

杭州电子科技大学

硕 士 学 位 论 文

题 目： 仪器设备状态智能监测系统设计

研 究 生 安 建 伟

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 3 月

杭州电子科技大学硕士学位论文

仪器设备状态智能监测系统设计

研 究 生： 安建伟

指导教师： 陈科明 副教授

2021 年 3 月

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

Design of Intelligent Monitoring System for Instrument and Equipment State

Candidate: An Jianwei

Supervisor: Associate Prof. Chen Keming

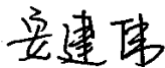
March, 2021

杭州电子科技大学

学位论文原创性声明和使用授权说明

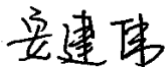
原创性声明

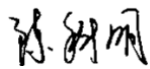
本人郑重声明： 所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。申请学位论文与资料若有不实之处，本人承担一切相关责任。

论文作者签名：  日期： 2021 年 3 月 16 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各单位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：  日期： 2021 年 3 月 16 日

指导教师签名：  日期： 2021 年 3 月 16 日

摘要

随着国家实施科技强国的战略以来，高校实验室扩展迅速，大型实验仪器资产也不断增加。同时也暴露了很多方面的问题：对仪器设备的使用情况，还是大多采用非常低效的纸质人工记录方式，不能实时的进行监测；仪器设备的工作阈值不尽相同，区分各个工作状态较为繁琐，缺乏有效自适应阈值算法；海量仪器设备的电流数据上报容易阻塞网络，影响数据的准确性；仪器闲置率较高，有效工作时间短，仪器重复购进的现象，缺乏有效的共享机制。

本文设计了一种仪器设备状态智能监测系统，通过基于 NB-IOT 的智能电流数据采集终端采集数据，对数据进行加密处理后传输到云平台，并且设计错峰上报机制，优化数据上报的准确性；改进的 K-means 自适应算法对仪器设备使用的历史电流进行自适应工作阈值分析，获取其关机、待机、运行电流状态值；然后下发阈值到智能采集终端，从而有效对流量上报进行限制；在可视化终端上展示仪器设备状态的使用信息以及在线工作时间，实现了系统的状态实时监测，并且通过客户端、Web 端实现了智慧预约仪器使用。

针对目前大型设备的种类繁多以及不同设备的状态阈值都不尽相同，需人工采集一一分类对比、效率低下等问题，自主设计了一种改进 K-means 自适应阈值的分析算法，可对大量设备使用的工作电流信息进行数据分析，得到不同大型设备仪器对应的各种工作状态阈值。其中从选取 K 值、初始聚类中心、孤立点、三个方面优化改进算法，设计了基于 GSA 自适应 K 值算法，实现了当仪器设备工作周期不完整的情况下，也可以自动识别输出最佳 K 值。设计了基于 Huffman 二叉树算法，实现除去孤立的相异点，完成初始聚类中心的选择机制，有效避免了局部最优。经过对聚类模型仿真测试及参数分析，与原始 K-means 算法对比发现，改进算法的准确率提升了 12%，在算法耗时降低了 60%，算法时间复杂度也大大缩小了。

在 NB-IOT 无线传输网络下，大量设备对应的海量电流数据上传，依据 K-means 得到的电流阈值结果进行数据的清洗。降低了系统 NB-IOT 通信中 60% 的数据流量上传。从而降低云平台的数据存储压力，同时提升了系统的稳定性、准确性、高效性。

针对于仪器设备中那些昂贵且数量较少，当使用时出现多人并发预约时等缺点，设计了对基于贪心策略算法预约机制，实现预约的最大有效性，更好的完善了仪器共享平台的预约机制。测试表明，仪器设备预约共享机制相比之前，仪器有效预约使用提升了近 40%。

关键词：仪器设备、状态监测、K-means 分析、NB-IOT、贪心预约

ABSTRACT

With the implementation of the strategy of strengthening the country through science and technology, university laboratories have expanded rapidly, and large-scale experimental equipment assets have also increasing. At the same time, many problems have been exposed: For the use of instruments and equipment, most of them use very inefficient paper-based manual recording methods, which cannot be monitored in real time; the working thresholds of instruments and equipment are not the same, and it is cumbersome to manually distinguish their working status one by one, and there is no effective self-control. Adapt to the threshold algorithm; the current data report of massive instruments and equipment is easy to block the network and affect the accuracy of the data; the idle rate of the instrument is high, the effective working time is short, the phenomenon of repeated purchase of the instrument, and the lack of an effective sharing mechanism.

This paper designs an intelligent monitoring system for the state of instruments and equipment through the NB-IOT-based intelligent current data acquisition terminal to collect data, encrypt the data and transmit it to the cloud platform, and design a peak reporting mechanism to optimize the accuracy of data reporting; The improved K-means adaptive algorithm performs adaptive working threshold analysis on the historical current used by the instrument and equipment, and obtains its shutdown, standby, and running current state values; then sends the threshold to the intelligent collection terminal to effectively limit the flow report; Display the use information of the equipment status and online working hours on the visual terminal, realize the real-time monitoring of the system status, and realize the smart appointment of the use of the instrument through the client and Web.

In view of the wide range of large-scale equipment and the different state thresholds of different equipment, manual collection of one-by-one classification and comparison, low efficiency, etc., we independently designed an analysis algorithm to improve the K-means adaptive threshold, which can be used for a large number of Perform data analysis on the working current information used by the equipment to obtain various working state thresholds corresponding to different large-scale equipment and instruments. The algorithm is optimized and improved from the three aspects of selecting K value, initial clustering center, outlier, and the design of an adaptive K value algorithm based on GSA, which can automatically identify the best output even when the working cycle of the equipment is incomplete K value. A binary tree algorithm based on Huffman is designed to remove isolated dissimilar points and complete the selection mechanism of initial clustering centers,

effectively avoiding local optimal. After the clustering model simulation test and parameter analysis, compared with the original K-means algorithm, it is found that the accuracy of the improved algorithm has been increased by 12%, the algorithm time consumption has been reduced by 60%, and the algorithm time complexity has also been greatly reduced.

Under the NB-IOT wireless transmission network, massive current data corresponding to a large number of devices are uploaded, and the data is cleaned according to the current threshold results obtained by K-means. Reduced 60% of the data traffic upload in the system NB-IOT communication. Reduce the data storage pressure of the cloud platform, while improving the stability, accuracy and efficiency of the system.

Aiming at the disadvantages of expensive and small number of instruments and equipment, and when multiple people make concurrent appointments when used, the appointment mechanism based on the greedy strategy algorithm is designed to achieve the maximum effectiveness of appointments and better improve the instrument sharing platform. Reservation mechanism. Tests have shown that, compared with the previous sharing mechanism of equipment reservation, the effective reservation use of equipment has increased by nearly 40%.

Keywords: equipment, condition monitoring, K-means, NB-IOT, greedy appointment

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 设备状态获取技术研究现状.....	2
1.2.2 设备共享技术研究现状.....	4
1.3 仪器设备状态智能监测系统的技术分析.....	4
1.4 本文主要工作及章节安排.....	6
第 2 章 仪器设备状态智能监测系统.....	7
2.1 系统总体设计.....	7
2.1.1 需求分析.....	7
2.1.2 总体框架.....	7
2.2 智能采集终端设计.....	9
2.2.1 硬件设计.....	10
2.2.1.1 单相电流模块.....	11
2.2.1.2 三相电流模块.....	12
2.2.2 软件设计.....	12
2.2.2.1 单相电流模块.....	13
2.2.2.2 三相电流模块.....	14
2.2.3 终端接入方式和展示.....	15
2.2.3.1 单相智能终端.....	15
2.2.3.2 三相智能终端.....	16
2.3 网络层设计.....	17
2.3.1 常用的无线通信技术选择.....	17
2.3.2 网络基本架构.....	19
2.3.3 NB-IOT 通信协议.....	19
2.3.3.1 设备联网信息上报.....	20
2.3.3.2 设备周期性数据上报.....	20

2.3.3.3 异常情况报警上报.....	21
2.3.3.4 电信云 MCU 远程升级通信.....	22
2.3.4 关键机制.....	24
2.4 云平台.....	25
2.4.1 设备云设计.....	26
2.4.2 业务云设计.....	27
2.5 应用层设计.....	28
2.5.1 Web 设计	28
2.5.1.1 Web 框架设计	28
2.5.1.2 Web 端仪器管理功能设计	29
2.5.1.3 Web 端系统展示	29
2.5.2 APP 设计.....	32
2.5.2.1 设计模式.....	32
2.5.2.2 网络层设计.....	32
2.5.2.3 功能设计.....	33
2.5.2.4 APP 结果展示.....	34
2.5.3 安装运维小程序设计.....	34
2.5.3.1 功能设计.....	34
2.5.3.2 处理机制设计.....	41
2.5.3.3 小程序结果展示.....	42
2.5.3.4 现场安装图展示.....	43
2.6 本章小结.....	44
第 3 章 改进 K-means 的自适应阈值算法	45
3.1 K-means 聚类算法应用特点分析	46
3.1.1 K-means 聚类算法原理以及应用场景	46
3.1.2 K-means 聚类算法流程	47
3.1.3 K-means 算法存在的问题以及优化	47
3.1.3.1 自适应聚类 K 值的策略.....	47
3.1.3.2 初聚类中心的选择策略.....	49
3.2 改进 K-means 自适应阈值算法	50
3.2.1 改进算法实现的流程框架.....	50
3.2.2 数据源模型构建.....	51
3.2.3 数据预处理.....	52
3.2.3.1 数据归一化处理.....	52

3.2.3.2 数据 PCA 降维处理.....	52
3.2.4 基于 GSA 自适应 K 值算法.....	52
3.2.4.1 原始 GSA 算法流程.....	53
3.2.4.2 手肘法判定 K 值.....	53
3.2.4.3 改进 GSA 自适应 K 值算法实现步骤.....	54
3.2.4.4 数据仿真测试.....	55
3.2.5 初始聚类中心以及孤立异常点去除.....	56
3.2.5.1 数据仿真测试.....	57
3.3 流量上报限制机制应用.....	58
3.3.1 流量上报限制应用测试.....	60
3.4 改进 K-means 算法数据测试应用	60
3.4.1 测试方案.....	60
3.4.2 测试结果.....	62
3.4.3 结果分析.....	63
3.5 本章小结.....	63
第 4 章 基于贪心策略算法的智慧预约.....	64
4.1 贪心算法的特点.....	64
4.2 贪心算法的应用场景.....	65
4.3 模型建立.....	65
4.4 基于贪心策略算法的智慧预约机制.....	66
4.4.1 预约机制指标.....	66
4.4.2 仪器设备预约模型实现.....	67
4.5 测试分析.....	68
4.5.1 预约算法数据测试.....	68
4.5.1.1 测试方案.....	68
4.5.1.2 测试数据.....	69
4.5.1.3 测试分析.....	69
4.5.2 仪器共享数据对比.....	70
4.6 本章小结.....	70
第 5 章 总结与展望.....	71
5.1 本文工作总结.....	71
5.2 展望.....	71
致 谢.....	73
参考文献.....	74

附 录.....	77
----------	----

第1章 绪论

1.1 研究背景和意义

大型科研仪器设备常用于科技最前沿，解决人类科学中的难点，提高我们社会的经济效益，具有专业性强、精准度高的特点。随着智能制造的不断推进，智慧物联蓬勃发展。人们的生活节奏也越来越快，社会也慢慢加快了向数字化转型发展，利用物联网技术解决生活的问题也逐渐增多，俨然成为一种趋势。

随着国家鼓励科技创新发展，实施科技强国战略，对教育的投入也渐渐增多，对大型实验仪器的引入至高校也逐步增多，这也导致高校实验室里大型科研仪器数量激增。从 2010 年至 2019 年间，国家支出投入教育领域经费增加了 3.6 倍，高校大型科研仪器设备数量增加了 6 倍之多^[1]。统计数据如图 1.1 所示：

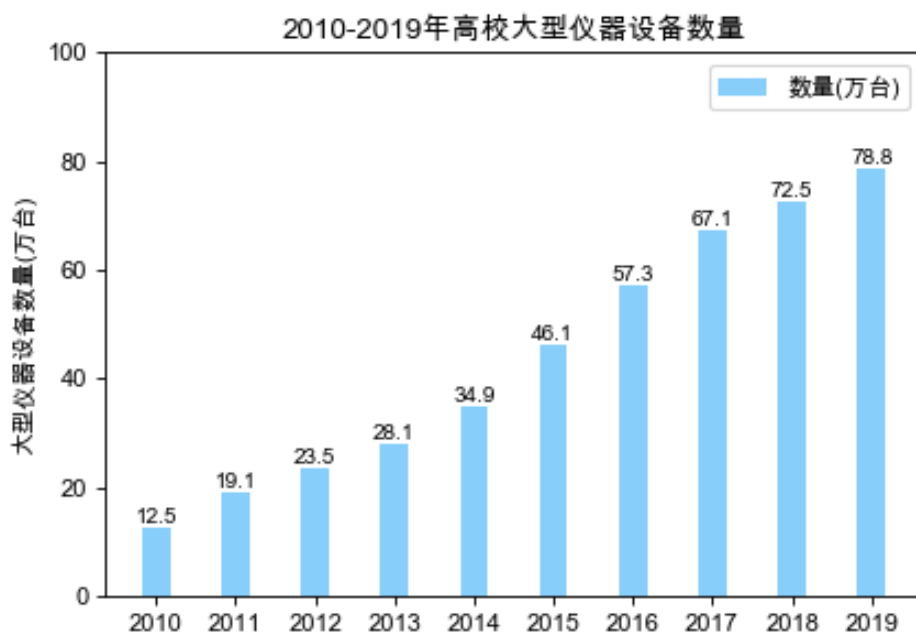


图 1.1 2010-2019 年高校大型实验仪器数量

现今，高校对于大型科研仪器设备的管理仍处在一个比较滞后的阶段，其管理采用的方式仍然是通过人工，纸质现场记录的方式。这种不仅时效性差，而且严重降低了这些大型科研仪器设备的使用效益，缺乏实时监测。

其次，各个高校间的大型实验仪器信息不互通，缺乏共享，跨校区使用还要联系对应管理人员较为麻烦，仪器重复购进也比较严重，而仪器设备随着时间慢慢年久失修，逐渐淘汰，

将造成极大资源浪费^[2]。并且大型科研仪器相对价格昂贵，比较稀缺，普通高校实验器材资源较少，相对集中在 985、211 高校^[3]，设备的闲置率较高，高校设备供需关系不平衡，资源分配不均。统计数据对比如图 1.2 示；

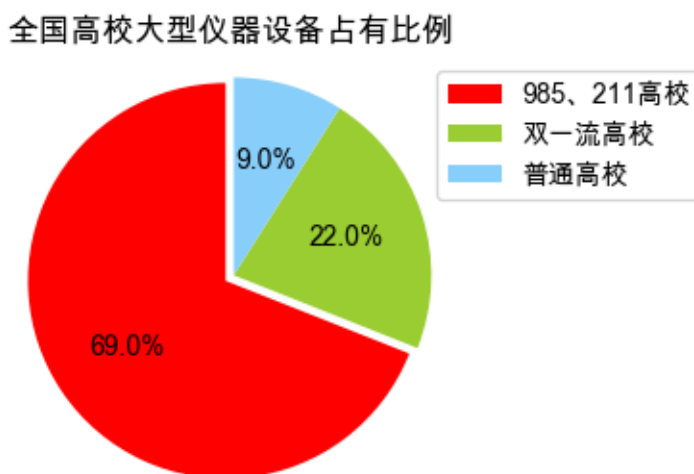


图 1.2 全国高校持有大型设备数量比例

对于设备共享管理方式亟需升级，并且为贯彻国家对大型科研仪器对外开放共享的政策^[4]，通过物联网传感设备联网技术对大型实验仪器进行加装赋能。科研探索往往需要昂贵的仪器设备支撑，若对仪器缺乏有效的管理导致大量仪器设备闲置，随着时间的发展，这类仪器将被新的仪器淘汰，这不仅导致国家资源的浪费，同时影响其他需要进行科研的机构的发展。而且在高校中还存在热门仪器，基本天天运转也不够使用，这在共享预约机制中^[5]，如何实现资源均匀分配使用的棘手问题。

因此，面对当前这种形式，本文设计实现了一种大型实验仪器设备状态智能监测系统，同时向两个方面解决问题。（1）提出一种能够自动记录大型实验仪器工作情况，实时监测设备的工作状态和使用情况，及时反馈到 APP 和 WEB 页面上，使科研人员、以及管理人员及时获知设备的实时状态。（2）提出一种智慧共享平台，完成大型科研仪器设备在线预约使用，通过基于贪心策略算法，解决了共享大型实验仪器的预约机制，满足在共享设备中大量用户并发预约，智能调度机制可以为科研人员提供决策，将大型实验仪器的利用率最大化。

1.2 国内外研究现状

1.2.1 设备状态获取技术研究现状

国内外学者利用电流分析设备工作状态，取得了许多成果，其中：国外学者提出对电器设备在工作和待机时，对设备的电流信号进行短态与暂态的特征信号提取，引入模糊原理^[6]建立判断因素集与识别模型，模糊匹配多项指标依次对应设备电流的特征信号，从而判定设备的运行状态。综合判断依据，最终实现监测。

国内学者在电力设备状态识别方法中，首先利用传感器获取设备的电流，并生成电流时序图，然后对比 BPNN、PNN、SVM（Support Vector Machine）以及改进后的 Faster-R-CNN^[7] 对不同状态设备电流时序图像识别结果，得到了较高的识别精度。通过图像对比，输出设备的工作状态。

在王珏、王晓玲等人的专利《一种基于仪器电流数据的大数据分析方法》中提出利用设备仪器管理系统采集包含设备所有状态的电流值，确定不同电流对应的设备工作状态之后，将设备状态、设备型号、设备电流作为一个数据集，由 SVM 进行分类^[8]，得到识别模型，对之后的实时电流数据进行分析对比。

利用设备电流进行分析得到设备状态已经初步应用实际场景中。如：基于大数据下的一些小电流系统的故障分析，其核心实现是根据获取的数据使用数据挖掘算法建立模型，进行训练，将实时获取的数据标记成故障特征向量^[9]。大多数，数据处理框架都采用了离线训练在线应用的一种模式，对原始数据进行预处理，依次生成电流的特征矩阵模型，最后通过分析模型得到最终结果。在国内外的研究中，大多数案例都偏向于统计学，使用 MATLAB 或者 SPSS 软件^[10]对数据进行挖掘。

目前，对于大仪设备状态监测并没有一个很完善的解决方案，目前通常采用的方式为对应不同的设备，将不同状态的电流阈值设置位固定值，这样做首先无法应对电流工作波动大的场景；其次，若对于每台设备都进行固定阈值是不现实的，我们无法做到对每台设备不同状态进行一一记录。利用 SVM 向量机对设备电流进行分类是可行的，但是并不适用于当前的场景。以 SVM 向量机为代表的分类模型首先需要对数据的标定，标定的过程需要人工执行，这实际上与固定阈值的方式类似^[11]，在不知道设备状态的前提下对设备上报电流划分是不切实际的。

由于一台设备在不同状态下运行的电流值是不相同的，因此可以实时将设备的电流值采集之后进行分析，从而分析得到设备的运行状态。同时，面对不同型号的设备，其不同工作状态的电流阈值也不相同，仅仅以固定的阈值是不能满足这种情况的。

因此，本文提出如下方案：基于 K-means 的自适应大型实验仪器的阈值方法，对于不同的设备，我们目前都可以将其状态分做两类，关机、开机，或者三类，关机、待机、工作。由于同一设备在不同状态中消耗的电流值不同，但是同一状态下的电流往往都趋近一个区间，因此可以使用聚类算法对电流值进行处理。设备需要上报完整的工作周期的电流数据，根据这些电流数据，采用 K-means 算出不同状态下电流的聚类中心。由于聚类中心通常都是处于某类中的均衡位置，因此电流阈值可以取不同的聚类中心的重点进行计算。从而动态计算出各类型号设备的关机、待机、运行等不同状态的电流阈值。通过上述方式计算后得到的电流阈值再对采集到设备电流进行分类，就能够实时掌握任意时间设备的工作状态。将仪器设备状态再反馈至 WEB 和 APP 端，实现简单化、透明化、高效化的管理目标^[12]。从而进一步有效整合高校的大型科研仪器设备资源，推动科研仪器的物联共享，提高利用效率，增强创新

能力。

1.2.2 设备共享技术研究现状

仪器设备资源共享是新时代科技创新发展的必然趋势，国家也倡导和贯彻共享的发展理念^[13]。设备仪器共享的概念，国外比较早就已经提出。在我们刚刚进入互联网信息时代时，欧美国家就已经在思考怎样将互联网和共享科学仪器结合起来，比较出名的是欧洲核粒子中心^[14](European Organization for Nuclear Research,CERN)，在 20 世纪 60 年代出现，由众多欧洲国家共同组建而成，建设史无前例，而且年年都会有数以千计的科研工作者来此科研与实验，CERN 负责良好的科研环境包括大型和超大型的精密科学仪器^[15]。其中全世界极少数的大型强子对撞机(LHC)也在此诞生，因此实验中心建立了协同共享平台系统^[16]，平台可以更加方便的为科研工作者提供服务，以及在网上预约仪器设备也更加高效方便，提高了仪器设备的使用率，也达到了共享理念。

我国仪器管理结合信息化相对国外发展滞后，此后国内慢慢意识到仪器共享和在线管理的重要性^[17]，目前国内相对较好的大学已经部分实现在线管理仪器设备，如北京大学、清华大学、浙江大学高校等。最初刚起步时是由长三角区域构建了大型科学仪器设备协作共用网^[18]。陆续有联合高校建立自身仪器共享系统平台，打破了部门、地区的使用局限，逐步整合了高校资源。而且区别于以教学为目的的社会服务也可以实施有偿使用，通过共享服务获得收益可以用于仪器设备的维护，实现了“以机养机”模式的可持续发展，也提升了自身利用效率，形成了良性循环。

我们在建立大型实验仪器共享平台中，在网络共享中出现的很多的问题：如何能充分提高仪器设备使用率，以及相对集中热门的仪器使用该如何合理分配，让科研工作者预约更加公平，使我们的共享平台更加人性化、智能化。平台在用户使用时遇到掉线和网络延迟带来的数据不同步等问题。

1.3 仪器设备状态智能监测系统的技术分析

本文针对高校的大型仪器实验设备的有效管理，落后的纸质记录仪器设备的使用状况，且不能实时进行在线监测，以及共享预约的问题。设计并提出了智能采集终端、设备云平台、业务云平台、iOS 客户端、Android 客户端以及 Web 客户端。在智能系统的设计基础上，本文将基于改进的 K-means 算法的自适应大型实验仪器阈值，实现在线监测设备使用状态，提升了仪器设备的有效使用时间；通过共享仪器调整了高校间的资源平衡，也降低了仪器设备的维护成本和采购设备成本。对智能系统进行了技术升级。

针对目前大型实验仪器管理的缺点以及数字化转型在线监测系统的实际需求，本文对所有技术设计都提出一致的要求：

1) 兼容性。根据大型实验仪器电源和功率不统一的情况下，不依赖大型仪器电流用电情况采集数据，我们设计单相电流计和三相电流计分别对大型实验仪器进行采集其使用时的工

作电流，并且不干扰原来仪器设备的使用。

2) 强鲁棒性。在低网络下，可以保证用户下发的指令能完整地、准确低延迟地到达业务云平台，并在可以完整的返回命令结果。对待大量学生和科研人员同时并发操作时，要求保证系统不会崩溃，可以正常维持服务。

3) 可扩展性。在系统技术设计时，我们会考虑到后期可能扩展的硬件、软件情况。对系统进行预存接口，确保系统的可扩展性。在软件设计时，为预防采集端会有功能更新需求，需设计智能远程更新智能采集终端，实现远程更新 CPU 程序；在云平台需要为智能采集终端和客户端提供请求接口以及可扩展的协议设计。

4) 安全性。对于物联网应用，大数据时代下，我们对大型实验仪器采集的一些基本信息参数，在数据传输时，必须设计为数据加密化、信息不涉及隐私、为脱敏数据。

5) 可靠性。智慧采集终端在封闭实验室内信号可能出现波动，这时我们设计了当采集终端发生网络异常中断、关机重启时，要求采集终端联网后，数据立即重新发送，保证了系统使用的可靠性。

6) 自适应性。改进的 K-means 自适应阈值算法应用。使大型实验仪器采集的电流数据在不具有完整的工作周期时，也可以自适应区分，设备的各个工作状态电流阈值。在数据处理下，也实现了自适应网络环境，调节数据错峰上报时机，缓解了云平台的存储压力，增强了系统的实时性。

7) 共享性。可以实时监测仪器的使用状况，有效的进行仪器设备预约使用。基于贪心策略算法应用，解决了大型实验仪器被并发预约的情况，最大化的使科研人员更好的均匀分配使用。增加了仪器共享的理念。

本文依据物联网技术为载体，由智能采集终端、云平台、客户端软件（APP 应用、PC 客户端、Web 应用）构成一个闭环传输数据的仪器设备状态智能监测系统。本文设计了基于 NB-IOT 的智能采集终端，通过实时获取大型实验仪器的信息数据，加密后上传至设备云平台存储，并且防止单个区域内，仪器设备数据集中上报，设计了错峰机制上报，降低了同一时刻多数据上报的压力和丢包率，而且保证了数据的准确性；并设计了基于改进的 K-means 算法对仪器设备的电流数据进行数据分析，自适应获取了仪器设备的工作阈值；基于贪心策略的预约仪器设备机制，实现了对有限资源最大化、最合理的预约共享机制；而在客户端和 Web 端上，实现仪器设备系统的可视化监测和有效预约管理机制，将仪器设备的各个使用情况展示在屏幕上，有效地提高仪器的管理与使用。

综上所述，仪器设备状态智能监测系统的设计可以有效的提升设备使用率、降低仪器设备的维护成本，提供有效的采购决策；为科研者、管理者提供了更智能、便捷的服务，对高校科研的数字化、智能化升级上做出技术性贡献。本文以大型实验仪器作为研究对象，对智能系统的设计、改进 K-means 算法的应用于仪器设备状态监测、贪心策略算法的应用于共享预约机制的实现等进行一一阐述。

1.4 本文主要工作及章节安排

本文主要研究的是仪器设备状态智能监测系统，实现大型实验仪器的远程在线状态监测以及智慧共享预约系统，通过基于 NB-IOT 无线技术将设备采集的数据上传至云平台，通过 K-means 聚类算法分析仪器设备的状态，再反馈至 WEB 和 APP 中进行在线监测。在智慧共享平台中通过基于贪心策略算法解决热门仪器设备的分配使用时间，使资源分配更加智能。使设备利用最大化。文章的内容安排如下：

第一章：绪论，详细介绍仪器设备状态智能监测系统的研究背景和意义，提出了当前大型实验仪器管理存在的问题，明确了论文的目的及意义，然后分别对设备状态获取技术现状以及设备共享技术研究现状做了分析，为本文设计提供指导。

第二章：仪器设备状态智能监测系统设计及实现，首先进行系统总体设计需求和总体框架，然后分别对系统各个部分进行设计，包括智能采集终端、网络层和应用层的设计。最后对整个系统进行了展示。

第三章：改进 K-means 的自适应阈值算法设计。首先介绍原 K-means 算法的应用特点、算法流程、以及存在的缺点问题，基于改进的 K-means 自适应阈值算法，将智慧采集终端采集仪器设备状态电流信息，进行数据分析，获取到仪器的设备的关机、待机、运行的状态阈值。其中从选取 K 值、初始聚类中心、孤立点、三个方面优化改进算法，设计了基于 GSA 自适应 K 值算法，实现了当仪器设备工作周期不完整的情况下，也可以自动识别输出最佳 K 值。设计了 Huffman 二叉树算法，实现除去孤立的相异点，完成初始聚类中心的选择机制。通过仿真数据测试，提高了聚类模型的测试。其次，利用获取的电流阈值进行采集终端的数据清洗上传限制，从而降低云平台的数据存储压力。最后，对此算法进行相关的准确测试与验证。

第四章：基于贪心策略算法的智慧预约平台研究。首先介绍了贪心策略算法的特点及应用事例，并实现通过贪心策略算法解决了多人并发预约时，仪器设备分配的算法流程机制。最后，对此算法进行相关的准确测试与验证。

第五章：总结与展望。对全文的系统设计和使用的关键技术研究加以总结，并对更深层次构建仪器状态智能监测系统提出建议和展望。

第2章 仪器设备状态智能监测系统

针对目前大型实验仪器状态不能及时反馈存在的缺陷，本文设计了一种仪器设备状态智能监测系统。系统包括智能采集终端层、网络层、云平台、应用层。实现了对大型实验仪器的可视化管理状态监测和智慧共享预约、服务于高校。

2.1 系统总体设计

2.1.1 需求分析

根据构建大型设备仪器智慧系统的关键要素以及目前仪器设备关键技术存在的不足，对于需要设计的仪器状态智能监测分析系统，主要需求有如下几部分：

(1) 改变传统人工登记使用仪器设备记录的信息以及只能到现场查看设备的工作状态的缺点，实现对大型设备仪器运行状态的实时监测，提高设备生产率和使用率。

(2) 通过构建智能采集终端将采集大型设备仪器的各个参数上传至云平台，实现本地资产数据库构建，将云平台的历史数据保存，再通过机器学习聚类 K-means 方法获得设备的工作阈值；反馈至智慧采集终端，降低无用数据重复上报，减少了云平台的存储压力；在同一区域智慧采集终端同时上报会造成网络拥堵和数据丢失、因此设计了错峰上报机制，降低了数据的丢包率。

(3) 通过改进的 K-means 自适应阈值的算法可以区分云平台的历史数据中仪器设备状态的 K 值，再用获得的阈值，反馈至智慧采集终端，实现边缘网动态调度限流机制，减少的对应网络请求响应延迟。降低了系统 NB-IOT 通信中 60% 的数据流量上传。同时提升了系统的稳定性、准确性、高效性。

(4) 开发对应 APP 软件服务，方便用户可通过客户端实现便捷的发现设备仪器是否空闲。APP 开发秉承便捷化、操作简单、服务更加人性化的原则，实现使用者和设备、管理者和设备的最便捷的操作。

(5) 通过贪心算法策略实现智慧共享，使大型实验仪器资源，可以更公平的被预约使用。

(6) 对客户端向服务器高并发请求时，通过使用 RabbitMQ 解决在用户遇到掉线和网络延迟带来的数据不同步等问题。

2.1.2 总体框架

设计智能采集终端通过采集大型设备仪器的各个参数上传至云平台，通过 NB-IOT 通信技术将采集的设备参数上传至云平台，并将云平台上传的历史数据保存，再通过机器学习聚类改进 K-means 方法获得设备的工作阈值。再反馈至 Web 端和 APP 页面显示设备实时的工作状态，以及相应的设备基本信息，实现云平台和 APP 的通信。另外 APP 可通过调用 API 请

求实现获取最新平台数据显示设备状态，根据设备状态实现对设备使用周期的相应调整。系统总体框架图如图 2.1 所示。

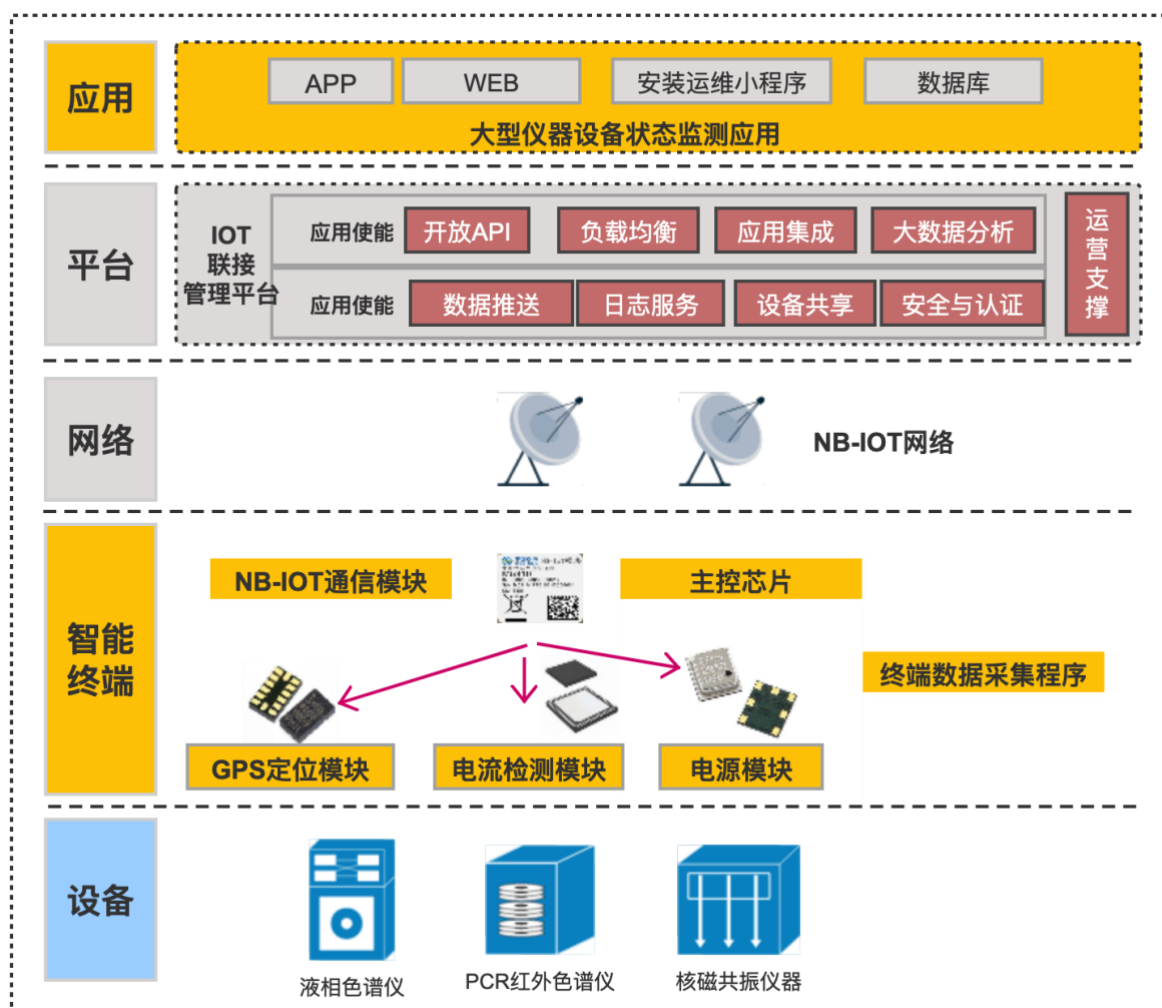


图 2.1 系统总体框架图

整个系统分为四个部分：智能设备采集终端、网络层、云平台以及可视化终端。

智能采集终端层分为主控芯片、电流检测、电源模块、NB-IOT 通信模块、GPS 位置采集模块五个模块。包括数据上报、数据清洗限制、错峰上报等功能。其中电流检测模块采集设备的实时电流和 GPS 定位模块采集的设备位置信息通过 NB-IOT 通信模块上传到云平台。

网络层是采用了电信云的 NB-IOT 的通信方式，实现讲智能采集终端的数据上传到云平台。

云平台分为设备云和业务云两个模块，包括数据存储、数据分析、数据格式解析、数据推送等功能。其中设备云与智能终端通过 NB-IOT 建立连接，实现大型实验仪器电流的采集，负责各种信息点的传输与存储；业务云主要是提供业务需求服务接口，通过 RESTful API(Application Programming Interface, 应用程序接口)对客户端软件提供调用服务接口。包括

仪器设备的基础信息查询、设备管理、设备共享预约、用户管理等。

客户端使用 API 来完成和云平台的业务交互。其中 APP 可以通过扫描高校资产二维码的信息，将设备的一些详细基本信息上传到云平台保存。并且提供在客户端和 Web 端上进行可视化管理服务。可以随时为科研工作者、管理人员提供仪器设备的相关服务。而小程序端设计了对接智能采集终端接入到仪器设备监测，为安装测试人员提供了更加高效、便捷的服务。

仪器设备状态智能监测系统的设计特点主要有以下四点：

1)实时性：智慧系统按照特定的时间间隔通过采集终端发起数据采集命令，采集的数据都是最新的仪器设备状态，而且实现在线实时反馈，让用户第一时间观察到仪器设备的最新状态。

2)准确性：在任何使用场景下，都需要保证用户操作能准确地转发到业务云服务器，对客户端向服务器高并发请求时，通过使用 RabbitMQ 解决在用户遇到掉线和网络延迟带来的数据不同步等问题。

3)稳定性：大型实验仪器所处实验室的环境网络波动大，对智能采集终端的通信能力要求很高，信号的抗干扰以及通信稳定性意义重大。

4)安全性：由于众多设备采集大量的设备信息，采用 base64 加密算法对数据加密化、信息不涉及隐私、为脱敏数据，保证了数据的安全。

2.2 智能采集终端设计

仪器设备的联网需要通过物联网传感设备的加装赋能。旨在采集获取仪器设备的使用电流等状态参数，通过设计物联网传感器设备，以便捷的方式配对安装到大型仪器上，实现仪器设备的统一状态联网接入。为保障设备的技术合理性、质量安全性、安装服务的可靠性、须建立设备入网的检测机制，以及设备安装维护的服务保障机制。

其中设备入网的检测机制要求对应硬件配置，包括设备的稳定性、安全性、兼容性等。传感器上传的信息进行加密，以防第三方信息的窃取和破解，与协作平台制定通信协议，设计数据解析模块，实现设备数据实时采集，快速接入平台。

由于高校实验室设备种类繁多，而且使用的额定功率和额定电压、电源供电方式不一样，以及一些实验室的精密仪器需要稳流器不能被电流所干扰。就此，针对以上情况，设计的智能采集终端，满足以下特点和功能：

特点：1、通过互感器非接触式获取电流信息，不影响仪器正常工作。2、传感器外挂，安装简易，高效、便于维护。3、免破线取电，独立、安全可靠。4、区分供电方式不一样，有单相、三相两种采集方式。

功能：1、实时采集仪器设备的电流参数，通过 NB-IOT 传输至云平台。2、云端联动，通过聚类算法，准确获知仪器设备的状态阈值，工作时长。3、错峰上报、边缘流量限制上报，极大降低了网络延迟以及云端存储压力。4、精确定位，快速查找，防止仪器丢失。

单相非接触方式如图 2.2 所示。

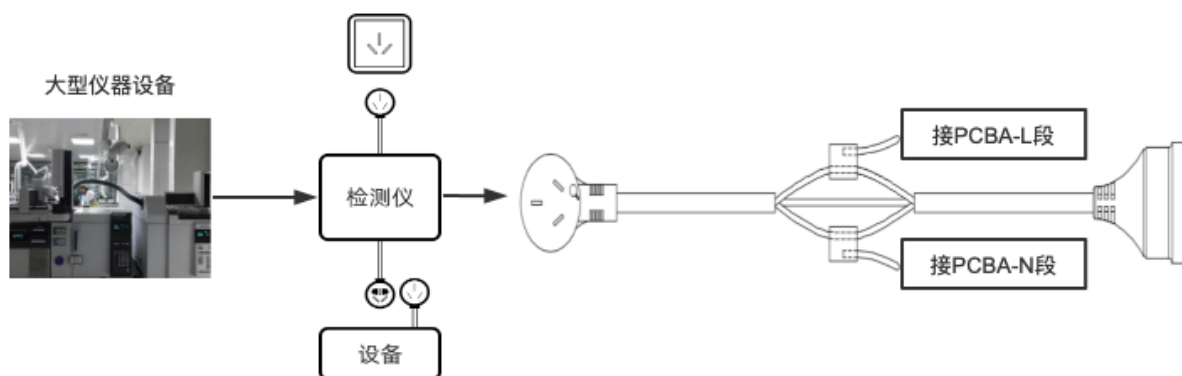


图 2.2 单相非接触示意图

三相非接触方式如图 2.3 所示。

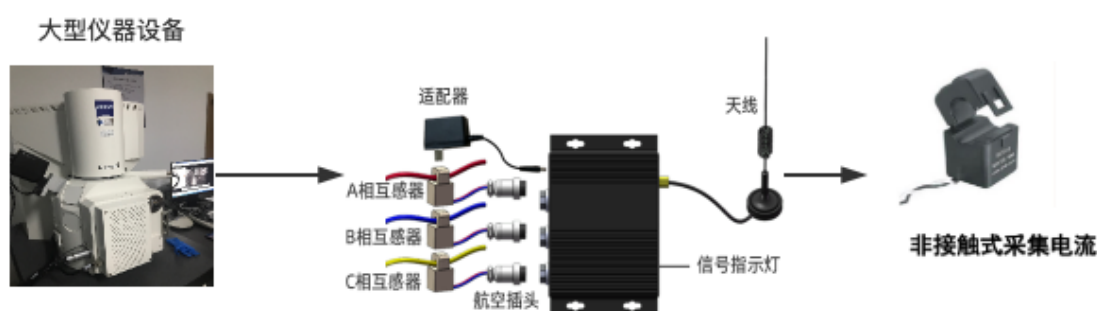


图 2.3 三相非接触示意图

2.2.1 硬件设计

智能采集终端硬件部分是基于 NB-IOT 技术的电流智能检测包括 MCU 处理模块、NB-IOT 联网模块、电流检测模块、电源模块、GPS 定位模块四个部分。其中智能采集终端主控硬件结构框图如图 2.4 所示：

主控 MCU 负责处理接口汇总过来的数据并且转发给所需要的其他模块以及对上报至云平台的数据进行数据清洗。MCU 主控模块主要包括主芯片 STM32F103C8T6 和晶振电路。STM32F103C8T6 具备高达 72M 的处理速度^[19]，使得该集中器具有很强的计算能力，拥有灵活的数据存储结构。并且具备 128KB 的 Flash ROM 和 20KB 的 SRAM，为整个系统提供程序存储空间和运行空间。而且该芯片内部支持众多的硬件接口，包括 DMA、SPI、TIM 以及 RTC 等。另外 STM32F103C8T6 还通过串口 UART 与 NB-IOT 模块进行通信。

NB-IOT 联网模块与云平台建立通信，将检测到的电流数据上报云平台。NB-IOT 模块选用的是移柯通信的 L620。L620 具有封装小，性能稳定的特点，在 3GPP 下 R13/R14 标准下涵盖的低频/中频/高频全频段使用^[20]；L620 以串口方式通信，由于 NB-IOT 模块 L620 与 MCU 芯片 STM32F103C8T6 的电压阈不同，所以在串口通信加入电平转换电路确保 NB 模块与

MCU 通信成功。NB-IOT 通信联网模块主要负责与云平台建立通信，将采集到的设备的工作电流实时上报。

电流检测模块通过电压型电流互感器采集设备电流并将采集到的数据输送给 MCU 处理模块处理；

GPS 位置采集模块通过获取到当前设备位置经纬度信息输送给 MCU。GPS 位置模块采用的是 Air530，具有体积小、精准度高、功耗低、可拓展软、硬接口的特点。因此，大大降低了用户的开发成本。

电源模块：供电并且负责各模块间电压源的稳定。

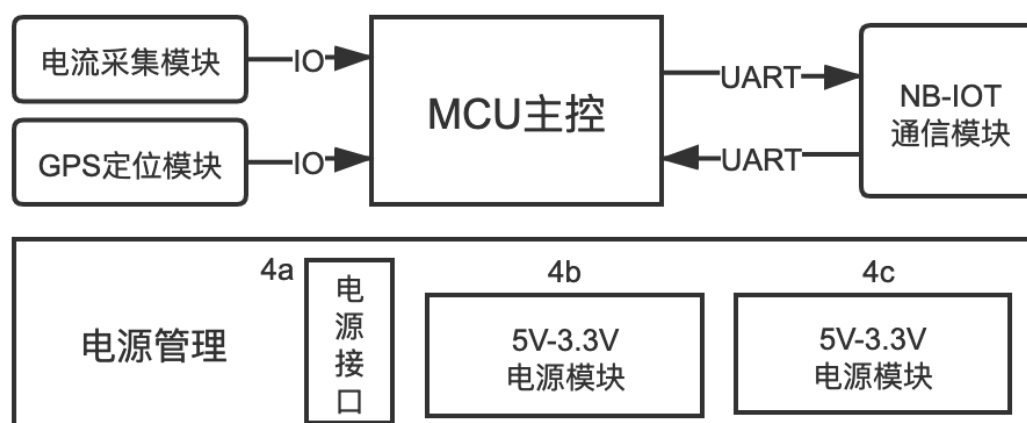


图 2.4 硬件结构框图

其中单相终端和三相的终端结构框图一样如图 2.2 所示，而不同的地方是电流采集模块以及电源供电方式。

单相终端里电源接口 4a 是对应 220V 转 5V 模块。而三相是所述电源接口 4a 与外接 5V 的适配器相连，5V-3.3V 电源模块。4b 和 4c 是相同的，分别为电流采集模块、NB-IOT 联网模块、MCU 供电。考虑到 NB-IOT 联网模块峰值电流可达 1A，所以电源模块分为 2 路 3.3V 供电，且 NB 模块供电电源选择输出电流值较大的电源芯片供电。

2.2.1.1 单相电流模块

单相电流采集模块的电路硬件框图如图 2.5 所示，由电流互感器输出采样电流，经过滤波网络，包括 LC 滤波、经过 RC 滤波。至 V9260S 芯片计算后输出 MCU。其中选用的是 V9260S 作电流检测芯片，主要负责被检测设备电流的检测，该芯片采用低功耗设计，并且拥有同时计量有功、无功电能的特点^[21]，且计量精度误差小于 0.1%。能很好的实现我们所需要的检测精度。

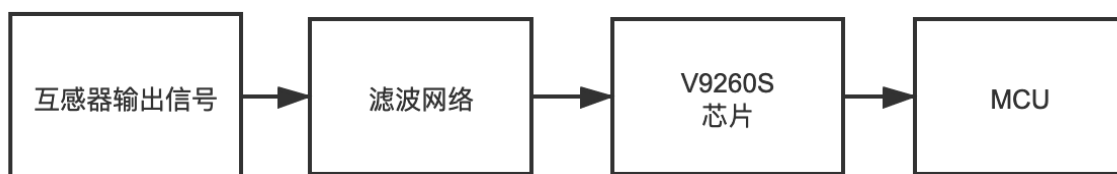


图 2.5 单相电流采集硬件框图

单相电流计芯片提供的采集信息包括相有功功率，设备电压，设备电流。设备的电量需要进行计算。计算如下：

$$PO = Kp \times P_{Freq0} \quad (2.1)$$

$$Q = \frac{PO \times 1}{1000 \times 3600} \quad (2.2)$$

式中其中 Kp ：有功功率转换系数， P_{Freq0} ：有功功率对应频率， PO ：有功功率， Q ：1 秒内的用电量 Kwh。

2.2.1.2 三相电流模块

三相电流检测模块的电路硬件框图如图 2.6 所示，主要是检测三相供电设备电流并将数据输送给 MCU 处理模块处理；TA1、TA2、TA3 为电压型电流互感器，分别检测 A、B、C 三相电流，由电流互感器输出采样电流，经过滤波网络，其中包括 LC 滤波、经过 RC 滤波。通过采样电阻转换成采样电压输出给后续电路处理；电压跟随以及电容滤波后送给 MCU 处理模块处理。

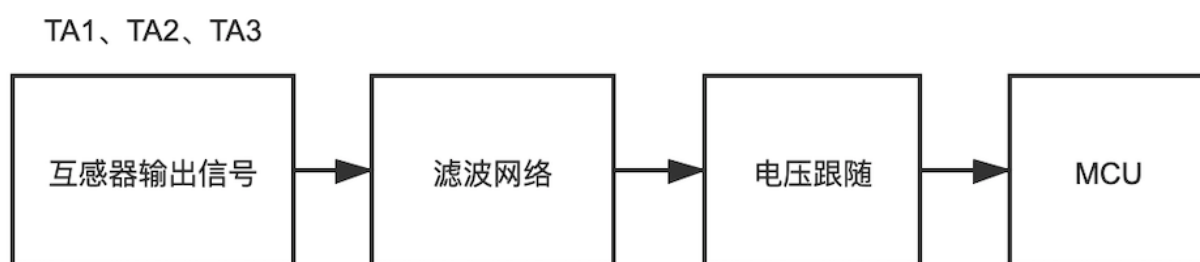


图 2.6 三相电流采集硬件框图

2.2.2 软件设计

智能终端软件设计的主要功能包括硬件驱动、NB-IOT 通信、数据处理、远程更新。其中

软件设计结构图如图 2.7 所示：

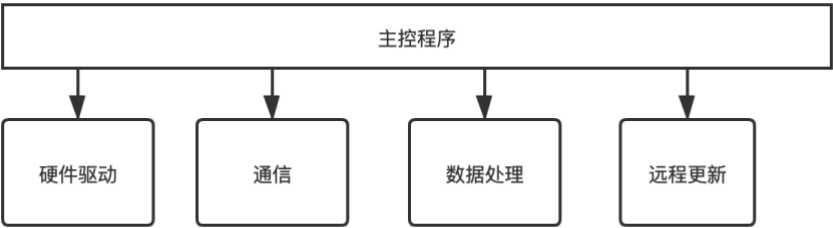


图 2.7 MCU 软件设计结构图

软件设计程序，主要负责外设初始化、电流计数据读取、GPS 位置数据读取以及数据上报。其中 MCU 数据处理程序流程如图 2.8 所示。



图 2.8 主控软件设计流程图

2.2.2.1 单相电流模块

单相电电流计的软件程序流程如图 2.9 所示：

其中单相电流计芯片读取的内容参数如表所示：

表2.1 单相电流计上报的数据内容

数据项序号	数据参数项名称
数据项 1	相电压
数据项 2	相电流
数据项 3	相有功功率

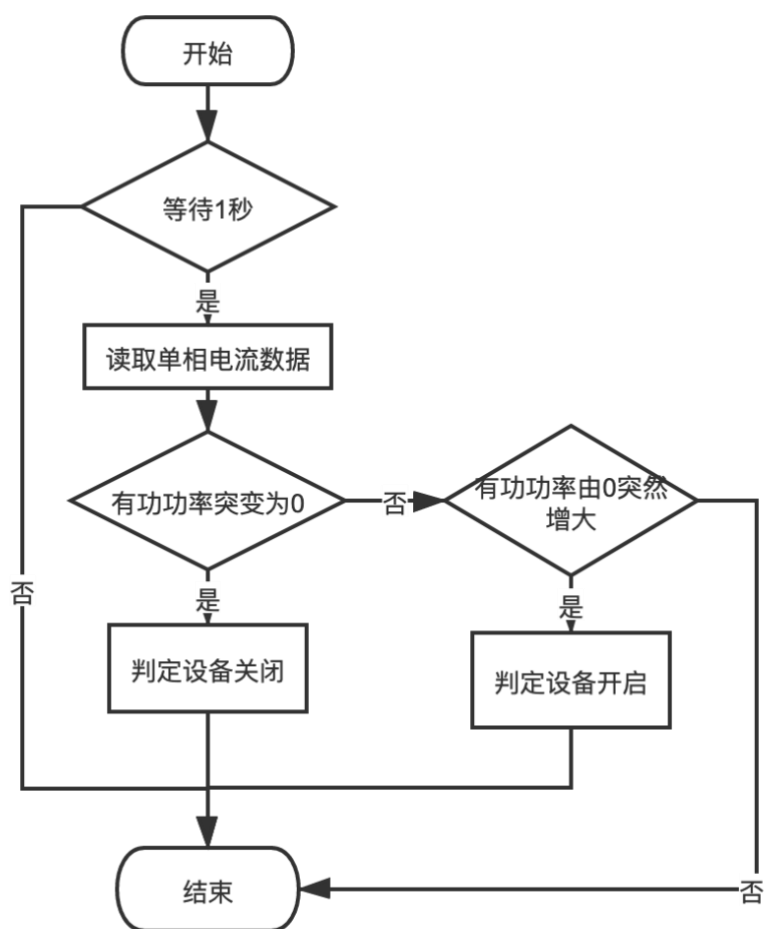


图 2.9 单相电流计软件设计图

2.2.2.2 三相电流模块

三相检测模块软件设计流程如图 2.10 所示：其中三相电电流计芯片读取的内容参数如下表所示：

表2.2 三相电流计上报的数据内容

数据项序号	数据参数项名称
数据项 1	A 相失压
数据项 2	B 相失压
数据项 3	C 相失压
数据项 4	有功总电量
数据项 5	总有功功率

由于三相电电流计芯片已经提供总用电量的数据统计，所以不需要再另行计算用电量。当

判断设备的开启或者关闭，需要将 A 相失压、B 相失压、C 相失压、有功总电量、总有功功率这 6 个数据进行上报。同时也会周期性的上报有功总电量。

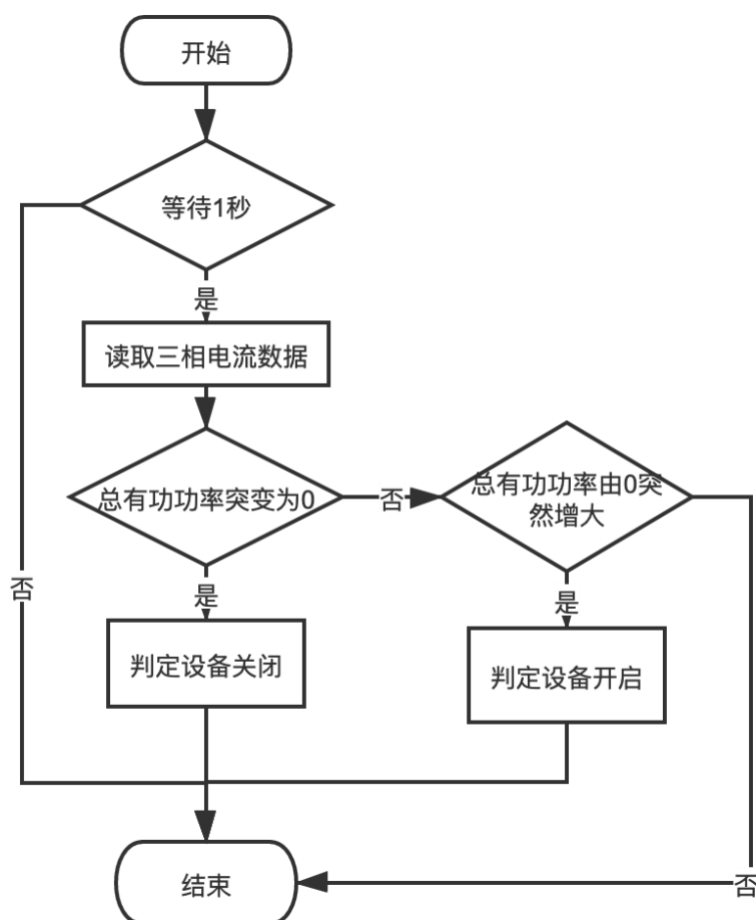


图 2.10 三相电流计软件流程图

2.2.3 终端接入方式和展示

2.2.3.1 单相智能终端

其中单相终端接入仪器设备方式。如图 2.11 所示。

将仪器设备的电源头插入单相终端检测仪的插座。单相检测仪连接 220V 市电插座后，蓝色指示灯保持 1 秒连续闪烁，闪烁时间根据 NB-IOT 网络信号强弱持续 3-10 秒，当蓝色指示灯不再闪烁并保持常亮，代表单相终端检测仪连入 NB-IOT 网络成功。此时可以在微信小程序端（2.5.3 小结）进行提交安装仪器设备的基础信息。



图 2.11 单相安装标示图

单相采集终端实物、单相采集终端主板实物如图 2.12，图 2.13 所示。



图 2.12 单相采集终端实物图

图 2.13 单相采集终端实物主板图

2.2.3.2 三相智能终端

其中三相终端接入仪器设备方式。如图 2.14 所示。三相检测仪主要是采用无接触式互感采集电流。优点是可以不干扰仪器设备的使用，保护线路安全，便于维护。

首先将三相检测仪适配器插头插入插座进行上电，此时蓝色指示灯保持 1 秒连续闪烁，闪烁时间根据 NB-IOT 网络信号强弱持续 3-30 秒，当蓝色指示灯不再闪烁并保持常亮，代表三相检测仪连入 NB-IOT 网络成功；其次，将三相检测仪引出的电流互感器 A、B、C 分别夹到仪器设备的 A、B、C 三相线上；最后，在微信小程序端（2.5.3 小结）进行提交，如果在小程序端进行状态学习中，不论是开机还是关机情况下，均出现学习状态不成功时，则拔下三相外壳中的小红帽，插上吸盘天线，重新进行状态学习。

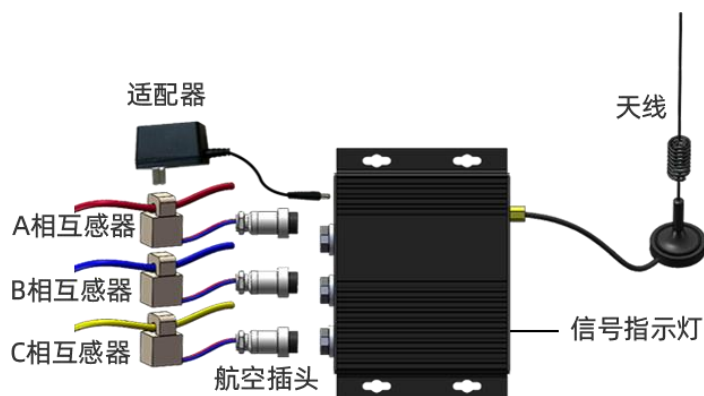


图 2.14 三相安装标示图

三相采集终端实物、三相采集终端主板实物如图 2.15，2.16 所示。



图 2.15 三相采集终端实物图 图 2.16 三相采集终端实物主板图

2.3 网络层设计

针对大型实验仪器使用的实时数据可以无差上传，通过基于 NB-IOT 的无线网络的设计实现设备及时上传，使仪器设备更好的接入互联网，从而直观展示大型实验仪器的实时状态，提升数字化智能工业化水平。本文设计了一种基于 NB-IOT 通信的无线局域网络，NB-IOT 无线网络首先采用蜂窝型结构，然后通过对通信网络协议采用数据点、周期上报的设计，提高 NB-IOT 网络通信的稳定性、可靠性。

2.3.1 常用的无线通信技术选择

在大型实验仪器共享系统中，设备间需要进行数据传输通信，采用无线方式的通信技术有很多，系统如何对应选取合适的通信方式应用在我们的系统中，尤为重要。下面从无线传输组网方式、功耗、网络节点、频段范围、传输速率、最大传输距离、网络延时、成本、应用场

景等参数对比。如下表2.3 所示：

表2.3 当前常用无线通信技术

	Bluetooth	ZigBee	WiFi	LoRaWAN	NB-IOT
频段范围 (Hz)	2.4G	2.4G	2.4G/5G	433M/470M/868M/915M	电信/移动/联通
功耗	低	低	高	低	低
网络节点	7	65536	32	5000~10000	5w~10w
组网方式	点对点	mesh	星型	星型	蜂窝型
传输速率 (bps)	720k	250k	11~54M	21.8k	200k
室外传输距离	50m	500m	50m	20km	35km
室内传输距离	10m	20m	15m	100m	300m
网络延时	2s	1s	30ms	TBD	5s
成本	中等	低	较高	较高	低
应用场景	资料传输	室内控制	网络通信	电表	物联网

当下，主要的无线通信技术是 Bluetooth、ZigBee、WIFI、LoRaWAN^[22]和 NB-IOT。其中 Bluetooth、ZigBee、WIFI 都应用于 2.4G 频段，会发生一定的同频干扰现象，其次 Bluetooth 主要应用的场景是文件传输领域^[23]，而 WIFI 一般是用在互联网网络通信领域。ZigBee 网络低成本、功耗较低，但是其主要应用于数据量少、距离短的场景^[24]。不能满足大量仪器设备数据上报的应用要求，所以也不能使用。LoRaWAN 一般在国内的工作频段是 470M，主要是在物理层引用了 LoRa 扩频调制方式，主要是 LoRa 用于多数是建立网关也就是需要构建复杂的星型结构，另外 LoRaWAN 部署成本较高，所以 LoRaWAN 也不适用本系统。而 NB-IOT 适用于低成本、低功耗、低带宽、广覆盖的使用场景^[25]，并且在 5G 技术大环境的发展，也一直被应用于智慧农业、智慧城市等。与 LoRa 比较，NB-IOT 在射频天线上都是重复利用的并且不用建网。综合对比选择了 NB-IOT 通信方式。而 NB-IOT 一般在国内主要是三大运营商，下表2.4 是各个运营商对应的频段、频宽。中国电信在对 NB-IOT 终端业务开展以及研发大力投入。全面部署了 800MLTE 网络覆盖。

表2.4 国内 NB-IOT 各个运营商的频段

运营商	上行频范围 (MHz)	下行频范围 (MHz)	频宽 (MHz)
中国电信	825-840	870-885	15
中国移动	890-900 1725-1735	934-944 1820-1830	10 10
中国联通	900-915 1745-1765	954-960 1840-1860	6 20

2.3.2 网络基本架构

NB-IOT 的网络体系架构，包含了终端、eNodeB、核心网、云平台以及服务器^[26]。其中体系架构如图 2.17 所示：

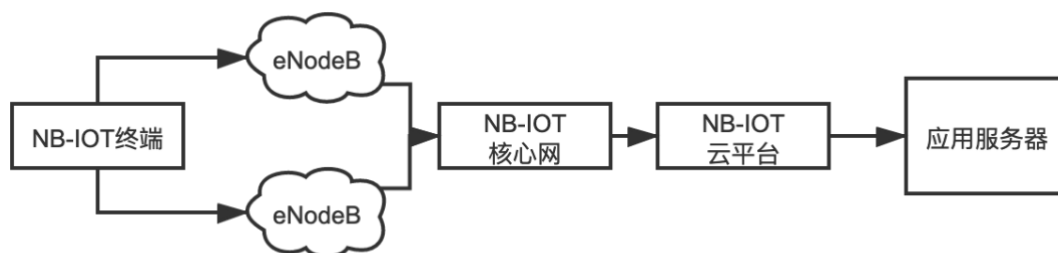


图 2.17 NB-IOT 网络基本结构

终端：接入基站通信，负责接入 NB-IOT 网络和数据采集上报以及执行命令。

eNodeB：基站。构成蜂窝小区进行移动通信，前接终端入网以及 UE 通信的功能，后接 NB 核心网负责数据转发。

核心网：负责基站和云平台的交互枢纽，还负责一些网络拥塞和流量控制。

云平台：负责存储终端上报的数据。提供不同的接口供应用服务器调用处理。

应用服务器：负责自定义应用场景。按照具体应用场景对数据处理分析等，通过 API 与移动端进行数据交互，实现可视化管理等功能。

2.3.3 NB-IOT 通信协议

在网络层中应用了 NB-IOT 通信模块，其中协议里需要通信的数据类型一共分为四种：设备联网初始化信息上报；设备周期性数据上报；异常情况报警上报；电信云 MCU 远程升级通信。按此四种数据类型上报格式进行描述。电信云平台的数据流是按地址码区分，故以上数据类型按地址码做区分，地址码一共 2 个字节（字符串格式）。

2.3.3.1 设备联网信息上报

设备通电后会主动接入网络，联网之后会自动初始化信息即刻上报。初始化信息主要是 4 个数据项如下表 2.5 所示。上报格式如下：

{"INIT_ID ":"<模组 IMEI 卡号>,<IMSI 卡号>,<设备版本号>,<信号质量>"}

表 2.5 设备联网信息上报协议数据内容

参数项	参数名称
1	模组 IMEI 号
2	IMSI 卡号
3	设备当前软件版本号
4	当前信号质量

参数说明：模组 IMEI 卡号，是设备的型号身份码，一般是 15 位数组成。

数据 IMSI 卡号，NB 模块内置卡号，一般是 15 位数组成。

设备当前版本号：MCU 远程更新中标注的代码版本号，固定为 4 个字节。

当前信号质量：以 rsrq 值大小为依据，rsrq 信号值大小对应着当前信号好坏。正常值是大于 10 都可以正常使用。

2.3.3.2 设备周期性数据上报

设备上电后，进行每隔一个小时上报一次数据。数据内容主要是 6 个数据项如下表 2.6 所示，包含总电量、定位信息状态、纬度说明、定位信息纬度、经度说明、定位信息经度。上报格式如下：

{"NORMAL_REPORT_ID ":"<总电流>,<定位信息状态>,<纬度说明>,<定位信息纬度>,<经度说明>,<定位信息经度>"}

表 2.6 设备周期性上报协议数据内容

参数项	参数名称
1	总电流
2	定位信息状态
3	纬度说明
4	定位信息纬度
5	经度说明
6	定位信息经度

参数数据说明：

总电流：表示的数值范围为 0-99999.999A。

定位信息状态：由一个字母表示，A/V 分别表示位置状态是否有效。

定位信息纬度：表示的数值范围为 0000.0000~9000.0000。

纬度说明：由一个字母表示，N/S 分别代表纬度是北纬还是南纬。

定位信息经度，表示的数值范围为 00000.0000~18000.0000。

经度说明：由一个字母表示，E/W 分别代表经度是东经还是西经。

2.3.3.3 异常情况报警上报

设备联网成功后，当设备发生设备开关的情况，就会主动上报该条数据。数据内容主要是 5 个数据项如下表 2.7 所示，包含设备开关、相电压状态、有功功率。上报格式如下：

单相电流计报警数据：

{"ALARM_ID": "<设备开关>,<电压状态>,<有功功率>"}

三相电流计报警数据：

{"ALARM_ID": "<设备开关>,<A 相状态>,<B 相状态>,<C 相状态>,<有功功率>"}

表 2.7 三相电流计异常情况上报协议数据内容

参数项	参数名称
1	设备开关
2	A 相电压状态
3	B 相电压状态
4	C 相电压状态
5	有功功率

参数数据说明：

设备开关，表示当前设备开关状态，0 表示设备开启，1 表示设备关闭

A 相状态，表示当前设备 A 相有无电压，1 表示无电压，0 表示有电压

B 相状态，表示当前设备 B 相有无电压，1 表示无电压，0 表示有电压

C 相状态，表示当前设备 C 相有无电压，1 表示无电压，0 表示有电压

有功功率，表示当前设备的有功功率，范围为 0000.0000~9999.9999Kw

以上之所以要带上 ABC 三相状态和有功功率，是由于当设备处于关闭情况下，有两种可能性，一个是设备断电，一个是设备没启动，这两种情况分别对应情况如下表 2.8 所示：

表2.8 设备关闭时情况

设备断电	1) ABC 三相至少一相无电压 2) 有功功率接近 0
设备没启动	1) ABC 三相都有电压 2) 有功功率接近 0

2.3.3.4 电信云 MCU 远程升级通信

在电信云平台创建软件升级任务后，若平台上产品为 PSM 类型，当相应设备上报任意数据到平台后，平台才会下发指令启动 SOTA（Software-Over-The-Air）过程。反之平台上产品为非 PSM 类型，平台会立即下发指令启动 SOTA 过程。指令启动实际为版本查询指令，得到设备的版本回复后，平台才会继续后面流程。同样在设备 MCU 侧也有版本比较，只有不同时才会进行更新。

后续流程基本就是设备向云平台请求数据包，则平台对应下发一包。为防止网络原因丢包，设备具有重发机制，在远程升级时间段可以重发 10 次。若设备重发次数已满，仍未收到云平台数据，则设备判定更新失败，退出远程升级模式。

同样，在更新过程中，平台侧若一段时间未接收到设备的上报指令，则会判断更新失败，具体时间根据进展的流程而定。当设备接收完全部的数据包后，会向平台发送接收成功帧以及新版本序号等，平台收到并确认后，会显示软件升级任务成功。具体界面图 2.18 所示。其 MCU 的整体远程更新流程图如图 2.19 所示。



图 2.18 升级成功在平台显示图

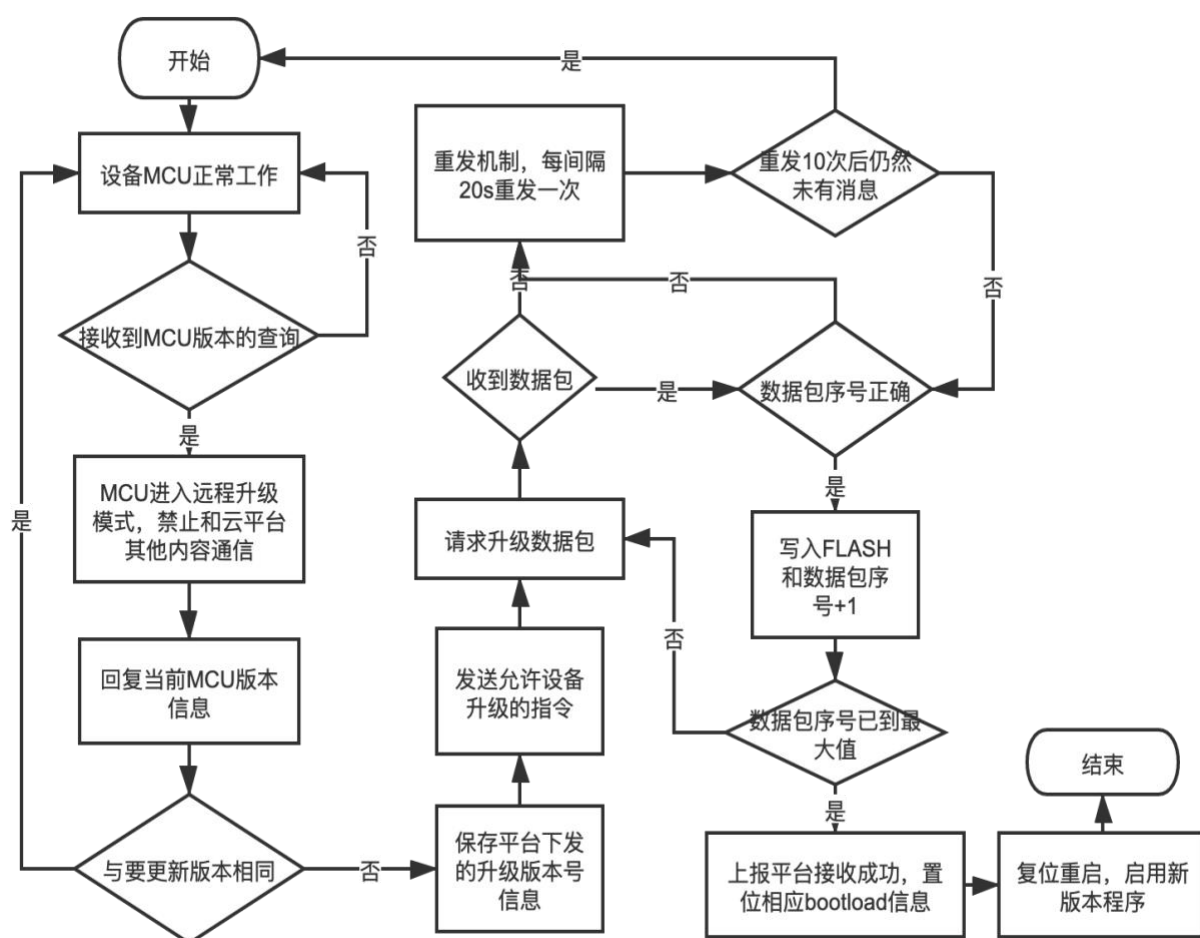


图 2.19 MCU 远程更新流程图

1) 步骤一：上报当前 MCU 版本号。

收到查询帧后上报当前 MCU 版本号。传输方向：从平台下发至设备查询设备版本；当设备接收到平台的查询软件版本后，根据协议内容格式对应从设备端返回至云平台上报当前 MCU 的版本号信息。

2) 步骤二：新版本通知。

从平台对设备端下发更新版本通知，设备端先根据协议内容规定进行判断是否允许升级。如果可以升级则进行第三步操作，反之停止更新命令。

3) 步骤三：更新包下载。

当确认可以升级时，设备端开始向平台请求升级包的发送。为避免数据丢失，加入重复下发的机制。这样可以有效的确认数据下发成功。但是，也会出现另一个沾包问题。因此，也在发送协议中增加了帧标示序列号，对数据的先后进行区分，有效的避免了沾包问题。然后，升级包的请求依次按照升级包序列号进行逐步操作即可。

4) 步骤四：上报升级包下载状态。

依据协议规定，在请求更新包下载后，设备端需实时上传查询信息流，通知平台更新下载包的进度。之后步骤五中设备端会确认进行升级。

5) 步骤五：执行升级。

在设备端处于升级包下载状态时，平台下发执行升级命令。

6) 步骤六：上报升级结果。

设备升级结束，需要上报此刻设备的版本以及更新结果。

2.3.4 关键机制

(1) 超时重发机制

超时重发机制的处理流程如图 2.20 所示，在当需要进行数据发送时，都会遇到网络拥堵、网络中断等，导致服务器没有有限及时的反馈至请求端。并且为了防止长时间等待服务器的的应答信号，我们需要设计超时重发机制，确保消息能够准确的到达服务端。而且也设计了三次重发机制，如果重发三次还是没有回复，确定设备与服务器断开连接了，停止重发。

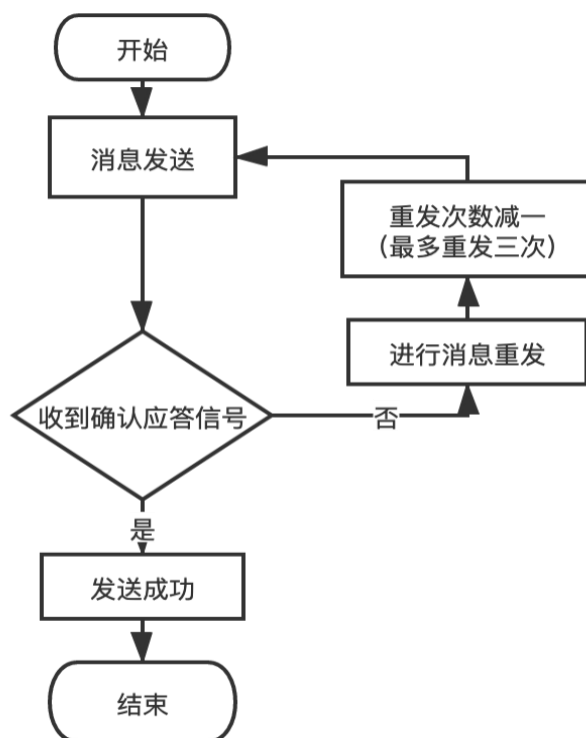


图 2.20 超时重发机制流程图

(2) 错峰上报机制

基于 NB-IOT 的通信模块在上报数据时，基本都需要设定周期性上报、仪器设备的状态的数据等。但是相对在同一区域、同一时刻下，每个基站的数量是有限的，当我们同时将采集的数据上报时，会导致网络阻塞、使数据丢失，而在下一时刻，基站又处于空闲状态，有效的资源没有合理的利用。因此，我们设计了一种错峰上报机制算法：

$$T = \left(IMSI \bmod \left(\text{int} \left(\frac{3600}{F} * |t_1 - t_2| \right) \right) \right) * F + x \quad (2.3)$$

式中：IMSI 是 NB-IOT 设备的号的后八位、 F 为错峰的系数、 t_1 是起始时间、 t_2 是结束时间、 x 是在 $0 \sim F$ 间的随机数，并且要求是整数。其中 $\text{int}()$ 、 $\text{mod}()$ 分别是取整函数和取模函数。

(3) 加密机制

智能采集终端在上报信息至云平台时，系统采用了 base64 对数据流字符串进行加密，保证了数据的安全性。加密机制流程如图 2.21 所示。

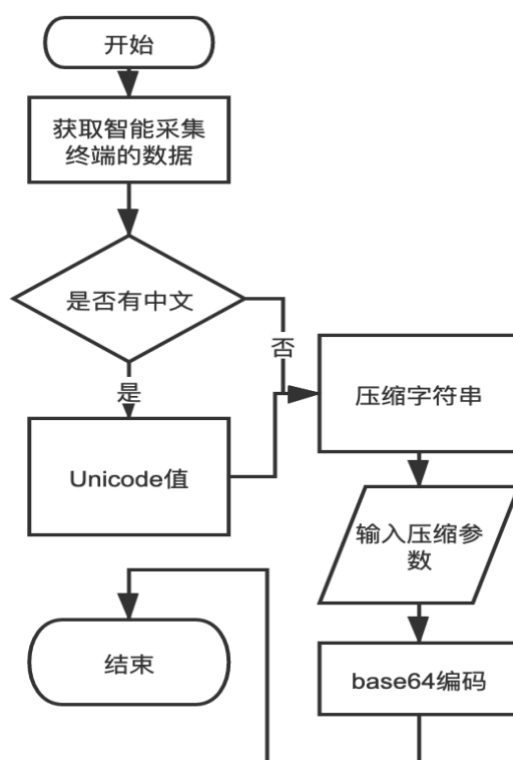


图 2.21 加密机制流程图

2.4 云平台

云平台设计分为设备云和业务云两部分。云平台的主要原则是确保客户端与设备之间的交互枢纽，完成可视化数据流状态管理、数据自动采集等功能。其次云平台也必须保证数据不会发生泄漏、降低丢包率、可扩展、实现更加稳定、安全的云平台设计才能保证整个系统的高性能。该智慧系统同时搭载了设备云服务器和业务云服务器：其中设备云为 IoT 设备提供了的接口通信协议，可以实现设备方便快捷地通过协议接口对云端传输采集数据；而业务云这负责提供 API 接口，向客户端提通过 API 查询设备相关的信息，并且业务云通过设备云的提供的 API 实现两者的交互。

2.4.1 设备云设计

设备云平台是采用中国电信云平台，可以解决 NB-IOT 将采集数据上报，按照约定好的通信协议解析设备上报的数据，而且可以进行远程设备更新，并自动存储设备上报的数据。我们系统的设计目标需求接入大量的仪器设备，因此设备云平台的设计就必须是高可用、高并发的大型平台，这样的设计需要极高的运行成本，个人或者小公司的实力无法实现这样一个平台，而且电信云有庞大的用户群体，服务器有接入海量大数据的能力，而且技术较为成熟。符合本系统开发需求。

使用第三方物联网云平台可以降低设备入网的难度，将工作重心放在自身应用的开发，降低了系统研发的开发周期。这也是我们设计所需求的。AEP 云平台是第三方物联网开发服务平台，由巨头电信公司开发并运维，与 OneNET、WASTON 等平台类似^[27]，都是为物联网客户端应用与设备之间提供优秀的通信桥梁。

其中设备云平台架构如图 2.22 所示：

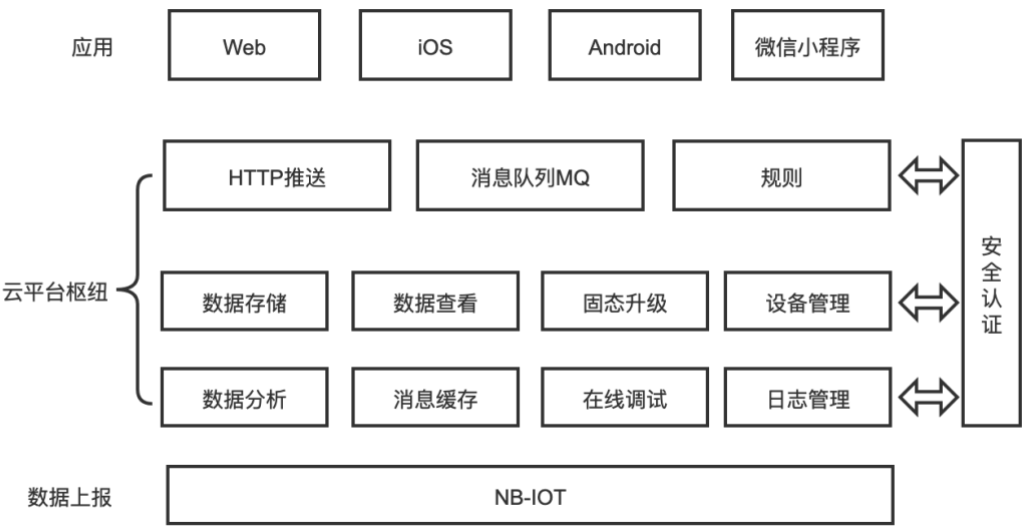


图 2.22 设备云平台架构图

AEP 电信云可以为设备提供数据存储、远程升级、数据分析、消息队列、数据查询、日志管理等功能并有完善技术支持体系。AEP 为智能系统提供 NB-IOT 的协议接入，实现智能采集终端设备的快速接入；同时对客户端提供了基于 RESTful 的高可用接口和消息队列 MQ 推送的交互服务等^[28]。并且客户端可以节省很多的系统资源，遗弃轮询查询方式等低性能低效率的操作。

其中智慧采集终端和设备云平台之间的模型如图 2.23 所示：

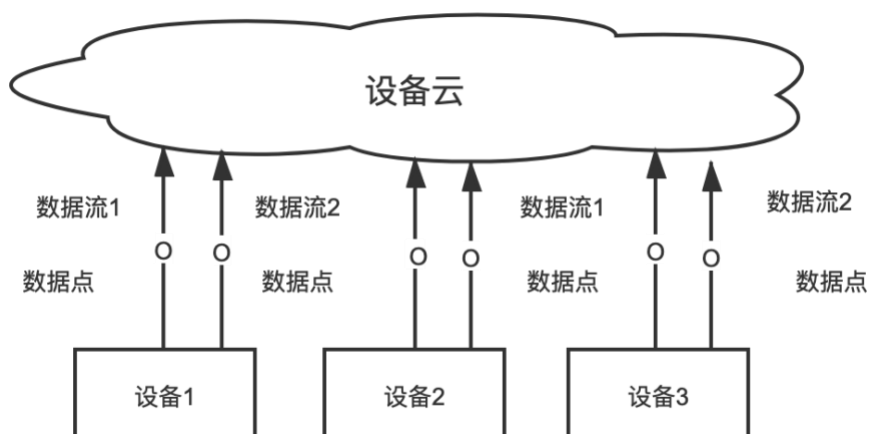


图 2.23 设备与设备云的交互模型图

在 AEP 中的 API，我们选择的是基于 RESTful 方案设计的，它可使设备通信相关的数据模型命名为多个关键名词：用户、设备、数据流、数据点、token、APIkey 命令等^[29]。其中数据流一种较为特殊数据类型，比如电流值、时间，都是一种特殊数据流。当然其中数据点也对应数据流类型，数据流传输在 IOT 应用场景较多，因为物联网下，设备众多，采集的数据量大，通过实时数据流可以提高物联网设备传输的数据量，而且也有利于工业物联网下的仪器设备数据进行在线分析^[30]。

2.4.2 业务云设计

业务云平台主要是由第三方阿里云提供的 ECS 云服务器上搭建的业务服务器集群，负责了智慧系统在客户端软件方面交互时的一些复杂问题。业务云平台的整体架构如图 2.24 所示。

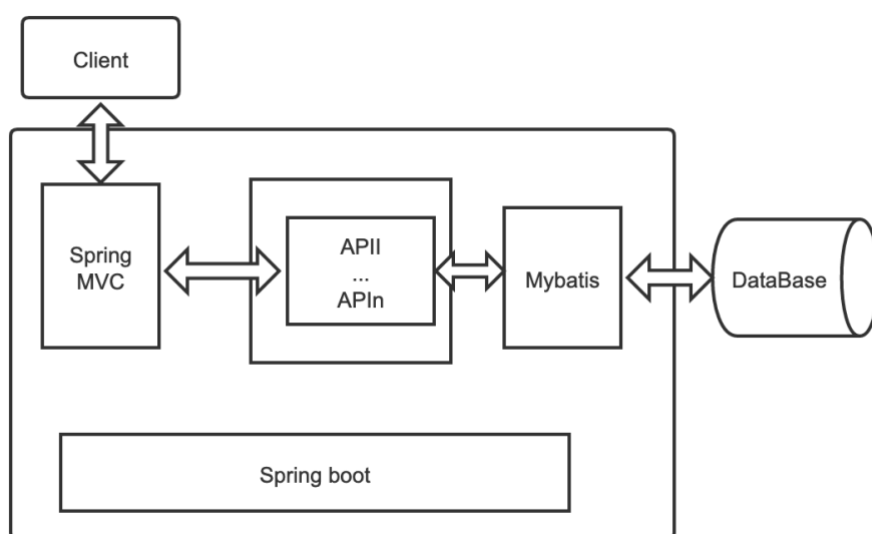


图 2.24 业务服务器架构图

在负责客户端的后端服务平台，采用的是 Spring+SpringBoot+Mybatis 技术栈进行架构开发。满足可扩展性、准确性、存储调度的功能等^[31]，而且提供高可用的 API，供客户端获取后端服务器的设备信息。主要的功能包括、设备负载均衡、设备管理、用户信息管理、设备信息存储等模块功能。并且客户端通过使用 HTTP 协议，去后台服务器进行数据增删改查等逻辑操作。其中通过简单的操作 Spring Boot^[32]，就可以高效的动态搭建一个生产级别的 Spring 网站，可以直接运行使用，这也大大提高了我们开发的效率。

业务云中数据库，选择了 Mysql，它可以把数据库中表内容和业务服务器中的 DAO 逐层对应，并通过 Mybatis 框架来实现 Mysql 数据库的增删查改的调用，可以简化向面向对象引用代表复杂的数据库操作，极大地降低了研发的项目周期。

2.5 应用层设计

2.5.1 Web 设计

2.5.1.1 Web 框架设计

目前 Web 客户端技术发展已经非常成熟，具体丰富的生态开发环境，热门的主要框架是 React、Vue 等技术，其中第三方库类型众多，功能也比较完善^[33]。为本系统开发提供了很多技术支持。本文选择了更加轻量化的 Vue 框架，因为相对适合于物联网环境下，而且开发更灵活、扩张性也很强，满足本系统的实现需求。

利用 Vue 框架中的 npm 开发管理工具，通过下载导入一些第三方的功能模块，降低周期的开发，并且可以实现状态管理、路由管理、异步请求等技术。可以快速灵活的迭代业务。利用高效的 HTTP 请求通信方式，可以实现与设备云平台、业务云平台的交互。Web 客户端架构如图 2.25 所示：

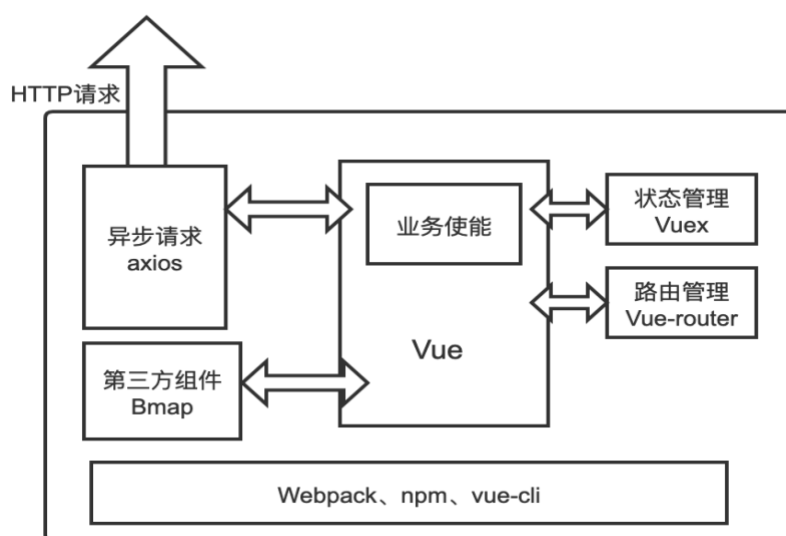


图 2.25 Web 客户端架构图

2.5.1.2 Web 端仪器管理功能设计

Web 管理服务平台基于浏览器开发，是用户运维整个大型实验仪器设备状态监测系统的 PC 终端，并且包括信息门户、设备管理、数据监控、历史信息查询、仪器预约等功能模块，如图 2.26 所示。通过利用 AEP 提供的第三方接口，在 Web 端实现功能交互。其中 Web 端一般通过 API 向业务云平台下发对设备的命令帧，从而获取查询仪器的设备的数据流相关信息，并且在 Web 端管理员可以审核一些预约请求。

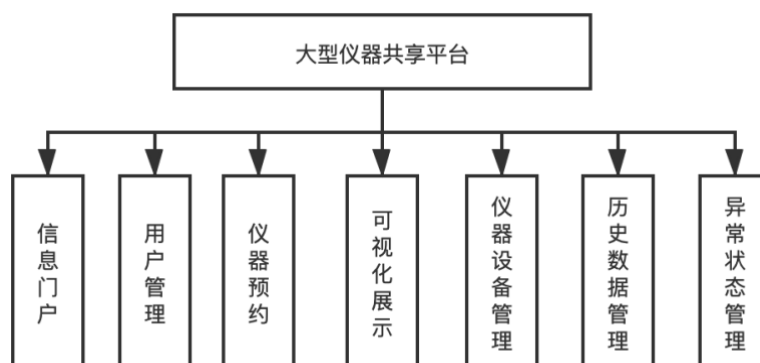


图 2.26 Web 端功能模块

详细功能概况如表2.9 所列出：

表2.9 大型仪器共享平台功能表

条目	内容
信息门户	通知公告、平台简介、仪器使用规范制度、可共享的仪器列表
用户管理	校内登录、校外手机注册登录、管理人员注册登录
仪器预约	仪器查询、预约仪器查询、仪器预约、审核预约单
可视化展示	仪器使用情况的基本信息展示（使用率、使用时间等）
仪器设备	仪器的添加、仪器的修改、仪器的删除、仪器的使用记录、仪器使用的反馈意见
历史数据管理	个人的预约记录、仪器使用情况、不良的仪器使用记录
异常状态管理	记录仪器设备出现故障信息

2.5.1.3 Web 端系统展示

Web 客户端是大型实验仪器共享系统的主要用户终端之一，同时也是管理者的运维终端。Web 端是大型实验仪器系统数据的可视化重要平台，将仪器使用的情况的实时数据显示在屏幕上，为和科研人员、管理人员提供服务。其中 Web 客户端包括信息门户、用户管理、仪器预约、可视化展示、仪器设备管理、历史数据管理、异常状态管理等多种模块，如图 2.27、图 2.28、图 2.29、图 2.30、图 2.31 所示。



图 2.27 高校仪器共享平台登录页面图



图 2.28 高校地图大型实验仪器分布图



图 2.29 高校大型实验仪器实时数据统计图



图 2.30 仪器使用机时情况统计界面



图 2.31 人员实验的情况统计界面

2.5.2 APP 设计

2.5.2.1 设计模式

智能系统移动端 APP 应用了单例模式、多线程模式、命令模式、生产模式等设计模式。在 APP 的架构上采用了 MVC 的分层架构。MVC 中的字母 M 代表 Model 负责数据管理；字母 V 代表 View 负责将 M 的数据以视图形式展示出来，以及负责处理响应事件；字母 C 指视图控制器 Controller。MVC 模式中交互管理如图 2.32 所示：

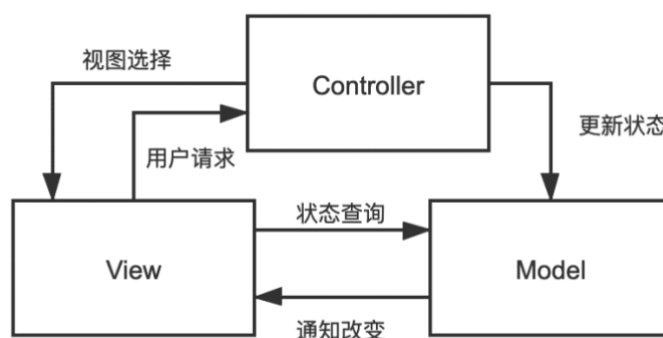


图 2.32 APP 客户端 MVC 架构

1) Model: 模型，数据库核心内容，保存用户和设备的数据以及进行传递设备的一些状态，处理 View 的状态的查询事件，一般可以利用消息通知、委托代理等机制将数据或状态变化传递给 View。

2) View: 视图，移动端 APP 的展示层，一般是提供响应用户的触控操作并且对用户的操作做出对的视图选择。当用户在 View 层操作时，代理 Controller 层会向 Model 层作数据请求。

3) Controller: 控制器，是模型与可视界面之间的交互枢纽，因为遵循设计的单一指责，Model 层与 View 层不能直接交互数据，必须通过相当于第三方的控制器进行双方数据交互，从而响应用户的操作，符合程序的单一原则，大大提高执行效率。

2.5.2.2 网络层设计

网络层是 APP 与业务云平台通信的关键，是与外界交互重要的方式。本文对大型实验仪器智慧共享系统客户端着手于三个技术点：

1) 命令请求方式封装：用户某一时刻操作 APP 发起命令请求，要求返回请求的结果展示在 APP 中，而 URL 和请求参数构建这一类业务，可以将这些相同的命令请求代码出来，形成一个基类封装，暴露出接口，方便调用，也简化了代码重复使用，实现了离散型请求架构。

2) 命令请求返回数据解析：每一次用户命令请求操作后，会返回在完整的 request & response，命令请求返回的数据可以分四种状态，如图 2.33 所示，需要将数据转模型化，实

现面向对象编程思想。

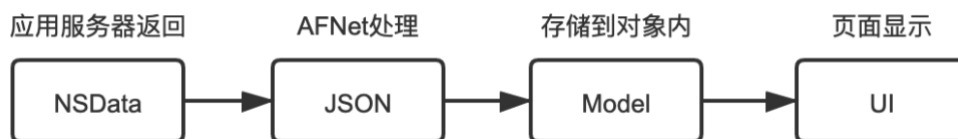


图 2.33 命令请求回复解析过程

3) RabbitMQ 订阅推送模式: 客户端通过订阅 RabbitMQ, 当客户端向服务器高并发请求时, 通过使用 RabbitMQ 推送解决在用户遇到掉线和网络延迟带来的数据不同步等问题。推送过程如图 2.34 所示:

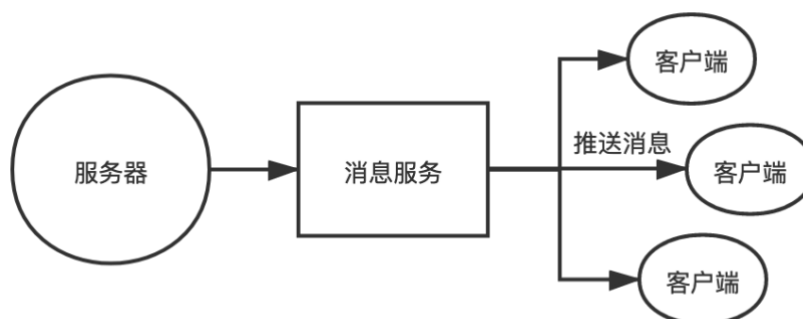


图 2.34 RabbitMQ 订阅推送过程

这些全局的命令, 本文都把他们放入单例模型一个特殊 manager 类进行管理、控制中, 方便全局使用, 将离散型请求的代码逻辑抽离为统一行为接口, 通过调用使用, 减少了大量的重复性代码, 设计了重发机制和同步机制提高了命令请求命中率, 优化了数据的准确性。

2.5.2.3 功能设计

APP 客户端运行于移动设备上, 并且屏幕大小都与 PC 端有所区别, 目前在移动端的功能设计少于 Web 端, 比较复杂的功能还是需要用户去 Web 去执行。APP 常用的:

1) 用户登录: APP 登录, 主要针对校内人员, 数据对接学校的数据库。学生可以用学号直接登录。

2) 仪器列表: 学生可以查询仪器设备列表, 获取仪器设备是否可以预约等。

APP 也支持二维码扫描仪器设备的资产二维码准确查询该仪器设备。

3) 历史数据管理: 用户在当前页面可以查询自己之前, 使用的仪器设备记录。

4) 仪器预约功能: 学生可以选择自己预约仪器设备的使用时间, 以及查看管理人员的审批流程。

5) 仪器管理: 学生在使用完仪器设备需要确认, 并且可以在里面上报仪器设备的故障问

题。

2.5.2.4 APP 结果展示

大型实验仪器共享系统的 APP 客户端的功能有，登录功能、仪器列表、仪器管理、仪器预约等功能。其中实现效果如图 2.35，图 2.36 所示。

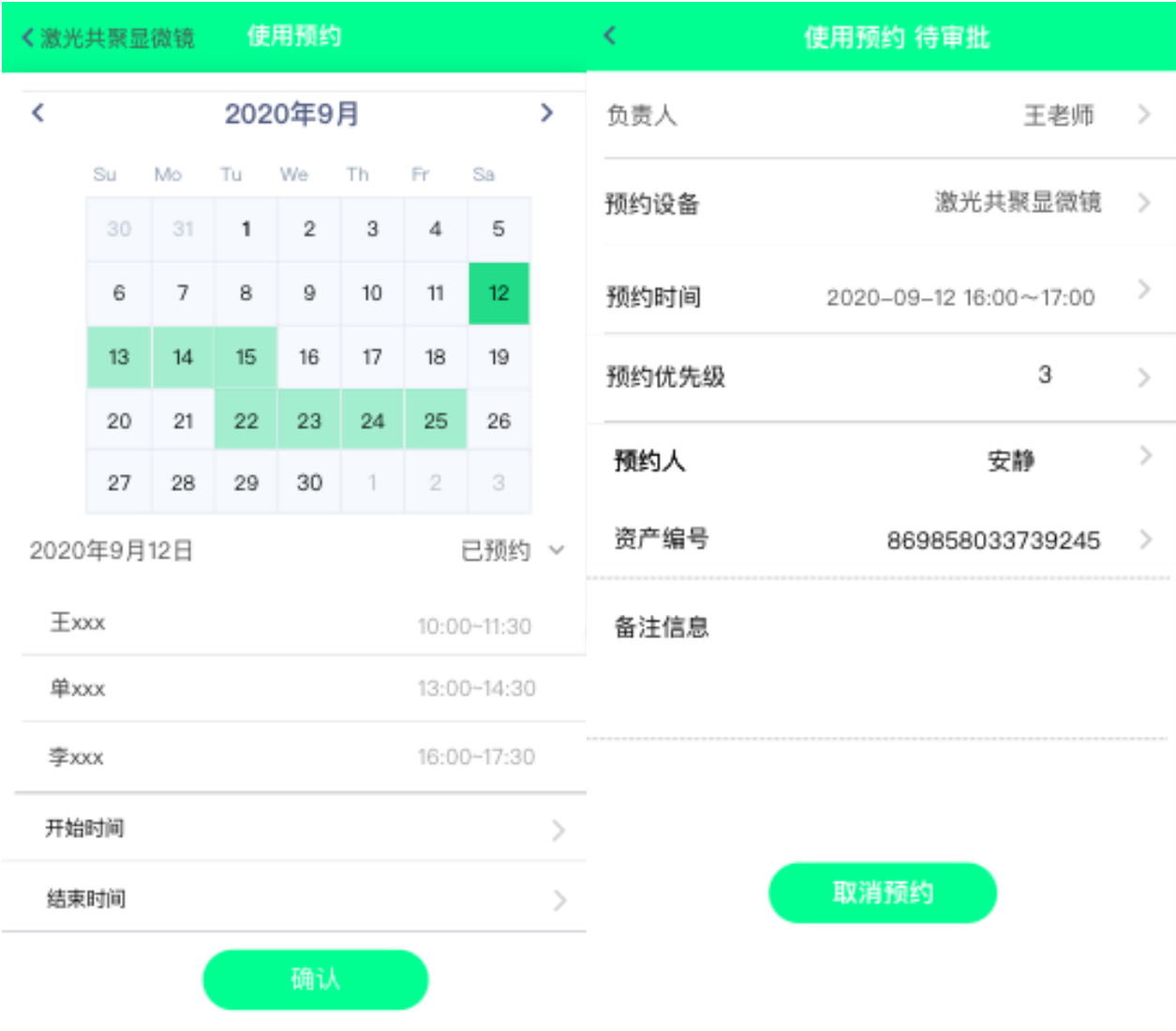


图 2.35 仪器设备使用预约

图 2.36 仪器设备待审批页面

2.5.3 安装运维小程序设计

针对智能采集终端接入仪器中监测电流，设计了安装运维小程序，方便安装人员操作。主要实现智能采集终端接入仪器的安装，仪器监测，仪器资产的基础信息录入等。

2.5.3.1 功能设计

安装运维管理服务端基于微信小程序开发，是用户运维整个智能采集终端接入大型仪器

的移动终端，并且包括安装人员登录、资产信息统计、生产检测、故障问题、设备入库、出库等功能模块。如下图 2.37 所示。

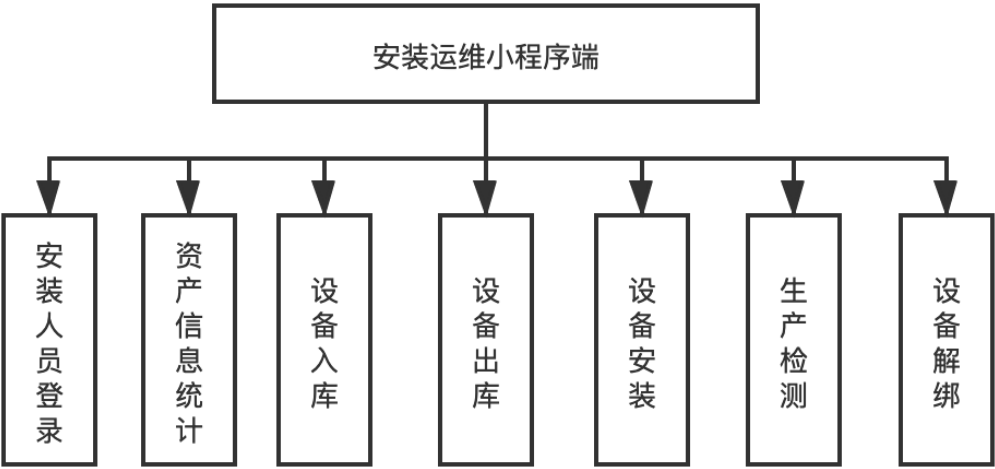


图 2.37 小程序端功能模块

详细功能概况如表 2.10 所列出：

表2.10 安装运维小程序功能表

条目	功能内容
注册、登录	安装人员、运维管理注册、登录。
安装列表	安装列表查询、安装详情查看、设备解绑
安装设备页	设备测试、设备安装
仪器设备统计	仪器设备资产基础信息录入
个人信息	安装人员、管理人员信息
设备巡检页	设备巡检提交
解绑列表页	设备解绑查询

1) 安装人员登录：微信小程序注册、登录。主要针对运维人员、安装人员。其中软件设计流程如图 2.38 所示。

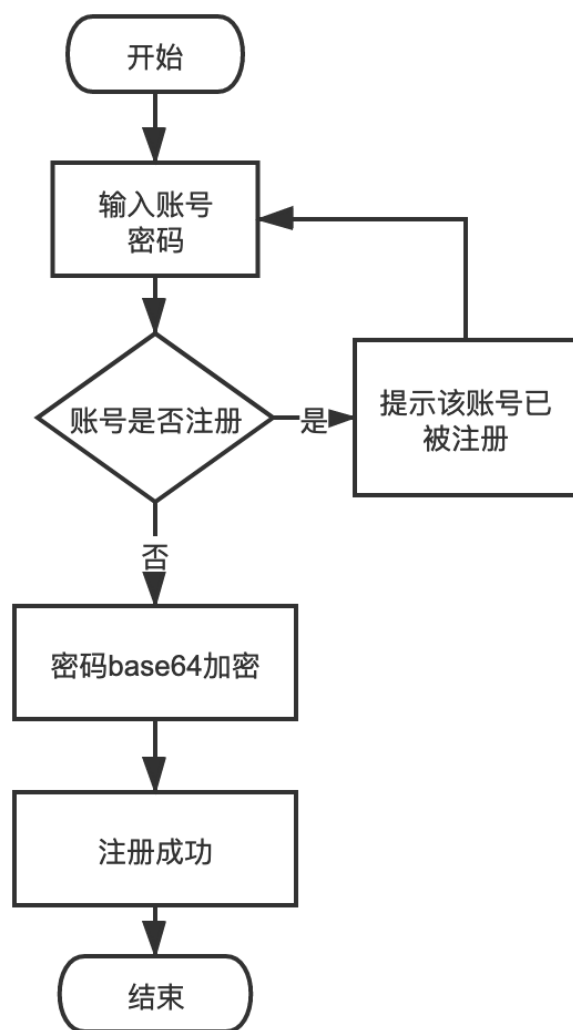


图 2.38 注册功能软件流程图

2) 资产信息统计：接入智能采集终端时，通过扫描仪器设备的资产二维码将仪器设备的基础信息导入至资产数据库。主要包括仪器设备的名称、存放位置、仪器使用年限等。

3) 设备生产测试：在进行设备生产检测的时候，程序调用接口，传入 IMEI、电信 AEP 相关配置、检测数据、状态等。其中软件设计流程如图 2.39 所示。

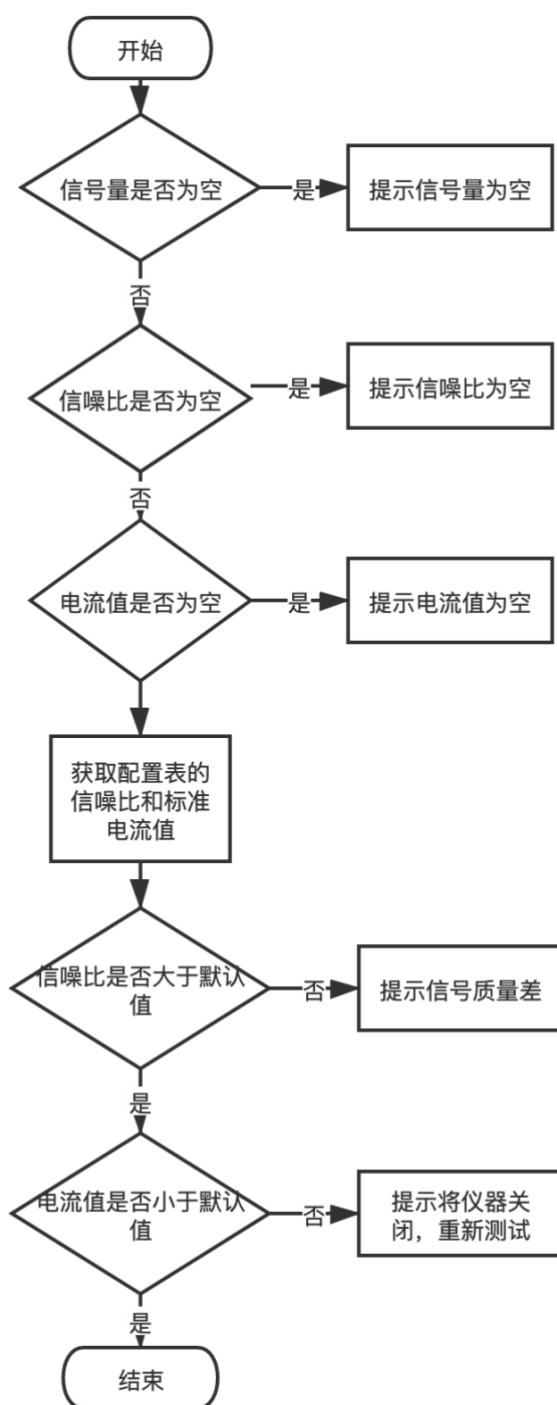


图 2.39 设备生产测试软件流程图

4) 设备入库：管理员扫描箱子编码或者 IMEI 实现设备的批量入库，如果入库失败会返回失败原因。以下两种情况不能入库：设备未检测；设备已安装。

考虑到管理员误操作和设备出库后未安装想重新入库的情况，程序设计如图 2.41 所示：如果 IMEI 已存在并且设备未安装，则进行更新操作，不存在进行新增操作。

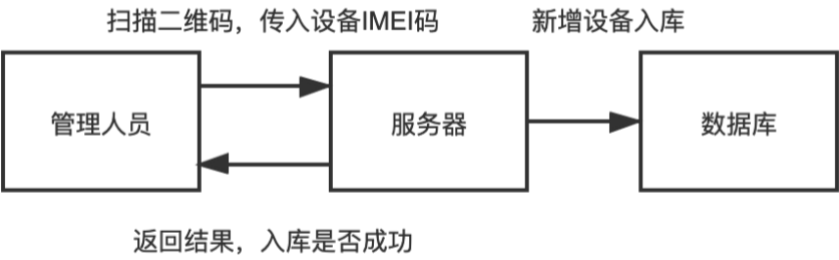


图 2.40 采集设备终端入库逻辑关系图

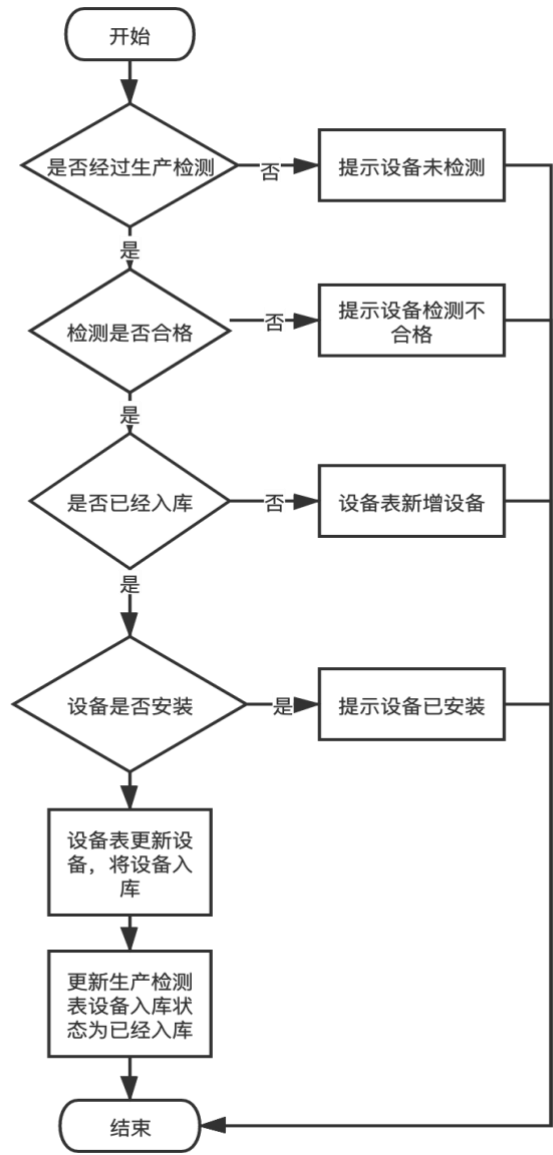


图 2.41 设备入库软件流程图

5) 设备出库：管理员扫描箱子编码或者 IMEI 实现设备的批量出库，将设备状态变为出库状态。

6) 设备安装：安装人员通过选择学校和扫码输入资产编码，获取后台返回的资产信息。当扫码输入 IMEI，调用设备测试接口，获取电流值、信号量和信噪比。当设备满足设备出库并未被安装、资产存在并未被绑定时，可以进行安装操作。调用安装接口时，除了数据库操作外，还会去电信 AEP 修改设备的名称，设备名称格式为系统针对不同学校配置的简称+资产编码。其中软件设计流程如图 2.42 所示。

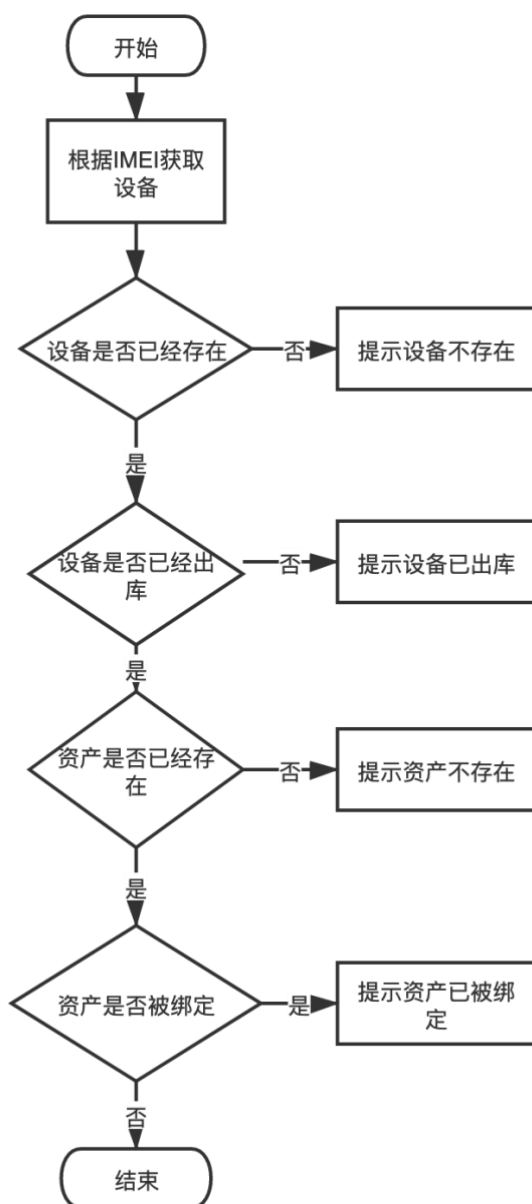


图 2.42 设备安装软件流程图

7) 设备解绑：为防止安装人员，将采集终端设备与仪器配对错误，造成仪器对应基础信息绑定错误，增加设备解绑功能模块。其中软件设计流程如图 2.43 所示。

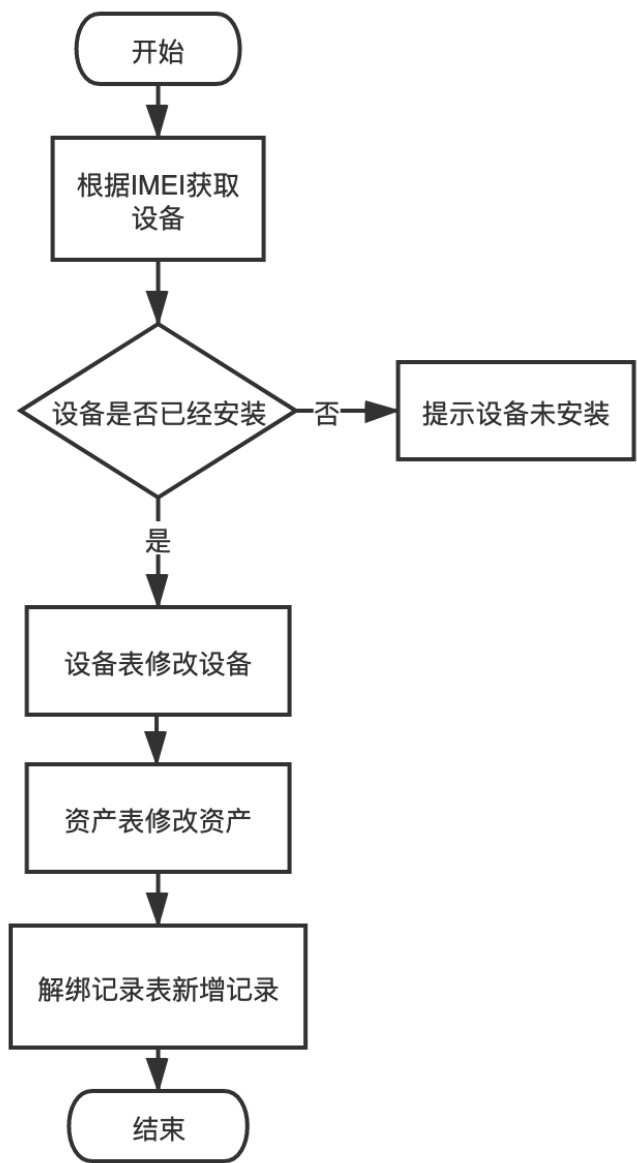


图 2.43 设备解绑软件流程图

8) 设备在线同步：当设备状态从离线变化为上线时，设备需要更新同步信息。其中软件设计流程如图 2.44 所示。

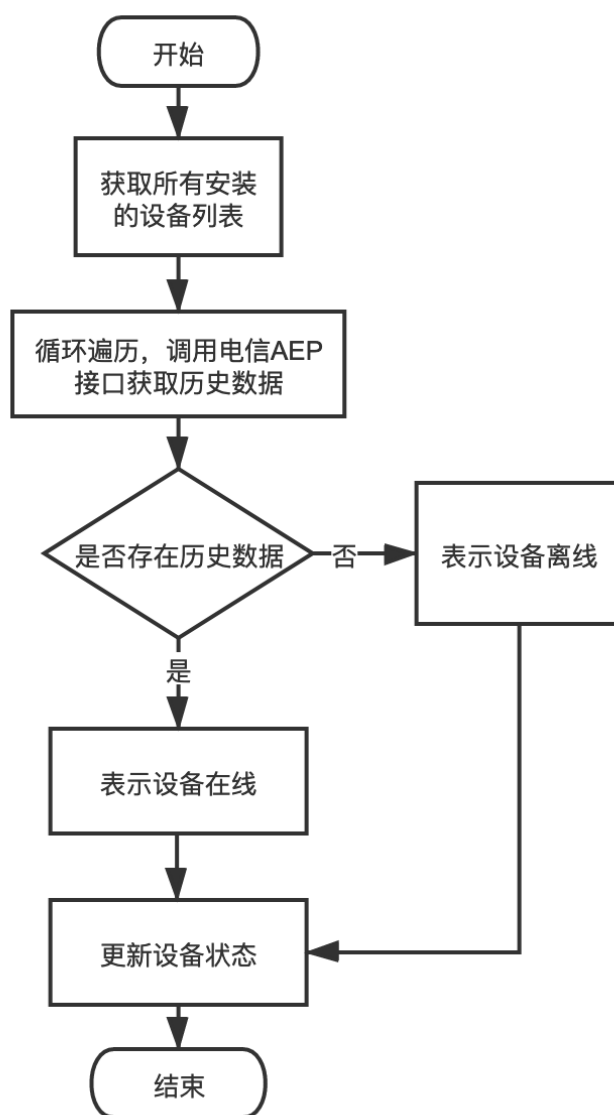


图 2.44 设备在线同步软件流程图

9) 安装备注问题: 如果仪器设备已经报废、年久失修等记录上报。

2.5.3.2 处理机制设计

1) 日志处理

日志处理的框架采用 logback, 也是扩展于 Spring 的日志框架, 比原先采用的 log4j 更快更稳定。日志的级别分为 error, warn, info, debug, trace 五个级别, 可以利用 log 对象的五种不同的日志级别方法来输出日志信息, 方便在程序各个地方输出不同的信息。通过配置文件, 可以将日志信息导出至文件中, 配置日志自动打包备份等功能。

2) 异常处理

异常主要分成数据库操作异常和 HTTP 请求异常。

针对数据库操作异常，程序中需要对 Spring 框架进行配置，使其开启事务的支持。作用是保证数据访问的原子性，当操作数据发生异常时，可保证操作实现回滚，也即回到操作数据之前的状态，保证了数据操作的原子性。

除了开启事务支持，还需要在程序中针对每一个数据库运行过程进行异常捕获，捕获后输出异常信息日志，方便程序的调试。

针对 HTTP 请求异常，主要是创建异常处理响应类，然后通过配置 Spring MVC 的 Filter（Filter 的位置处于 Controller 层之前，也就是 HTTP 请求会先到达 Filter，经过过滤后再到达 Controller），让其将异常响应信息返回给调用者，从而使调用者能够正确调用 API。

2.5.3.3 小程序结果展示

仪器设备智能采集终端安装运维小程序端的功能有个人信息、设备安装信息、仪器资产列表、设备解绑、异常问题记录等功能。其中实现效果如图 2.45，图 2.46，图 2.47，图 2.48 所示。



图 2.45 仪器设备列表页面

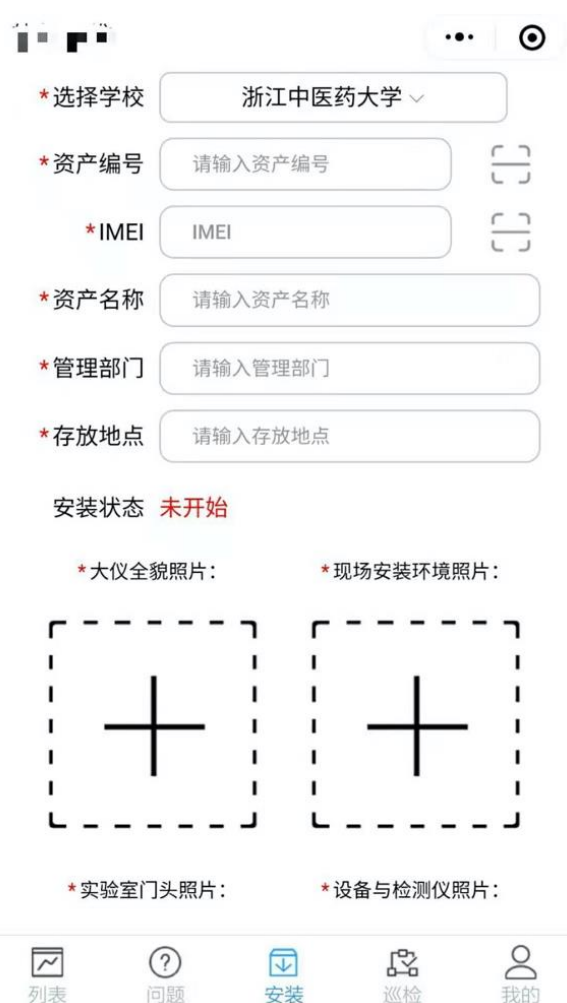


图 2.46 设备安装页面



图 2.47 设备安装机构选择页

图 2.48 安装备注信息页面

2.5.3.4 现场安装图展示

仪器设备状态智能监测系统整体设计中,安装采集终端接入大型实验仪器的现场如图 2.49 所示。



图 2.49 智能采集终端接入大型实验仪器的现场图展示

2.6 本章小结

本章为介绍了仪器设备状态智能监测系统的设计，首先进行了系统总体设计包括需求分析和总体框架，然后对系统各个部分进行分层介绍，包括智能采集终端设计、网络层的设计、云平台设计、应用层设计以及各部分结果展示。

第3章 改进 K-means 的自适应阈值算法

加快推进“互联网+政务服务”大背景下，需对传统仪器设备资产进行监控使用管理，实时在线查看设备状态，统计设备在线服务使用时间。因此，必须寻找可以有效获取设备状态的方法。

为了对仪器设备进行状态监测，以及分析其使用的时间。一般是通过获取它的工作阈值从而判定设备的开机、待机、工作的状况，进而分析在线工作的时间。但是，目前设备的种类繁多，以及不同设备的状态阈值都不尽相同，人工采集对比监测较为繁琐，效率低下，不切实际。目前获取设备状态的常用两种方式：

1) 固定阈值，通过 SVM 向量机算法进行数据标签一一对比。

以 SVM 向量机为代表的分类模型首先需要对数据的标定，标定的过程需要人工执行，在不知道设备状态的前提下对设备上报电流划分是不切实际的。利用 SVM 向量机算法对设备电流进行分类是可行的，但是并不适用于当前的场景。对应不同的设备，将不同状态的电流阈值设置位固定值，这样做首先无法应对电流工作波动大的场景；若对于每台设备都进行固定阈值是不现实的，我们无法做到对每台设备的不同状态进行一一记录。

2) 图像处理，通过模糊匹配算法进行电流图像一一对比。

对电器设备在工作和待机时，产生的工作和待机的电流信号进行特征信号提取，画出对应电流状态曲线图，利用模糊匹配算法建立图像集与识别模型，模糊匹配多项指标依次对应设备工作、待机的电流曲线，从而判定设备的运行状态。但是此算法适合应用在同类设备中，阈值跨度较小的场景，而且该方法不能区分设备运行是否为完整的工作周期。

当对仪器设备完整的工作周期电流进行分析时，观察发现，其仪器电流值的大小正向分为关机值、待机值、工作值。正好可以使用 K-means 算法进行划分不同类簇。但是，在执行过程中发现，原始 K-means 算法存在几个问题：

1) K 值的功能是确定数据中所要分配聚类中簇的数量。当采集电流数据的工作周期不完整时，K 值选取会出现问题。对于选取不同的 K 值，会影响聚类算法的最佳结果；

2) 原始 K-means 算法的初始聚类中心选取都随机默认确定的，因此并不能够确保算法是解代表全局结果，随机性会对之后结果产生较大的误差，容易陷入局部化求解。这也是很多聚类算法常有的缺陷。

3) 容易收到一些异常点的干扰，如果选择中心点的距离是依据其中几个少量的异常点算出的，就直接干扰算法结果，造成样本数据时间复杂度增加，还影响确定初始中心偏离正常值，从而使平均值发生偏离预期结果。

因此，本章设计了一种改进 K-means 的自适应阈值分析算法。其中从选取 K 值、初始聚类中心、孤立点、三个方面优化改进算法，设计了基于 GSA 自适应 K 值算法，实现了当仪器设备工作周期不完整的情况下，也可以自动识别输出最佳 K 值。设计了基于 Huffman 二叉树算法，实现除去孤立的相异点，完成初始聚类中心的选择机制，有效避免了局部最优。因而，获取的设备状态阈值，可以有效的统计设备的在线工作的时间。经过对聚类模型仿真测试及参数分析，证实设计的算法适用于该系统。

本章算法的两个目的：对大量设备使用的工作电流信息进行数据分析，得到不同大型设备仪器对应的各种工作状态阈值，从而反馈至智能终端进行数据清洗；其次是下发至可视化终端显示对应的仪器设备的使用信息以及在线工作时间。关系框图如图 3.1 所示：

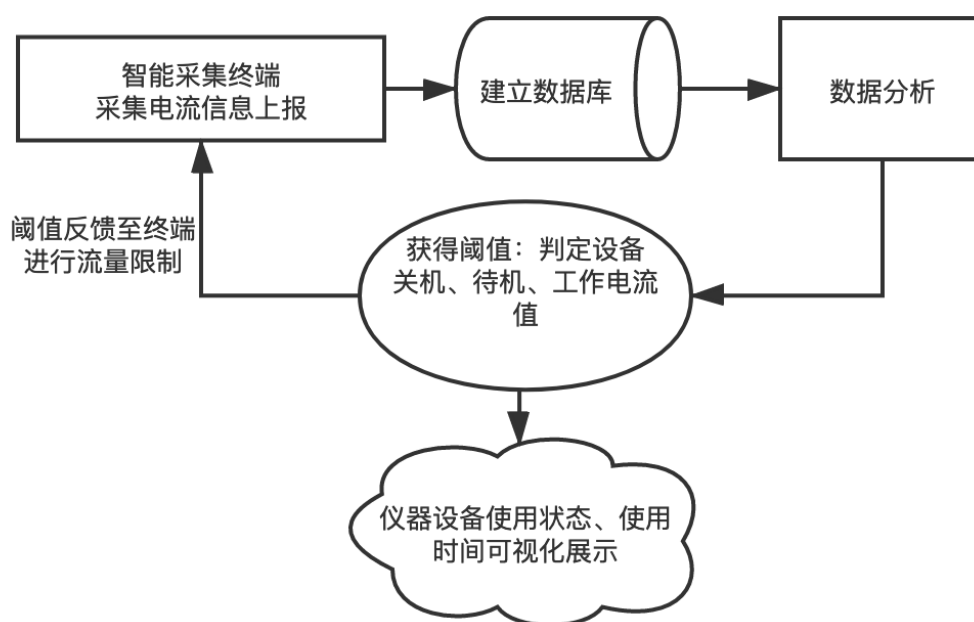


图 3.1 数据分析关系框图

3.1 K-means 聚类算法应用特点分析

K-means 算法是一种通过人为先指定聚类数再对数据集进行聚类分析^[34]。首先将样本点视看作是在 k 维空间中的数据点，并通过每个点之间距离的大小看作是判断数据间相似的指标。其中 K-means 算法的根本思想是：先将各个数据点汇聚到最近的簇中，然后对数据集进行初始的划分，然后进行不停的迭代划分，最终获得数据的聚类的最佳结果。

3.1.1 K-means 聚类算法原理以及应用场景

数据挖掘是一门针对互联网积累的数据进行挖掘、分析、可视化分析的技术。数据挖掘中聚类算法是尤为关键的技术之一，一般聚类算法形式上分为大约五类：划分类、网格类、密度类、层次类、模型类的聚类算法^[35]。其中 K-means 是通过不同的类簇间的相异性的大小

来划分的一种算法。即衡量标准是：在同一类簇数据特征基本一致，非同一类数据特征差异度较大。

K-means 应用场景大部分都是在所包含的数据有很多不同的类型和状态组成^[36]。一般都有文档分类器、依据标签、文档主题的内容不同进行分类，将文档用矢量来表示，并通过常用术语作为标签识别对文档代表向量分类。进而对文档向量进行聚类分析，然后辨认文档中的相似性^[37]；应用于用户分类，在这个大数据“杀熟时代”，对用户的购买力、兴趣、等数据通过聚类分析，对客户推送一下用户喜欢购买的东西，增加销售收入；应用于 IT 仪器设备自动化状态分类，由于设备仪器对应的不同状态对应不同模式，就如仪器设备在开机、关机、待机、运行等状态下对应的电流和功率都不相同，如果我们拥有了这些数据之后，我们通过聚类分析就可以判断出此时此刻仪器设备的状态，有助于对设备状态信息的掌握，提供仪器设备的使用情况以及也对未来仪器设备维护和故障进行提前预防。

3.1.2 K-means 聚类算法流程

K-means 算法是数据分析中常被人应用于解决局部收敛的聚类模型^[38]。比如，**K-means** 算法样本数据集合数量为 n ，每一个样本点的代表属性为 m 维，再将类别数设置为 K 。**K-means** 具体执行步骤是按照欧氏距离将数据划分到 K 个簇中，其中 K 需要满足小于 $K < n$ ，在相同的簇中包含的样本数据对象都是基本相似的，而簇之间的样本数据对象的特征差异度尽可能的保持最大。原始 **K-means** 算法执行流程步骤是：

一是为 **K-means** 算法随机选取 K 个数据作为初始聚类中心；二是计算样本中每一个数据到各初始聚类中心的距离，排序之后将样本数据划分到 K 个簇中；三是循环计算 K 个簇的中心；四是重新计算样本数据到 K 个中心的距离；五是更新划分样本数据，一直到样本数据的划分不再发生变化或者满足预先设置的条件。

3.1.3 K-means 算法存在的问题以及优化

K-means 算法是属于区分不同类型数据中效率最高效的一种聚类算法，典型的划分聚类算法^[39]，具有较强的兼容性，并且能够应用于海量的数据集处理。但是原方法也有部分缺陷，本节针对 **K-means** 算法提出了优化策略。

3.1.3.1 自适应聚类 K 值的策略

K-means 算法里， K 值的功能是确定数据中所要分配聚类中簇的数量^[40]。因此，**K-means** 算法的起始值选取尤为重要，对于选取不同的 K 值，会影响聚类算法的最佳结果。下面通过一组数据说明，其中下图 3.2 所示，图像表明，该数据由 4 类数据组成。

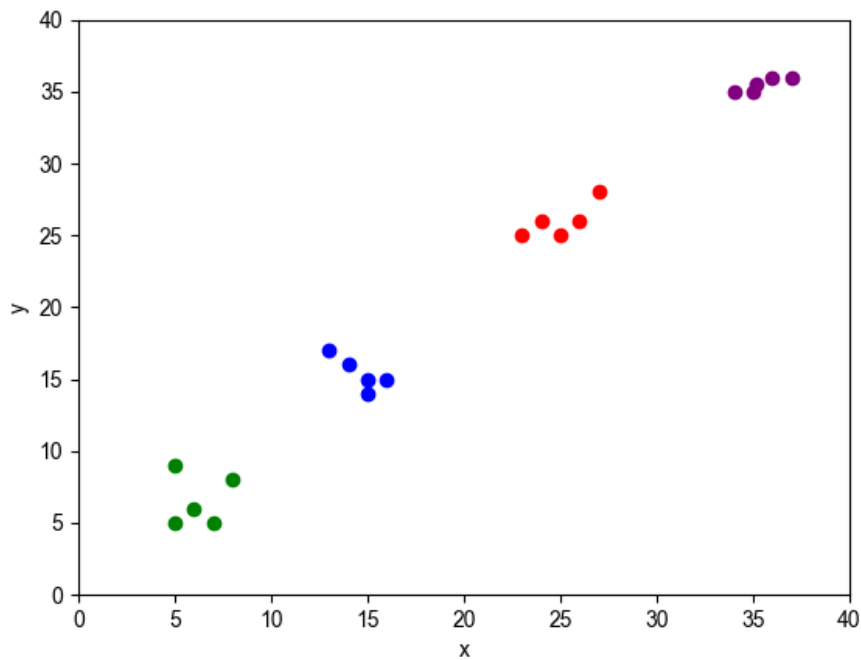


图 3.2 初始样本数据分布图

我们先开始对此数据进行分析，按 $K=4$ 和 $K=3$ ，分别对样本数据进行初始聚类中心，选择欧式距离计算数据离散度，然后通过聚类迭代，其中中心点的变化过程如下图 3.3 所示：

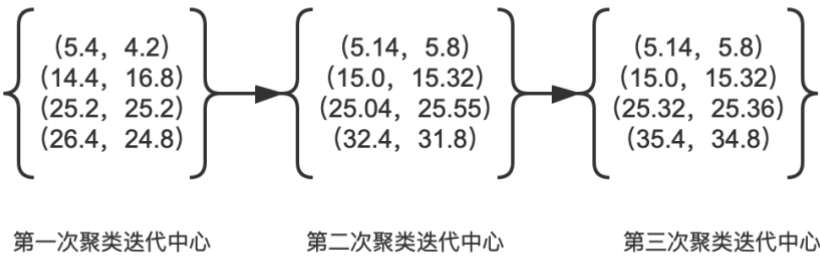


图 3.3 $K=4$ 聚类中心迭代过程图

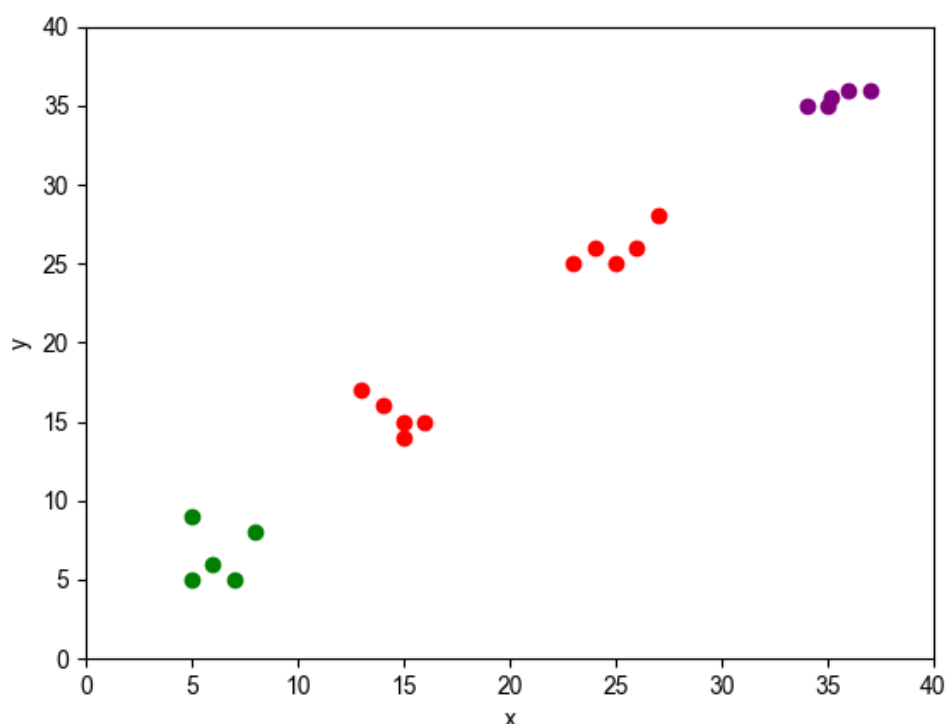


图 3.4 K=3 聚类数据分布图

其中数据结果表明，聚类的最佳结果符合我们的定为 $K=4$ 的类中心预期，和最初样本数据比基本一致。但是我们也发现如果最初我们将 $K=3$ 时也对应能获得一个聚类结果分别是每一类中心点 $((5.12, 5.7), (20.23, 20.38), (35.73, 35.0))$ ，但是并不能代表数据的分类结果，如图 3.4 所示。因此，当我们不清楚自己的样本数据是某几种分布类型时，盲目的选取 K 值将会影响数据准确性，所以需提出一种可以自适应选取样本数据 K 值的策略方法。

3.1.3.2 初聚类中心的选择策略

原始 K -means 算法的初始聚类中心选取都随机默认确定的，因此并不能够确保算法是解代表全局结果，随机性会对之后结果产生较大的误差，因此原始 K -means 算法的聚类结果与初始中心选取结果息息相关。这也是很多聚类算法常有的缺陷；当 K -means 算法按照 K 值在样本数据中随机了 K 个中心坐标值，之后按照此类结果不停循环选取直到其中心坐标不在变化，得到结果。此时，如果是选取中心只能是局部代表中心，就会陷入局部最优解的循环当中，可能难以获取到最佳的聚类结果，算法时间复杂度增加，甚至可能造成聚类算法产生严重的错误。

其次， K -means 聚类算法执行过程还容易受到异常点、噪声点、孤立点，所谓利群点的影响^[42]，不适合处理样本中数据差异过大的簇。并且 K -means 聚类算法计算中心时，容易收到一些异常点的干扰，如果选择中心点的距离是依据其中几个少量的异常点算出的，就直接干扰算法结果，造成样本数据时间复杂度增加，还影响确定初始中心偏离正常值，从而使平均

值发生偏离预期结果，容易导致聚类算法不能够具有较强的鲁棒性或准确性。

针对上述问题，提出了一种初始聚类中心的选择策略。首先，需获得去除了孤立点、离群点的样本数据集合，再重新计算样本数据中，每个属性纬度的取值范围，依据上一节小获得的 K 值，为该样本的簇类数目，再将每个属性纬度的数据范围平均划分为 K 个数目的数据，依据本次的数据，再选取下一次的初始聚类中心^[43]。该优化方式有效的避免了传统随机选择的聚类中心思想会陷入局部最优的问题，提升了算法的执行效率和聚类效果，增强了数据分析的准确性。

3.2 改进 K-means 自适应阈值算法

本小节针对现有技术的不足，提出一种能够自适应不同设备状态阈值的方法。针对在不同设备，在各种工作状态下，获得其对应的工作阈值，从而了解设备的运行状态，提升设备使用效率。在自适应不同设备状态阈值的方法中，数据分析主要基于改进的 Kmeans 聚类算法实现：

(1)通过电流计传感器，采集设备在使用时的电流值，通过内嵌的 NB-IOT 通信模块上传到后台服务器。收集到仪器状态电流的历史数据进行分析。

(2)执行预处理，包括归一化和 PCA 方法处理。

(3)对数据预处理后，通过基于 GSA 自适应 K 值算法来确定 K 值，进行孤立异常去除之后，再初始化聚类中心，至中心向量值不变，最后输出对应的聚类结果。

(4)将送入 K-means 聚类算法中获得的对应阈值再进行反归一化输出仪器设备对应不同状态的开机、关机、待机、运行的电流阈值，获得的设备状态的聚类结果和设备一一对应。

(5)根据设备状态样本评估设备的当前状态，反馈至 APP 端和 Web 页面显示。反馈至智能采集终端可以限制流量上报。

3.2.1 改进算法实现的流程框架

优化改进 K-means 流程如图 3.5 所示，从数据库读取数据进行优化分析处理，最终得到聚类结果。

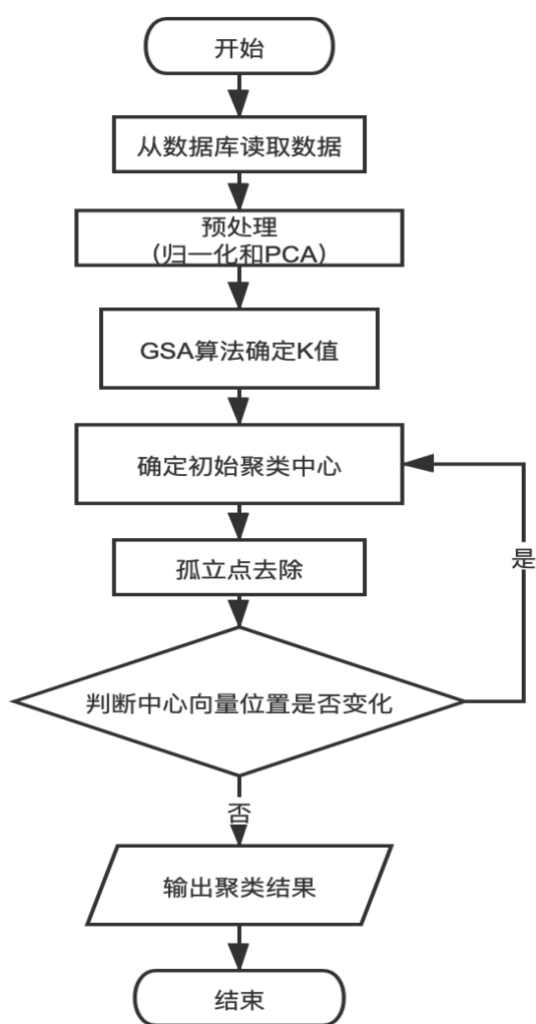


图 3.5 改进 K-means 算法流程图

3.2.2 数据源模型构建

经过设备采集模块将仪器设备的工作周期对应的电流上传到云平台后，再进一步对设备使用的电流历史信息进行分析。我们目前采集的仪器设备有两种型号，一种是单相电流计，另一种是三相电流计。其中电流计上传到云平台我们对数据协议进行解析封装成适合我们分析的数据格式，IMEI 是代表不同设备的唯一码。表3.1 中是对仪器设备分析数据的格式：

表3.1 仪器设备模型对应的数据格式

时间区间	IMEI（仪器设备资产编号）	电流值
一个月	某个具体的设备	单相：[单相电流] 三相：[A 相，B 相，C 相]

我们建立数学模型，将仪器设备电流数据用矩阵进行表示。本文使用 $n * m$ 矩阵对电流数据进行表示， n 表示有 n 个仪器设备使电流的模型聚类对象。每个仪器设备的电流值向量含 d

维数据， d 为数据采集频率，本文使用的仪器设备电流采集是 30 分钟上报一次，一天总共是 30×48 个电流数据点。其中数据电流为 float 型，用矩阵表示如式(3.1)所示

$$\begin{bmatrix} x_{11} & \cdots & x_{1m} & \cdots & x_{1d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & \cdots & x_{im} & \cdots & x_{id} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{nm} & \cdots & x_{nd} \end{bmatrix} \quad (3.1)$$

3.2.3 数据预处理

数据预处理在数据分析中，起着至关重要的作用。我们一般直接获取的数据通常是不完整的、格式不一致、极易收到其他错误干扰，所以我们一般都会在数据分析之前，进行预处理，以确保数据进行算法分析的统一性、准确性。

3.2.3.1 数据归一化处理

数据预处理方法中，首先执行的是归一化，因为数据归一化后，可以很大程度的兼容其它算法。其中对于一般聚类算法通过规范化数值，可以减小数据域中相对较大和较小数值之间所占距离比重的差距。为了保证数据变量间的可比性，提升算法的效率，常常会选取的样本变量做归一化处理，将样本数据映射到 0 到 1 之间，其中归一化公式如下：

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad i = 1, 2, \dots, n \quad (3.2)$$

式中： x_i 为选取的样本数据变量， y_i 为归一化后的变量， n 为数据长度， $x_{\min}$ 为选取的样本变量的最小值， $x_{\max}$ 为选取的样本变量的最大值。

3.2.3.2 数据 PCA 降维处理

PCA (principal components analysis) 是一种无监督数据的处理算法，经常应用于数据压缩的方法之一 (降维技术)。PCA 将数据主要特征提取之后进行降维。

其中三相电流计样本需要数据进行降维。对样本集 $D = \{x_1, x_2, x_3, \dots, x_m\}$ ，具有 m 个特征，每个特征有 n 个样本。降低空间的维度。 d' 。对其进行 PCA 的操作步骤：

(1) 去中心化处理 $x_i = x_i - \frac{1}{m} \sum_{i=1}^m x_i$

(2) 计算协方差矩阵 XX^T

(3) 计算 XX^T 的特征值与特征向量。通过 XX^T 的矩阵分解求特征值，再将每个特征值依次代入，获取每个特征值对应的特征向量。

(4) 对特征值从大到小排序

(5) 保留最大的 d' 个特征值所对应的特征向量 $I_a, I_b, I_c, I_{d'}$;

(6) 输出投影矩阵 $W = (I_a, I_b, I_c, I_{d'})$

3.2.4 基于 GSA 自适应 K 值算法

在 K-mean 算法确定 K 值的方法中，一般也可以使用手肘法^[44]。手肘法的缺点是不够自

动化，需要将数据导入通过观察可视化选择 K 值。而原始 GSA 间隙统计算法(Gap Static Algorithm)算法不再需要“手肘法”式的可视化判断，适用于批量化大数据应用，但是不能有效进行区分 $K>1$ 的数据集。因此，本小结设计了基于 GSA 自适应 K 值的算法，是由 GSA 方法与肘形判据的方法相结合，用于确定 K 值，即聚类算法中数据分布的聚类数，可应用于所有聚类方法。此算法不仅提升了计算速率，而且准确性高。通过应用基于 GSA 的自适应 K 值算法自动获得最优聚类值。

3.2.4.1 原始 GSA 算法流程

对样本数据集 $D = \{X_i\}$ ，利用聚类算法获取 K 个簇类，为 I_1, I_2, I_k ，对于单个类 I_a 。执行流程是：

1)先执行遍历样本 X_i ，通过计算方差和，度量样本的密集度，计算簇类和样本中 X_i 的距离平方和 D_a 。

$$D_a = \sum_{x_i \in I_a} (x_i - Z_a)^2 \quad (3.3)$$

其中公式(3.3)中： Z_a 是类 I_a 的均值中心向量。

2) 计算每个 K 值对应的聚类离散度。

$$W(K) = \sum_{a=1}^K D_a \quad (3.4)$$

3) 计算每个 K 值对应的聚类之间的间隙值。

$$g_{ap} = E \ln[W_r(K)] - \ln[W(K)] \quad (3.5)$$

其中公式(3.5)中 r 是对比的数据样本， E 是对比数据样本的数学期望。

4) 在生成 F 组对比的数据集，需要计算每组对应的离散度均值 $E \ln[W_r(K)]$ 。

$$E \ln[W_r(K)] = \frac{1}{F} \sum_{j=1}^F \ln W_{r,j}(K) \quad (3.6)$$

$$sd(r) = \left(\frac{1}{F}\right) \sum_{j=1}^F [\ln(W_{r,j}) - E(\ln(W_r))] \quad (3.7)$$

式中 $sd(r)$ 是样本数据，在生成中模拟的随机误差。最后获得的样本分布误差为：

$$\sqrt{1 + 1/F} (sd(r)) \quad (3.8)$$

再根据公式(3.5)计算 $g_{ap}(K)$ ，获取最优 K 值是：

$$g_{ap}(K) \geq g_{ap}(K + 1) - S_{k+1} \quad (3.9)$$

最后得出最大 $g_{ap}(K)$ 值。对应的 K 值即为目标值。

3.2.4.2 手肘法判定 K 值

手肘方法获得聚类个数的原理：判断聚类个数增加的同时对应聚类离散度会逐渐减小。一般是由于一个聚类对象分裂为两个，聚类离散度的变化较小。因此，可以通过观察离散度- k 变化图的拐点，横轴 k 对应的就是聚类数。如图 3.6 所示。

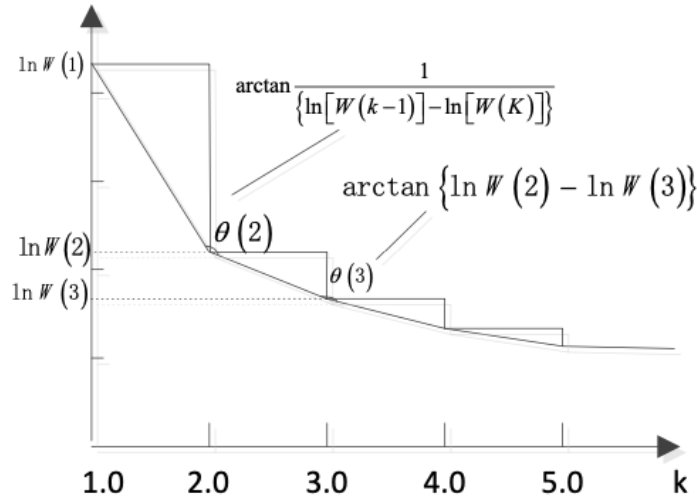


图 3.6 肘形 K 值判断图

所述 GSA 算法分析对比, 计算最大间隙值 $g_{ap}(K)$ 的位置, 其实就是在图中 $\ln[W(K)] - K$ 里 k 值逐渐增大, 曲线逐步变缓的点。

依据图 3.6 中观察在不同 K 值对应的肘形角 $\theta(K)$, 前后 K 值间对应的 $\ln[W(K)]$ 组成的夹角, 计算夹角公式(3.10):

$$\theta(K) = \frac{\pi}{2} + \theta_u + \theta_r \quad (3.10)$$

其中式中:

$$\theta_u = \arctan \left\{ \frac{1}{\ln[W(K-1)] - \ln[W(K)]} \right\} \quad (3.11)$$

$$\theta_r = \arctan \{ \ln[W(K)] - \ln[W(K+1)] \} \quad (3.12)$$

根据公式可知想要求夹角 $\theta(K)$ 值, 必须先得到 $\ln[W(K-1)]$ 和 $\ln[W_j(K)]$ 的值, 因此可知若求 $\theta(1)$ 需要计算 $\ln[W(0)]$ 值, 但是, $\ln[W(0)]$ 是无法获取的数值。

综上所述, 当 $K=1$ 也就是样本数据都是同一类数据时, GSA 算法不能有效的解决此时的问题。

3.2.4.3 改进 GSA 自适应 K 值算法实现步骤

基于 GSA 自适应 K 值算法的执行步骤: 算法参数输入: 数据集、聚类最大个数; 有效输出: 聚类 K 值。

1) 依次计算初始聚类中心 $K=1, 2, \dots$ 值所对应的聚类离散度 $\ln[W(K)]$;

2) 通过公式(3.4)分别计算均匀分布生成 F 组中, 对应的每个组的聚类离散度 $\ln[W_j(K)]$ 。

3) 通过公式(3.5)计算聚类间隙值 $g_{ap}(K)$, 其中 $j=1, 2, \dots$; $K=1, 2, \dots$ 。

4) 通过比较 $g_{ap}(1)$ 和 $g_{ap}(2)$ 对比最优个数, 若 $K=1$ 算法流程结束; 当 $K \geq 2$, 使用公式(3.12)得到肘形夹角 $\theta(K)$ 。

$$\theta(K) = \frac{\pi}{2} + \arctan\left\{\frac{1}{\ln[W(K-1)] - \ln[W(K)]}\right\} + \arctan\{\ln[W(K)] - \ln[W(K+1)]\} \quad (3.13)$$

5)最后，输出 K，满足 $\min(\theta(K))$ ；

3.2.4.4 数据仿真测试

将前一小节的聚类数据进行基于 GSA 自适应 K 值计算，依次对比从 $K=1,2\dots 8$ 的间隙值 $g_{ap}(1) < g_{ap}(2) - S_2$ ，如表3.2 所示：

表3.2 聚类数据集肘形结果

聚类 K 值	离散度 $W(K)$	$\ln[W(K)]$	角度大小
1	9872.87	9.10	/
2	2389.44	7.80	183.667
3	558.22	6.30	196.815
4	37.78	3.63	151.845
5	35.33	3.57	179.734
6	34.17	3.53	180.445
7	33.22	3.50	181.589
8	31.76	3.45	/

仿真肘形K值图

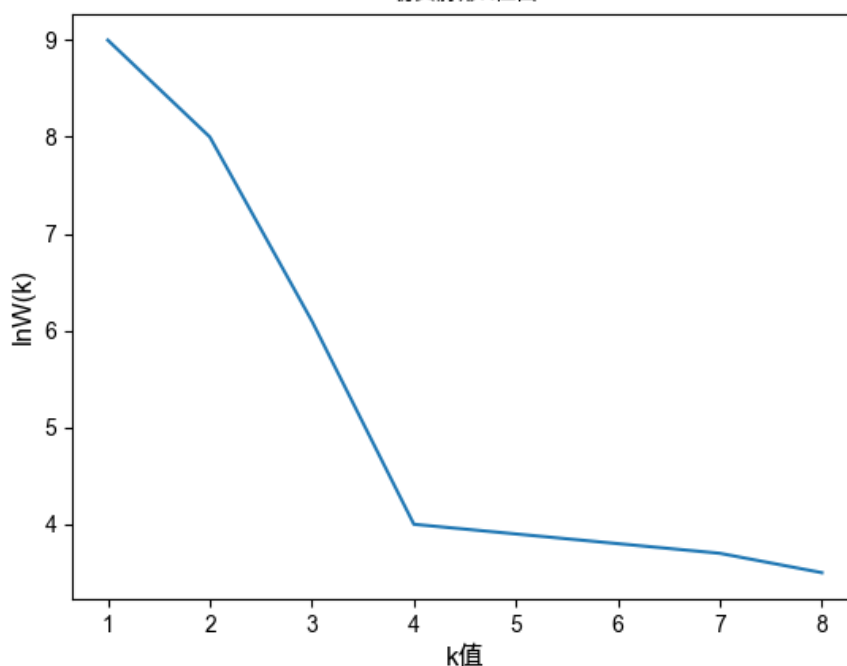


图 3.7 仿真肘形 K 值判断图

由表和曲线图 3.7 都可以准确获得聚类 K 值。 $K=4$ 。经过多次重复训练样本数据，结果符合样本数据的特征分布，该算法可以准确验证聚类的 K 值。

3.2.5 初始聚类中心以及孤立异常点去除

本小结针对 K-means 算法常被聚类中心、孤立异常点干扰的缺点，设计了一种基于 Huffman 二叉树算法^[46]，可以提高聚类样本的聚类中心的准确性，极大地提高了聚类质量和运行效率。因为影响聚类中心的关键属性就是数据密度^[47]，聚类质量的优劣很大程度上决定数据的准确度。数据密度一般是指在数据中代表球形区域内，指定半径内分布的数据量的多少，而其中数据的相异性矩阵代表数据间的距离；依据数据密度和相异性矩阵了构建 Huffman 二叉树^[48]，再反向移除 K 个节点，获得 K 个初始中心点。

判定样本数据的高质量中心的依据之一，样本内点的密度越高，样本中数据对象和其周围相邻的数据对象越密集，说明是样本中的代表中心点，其一般可以作为样本中 K 其中的一个高质量的聚类中心点。

算法执行步骤流程是：

1)先计算样本的数据密度：设定样本集合 $U = \{p_i \in R, i = 1, 2, \dots, n\}$ ，以样本数据 p_i 为球心，指定距离 R 内数据点称为样本 p_i 的数据密度。记作 $\text{den}(p_i)$ ：

$$\text{Dis}(i, j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (3.14)$$

$$R = \alpha * \frac{\max(\text{dis}(p_i, p_j))}{k} \quad (3.15)$$

$$\text{Den}(p_i) = \sum_{j=1}^n u(R - \text{dis}(p_i, p_j)), u(z) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (3.16)$$

其中公式里， k 是聚类数值， α 是动态参数，默认是 1， $\text{dis}(p_i, p_j)$ 是点 p_i 与点 p_j 的欧式距离。

2)构建高密度集 U_1 ，一般在处理样本数据集密度时，通常的判断依据时数据集对象是否是高密度的对象。代表参数是平均密度 D_{av} 的计算为式(3.17)，高密度集 U_1 为式(3.18)：

$$D_{av} = \beta * \frac{1}{n} \sum_{i=1}^n \text{Den}(P_i) \quad (3.17)$$

$$U_1 = \{p | \text{Den}(p_i) \geq D_{av}, p_i \in U\} \quad (3.18)$$

式(3.17)中， β 为动态调整参数(默认为 1)。一般当 U_1 样本数据集数目小于 k ，则逐渐减小 β ，来调整高密度集的大小，以降低下一步构造 Huffman 树的时间复杂度。

3)构建 Huffman 二叉树，计算出高密度矩阵数据集 U_1 的相异性矩阵：

$$ad_{ij}^k = \frac{|x_{ik} - x_k| - |x_{jk} - x_k|}{x_{\max k} - x_{\min k}} \quad (3.19)$$

其中公式(3.19)， x_{ik} ， x_{jk} 分别是是 x_i ， x_j 在维度 k 中的数值， x_k 是所有样本数据集在第 k 维里的平均值， $x_{\max k}$ ， $x_{\min k}$ 分别是所有样本数据集第 k 维里的最大值、最小值。

为了保持样本数据的准确，对相异性的定义进行优化。对于样本数据中任意两个数据对象 $I_i = \{I_{i1}, I_{i2}, I_{in}\}$ ， $I_j = \{I_{j1}, I_{j2}, I_{jn}\}$ 的相异性重新优化为：

$$D_{ij} = \frac{\sum_{k=1}^n ad_{ij}^k}{n} \quad (3.20)$$

然后，样本数据中的所有的对象，由此两两之间构成相异性矩阵，记作 P_u ：

$$P_u = \begin{vmatrix} D_{12} & D_{13} & \dots & D_{1(n-1)} & D_{1n} \\ D_{23} & D_{24} & \dots & D_{2n} & \dots \\ & \dots & & \dots & \\ D_{(n-2)(n-1)} & D_{(n-2)n} & & & \\ D_{(n-1)n} & & & & \end{vmatrix} \quad (3.21)$$

构造 Huffman 二叉树：拥有权路径长度的最优二叉树^[49]。下面是构造 Huffman 树的过程如图 3.8 所示。

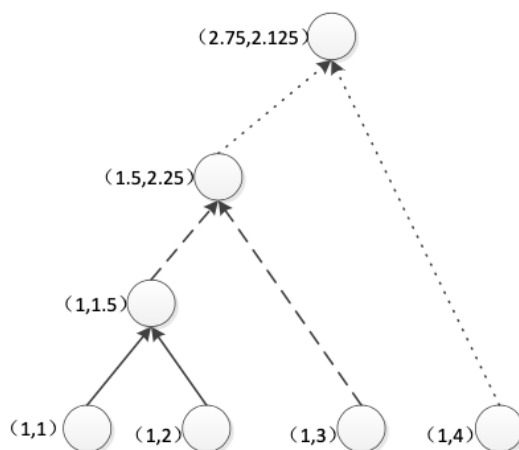


图 3.8 构造 Huffman 树过程图

由图 3.8 可以看出，Huffman 树中的 K 个点就对应 K 个子树的均值。在算法步骤第一步和第二步的两个数据集再结合 Huffman 树构造最优的二叉树，最后反向地移除 K 个点，可以保证聚类中心结果的正确性，会更加符合聚类数据的分布，同时也不会使数据解陷入局部最优，可以有效应用求解权值。

3.2.5.1 数据仿真测试

本小结应用了机器学习里的 UCI 数据集中非常著名聚类数据集 iris 、 wine。其中 Iris 数据集由 4 维向量包含了 750 个数据，设定为 3 类；而 wine 数据集包含了 580 个数据，每个数据由 8 维度向量组成，也设定为 3 类。而在 wine 数据集中，采用的是数据点顺序选取中心点。算法的结果如表3.3 和表3.4 所示。

表3.3 Iris 模拟样本数据集聚类结果对比

	初始聚类中心	准确率	运行 次数	平均的准确率
--	--------	-----	----------	--------

续表 3.3

原始随机选取聚类中心	(5.4 2.9 1.4 0.2) (6.6 2.2 5.4 1.2) (7.1 3.5 6.4 2.2)	68.67%	18	80.67%
	(4.8 3.5 4.5 0.2) (2.1 2.5 1.4 0.1) (5.6 3.5 4.0 0.2)	92.00%	19	
	(5.1 3.5 1.4 1.2) (5.5 4.5 1.3 2.3) (6.3 3.5 1.4 5.2)	92.67%	18	
	(5.8 6.5 3.4 0.2) (4.1 3.5 4.6 1.1) (8.1 5.5 1.4 4.2)	69.33%	20	
本文优化算法选取聚类中心	(5.9 3.5 4.4 1.2) (5.1 3.4 2.4 0.6) (6.8 3.5 5.5 2.5)	95.67%	17	95.67%

表3.4 Wine 模拟样本数据集聚类结果对比

	初始聚类中心	准确率	运行次数	平均准确率
数据顺序选取聚类中心	24,6,156	71.46%	8	71.17%
	95,35,120	70.26%	7	
	39,131,71	57.87%	9	
	165,125,128	85.12%	9	
本文优化算法选取聚类中心	(12.8,3.37,2.3,19.5,88.0,1.48,6.6,0.4,0.97,10.2,0.72,1.75,685) (11.8,2.14,2.78,26.0,92.5,2.13,2.24,0.54,1.76,3.0,0.97,2.44,462) (12.2,3.43,1.94,16.5,81.2,1.63,1.25,0.43,0.83,3.4,0.75,2.16,396)	85.71%	7	85.71%

从表中的结果可以得知，本文设计的初始聚类中心优化选取算法可以有效地增加聚类效果，确保了结果的准确性，避免了数据解陷入局部最优的情况，通过优化算法获取的初始聚类中心更加趋近于最优聚类结果，同时降低了运行次数，提升了算法运行效率，验证了该算法有效准确性。

3.3 流量上报限制机制应用

针对本文智能采集终端与设备云平台通信过程中的数据延迟高、流量效率低的问题，提出了一种高智能采集系统，通过降低智能终端发出的数据流量达到目的需求。设计了智慧采集终端实现流量限制方案。

目前，在多个采集终端对应的多个大型实验仪器，每天上报海量的数据中，存在大量的无用数据，不仅浪费网络上传流量而且对云平台存储数据造成压力。现在我们可以通过改进 K-means 算法获得的仪器设备对应的工作阈值，进行上报限制。

智能采集终端中数据上报流程如图 3.9 所示：

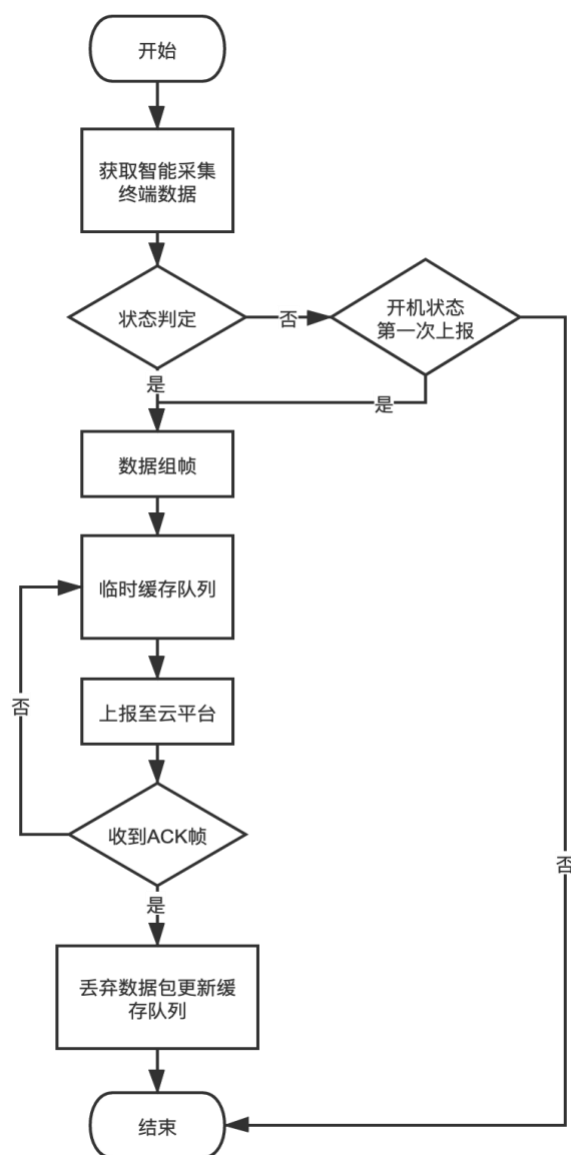


图 3.9 智能采集终端上报数据流程

边缘流量限制机制核心思想是对数据采集端上报至云平台的数据进行限制。与最近的一次数据进行对比，如果仪器设备电流变化了再上报，没有变化就不进行上报，半个小时进行周期上报一次，包括设备刚开机启动时。由图 3.9 中可以看出，流量限制上报机制主要从三个方面做出改进：

1)首次设备开机上报，当智能传感器判定此时是仪器刚刚开机，需要上报一次当前的状态电流。

2)设置数据缓存队列防止数据丢失，当在判定仪器设备状态变化后先将此时的状态电流存入缓存队列中，等待数据已经成功上传至云平台后再丢弃本地数据。

3)ACK（应答回复），当云平台端没有响应回复 ACK 应答帧，那本地缓存的数据不更新，重新进行上报一次。如果成功则移除本地存储数据。

3.3.1 流量上报限制应用测试

测试智能采集终端在流量限制前后数据帧进行对比，时间周期取 4~20h 内间隔 4h 对比上报数据帧。

表3.5 数据量对比

测试时间 (h)	未限制上传数据帧 (byte)	流量限制后数据帧 (byte)
4	384	384
8	768	384
12	1536	768
16	3072	1536
20	6144	3072
总数据	11904	6144

从表3.5 可以看出，通过流量限制上报机制，数据帧在短时间内上报数据急剧下降，流量上报降低了 60%，可以有限的避免上报数据的压力，也使云端服务存储减轻了负担。同时提升了系统的稳定性、准确性、高效性。

3.4 改进 K-means 算法数据测试应用

3.4.1 测试方案

本章阐述了改进的 K-means 算法获取仪器设备状态的电流阈值，基于 GSA 自适应 K 值的状态机制，以及基于 Huffman 二叉树算法来确定数据的初始化聚类中心，都旨在降低数据分析的容错率、时间复杂度、减少孤立点的干扰，提升算法的效率和数据准确度。

本小结设计了如下测试方案：在系统中选择实验室里六台仪器设备，并且这六台仪器设备采集的电流值数据时间周期都是一个月，其中按智能采集终端半个小时上报 30 数据点，电流数据点是 1 万到 10 万。对其进行电流阈值数据测试，并对数据准确度和算法性能进行对比分析。



图 3.10 仪器设备现场测试环境一



图 3.11 仪器设备现场测试环境二

表3.6 仪器设备数据测试表

IMEI	测试仪器设备名称	数据量（电流数据点）	采集电流值的曲线图
869858032091655	BH 测试仪	11293	
869558033720005	元素分析仪	18501	
869858033739245	双通道虚拟现实系统	55200	
869858032095482	液相色谱仪	68692	
869858032096688	荧光光谱仪	79829	
869858032100399	超高效液相色谱仪	104492	

3.4.2 测试结果

表3.7 仪器设备状态阈值数据测试结果

IMEI	仪器设备名称	原始算法 聚类中心	改进算法 聚类中心	原始仪器状态阈值 (电流值 A)	改进仪器状态阈值 (电流值 A)
86985803 2091655	BH 测试仪	(0.48、 0.43、0.20)	(0.34、 0.33、0.10)	工作阈值：0.45 待机阈值：0.26 关机阈值：0.20	工作阈值：0.37 待机阈值：0.24 关机阈值：0.10
86955803 3720005	元素分析仪	(4.57、 1.10、0.10)	(4.47、 1.03、0.10)	工作阈值：2.83 待机阈值：0.60 关机阈值：0.10	工作阈值：2.75 待机阈值：0.56 关机阈值：0.10
86985803 3739245	双通道虚拟 现实系统	(0.40、 0.36、0.15)	(0.38、 0.37、0.10)	工作阈值：0.40 待机阈值：0.25 关机阈值：0.15	工作阈值：0.37 待机阈值：0.24 关机阈值：0.10
86985803 2095482	液相色谱仪	(0.20、 0.15、0.07)	(0.17、 0.15、0.07)	工作阈值：0.17 待机阈值：0.11 关机阈值：0.07	工作阈值：0.16 待机阈值：0.11 关机阈值：0.07
86985803 2096688	荧光光谱仪	(1.32、 0.66、0.10)	(1.42、 0.76、0.10)	工作阈值：1.23 待机阈值：0.33 关机阈值：0.10	工作阈值：1.10 待机阈值：0.43 关机阈值：0.10
86985803 2100399	超高效液相 色谱仪	(0.66、 0.35、0.10)	(0.66、 0.25、0.10)	工作阈值：0.55 待机阈值：0.17 关机阈值：0.10	工作阈值：0.45 待机阈值：0.17 关机阈值：0.10

表3.8 算法测试性能对比

		原始 K-means 算法结果			改进 K-means 算法结果		
仪器	数据量	准确率	迭代次数	耗时 (s)	准确率	迭代次数	耗时 (s)
1	11293	71.5%	199	1.48	87.5%	109	0.58
5	18501	75.1%	210	4.55	88.3%	115	1.86
2	55200	76.2%	239	9.38	89.2%	126	5.59
6	68692	78.2%	240	11.59	88.3%	136	8.82
3	79829	78.6%	255	18.88	88.8%	148	12.15
4	104492	80.4%	528	88.44	92.3%	369	43.52

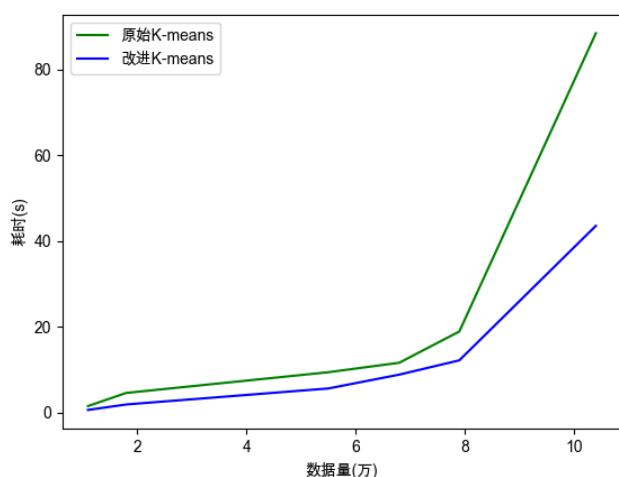


图 3.12 算法耗时对比图

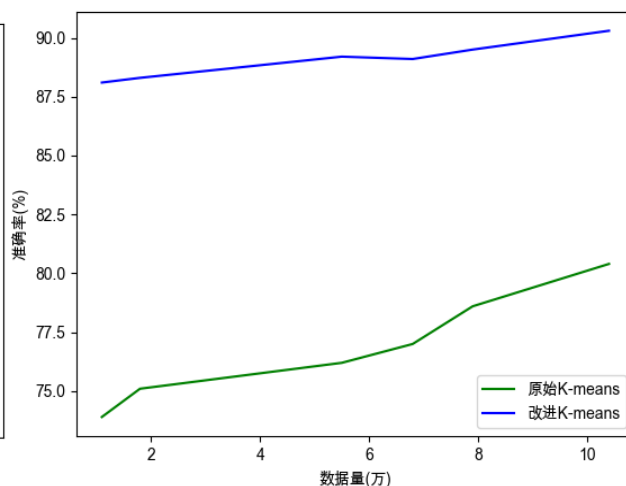


图 3.13 算法准确率对比图

3.4.3 结果分析

完成测试之后，由表3.7 可以看出，改进后的 K-means 算法的初始聚类中心准确度更高，状态阈值的误差也比较低。从图 3.12 和图 3.13 看出，改进优化算法的性能大大提升。准确率提高了 12%，算法耗时降低了约 60%，提升了系统使用的效率。

3.5 本章小结

本章主要从多个纬度介绍聚类算法的选取和优化改进方法。首先介绍了 K-mean 算法的原理、应用场景以及算法流程等。然后指出了 K-means 算法在使用时的存在问题，并且提出了对样本数据在初始化聚类中心的选取、去除异常点、孤立点的干扰等优化的方案；然后从实际问题进行数学建模和确定了数据格式。并且在模型数据进行了预处理，包括归一化和 PCA 降维处理，提升了算法计算速率，并依次详细介绍了方法执行步骤；设计了基于 GSA 自适应 K 值的算法，结合了手肘形的算法优点的进行对比，解决了样本数据自适应 K 值的问题；对聚类中心、孤立异常点敏感的缺点，设计了一种基于 Huffman 二叉树算法，可以提高聚类样本的聚类中心的准确性，极大地提高了聚类质量和运行效率；在采集端实现对数据流量上报的限制机制，优化了上报数据，减少了云平台数据的存储压力，提升了后面数据分析的准确性。

最后，对改进的 K-means 算法进行数据测试和测试结果进行分析，该算法可以自适应样本数据 K 值、初始聚类中心的获取，以及准确的获取到仪器设备的状态阈值电流；其此对原 K-means 算法和改进算法在性能上对比，有效地确认了改进算法，在准确率上提升了大约 12%，在耗时上降低了原来的 60%，同时时间复杂度也大大缩小了，提升了算法的性能。

第4章 基于贪心策略算法的智慧预约

为了贯彻国家对大型科研仪器对外开放共享的政策,针对于仪器设备中那些昂贵且数量较少,当使用时出现多人并发预约时等缺点。并且当仪器共享平台需要面对省、市的众多高校大学、以及全面向国用户开放的话,必须要面对的一个问题是:在有限的大型精密科学仪器设备下,并且使用需求非常之高,为了解决在有限的大型仪器下使我们仪器设备资源的社会价值最大化^[53],平衡供需关系。

完善在线预约共享平台需解决的问题有:

- 1)大型实验仪器在预约使用时,必须要规范其使用的时间段,而且预约借用仪器的设备,在同一时间段只能被同一人预约使用;
- 2)使大型实验仪器尽量满足所有预约申请借用的科研人员,并在此基础上使实验设备发挥最大的效益;
- 3)为了良性发展仪器预约机制,需加入一些预约限制机制,如有不良使用记录,预约则相对较难审核通过,还有一些爽约、迟到等。

因此,本文设计了多用户并发预约的模型,采用基于贪心策略的算法使效益最大化,实现预约的最大有效性,更好的完善了仪器共享平台的预约机制。

4.1 贪心算法的特点

贪心算法的核心思想是找寻局部最优解,不从整体最优出发,而是结合全部求得的局部最优解,再组合为整体上的一个近似最优解^[50]。

贪心算法对问题求解时,往往是逐步的进行,以现有的情况为基准依据其中一个指标作最佳的选择,一般不是考虑整体的求解。省去了很多算法去找最优解的时间复杂度,经过一直迭代循环找到局部最优解,提升算法性能。因此,在应用贪心算法时必然符合下面的性质:

- (1) 最优子结构性质:一般将所需解决的问题通过划分成小的子问题,再从众多的子问题中寻求这些子问题的最优解。
- (2) 贪心选择性质:在对问题求解时,依据选取局部最佳结果,推出整体结果,并且不依赖之后的选择,这就是贪心选择性质。

贪心策略算法的基本步骤如下:

- (1) 把所解决的问题的其中一个初始解作为起点;
- (2) 应用局部最优解的策略,经过不停循环选取局部最佳结果,最后从局部到整体再逐一求解。
- (3) 最终将局部的最优解集合看作是整体的最优解。

贪心策略算法解决问题的核心是寻找度量标准。因为是线性二次式,所占用的空间也较少,因此,求解速度相对较快,是解决实际问题常用的有效算法之一。

一般可以应用贪心算法，第一先看是否满足其描述的性质，第二是否满足贪心算法饿依据标准。其中贪心算法常用的事例，一个任务安排问题。描述如下：

多个应用场地被多个活动对应一一占用。其中每一个活动都需要使用某个资源，一般比如举行晚会，需要用到大礼堂，而体育活动则需要用到操场等，但是，这些活动内容都需要保证在其使用的起始时间和一个结束时间内，只能安排某一个活动。活动期间内，完全占用此资源，其它活动将不能使用。贪心算法中常以时间活动范围的选择为依据，提出有效的决策方案。

4.2 贪心算法的应用场景

1) 最短路径：一般是不断迭代求出多个路径和最小的值。

我们一般如果是存在 $1 \sim n$ 个城市，而且每个城市想通，若从其中某个城市出发，经过各个城市，最短路径到达目的地等问题。

2) 解决找零钱问题：一般是不追求最优解，只是去找到局部满足的最优解。

当有不同数量面额的纸币，现在要购买一个价值 K 元的东西，请问怎么才能适用最少的纸币？限制条件是每张纸币的个数，目的期望是选用最少的纸币。需要不断以最大的面值进行依次循环找到满足 K 元的东西等问题。

3) 执行任务调度问题：一般是求解执行任务集合中选出最大的相容任务子集合。

执行任务问题一般是在大集合中拥有多个活动任务需要安排，不同任务活动的时间区间不经相同，需要求出它们间互相不冲突的最大子集合^[51]。一般应用贪心算法可以有效解决，可以高效的任务安排每个活动的时间，最大程度有效兼容资源使用最大化^[52]。

其中，第三种的应用场景与本文预约机制安排问题一致，可以建立模型进行解决优化问题。

4.3 模型建立

共享预约问题，相似于一种执行任务活动 **NP-hard** 的问题^[54]。其中在执行任务安排中往往要安排最多相容活动或者是其中最大的相容活动子集。因此，本文将该问题的活动安排在某共享的资源上，求得结果满足最大相容任务子集。问题描述如下：

其中设定任务安排列表 $Y=\{1,2,\dots,n\}$ ，而任务 1 到任务 n 为同一资源，如会议室开会、餐厅吃饭等，其中任何一段时间内，都只能安排其中一个任务使用这个资源。条件还要每个任务都提供完整使用该资源的开始时间和结束时间。当前就称两任务安排是相容的，因为它们两个任务活动占用资源在使用期间不冲突。

而应用于多用户并发预约模型问题描述如下：

在多名科研人员并发预约实验活动中，大量科研人员通过预约平台预约某台仪器设备，系统会为每一台设备设置有效的预约时间段，科研人员选择在该时间段内满足条件的预约实验

分配仪器设备，每个实验都有一个要求使用仪器设备的开始时间和结束时间，因此在每个实验之间会存在时间上的交叉冲突，一般把时间上不冲突的两个实验称为相容实验。在该时间段内每个实验也可以选择是否服从时间调剂。每个实验都要求使用同一台仪器设备时，而在同一时间段只有一个实验能使用该台设备并且使用过程中不能其他人员预约。

4.4 基于贪心策略算法的智慧预约机制

4.4.1 预约机制指标

仪器设备预约使用过程中，需注意在申请预约前，科研人员要查询，当前预约仪器的实时状态，如果是在不在预约使用状态下，是无法预约成功的。其中只有当大型实验仪器处于非维修状态时，科研人员才可以进行预约申请。下面是仪器设备具体状态如图 4.1 所示

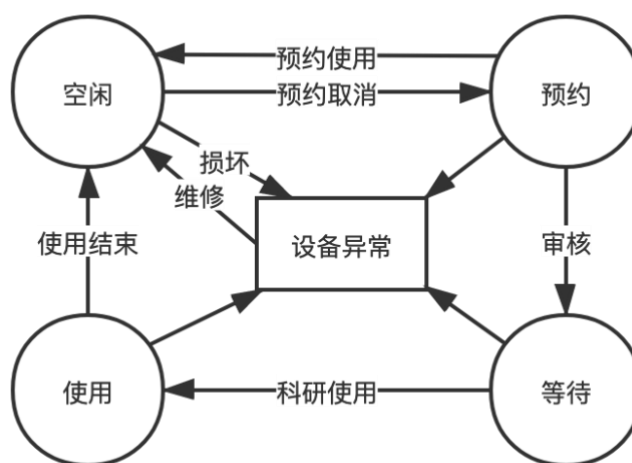


图 4.1 仪器共享平台预约流程图

当下为了科研人员在申请中的预约时，优化合理利用资源，本文设置了三个属性，来衡量社会价值的体现，为预约机制重要参数指标。

1) 设置时间属性。科研人员在仪器共享平台上提前预约某台仪器时，系统将会给该设备分配其可以使用和预约的时间段。每个科研人员，在做每个实验时，要对应选择自己的开始使用时间和结束时间；

2) 设置信用属性。其次我们增加的大型实验仪器预约使用信用值，如果在使用仪器发生爽约、损坏仪器、等不良记录，会降低预约；

3) 设置科研价值属性。系统也增加了一项重大社会科研项目优先级，如果当前科研项目是重大的科研项目、社会价值较大，可以设置第一优先级。

仪器预约流程如图 4.2 所示：

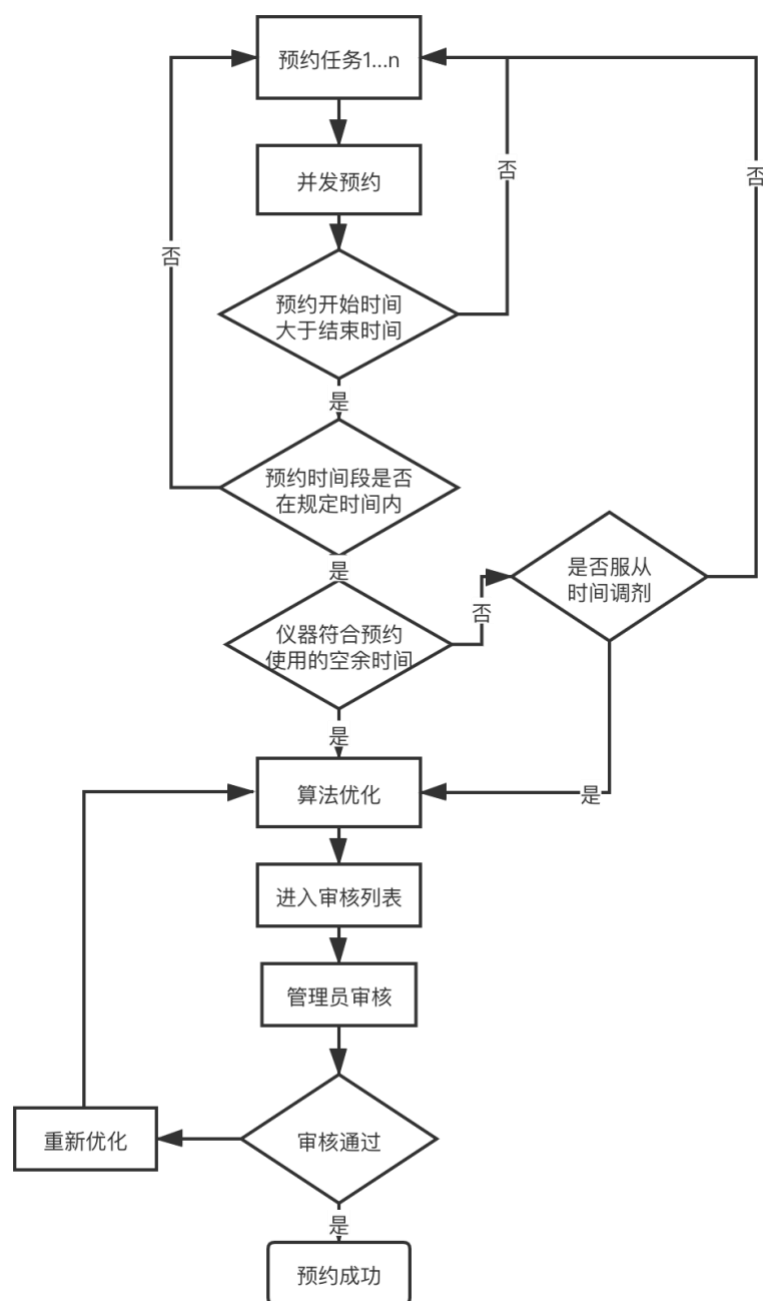


图 4.2 仪器共享平台预约流程图

本文设计的智慧共享预约机制，对比之前传统的先来先用的机制，提升了整体的智能资源共享，解决了在传统模式下，使用者因迟到、爽约等造成仪器浪费现象^[55]。提高了仪器管理的整体机制。

4.4.2 仪器设备预约模型实现

大型实验仪器预约的功能，要解决多个人并发申请同一台设备资源是，合理调度仪器设备的预约时间，和整个申请使用流程，而且是大型实验仪器的资源利用率最大化。

因此在仪器设备预约问题中，基本确认和活动安排问题标准机制基本一致，可以使用时间区间作为科研人员预约的使用仪器设备时间范围，利用贪心算法解决方案如下：

首先构建样本集合 $X = \{x_i | x_i \in R, i = 1, 2, 3, \dots, m\}$, $Y = \{y_j | y_j \in R, j = 1, 2, 3, \dots, n\}$ 。X是有 m 人，Y是有 n 台仪器。当有 m 个人并发预约申请同一个仪器设备时，设置时间区间：在集合 X 里的元素 x_i 都要使用这台仪器设备时，由于它们申请使用的时间段有概率会发生冲突，因此标记一个当前使用的区间，给每一个使用者。每个集合 X 里的 x_i 都有一个要求使用仪器设备的时间段，起始时间设为 s_i ，则结束时间设为 f_i ，同时要求 $s_i < f_i$ 。在时间区间 $[s_i, f_i]$ 内，X 集合中的 x_i 在使用当前的实验仪器。当发现区间 $[s_i, f_i]$ 和区间 $[s_j, f_j]$ 没有交集时，则称为元素 x_i 和 x_j 相容。因此，此活动安排仪器设备使用，就转化为求解样本集合 X 中最大的相容子集合，其中子集合中的每个人都可以被顺利安排使用仪器设备，就可以提高利用率，使预约效果最佳。

其中具体实现方法如下：前提是该仪器设备处于非维修状态下，科研人员在从仪器设备借用时。需确定输入：预约的仪器设备名称、预约时间段选取、预约优先级。输出：预约结果。包含，预约成功仪器设备的实验室位置、使用的时间段。

$$Y_{ij} = \max (\sum_{j=1}^m \sum_{i=1}^n v_i x_{ij}) \quad (4.1)$$

式(4.1)中： Y_{ij} 表示最大的科研价值属性，为目标函数。 v_i 科研价值，

$$\sum_{j=1}^m x_{ij} \leq 1, \quad i = 1, 2, 3, \dots, n \quad (4.2)$$

式(4.2)中：表示当前预约的仪器只能最多被一个人使用。

$$\sum_{i=1}^n d_i x_{ij} \leq T_j, \quad j = 1, 2, 3, \dots, m; \quad d_i = F_i - S_i, \quad i = 1, 2, 3, \dots, n; \quad (4.3)$$

式(4.3)中：表示科研人员预约的时间段总和不大于 T_j （仪器支持使用的总时间）。

在客户端科研人员申请窗口中，输入各设备借用申请后，后台会第一时间生成一个预约列表，待各个预约者科研价值属性、和预约时间符合分配之后，会自动将列表进行升序排列。然后在我们大型仪器平台中所有的仪器进行分配，一一对应。待管理员审核通过后，即可通知使用。否则，则取消使用预约资格。

4.5 测试分析

4.5.1 预约算法数据测试

4.5.1.1 测试方案

本节测试方案是：样本集合 $Y_i = \{y_i | y_i \in R, i = 1, 2, 3, \dots, 50\}$ 首先，先以 50 人同时预约一台仪器设备做贪心策略算法测试，仪器设备一天中可供连续工作的时间为 14 小时，因此，平台设定有效的仪器可预约仪器设备的时间范围是 08:00~22:00；每一个科研人员预约实验时间大部分不超过 4 小时，小部分特殊实验，可为 4~14 小时。

4.5.1.2 测试数据

下表4.1 是 50 人预约实验的价值数据表。而 50 个预约实验的申请使用的时间分布如图 4.3 所示，其中经过算法优化等到的分布如图 4.4 所示。

表4.1 50 数据量实验价值表

标号	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13
价值	4	1	6	9	10	8	8	3	4	10	3	10	2
标号	Y14	Y15	Y16	Y17	Y18	Y19	Y20	Y21	Y22	Y23	Y24	Y25	Y26
价值	4	6	6	3	3	5	8	10	3	8	6	3	7
标号	Y27	Y28	Y29	Y30	Y31	Y32	Y33	Y34	Y35	Y36	Y37	Y38	Y39
价值	6	10	1	1	2	10	7	7	7	5	9	4	2
标号	Y40	Y41	Y42	Y43	Y44	Y45	Y46	Y47	Y48	Y49	Y50		
价值	5	1	8	9	1	5	5	8	7	8	4		

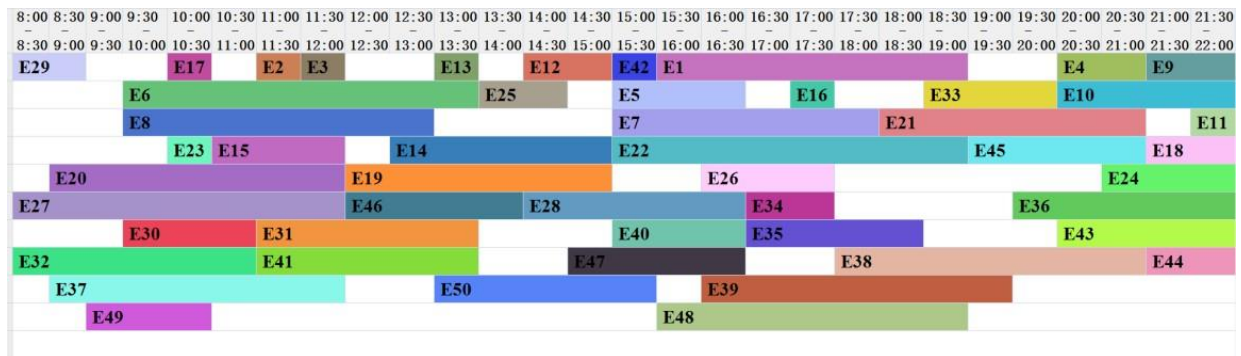


图 4.3 预约仪器时间分布甘特图



图 4.4 使用贪心算法优化结果甘特图

4.5.1.3 测试分析

完成测试后，从表4.2 可以看出，在实现基于贪心策略算法设计优化预约实验的算法后，在一台仪器设备相同时间内仪器高效预约整合度非常好，仪器有效预约使用提升了近 40%。

表4.2 50 数据量算法结果对比表

对比方式	预约方式整合度	仪器有效使用时间比例	最优值（价值）
原始预约方式	较差	54%	60.00
算法优化方式	非常好	94%	89.00

4.5.2 仪器共享数据对比

仪器设备预约共享平台运行一年，提升仪器设备的使用率超 50%。下面我学校中的主要几个仪器设备，在一年中统计使用的时间，人工和在共享平台下使用时间的数据对比，如表 4.3 所示。其中共享预约算法，时间复杂度低，运行效率高。对预约进行优化，使用户在预约中，也提升预约的速度，提升了实验仪器的使用率。其中使用率的计算如公式（4.3）。设定仪器设备年内的总时间 1600h，借助里仪器设备智慧共享平台，明显使用率较之前人工方式，提高很多。

$$\text{使用率} = \frac{\text{使用时间}}{\text{年内总时间}} * 100\% \quad (4.3)$$

表4.3 仪器设备年使用率对比表

仪器设备名称	使用时间 (人工方式)	使用率 (人工方式)	使用时间 (共享平台)	使用率 (共享平台)
PCR 红外线衍射仪	500	31.3%	1280	80%
激光共聚显微镜	486	30.4%	1350	84.4%
核磁共振仪	678	42.3%	1440	90%
超高效液相色谱仪	510	31.8%	1280	80%
元素分析仪	410	25.6%	1360	85%
BH 测试仪	450	28.3%	1410	88.1%
荧光光谱分析仪	350	21.9%	1320	82.5%

4.6 本章小结

本章主要介绍了仪器共享平台系统中对基于贪心策略的智慧预约算法技术的深入研究。介绍了贪心策略算法的特点以及应用场景。依据贪心策略解决问题的常用方式，我们建立了数学模型，进行了模型预约机制测试，完成了仪器设备活动预约的效果，提升了仪器设备在平台上预约的优化提升。最后，也对仪器共享平台进行一年仪器设备使用率的数据进行对比，该智慧系统的设计，有效的提升了仪器设备的使用率，达到了设计目标。

第5章 总结与展望

5.1 本文工作总结

物联网的发展,是我们的生活变得更加便捷。大型实验仪器的管理和使用也许凭借着信息化的建设,使仪器、科研人员以及仪器管理人员更加紧密的结合在一起,为实验室管理提供了方便、快捷、先进、可视化、数字化和智能化转变。需要不断的提升高校科研的服务质量,扩大高校科研仪器的资源共享使用。因此,本课题利用先进的物联网技术和电子技术,结合目前大型实验仪器管理使用的缺点,设计了仪器设备状态智能监测系统,本文主要工作如下所述:

1) 对目前仪器设备状态获取和仪器设备共享的现状进行分析,在物联网技术的飞速发展下,设计构建仪器设备在线监测智慧共享系统的方案,使高校大型实验仪器向数字化、智慧化转型提供基础。

2) 对仪器设备状态智能监测系统各个部分进行设计。系统包括智能采集终端、网络层、云平台、可视化终端四大部分。设计了通过 NB-IOT 的智能采集终端,采集数据传输至云平台,数据以 base64 进行加密传输,保证安全;并且设计错峰上报机制,优化了数据上报的准确性;数据请求加入重发机制,也大大确保了数据传输的稳定性。设计了远程更新 CPU 程序,扩展了设备的功能性。最后,仪器设备的状态信息,在可视化终端上进行了数据展示,实现了系统的状态实时监测,并且通过客户端、Web 端实现了智慧预约仪器使用,提升了仪器设备的使用率。

3) 针对大型实验仪器的状态分析,先介绍原 K-means 算法的应用特点、算法流程、以及存在的缺点问题,基于改进的 K-means 聚类算法,将智慧采集终端采集仪器设备状态电流信息,进行数据分析,获取到仪器的设备的关机、待机、运行的状态阈值。其中从选取 K 值、初始聚类中心、孤立点、三个方面优化改进算法,设计了基于 GSA 自适应 K 值算法,实现了当仪器设备工作周期不完整的情况下,也可以自动识别输出最佳 K 值。设计了 Huffman 二叉树算法,实现除去孤立的相异点,完成初始聚类中心的选择机制。经过对聚类模型仿真测试及参数分析,与原始 K-means 算法对比发现,改进算法的准确率提升了 12%,在算法耗时降低了 60%。

4) 对基于贪心策略算法智慧预约的改进研究,针对目前珍贵稀少的大型实验仪器,在使用时,出现多人并发同时预约时等缺点,贪心算法实现预约的最大有效性,更好的完善了仪器共享平台的预约机制。

5.2 展望

本文设计的仪器设备状态智能监测系统中,通过构建了物联网体系,解决目前大型实验仪器管理效率低下,仪器使用率低,共享机制不完善的缺点。也继续将大型实验仪器

向智能化和数据化转型，值得继续深入研究：

1) 构建的大型实验仪器监测共享系统中，主要目的还是提升仪器管理的合理化，降低人工成本，可视化展示仪器设备的所有情况，目前还是在 Web 端功能较多，随着移动化越来越发展，应该将可视化功能逐步实现到移动端。

2) 针对仪器设备的电流数据信息进行数据分析优化，模型的建立应该越来越完善，算法的时间复杂度、空间复杂度的降低和优化还需继续研究，而且对后期大量科研人员使用的仪器设备信息再继续进行分析，使用率较高的，可以分配较多的经费，提供采购决策。

3) 对仪器设备共享环境下，目前仪器共享平台用户量还相对较小，在逐步扩大时，在处理并发预约时，应该还许进行设计动态请求限制优化、负载均衡，防止服务器同事间服务量剧增导致，数据丢失以及宕机。

致 谢

人生天地间，若白驹过隙，时间转瞬即逝，在杭州电子科技大学电子信息学院，三年的时光，我时刻着“笃学力行、守正求新”的校训，真切的体会到了在学习上应该专心致志，潜心求学，努力实践，不断创新。学校浓厚的学术氛围、人文底蕴，也深深地熏陶着我们。过去的三年是学生生涯中是我人生最宝贵的财富，感谢给予我帮助的那些人，也感谢在杭电的学习生活带给我美好的一切。

首先，由衷感谢我的导师陈科明副教授，在老师门下三年的循循善导，老师不仅指引在工作上踏踏实实、勤勤恳恳、而且学习生活上让我们潜心专研，不断创新，提升能力。督促自己不断进步。老师严谨的教学态度、渊博的学识积累及勤勉的工作精神都让我敬佩不已。师恩门下，不悔而已。

感谢我的同门吴新虎、司端杨、郑杭明、杜坤和王奕怡同学，以及我的室友俞锋、王彪、檀宇，在学习、工作和思想上都给予了我莫大的帮助和鼓励，也感谢我的师弟师妹，在学校为我提供的一切帮助，能够和他们成为同学和朋友我感到非常荣幸。

感谢我的父母及家人，感谢您们这三年来对我的理解和支持，在我人生感到彷徨、迷茫时，为我点燃心中的灯，指引我不断前进，使我能够顺利完成学业。

最后，感谢论文评阅老师和答辩委员会的各位老师！再一次向帮助过我的老师和同学们表达我最诚挚的感谢和问候，这几年的生活因你们的存在而精彩。

参考文献

- [1] 陈思.高校实验室仪器设备管理现存问题及对策研究[J].教育教学论坛,2019(48):13-14.
- [2] 赵玉茹,冯建跃,赵月琴.浙江省高校大型实验仪器开放共享现状调研与思考[J].实验技术与管理,2017,34(10):255-258+266.
- [3] 王莹瑞,孔翠,张峰君.非 985、211 高校大型实验仪器共享平台研究与建议——以华东地区为例[J].科技视界,2020(25):163-165.
- [4] 何凯.高校大型实验仪器管理绩效评价研究[D].西安建筑科技大学,2017,28(17):6-9.
- [5] 张三军,洪家祺,赵杭美,薛云,黄岚兰.高校大型实验仪器开放共享管理的探讨与实践——以华东师范大学为例[J].实验技术与管理,2020,37(10):1-5.
- [6] 严晓光.电器设备运行状态的在线监测[D].华北电力大学(河北),2004,27(17):6-9.
- [7] 刘克冰.基于深度学习的电力设备状态识别方法与应用研究[D].华北电力大学,2018,47(27):16-19.
- [8] 刘艳红,张骥,余孝其,李静.大型实验仪器设备异常运行状态辨识方法研究[J].电子设计工程,2019,27(17):6-9.
- [9] 邵政.基于大数据的小电流接地系统单相接地故障选线方法研究[D].山东大学,2018,30(24):28-32.
- [10] Engineering; Researchers from Harbin Engineering University Report Recent Findings in Engineering (Image Object Extraction Based On Semantic Detection and Improved K-means Algorithm)[J]. Journal of Technology & Science,2020,27(17):22-29.
- [11] 王浩.高校大型实验仪器的管理现状及对策[J].当代化工研究,2020(19):157-158.
- [12] 黎惠.大型科学仪器共享监控平台设计[D].电子科技大学,2012,27(17):6-9.
- [13] 徐浩.高校实验室建设与管理共性问题的研究与实践[D].合肥工业大学,2014.
- [14] 董诚.欧洲核粒子中心的发展及其对我国的启示[J].中国科技资源导刊,2008(01):52-55.
- [15] 高凤新,卢士香.高校实验室仪器设备管理和使用的信息化建设[J].实验室研究与探索,2014,33(05):272-274+282.
- [16] 李春梅,何洪,程南璞,李元.高校大型实验仪器共享管理模式和运行机制探讨[J].西南师范大学学报(自然科学版),2018,43(02):83-88.
- [17] 宋霖婧,郝旭东,刘江宁. Discussion on How to Promote the Management Reform of Promoting the Sharing of Scientific Instruments in Scientific Institution[J]. 仪器与设备,2020,08(01).
- [18] 马骁.区域性大型科学仪器设备共享平台建设[D].山东大学,2008,27(17):6-9.
- [19] 苏俊盼.基于 NB-IOT 技术的远程采集终端设计与实现[D].西安科技大学,2019,37(17):6-9.
- [20] 王明.基于 NB-IOT 的监控系统的设计与实现[D].中国科学院大学(中国科学院沈阳计算技术研究所),2019,25(24):20-22.

- [21] 吴晓斌.基于 NB-IOT 的窖井盖及井下工况远程监控系统设计[D].中北大学,2018,35(34) :28-32.
- [22] 许杰.物联网无线通信技术应用探讨[J].无线互联科技,2018,15(14):19-20.
- [23] 钱志鸿,刘丹.蓝牙技术数据传输综述[J]. 通信学报, 2012, 33(4): 143-151.
- [24] 杨静.短距离无线通信技术对比及其应用研究[J].无线互联科技, 2016,12(13): 12-13.
- [25] Yuanyuan Liao,Liao Yuanyuan,Liu Shuxian. Research on Key Technologies of Smart Logistics Dynamic Positioning Based on NB-IOT[J]. Journal of physics. Conference series,2020,1634(1):43-51.
- [26] Caiyun Wen, Qi Jiang. Design of Public Garbage Can Overflow Alarm and Positioning System based on NB-IoT[J]. International Core Journal of Engineering,2020,6(7) :76-79.
- [27] Sequans Communications S.A.; Sequans' Monarch SiP and GM01Q Module for LTE-M and NB-IoT Networks Now Certified by Japan's Three Largest Mobile Operators[J]. Journal of Engineering,2019,33(4): 143-151.
- [28] 盛国林.基于物联网技术的大型实验仪器管理系统的设计与实现[D].山东大学,2011,15(24) :18-22..
- [29] 吴芳.大型仪器共享管理平台服务器端的设计与实现[D].东南大学,2015,55(34) :38-42.
- [30] 杨娟.基于 web 的仪器共享平台的研究与设计[D].北京邮电大学,2015,35(23) :50-52.
- [31] 李春红,高建华.使用分层模型改进 MVC 设计架构[J].计算机工程与设计,2007(04):766-769.
- [32] 饶绪黎,张美平,马现虎.基于物联网技术的实验设备运行监控管理系统设计[J].山东大学学报(理学版),2012,47(11):74-77.
- [33] Serdar Yegulalp. 8 great little Python web frameworks[J]. InfoWorld.com,2020,55(24) :28-32.
- [34] V. Santhana Marichamy,V. Natarajan. Big Data Performance Analysis on a Hadoop Distributed File System Based on Geometric Data Perturbation Technique[J]. Procedia Computer Science,2019,165(17):86-89.
- [35] Ouhda Mohamed,Ouanan Mohammed,Aksasse Brahim. An improved Smoothing Method Based on Diffusion Equation and K-means Clustering[J]. International Journal of Engineering and Future Technology,2018,15(4).
- [36] 杜青青.基于深度学习理论的设备健康状态识别研究[D].西安科技大学,2018,3(9-10).
- [37] 王春风,唐拥政.结合近邻和密度思想的 K-均值算法的研究[J].北京:计算机工程应用, 2011,47(19):147-149.
- [38] 罗军锋,锁志海.一种基于密度的 k-means 聚类算法[J].微电子学与计机,2014,31(10):28-31.
- [39] Search Heuristics for Resource Constrained Project Scheduling[J]. The Journal of the Operational Research Society,1996,47(5):6-9.
- [40] Michael Mascagni. Book review: Simulated annealing and Boltzmann machines: A stochastic approach to combinatorial optimization and neural computing[J]. Journal of Statistical Physics,1989,56(3-4).
- [41] 韩凌波,王强,蒋正锋等.一种改进的 K-Means 初始聚类中心选取算法[J].计算机工程与应用,2010,46(17):151-152.
- [42] 庄瑞格,倪泽邦,刘学艺.基于拟蒙特卡洛的 K 均值聚类中心初始化方法[J].济南大学学报(自然科学版),2017,31(01):35-41.

- [43] Pooya Tavallali, Peyman Tavallali, Mukesh Singhal. K-means tree: an optimal clustering tree for unsupervised learning[J]. The Journal of Supercomputing, 2020(prepublish) 27(17):6-9.
- [44] 陈杰, 朱娟. 基于簇间相似度判定的自适应 K 均值算法[J]. 计算机工程与设计, 2010, 31(10): 2270-2272+2375.
- [45] Meihong Wang, Yuhui Zheng, Ming Qiu, Yan Zheng. Research on schedule-based user recommendation model based on improved K-means algorithm[J]. Journal of Computational Methods in Sciences and Engineering, 2016, 16(3):56-59.
- [46] 刘洋, 张秋余. 基于 Huffman 树的模糊多类支持向量机[J]. 辽宁工程技术大学学报(自然科学版), 2008(01):96-99.
- [47] Yang Xu, BaoZhan Chen, ZhiChen Hu. Research for multi-sensor data fusion based on Huffman tree clustering algorithm in greenhouses[J]. Int. J. of Embedded Systems, 2016, 8(1):175-182.
- [48] Visalakshi N K, Thangavel K. Impact of normalization in distributed k-means clustering[J]. international Journal of Soft computing, 2009, 4(4): 168-172.
- [49] 王彤, 晁爱农, 张红梅. 一种动态 Huffman 优化算法的设计与应用[J]. 空军工程大学学报(自然科学版), 2005(02):76-78.
- [50] 王崚灏. 基于贪心算法的云计算资源调度策略[D]. 延安大学, 2018.8(08):86-88.
- [51] 李新艳. 基于贪心算法的仪器设备共享平台的设计与实现[D]. 燕山大学, 2018, 3(9-10).
- [52] 李胜成. 基于贪心算法的智能 RGV 的动态调度策略[J]. 科技创新与应用, 2020(06):16-17.
- [53] 邵科峰, 刘连臣, 吴澄. 设备共享环境下的并发预约问题研究[J]. 计算机集成制造系统, 2006(07):1117-1121.
- [54] Cheng Wang, Junqing Bai, Jiaojiao Li, Longlei Dong, Guirong Yan, Zengguo Sun. Max-min K-means clustering algorithm and application in responsesignal feature extraction[J]. International Journal of Applied Electromagnetics and Mechanics, 2012, 39(1-4).
- [55] K. Srinivasa Rao, K. Kiran Kumar, P. Srinivasa Rao. An Efficient K-Means Clustering Algorithm for Large Data[J]. Data Mining and Knowledge Engineering, 2011, 3(9-10).

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] **大型仪器电流智能检测与定位设备开发，横向课题
- [2] **仪器智慧预约共享平台系统研发，横向课题
- [3] **智能安装运维小程序设计研发，横向课题

专利：

- [1] 陈科明,安建伟,刘剑,徐志杰.一种自适应不同大型设备仪器状态阈值的方法[P]. 中国专利:CN112307671A,2021-02-02