 改进的 LTTB 算法流程图

4.2.3 ECharts 的数据可视化优化设计

数据可视化是将抽象的数据转化为直观的视觉表达，并且允许人进行一些基本的交互，最常见的就是各种图和表。近年来开发人员和用户期望的可视化产品不再仅仅是制作图形的工具，对交互和性能有了更高要求。

由百度开发和开源的 ECharts 在交互、性能方面有很大的优势，其自研的 ZRender 渲染引擎可以自由切换 Canvas/SVG 的渲染方式，开发体验流畅。目前 ECharts 在图表的动态设计、视觉、交互性能、开发体验方面进行了大量优化，这也是本章选取该可视化技术的原因。

4.2.3.1 内部渲染优化

ZRender 内部渲染引擎进行了多个渲染性能方面的优化，主要有两点：

(1) 使用定型数组做静态存储

内部图形数据的存储从常规的动态数组改为使用定型数组，用这种类型来分配、读取、写入数组，得到的数据数组（二进制数据）可以直接传给底层图形驱动程序 API，与显卡直接通信交互来驱动绘图，从而降低绘制时的内存占用和开销。这一优化在数据量很大的时候，优化提升越明显。

(2) 脏矩形渲染优化

ECharts5 新增的脏矩形渲染，提升了 Canvas 的大数据量重绘性能。脏矩形渲染算是一种懒加载的渲染手段，目的是只更新有改变的区域，而不是去更新全屏的区域。在 Web 领域，浏览器本身就实现了脏矩形渲染，比如底层的 Canvas/WebGL 为主的绘图接口，ECharts 在使用 Canvas 进行渲染时就可以选择是否开启脏矩形渲染。

Canvas 这样的 2D 图形渲染中，显示列表是常用的组织方式，它是一个树状的节点结构，每个节点称为显示对象，如下图 4.5 所示。

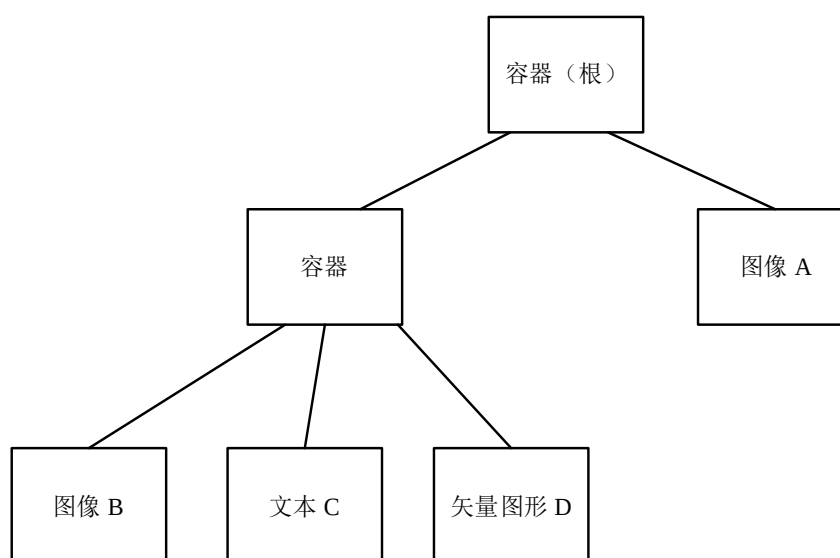


图 4.5 显示列表的组织形式

容器只用于位置定位，末端叶子节点的普通显示对象用于显示，比如图片、文本、矢量图位置，其位置和大小取决于父容器的位置和大小。

把所有的容器和显示对象渲染完毕后，后续产生更改要进行重绘。通常采用的是两种方式：全屏刷新渲染和局部刷新。全屏刷新渲染比较简单，但是每次刷新都必须全部刷新，没有必要。脏矩形渲染是局部刷新，可以先设定一个时钟频率如 60hz，每次时钟到来不直接清空刷新，而是先计算产生更改的区域，称为重绘区。找到所有重绘区，然后只需要重新渲染与该区域有交集的显示对象。关键步骤有两个：

获取重绘区，常用的是矩形选区方案，就是获得显示对象相对于屏幕的矩形位置信息。可以以屏幕左上端点为原点，计算对象的矩形数值，即相对于原点的上、右、下、左边框，由这四条边组成的矩形框就可以表征这个区域的位置和大小。每当该区域发生改变，就重新计算新矩形位置。前后两个矩形就成为该变更内容的重绘区。比如，图片从左向右平移后，就会产生两个重绘区，如图 4.6 所示。

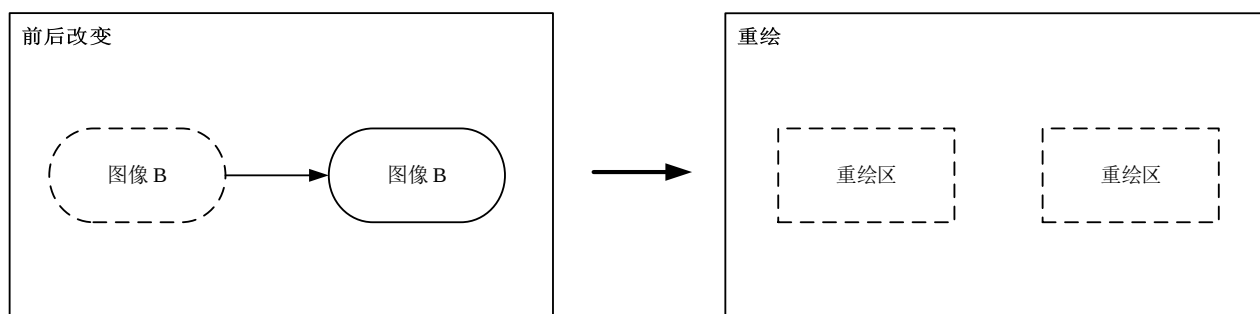


图 4.6 重绘区域示意图

合并重绘区，是把所有零散的重绘矩形合并成几个大区，避免太多次的重刷新。原则上如果两矩形合并后总面积小于各自面积之和，就说明二者有交集，允许合并。最终，所有的零散重绘区域会合并得到几个较大重绘区域，如图 4.7 所示。

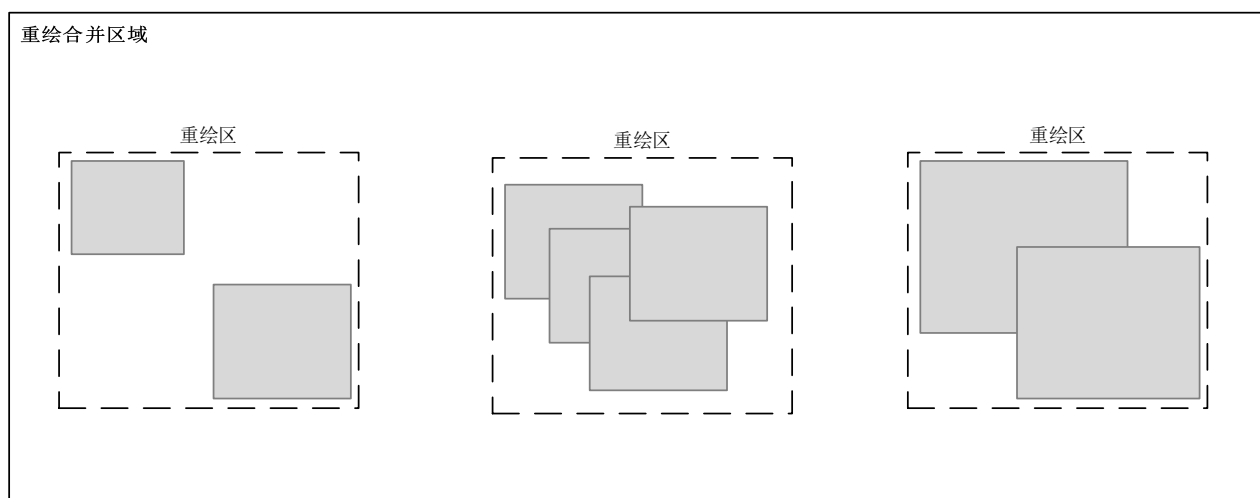


图 4.7 重绘合并区域示意图

4.2.3.2 实例拆分优化设计

本设计针对降采样后仍然可能数据过大的问题，设定阈值，把数据进行进一步的切片，对每一个切片数据单独进行渲染绘制。同时针对 ECharts 绘制大数据时耗时剧增的问题，提出将整体数据切分，而后分别创建实例进行渲染，最终将子图表定位叠合为一个完整的图表，完成绘制。具体实现步骤如下：

步骤 S1：数据切片。设定阈值，对于降采样后仍然较大的数据，进行切片操作，然后进入下一步骤。实验表明，同时进行的绘图实例不宜过多，而且数据量在百级以下绘图消耗几乎持平，所以本设计设定：切分的数据片个数最大阈值为 10、单个数据片容量最小阈值为 200。

步骤 S2：绘制渲染。本设计以 ECharts 绘图为例，最主要的耗时操作是绘制数据点。不过本设计由于进行了降采样算法步骤，在考虑实际的效果时，还要考虑到对数据进行处理算法耗时。该绘制过程主要包括：

S2-1：创建 ECharts 实例，对给定的每一个切片数据单独进行渲染绘制。

S2-2：使用层叠样式表语言，定位叠合每一个子图表，形成一个完整的图表。具体而言，将父容器设置为相对定位，将每一子图表绝对定位到容器内的同一位置，然后把图表的轴线刻度去除多余仅保留一份，就可以进行叠合。虽然是多个子图表，但在视觉上图表是完整的。

最终的流程图如图 4.8 所示。

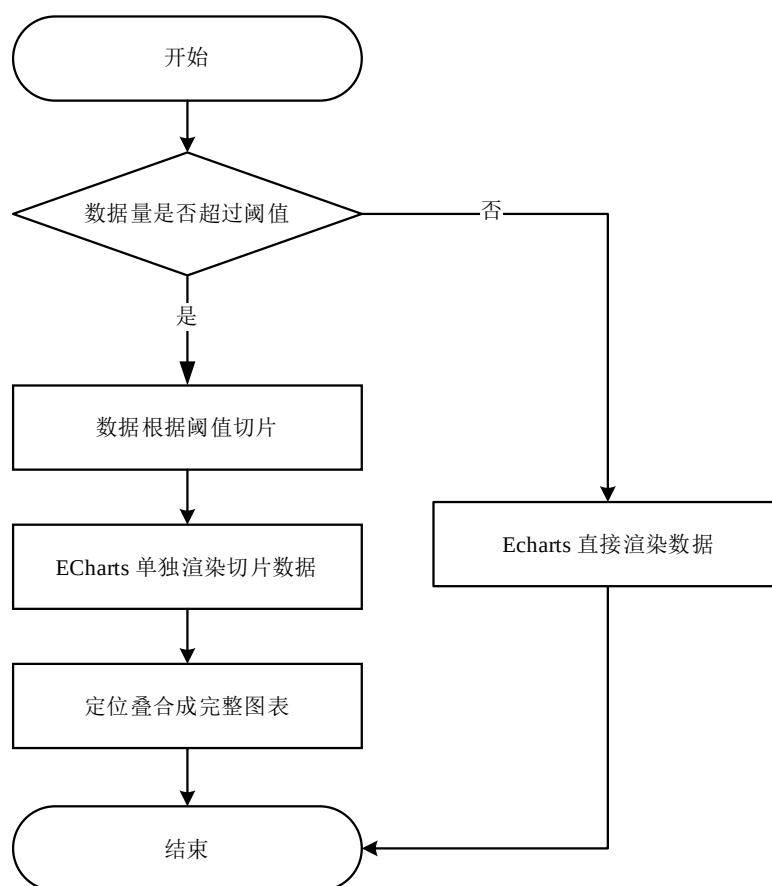


图 4.8 实例拆分优化流程图

4.3 实验测试和结果分析

4.3.1 改进的 LTTB 算法降采样测试

测试一：对比测试 LTTB 算法改进前后的数据点保留效果，尤其是对峰点的覆盖效果。

(1) 数据采样效果

模拟生成 5000 个频繁随机变化的“恶劣”数据作为原始数据，绘制波形曲线，图例以蓝色表示。分别使用 LTTB 算法和本文改进的 LTTB 算法，对原始数据降采样至 1000，绘制降采样后的波形曲线，图例以红色表示。最终得到对比图如图 4.9、图 4.10 所示。观察红色曲线对原始的蓝色曲线的覆盖情况，覆盖得越多、保留下来的尖峰越多，说明降采样效果越好。

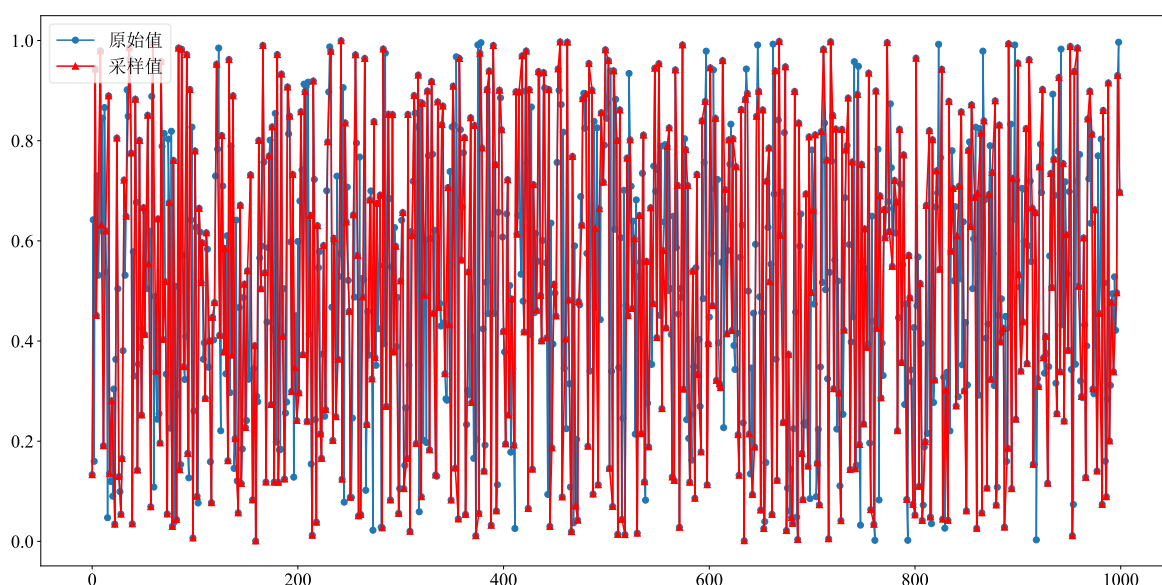


图 4.9 改进的 LTTB 算法降采样前后的数据点对比

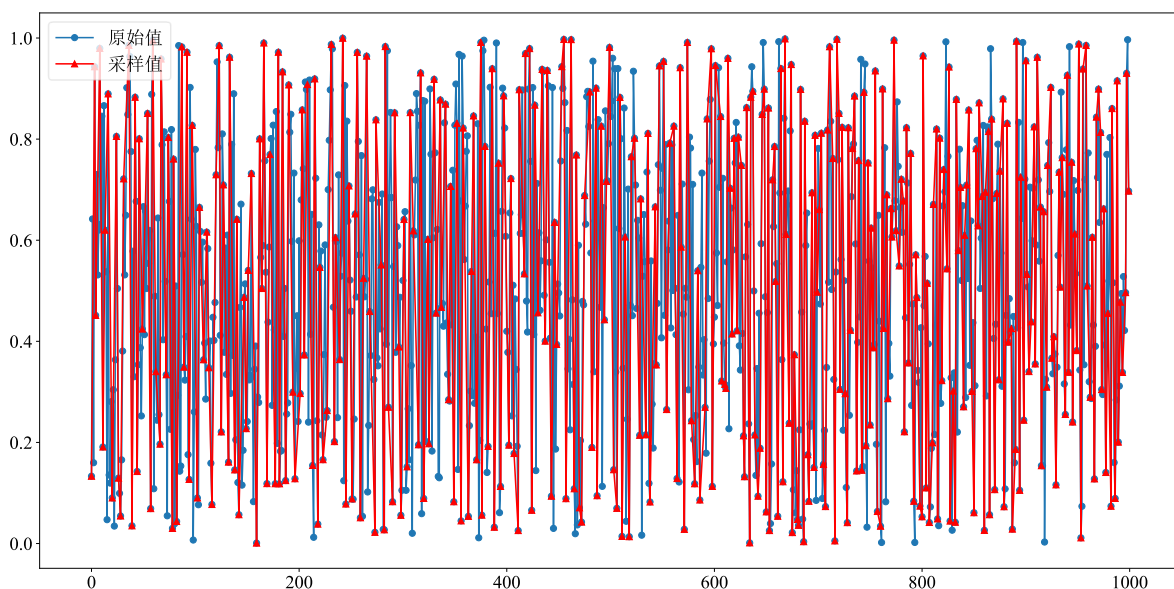


图 4.10 原始的 LTTB 算法降采样前后的数据点对比

上面两个图中可以看出，对同一段数据采用同样的采样倍率，改进后的 LTTB 算法降采样后对原始数据曲线的覆盖率更高，效果更好。尤其是原始算法（第二张图）的前段、中后段的大量下部尖峰，采样后都被抛弃掉了，对于原始数据的特征细节破坏较大，远不如第一张图的效果好。

(2) 峰点覆盖情况

为了直观量化对尖峰顶点的保留效果，使用 SciPy 库的寻峰函数 findPeaks 找出数据的所有峰点，其中，定义峰点为与左右两点绝对值相差均超过 0.2 的点。

分别对上述两图中，原始数据曲线（图 4.9 和图 4.10 的蓝色数据曲线）、改进 LTTB 的采样数据曲线（图 4.9 的红色数据曲线）、原始 LTTB 的采样数据曲线（图 4.10 的红色数据曲线）进行寻峰测试，得到的测试数据如下表 4.1 所示。

表 4.1 采样前后对峰点的保留程度

算法	原始峰点	采样后峰点	覆盖率
LTTB	151	112	74.1%
改进 LTTB	151	132	87.4%

可以看出，三条曲线寻峰得到的峰点个数分别为 151、132、112。量化来看，改进 LTTB 的采样数据对原始峰点的覆盖率为 87.4%，原始 LTTB 的采样数据对原始峰点的覆盖率为 74.1%，相比之下提高了 13.3%。

而且，由于本测试所用数据变化十分频繁，这种频繁变化的数据本就不易采样，而改进的 LTTB 算法降采样至 1000 个数据点后，仍然能保留相当程度的细节特征，这更加能说明本算法的优越性。

4.3.2 降采样前后的 ECharts 绘图测试

测试二：对比测试使用改进 LTTB 算法降采样前后的 ECharts 绘图效果。

对于大量数据的可视化绘图，实际开发过程常常不需要直接绘制信号波形，而是绘制其某种特征或趋势，不需要采用类似测试一的频繁变化数据。为了使测试数据效果更好，本测试使用 100000 个时序数据，并且采用平缓趋势变化的曲线。这与实际的页面应用绘图场景比较接近。

(1) 数据采样效果

基于过去某一毫秒时间戳，每一毫秒随机生成数值，遍历 10 万次就得到原始时序数据。对 10 万个数据采取不同的采样倍率：1（也就是原始数据不进行采样）、10、100、500、1000、10000，使用改进的 LTTB 进行采样，并按照上节的后续绘图方法进行绘制，得到的效果如图 4.11 所示。

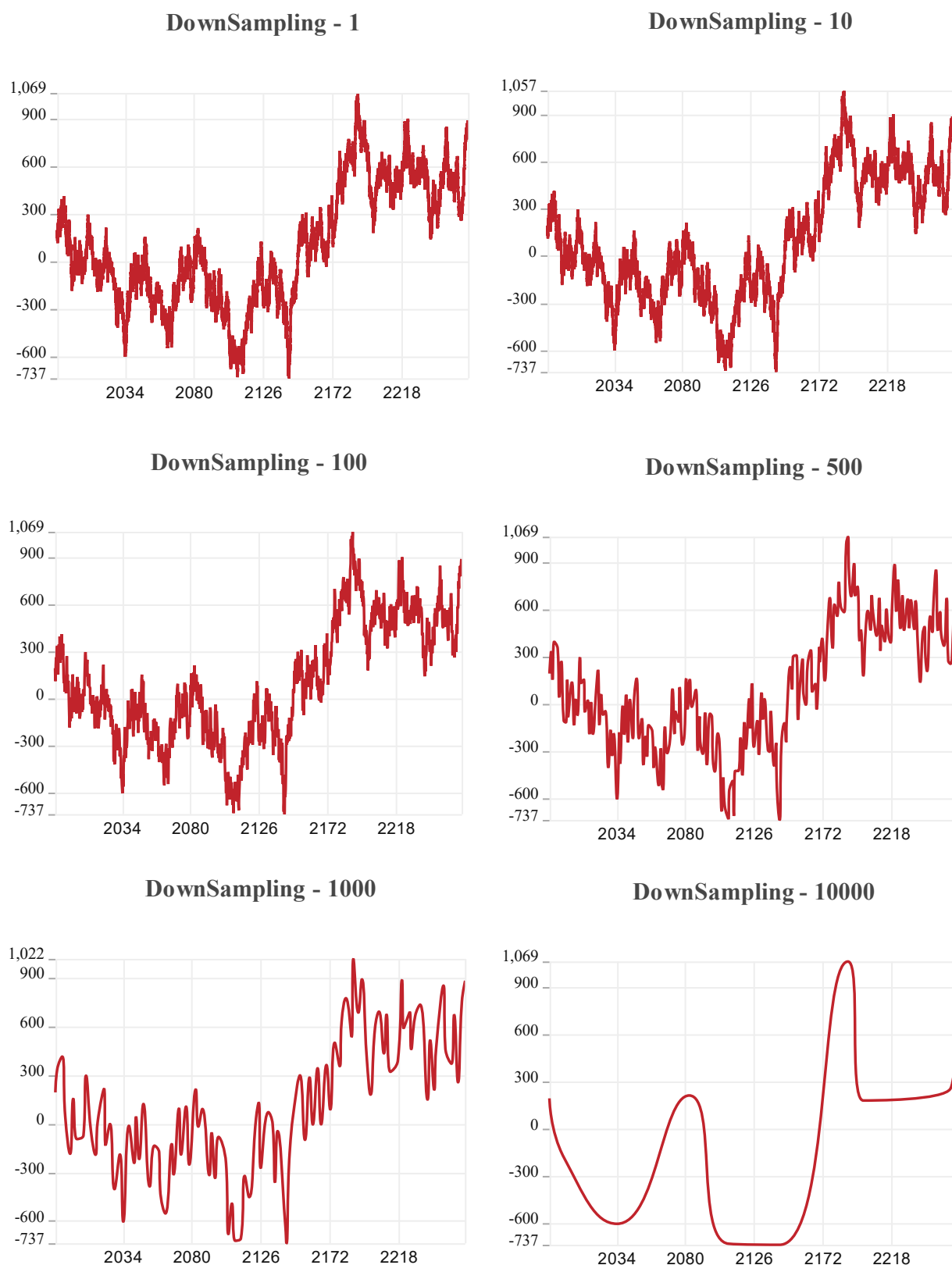


图 4.11 不同倍率下采样数据的绘图结果

上图中的六个折线图，第一张图是原始数据，其余五张均是采样图。其中第一、第二张图片的交互操作十分卡顿，几乎无法直接使用，比如选择数据点时鼠标的移动操作困难，所

以使用时进行降采样是很有必要的。第三到第六的四张图肉眼观察来看交互十分流畅。

而就图形的趋势特征来看，除了最后一张图采样倍率过高导致变形严重，几乎不能直接使用之外，第三、第四张图的细节效果几乎没有损失，波峰和波谷点保留完整，完全可以用于替代原始图像。第五张图虽然有些“毛刺”丢失，但是形状和趋势仍然保留得很好，且降采样了 1000 倍，仅仅用到 100 个数据点，就可以展现数据曲线的趋势，非常适合对性能和交互要求严格的场景。

(2) 测试绘图性能

除了绘图后的交互效果，还要考虑绘图过程的耗时和内存占用情况。耗时越小、占用越小，绘图性能越好。在页面绘制过程中，打开浏览器控制面板里的性能选项卡，可以查看页面绘图的性能数据，汇总如表 4.2 所示。

表 4.2 绘图性能数据

采样倍率	采样后数据个数	降采样耗时/ms	绘图耗时/ms	绘图内存占用/MB
1	100000	0	501.68	134
10	10000	14.62	132.25	76
100	1000	9.86	33.35	56
500	200	6.03	7.85	37
1000	100	1.32	7.27	26
10000	10	1.27	6.01	21

上表中记录了各采样倍率下绘图过程具体的性能数据，可以看出，采用本设计中的降采样及数据切分的绘图优化方法，对数据每降采样一个数据量级后，可以减小大约 70%-80% 的绘图耗时，同时降低大约 20%-30% 的内存占用，效果很好。

结合图 4.11 和上文的分析，在降采样 100~1000 倍时，即降采样到百至千级别的数据点个数时，细节特征几乎没有损失。同时绘图总耗时也可以保持在十级别，内存占用没有大幅增加，各方面指标都在一个比较好的水平，实现了海量数据可视化绘图的优化目标。

4.3.3 降采样前后的 SVM 模型数据处理测试

测试三：使用改进 LTTB 算法降采样，对比前后数据在 SVM 锯条寿命预测模型的表现。

在第三章的锯条寿命预测模型的研究过程中，对前期采集数据的准确性要求较高，为了控制变量和使研究结果尽量精确，并未直接在数据处理时进行降采样实践。所以本节对此进行测试，将降采样算法应用于第三章预测模型的前期数据处理，验证其效果。

实际上，第三章中前期数据的相关处理过程中，算法处理困难、需要优化处理的主要是信号分解和峭度值提取的操作。后续的计算过程是以峭度值为特征、以磨损数据为样本，量级不大，不需要特殊处理。所以本测试只针对数据的信号分解和提取峭度操作，考察二者的处理速度、得到的峭度值之间的差异。其中，采样后数据的处理速度越快、提取到的峭度值

越接近原始数据，说明越优秀。

(1) 数据采样

仍然取第三章中的 c1 刀具的样本数据，它共有 315 个采集数据文件，前中后三个阶段分别随机选取 2 个进行测试，共得到 6 个测试文件，每个文件中有 10 万条采集数据，以此为原始数据。结合上述两个测试结果，为了保留原始数据得准确性，初步设定采样倍率为 10，分别对 6 个测试文件的数据进行降采样，得到采样数据。

(2) 算法处理时间

根据 3.3.3 节的算法实现，分别对原始数据、采样数据进行 CEEMDAN 分解，接着对分解得到的 IMF 分量进行峭度值提取，记录并得到采样前后的算法处理时间对比如图 4.12 所示，其中纵轴代表每一个测试文件，横轴代表该测试文件数据的算法处理（包括信号分解和提取峭度）时间。

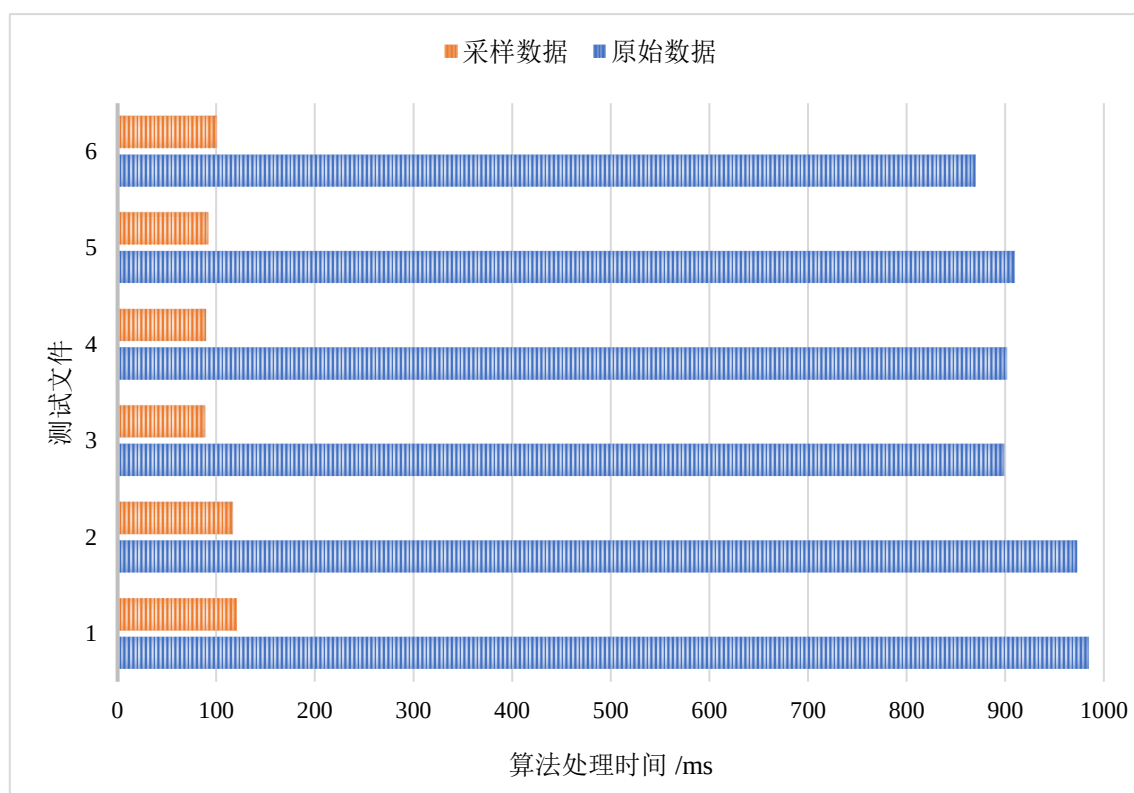


图 4.12 采样前后的算法处理时间对比

可以看出，降采样 10 倍后，算法处理速度提升得非常明显，有近乎一个数量级的提升，这是由于采样后数据量级的减小。不过在速度提升的同时，还要保证对后续处理不会产生较坏的影响，所以下面进行提取峭度值的结果测试。

(3) 峭度值结果

对步骤 2 中得到的峭度值结果进行统计分析，计算采样数据峭度值相比于原始数据峭度值的相对误差（取绝对值），得到采样前后的峭度值对比数据如下表 4.3 所示。

表 4.3 采样前后提取的峭度值对比

序号	原始数据的峭度值	采样数据的峭度值	相对误差
1	0.353	0.362	2.54%
2	1.125	1.167	3.73%
3	2.764	2.613	5.46%
4	3.167	3.256	3.09%
5	4.113	4.362	6.05%
6	3.864	4.025	4.16%

可以看出,相比于原始数据,降采样 10 倍后的采样数据提取得到的峭度值,相对误差普遍在 6%以下,是一个可以接受的范围,不会过多影响对磨损情况的判别。

综上所述,以上三个实验测试显示,本章设计可以大幅减少第三章 SVM 模型的前期处理数据,在峭度值结果误差较小的前提下,提升算法效率。还可以将第二章可视化绘图数据降采样到百级别数据点,且细节特征仍然保存良好。这说明本设计取得了预期的良好效果。

4.4 本章小结

针对海量监测数据量过大导致的数据处理开销大、可视化绘制性能差的问题,本章设计和实现了基于 LTTB 降采样算法的海量数据处理优化方案。首先,概述了传统 LTTB 算法的原理,并指出其问题和可行的改进方向。然后,着重进行方案的整体设计和实现:为了采样后尽可能保持原始数据曲线特征,对 LTTB 算法做出改进,通过基于距离和离散程度的自适应选取分段点的改进设计,减少 LTTB 因平分各段导致的“尖峰”丢失问题;为了克服海量数据可视化绘制的性能问题,对数据点设定阈值,拆分成多个绘图实例单独进行绘制,再定位叠合成完整图形。

最后,经实验测试对比,本方案可以在保留大多数细节特征的基础上,将算法处理数据至少减小一个数量级,相较于原始 LTTB,改进的 LTTB 算法对原始数据峰点的覆盖率提高了 13.3%,大幅提升了对原始细节特征的保留率。同时,将该改进算法应用于实践:对 SVM 模型的前期采集数据十倍降采样,大幅提升算法处理的速度,而峭度值的处理结果相对误差仅仅只有 6%左右;可视化绘图时,降采样后再结合实例拆分的优化设计,可以将绘图数据点减小到百至千级别数量级,和原始数据相比,绘图耗时减小数十倍,内存占用减小超过 30%,效果良好。

第5章 总结与展望

5.1 工作总结

针对当前传统的锯床产业存在的数字化程度不高的问题，本文基于物联网技术开发了一套功能完善的带锯床在线监测系统。物联网网关通过 Modbus 和锯床 PLC 控制器连接，实时获取锯床工作过程中的各类传感参数，并通过 MQTT 协议传输到设备云平台，业务云平台对其进行处理分析后，交付给用户侧的 Web 系统、小程序应用来进行展示。

进一步地，本文以监测系统为载体，对采集到的锯床数据进行处理和分析。从解决工业现实问题出发，在锯条剩余寿命预测模型、海量数据的优化处理方法上进行了研究。

总的来说，本文设计的系统和研究符合市场需求，有利于提高生产效率和缩减工厂成本，有较好的现实意义和创新性，其中主要工作和创新成果在于：

- (1) 满足了实际锯床工业场景的各种需求。丰富和完善带锯床监测系统的交互功能，比如视频监控、大屏设计、可视化效果。开发丰富的用户侧智能应用，包括 Web 系统、小程序等，根据企业用户需求，还可以一键移植开发 APP、工业 PDA 端的产品应用。
- (2) 针对带锯条健康状态无法直接有效获得的问题，对带锯条剩余寿命的预测方案做出详细研究，实现预测性维护。通过对磨损监测传感信号的相关性分析选择合适的信号输入、对磨损量的定量分析选择合适的模型输出，构造数据集使用 SVM 回归算法来训练磨损识别模型。针对振动信号干扰强的问题，使用 CEEMDAN 方法预先处理后再提取特征作为输入；针对 SVM 模型寻参过程的敏感问题，设计改进 ABC 算法，优化 SVM 的寻参过程，大幅提高模型泛化能力和识别精度。
- (3) 针对海量监测数据的处理开销大、可视化绘制性能差的问题，改进设计 LTTB 算法对数据进行降采样，尤其是在数据曲线尖峰位置做出基于距离和离散程度的改进设计，从而保留住尖峰点，更大程度地保留数据曲线的细节特征。采样后数据的量级减小，提高了后续数据处理算法的性能，同时经过实例拆分优化，可视化绘图操作性能得到大幅提升，从根本上得到了优化。

5.2 工作展望

本文设计和实现的带锯床在线监测系统功能完善，完成了实际需求，有助于锯床企业实现智能化转型。功能完善的在线监测系统提供各种人机交互功能，提高了锯切生产的效率，降低生产环节的故障。通过带锯条剩余寿命预测功能，识别和预测锯条的磨损状态，可以做到及时更换锯条、预测性维护设备。通过海量数据的降采样研究，提高算法处理速度，使得剩余寿命预测和数据可视化功能性能更优，更加贴合工业生产实况。但是由于本人水平有限，

论文仍然存在一些不足，后续可以进行以下完善：

- (1) 目前的在线监测系统中,由于实际生产的成本等限制,数据信号的采集和上报不连续,这抑制了数据平台的分析能力,需要后面的实际生产环境中加以克服。
- (2) 锯条真实磨损量的模型有赖于物理实验对刀具的监测,相关的权威试验还不够,这受限于研究实验条件,后续需要加强实验研究。
- (3) 目前对于大量监测数据的处理是先上报到云平台再处理,这极大浪费了数据传输带宽,将来边缘侧设备的算力提高以后,可以把本文的降采样处理、特征处理等内容交由边缘侧实时完成。

参考文献

- [1] 郭北涛. 国内外金切锯床的现状与发展趋势[J]. 制造技术与机床, 2004(02): 32-33.
- [2] 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.
- [3] 中国机械工业年鉴编辑委员会, 中国机床工具工业协会. 中国机床工具工业年鉴[M]. 北京: 机械工业出版社, 2017: 129-132.
- [4] 孙荣霞, 刘腾发, 刘畅, 等. 金属带锯床故障报警系统设计[J]. 科技资讯, 2019, 17(16): 21-23.
- [5] 刘佳. 金属带锯床远程无线监控系统研究与设计[D]. 武汉: 武汉理工大学, 2012: 12-28.
- [6] 刘腾发. 金属带锯床在线监测与故障诊断系统研究[D]. 保定: 河北大学, 2020: 14-28.
- [7] Potočník P, Thaler T, Govekar E. Multisensory chatter detection in band sawing[J]. Procedia CIRP, 2013, 8: 469-474.
- [8] Turchetta S, Sorrentino L. Forces and wear in high-speed machining of granite by circular sawing[J]. Diamond and Related Materials, 2019, 100: 107579.
- [9] 周东华, 魏慕恒, 司小胜. 工业过程异常检测, 寿命预测与维修决策的研究进展[J]. 自动化学报, 2013, 39(06): 711-722.
- [10] Le Son K, Fouladirad M, Barros A. Remaining useful lifetime estimation and noisy gamma deterioration process[J]. Reliability Engineering & System Safety, 2016, 149: 76-87.
- [11] 李烁, 陈震, 潘尔顺. 广义逆高斯过程的步进应力加速退化试验设计[J]. 上海交通大学学报, 2017, 51(02): 186-192.
- [12] Kalidass S, Palanisamy P, Muthukumaran V. Prediction and optimisation of tool wear for end milling operation using artificial neural networks and simulated annealing algorithm[J]. International Journal of Machining and Machinability of Materials, 2013, 14(02): 142.
- [13] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(03): 273-297.
- [14] 蔡波, 黄晋英, 等. 基于 MEEMD 多特征融合与 LS-SVM 的行星齿轮箱故障诊断[J]. 中国测试, 2021, 47(09): 126-132.
- [15] 张梅, 陈万利, 许桃. 基于 LS-SVM 的矿井提升机故障预测[J]. 电子测量技术, 2021, 44(12): 70-74.
- [16] Al-Hadeethi H, Abdulla S, Diykh M, et al. Adaptive boost LS-SVM classification approach for time-series signal classification in epileptic seizure diagnosis applications[J]. Expert Systems with Applications, 2020, 161: 113676.

- [17] Ali L, Wajahat I, Amiri Golilarz N, et al. LDA-GA-SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine[J]. *Neural Computing and Applications*, 2021, 33(07): 2783-2792.
- [18] 吴昊. 基于声发射的刀具磨损状态识别与预测[D]. 成都: 电子科技大学, 2017: 35-64.
- [19] Jingchao G, Anhai L, Rufeng Z. Tool condition monitoring in milling process using multifractal detrended fluctuation analysis and support vector machine[J]. *The International Journal of Advanced Manufacturing Technology*, 2020, 110(5-6): 1445-1456.
- [20] Singh A, Garg S, Batra S, et al. Bloom filter based optimization scheme for massive data handling in IoT environment[J]. *Future Generation Computer Systems*, 2018, 82: 440-449.
- [21] Mujahid T, Rahman A U, Khan M. GPU-Accelerated Multivariate Empirical Mode Decomposition for Massive Neural Data Processing[J]. *IEEE Access*, 2017, 5: 8691-8701.
- [22] Kun Z, Kang Z, Falin F, et al. Real-Time Massive Vector Field Data Processing in Edge Computing[J]. *Sensors*, 2019, 19(11): 2602.
- [23] 胡素娟. 基于云计算技术的分布式网络海量数据处理系统设计[J]. *科技与创新*, 2021(17): 84-85.
- [24] 杨晓云. 海量数据处理需求下的物联网实施数据库设计[J]. *电子制作*, 2020(Z1): 49-51.
- [25] 卢民荣, 郑建宁. 基于极值点特征识别的大规模时序数据压缩分析[J]. *统计与决策*, 2021, 37(20): 39-43.
- [26] Soni D, Makwana A. A survey on mqtt: a protocol of internet of things (iot)[C]//*International Conference On Telecommunication, Power Analysis And Computing Techniques (ICTPACT-2017)*. Bharath Institute of Higher Education and Research, 2017, 20: 173-177.
- [27] 俞锋. 基于机器学习的金属带锯条剩余寿命预测研究[D]. 杭州: 杭州电子科技大学, 2021: 29-44.
- [28] 潘柏松, 梁利华, 张元祥. 金属切削带锯床设计理论与方法[M]. 北京: 清华大学出版社, 2013: 29-30.
- [29] 李磊, 吴利平, 钱靖, 等. 考虑焊接缺陷对双金属带锯条锯齿切削性能的影响研究[J]. *南方农机*, 2022, 53(03): 18-22.
- [30] 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.
- [31] 王艺玮, 邓蕾, 郑联语, 等. 基于多通道融合及贝叶斯理论的刀具剩余寿命预测方法[J]. *机械工程学报*, 2021, 57(13): 214-224.
- [32] Mei W, Jie W. CHMM for tool condition monitoring and remaining useful life prediction[J]. *The International Journal of Advanced Manufacturing Technology*, 2012, 59(5-8): 463-471.
- [33] Lotfi M, Amini S, Aghaei M. Tool Wear Prediction and Surface Improvement in Vibration Cutting[J]. *Tribology Transactions*, 2018, 61(03): 414-423.

- [34] 武滢. 基于主轴电流信号多特征融合的刀具磨损状态监测[J]. 制造技术与机床, 2022(03): 44-48.
- [35] Chuangwen X, Jianming D, Yuzhen C, et al. The relationships between cutting parameters, tool wear, cutting force and vibration[J]. Advances in Mechanical Engineering, 2018, 10(01): 168781401775043.
- [36] 董伟航, 胡勇, 田广军, 等. 基于离散小波变换与遗传 BP 神经网络的木工刀具磨损状态监测[J]. 中南林业科技大学学报, 2021, 41(06): 157-166.
- [37] Huang N E. Review of empirical mode decomposition[C]//Wavelet applications VIII. SPIE, 2001, 4391: 71-80.
- [38] Xinhua S, Ran W, Quntao C, et al. Cutting sound signal processing for tool breakage detection in face milling based on empirical mode decomposition and independent component analysis[J]. Journal of Vibration and Control, 2015, 21(16): 3348-3358.
- [39] 周余庆. 立铣刀状态监测与剩余有效寿命预测方法研究[D]. 杭州: 浙江工业大学, 2020: 8-10.
- [40] 张宇. 高速列车轴承故障诊断系统的研究[D]. 大连: 大连交通大学, 2017: 18-19.
- [41] 李宁, 杨建国, 周瑞. 基于集合经验模式分解的汽油机爆震特征提取[J]. 机械工程学报, 2015, 51(02): 148-154.
- [42] Yang X, Mingzhang L, Tao L et al. ECG Signal De-noising and Baseline Wander Correction Based on CEEMDAN and Wavelet Threshold[J]. Sensors, 2017, 17(12): 2754.
- [43] 刘雪, 田云娜, 田园. 群智能算法研究综述[J]. 信息与电脑(理论版), 2021, 33(24): 63-69.
- [44] Karaboga D, Basturk B. On the performance of artificial bee colony (ABC) algorithm[J]. Applied soft computing, 2008, 8(01): 687-697.
- [45] 孔德同, 刘庆超, 雷亚国, 等. 一种改进的 EEMD 方法及其应用研究[J]. 振动工程学报, 2015, 28(06): 7.
- [46] 王圣杰, 彭珍瑞, 殷红. 基于多传感器信号处理的齿轮箱轴承故障诊断[J]. 组合机床与自动化加工技术, 2020(11): 5-10.
- [47] 舒悦, 谢传东, 何明, 等. 往复压缩机环状气阀振动信号的 LMD 故障特征提取方法研究[J]. 流体机械, 2019, 47(11): 13-18.
- [48] 韩清凯, 于晓光. 基于振动分析的现代机械故障诊断原理及应用[M]. 北京: 科学出版社, 2010: 34-42.
- [49] Schölkopf B, Smola A J, Bach F. Learning with kernels: support vector machines, regularization, optimization, and beyond[M]. Cambridge: MIT press, 2018: 47-52.
- [50] 吴晓松. 世界大数据技术进展与发展趋势研究[C]//海峡两岸服务贸易及大数据信息资源管理研究国际学术会议论文集. 云南财经大学, 2016: 161-176.

- [51] 邓建球, 赵建忠, 陈洪, 丛林虎. ABC 算法优化 SVR 的磨损故障预测模型[J]. 兵工自动化, 2018, 37(10): 60-64.
- [52] 董明, 陈慧艳, 伏晓, 朱世贤. 基于人工蜂群算法的大坝变形支持向量机预测模型[J]. 三峡大学学报(自然科学版), 2017, 39(02): 10-13.
- [53] 王延年, 程燕杰, 钟正, 等. 基于人工蜂群优化的自适应图像增强方法[J]. 工具技术, 2020, 54(08): 78-82.
- [54] 刘三阳, 张平, 朱明敏. 基于局部搜索的人工蜂群算法[J]. 控制与决策, 2014, 29(01): 123-128.
- [55] 孔德鹏, 常天庆, 戴文君, 等. 基于排序选择和精英引导的改进人工蜂群算法[J]. 控制与决策, 2019, 34(04): 781-786.
- [56] 黄英双, 曹辉. 改进人工蜂群算法优化支持向量机及应用[J]. 计算机应用与软件, 2021, 38(02): 258-263+277.
- [57] Steinarsson S. Downsampling time series for visual representation[D]. Iceland: Engineering and Natural Sciences University of Iceland, 2013: 21-23.
- [58] 陈键冬, 李旦. 时序数据的处理方法、时序数据处理装置及存储介质[P]. 广东省: CN113032461A, 2021-06-25.
- [59] 王立峰, 自明, 任成宾, 等. 一种用于时序数据降采样的方法和系统[P]. 山东省: CN112699165A, 2021-04-23.

致 谢

时光飞逝，转眼三年就过去了，研究生的生活即将画上句号。在杭州电子科技大学电子信息学院，我真切感受到了学院浓厚的学术氛围和高效的科研能力，也意识到自己在学术方面的不足。这三年来的求学历程，是我人生的宝贵财富，感谢在杭电学习和生活所经历的一切，感谢帮助过我的每一个人。

首先，感谢我的导师陈科明教授，老师严谨的求学态度、渊博的学识积累和勤勉的工作精神深深影响了我。无论是学习，工作还是精神上，陈老师都为我提供了宝贵的经验。这也为我提高专业的知识与技能打下夯实的基础。此次论文写作，无论是选题、立新还是定稿，陈老师都严格要求。正是这样的严格负责才让我能一直朝着目标前进，让我在学习上收获颇丰。三年的师生情谊让我耳濡目染，陈老师让我看到了一名专业人士应该具备的责任与态度。在今后的日子里，我也会牢记陈老师的教诲，争做一名优秀的杭电人。

感谢我的父母及家人，三年来他们对我的支持和鼓励，在我面临人生选择的迷茫之际，为我排忧解难，使我顺利完成学业。

感谢我的女朋友黄盼盼，在写论文期间默默支持我、鼓励我和帮助我，在生活、求学、求职过程中共同成长、共同进步。

感谢我的同门邢豪蔚、吴文楷、胡晓双、张百军、刘剑、徐志杰，他们在学习和工作上给予我莫大的帮助。能够和他们成为同学和朋友让我感到非常荣幸。

感谢论文评阅和答辩委员会的老师，感谢你们的耐心，我一定会悉心听取你们宝贵的意见并积极改正。

最后，衷心地祝愿所有人，身体健康，工作顺利，让我们一起迎接更美好的明天。

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] 面向锯切装备产业提升的“智能一代”技术研究及产业化，浙江省科技计划项目（KYZ044321014），2020.07-2022.12
- [2] 晨龙智能锯床远程运维系统，横向课题（202233010100），2021.01-2023.12

发表的专利：

- [1] 陈科明,蔡坤,黄盼盼,邢豪蔚. 一种基于最大三角形三桶算法的海量数据绘图优化方法[P]. 浙江省: CN114241086A, 2022-03-25. （已公开，2/4，导师第一发明人）
- [2] 陈科明,蔡坤,黄盼盼,齐玲玲. 基于改进人工蜂群算法的SVM的带锯条磨损识别方法[P]. 浙江省: 202210385044.5, 2022-04-13. （已申请，2/4，导师第一发明人）