

杭州电子科技大学

硕士学位论文

题目： 基于数据挖掘的光伏电站远程运维技术研究

研究生 王佳敏

专业 电子信息

指导教师 陈科明 教授

企业导师 曾佳

完成日期 2024 年 5 月

杭州电子科技大学硕士学位论文

基于数据挖掘的光伏电站远程运维技术研究

研 究 生： 王佳敏

指导教师： 陈科明 教授

2024 年 5 月

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

**Research on remote operation and
maintenance technology of photovoltaic
power station based on data mining**

Candidate: Wang Jiamin

Supervisor: Associate Prof. Keming Chen

May, 2024

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： 王佳敏 日期：2024 年 6 月 6 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： 王佳敏 日期：2024 年 6 月 6 日

指导教师签名： 陈利明 日期：2024 年 6 月 6 日

摘要

随着我国太阳能光伏装机规模不断扩大，如何智能高效地运营和管理光伏系统已成为业内备受关注的热门议题。伴随着物联网、云平台 and 大数据等技术发展，光伏产业逐渐从传统人工维护向智能决策和数据挖掘方向转型。目前工业界提出了许多光伏电站系统的运维方案，但是大部分系统依旧存在以下不足之处：不能远程实时获取光伏设备的运行数据；光伏电站清洁周期往往基于估计或静态设置，无法对组件表面积灰程度量化分析；系统预警功能薄弱，不能预测分析组件潜在风险，只能通过故障报警时的推送发现具体故障。

根据目前行业内光伏系统的需求和不足的现状，本文通过对基础数据的价值挖掘，结合物联网前沿技术设计了一个光伏电站远程运维系统：通过智能网关与光伏逆变器、气象站、辐照仪和红外抄表等设备进行数据交互，再将采集的实时信息上传至云平台进行解析处理。整个系统基于微服务架构来搭建后端业务算法集群，包含了远程升级、数据分析、清洁决策和故障预警等多个功能。最后将系统分析得到的数据通过 Web 端和小程序等进行可视化展示并且实现远程控制。

灰尘在光伏组件表面的大量积累会导致电站发电效率和收益的降低，严重时甚至会引发安全事故，因此一个合理可靠的清洁决策非常重要。本文提出一种采用无积灰理想发电量对损失发电量进行量化的算法思想，主要通过平台采集上来的数据筛选出电站历史多次清洁周期内最佳清洁效果的功率数据，并采用一种改进的网格搜索（Grid Search, GS）优化支持向量机回归（Support Vector Regression, SVR）算法构建无积灰理想功率预测模型。该算法通过参数搜索空间的二分查找操作以及步长的放缩操作对普通 GS 进行改进，大幅提升 SVR 模型中参数 c 、 g 的寻优效率，减少 SVR 的整体运行时间，算法的平均绝对误差达到 0.944，均方根误差达到 1.213。最后通过每日的实际功率和实际发电量可以计算出发电量损耗值并和清洁成本对比，从而确定电站的动态清洁周期。

预警与报警都是光伏电站中必要的安全机制，预警相对于报警提供了时间和机会来预防或减轻潜在的灾害。本文围绕双向长短时记忆网络（Bi-Directional Long Short-Term Memory, BiLSTM）搭建组串电流预测模型，并采用卷积神经网络（Convolutional Neural Networks, CNN）对特征提取进行优化。然后对晴天、阴天、雨天三种天气情况分别建模，并且与 LSTM、CNN-BiLSTM 两种预测模型进行实验对比。实验结果得出采用 CNN-BiLSTM 模型对光伏组串电流预测在三种天气情况下平均绝对误差最小，分别达到了 0.268、0.393、0.361，均方根误差也最小，分别为 0.331、0.475、0.458。最后引入偏差率指标结合滑动窗口方法建立报警阈值机制，实现光伏组件在故障报警推送之前的及时预警。

关键词：光伏系统，清洁决策，积灰清洗，发电量损失，故障预警，偏差率

ABSTRACT

With the continuous expansion of the installed scale of solar photovoltaic in China, how to operate and manage photovoltaic systems intelligently and efficiently has become a hot topic of concern in the industry. With the development of technologies such as the Internet of Things, cloud platforms and big data, the photovoltaic industry has gradually transformed from traditional manual maintenance to intelligent decision-making and data mining. At present, the industry has proposed many operation and maintenance schemes for photovoltaic power station systems, but most of the systems still have the following deficiencies: the operation data of photovoltaic equipment cannot be obtained remotely in real time. The cleaning cycle of photovoltaic power stations is often based on estimates or static settings and the degree of dust accumulation on the surface of components cannot be quantified. The early warning function of the system is weak and the potential risks of components cannot be predicted and analyzed. Specific faults can only be discovered through push notifications from fault alarms.

According to the current situation of the demand and deficiencies of photovoltaic system in the industry, this thesis designs a remote operation and maintenance system of photovoltaic power station by mining the value of basic data and combining with the cutting-edge technology of the Internet of things: through intelligent gateways, it communicates with photovoltaic inverters, weather stations, irradiance meter and infrared meter reading equipment, and then upload the real-time information collected to the cloud platform for analysis and processing. The whole system is based on microservice architecture to build back-end business and algorithm cluster, including remote upgrade, data analysis, clean decision, fault warning and other functions. Finally, the data obtained from the system analysis is displayed visually through the Web side and the small program, and remote control is realized.

The large accumulation of dust on the surface of the photovoltaic module will lead to the reduction of power generation efficiency and income, and even cause safety accidents in serious cases. Therefore, a reasonable and reliable cleaning decision is very important. This thesis proposes an algorithm idea that using the ideal power generation which is not affected by dust accumulation to quantify the power generation lost. It mainly uses the data collected by the platform to filter out the power data with the best cleaning effect during multiple cleaning cycles in the history of the power station. And it uses an improved Grid Search (GS) to optimize the Support Vector Regression (SVR) algorithm to construct Ideal power prediction model which is not affected by dust accumulation. This algorithm improves the ordinary GS through the binary

search operation of the parameter search space and the step size scaling operation, which greatly improves the optimization efficiency of parameters c and g in the SVR model, and reduces the overall running time of SVR. The average absolute error of the algorithm reaches 0.944, and the root mean square error reaches 1.213. Finally, the power loss value can be calculated based on the daily actual power and actual power generation and compared with the cleaning cost, thereby determining the dynamic cleaning cycle of the power station.

Both early warning and alarm are necessary safety mechanisms in photovoltaic power stations. Compared with alarm, early warning provides time and opportunity to prevent or mitigate potential disasters. This thesis builds a string current prediction model around Bi-Directional Long Short-Term Memory (BiLSTM), and uses Convolutional Neural Networks (CNN) to optimize feature extraction. Then the three weather conditions of sunny, cloudy and rainy days were modeled respectively, and the experiment was compared with LSTM and CNN-BiLSTM prediction models. The experimental results show that the average absolute error of the CNN-BiLSTM model for the prediction of PV series current is the smallest under three weather conditions, reaching 0.268, 0.393 and 0.361 respectively. And the root mean square error is also the smallest, reaching 0.331, 0.475 and 0.458 respectively. Finally, this thesis introduces the deviation rate indicator and the sliding window method to establishes an alarm threshold mechanism, thereby achieving timely warning of photovoltaic modules before fault alarms are pushed.

Keywords: Photovoltaic system, Cleaning decision, Dust cleaning, Power generation loss, Fault warning, Deviation rate

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	3
1.2.1 光伏电站运维发展现状	3
1.2.2 光伏电站运维系统研究现状	3
1.2.3 光伏板清洁决策研究现状	4
1.2.4 光伏板故障预警研究现状	5
1.3 论文主要研究内容以及各章节安排	6
第 2 章 分布式光伏电站的远程运维系统设计	8
2.1 系统总体设计	8
2.1.1 需求分析	8
2.1.2 总体架构	8
2.2 网关设计	10
2.2.1 网关的通信方式与协议	10
2.2.2 网关硬件设计	11
2.2.3 网关软件设计	12
2.3 后端架构	14
2.3.1 后端组件选型	14
2.3.2 云平台	18
2.3.2.1 设备云	18
2.3.2.2 业务云	18
2.4 服务设计	19
2.4.1 基础服务设计	19
2.4.1.1 实时推送服务	19

2.4.1.2 远程升级及优化	21
2.4.2 业务服务设计	22
2.4.2.1 接口需求设计	22
2.4.2.2 数据库设计	24
2.5 应用层设计	26
2.5.1 Web 前端设计	26
2.5.2 小程序设计	26
2.5.3 系统展示	27
2.6 本章小结	30
第 3 章 基于发电功率数据挖掘的光伏电站清洁决策方案	31
3.1 积灰对发电量的影响	31
3.2 清洁决策方案分析	31
3.3 光伏功率与气象因素的关系	32
3.2.1 辐照度对功率的影响	32
3.2.2 温度对功率的影响	33
3.2.3 湿度对功率的影响	33
3.2.4 风速对功率的影响	34
3.2.5 其他因素对功率的影响	35
3.4 数据集定义与分析	35
3.4.1 相关性分析	35
3.4.2 数据集定义	36
3.4.3 数据集筛选	36
3.4.4 数据归一化	36
3.5 支持向量机理论	37
3.5.1 支持向量机	37
3.5.2 核函数	38
3.5.3 支持向量机回归	39
3.5.4 支持向量机回归模型	41
3.6 基于改进 GS 的 SVR 理想无积灰功率预测算法	42
3.6.1 K 折交叉验证	42

3.6.2 网格搜索法及其改进	42
3.6.3 基于改进 GS 的 SVR 理想无积灰功率预测算法具体实现	44
3.7 实验测试与结果分析	45
3.7.1 数据预处理	45
3.7.2 模型评估指标	45
3.7.3 实验仿真和分析	46
3.8 清洁决策制定与实例验证	49
3.8.1 无积灰理想发电量的计算	49
3.8.2 确认清洁时间	50
3.8.3 实例验证	51
3.9 本章小结	53
第 4 章 基于组串电流数据挖掘的光伏组件故障预警方案	55
4.1 预警方法选型	55
4.2 预测模型选型	56
4.3 CNN-BiLSTM 模型	57
4.3.1 LSTM 模型	57
4.3.2 双向长短时记忆神经网络原理	59
4.3.3 卷积神经网络原理	60
4.4 基于 CNN-BiLSTM 组合光伏组串电流预测模型的实现	62
4.5 实验测试与结果分析	63
4.5.1 数据预处理	63
4.5.2 模型评估指标	64
4.5.3 模型初始化	64
4.5.4 实验仿真和分析	64
4.6 故障预警规则制定与实例验证	68
4.6.1 偏差率	68
4.6.2 滑动窗口优化	70
4.6.3 实例验证	71
4.7 本章小结	73
第 5 章 总结与展望	74

5.1 总结	74
5.2 展望	75
致谢	77
参考文献	77
附录	83

第 1 章 绪论

1.1 研究背景和意义

根据国家能源局发布的 2023 年全国电力工业数据统计，结合图 1.1 可得：2023 年全国并网太阳能发电装机规模达到了 6.1 亿千瓦，同比增长 55.2%^[1]。

2016到2023年我国太阳能发电装机容量及增速情况

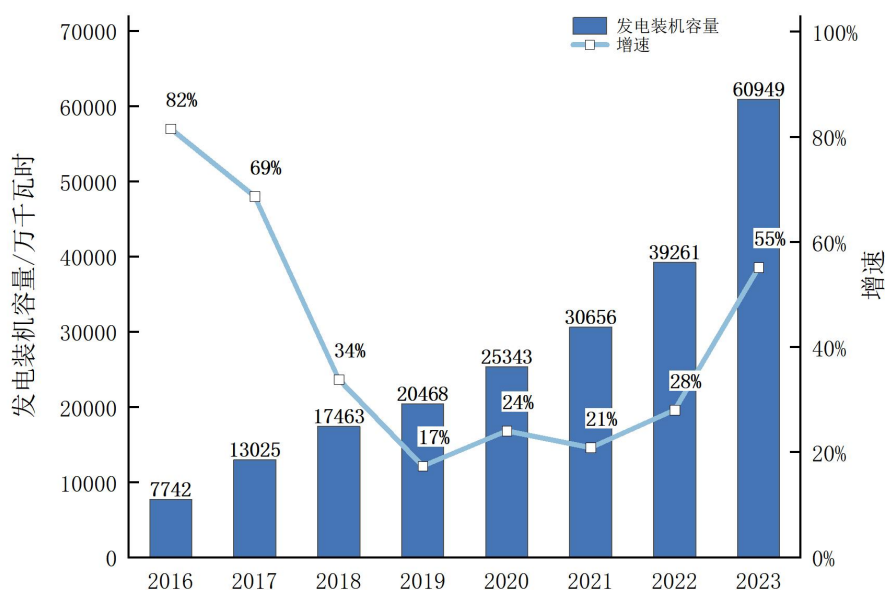


图 1.1 2016 年-2023 年我国光伏发电装机容量及增速情况

近年来随着世界新能源的开发规模不断扩大，其使用成本也快速下降，以风电、光伏为代表的新能源产业逐渐代替火电、水电等传统发电产业，成为全球能源的重要组成部分^[2]。图 1.2 为 2016 年和 2023 年不同电力生产方式累计装机容量占比对比图。图中太阳能发电在不同电力生产方式中的装机容量占比逐渐增大，说明中国光伏产业规模快速的扩大和市场的日益成熟。

2016年不同电力生产方式累计装机容量占比 2023年不同电力生产方式累计装机容量占比

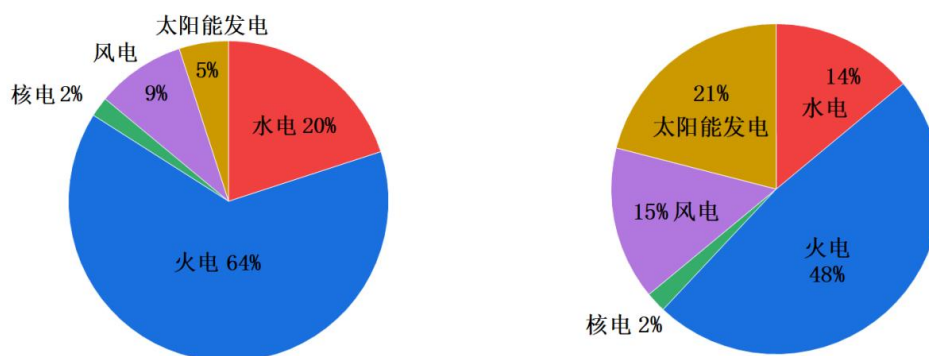


图 1.2 2016 年和 2023 年不同电力生产方式累计装机容量占比对比图

光伏发电主要可以分为分布式发电和集中式发电。集中式光伏发电大部分位于我国西部地区，其充分利用了矿山、荒漠等地区空旷的地理优势进行建设，并且西部地区具有用地面积广、海拔高、气候干燥、日照时间长且光照强度高等环境条件，因此更适合发展集中式光伏发电。但本地地区难以消纳这么大的电能产量，大多产生的电能都会接入高压输电系统远程输送至发达地区。另一类是本文主要讨论的发电场景：分布式发电，其更多应用于我国的东部地区，主要可以应用在商业建筑、市政等公共建筑物、工业领域厂房和农业设施等场所，其大多安装在屋顶、阳台、大棚等场合。分布式发电的光伏电源位于用户侧，从而可以就近解决用电问题，降低电力远程传输造成的线路损耗，同时利用并网技术进行电力差额的补偿和外送^[3]。

光伏板的清洁工作在电站运维中至关重要。目前，自然环境下的扬尘等细小颗粒物、工业废气、汽车尾气会随着风力气流和大气扩散等因素附着在光伏板表面并形成一层覆盖的灰尘，进而影响光伏系统的发电能力。通过使用清洁机器人对光伏组件进行周期性的清洗实验可以观察到：良好的天气条件下，该电站的日发电量有可能提升大约 20%。而在天气污染的情况下，通过组件清洁该电站的日发电量可以提高 40%左右^[4]。因此可以看出定期清洗光伏组件对提升发电量和发电功率是十分重要的。目前组件清洁方式也逐渐多样化，比如人工清洁、自动清洁系统^[5]、车载移动式清洁机、光伏清洁机器人等。但随着技术的发展和成本的降低，机器人清洁在大型光伏电站中的应用逐渐增多，而在小型和分布式伏系统中，人工清洁仍然是主要的清洁方式。而人工清洁中传统定期清扫方法依然被大部分光伏电站广泛采用，或根据人为经验进行适当调整，不能根据天气条件、积灰程度以及电站自身发电计划来实时调整清洁方案。清洁周期往往是基于估计或静态设置，而不是基于实际脏污程度变化，这样既没有对组件表面积灰程度进行实时量化分析，又缺乏对电站设备运行状态的监测和预防性清洁。因此，无法最大程度地实现清洁成本和清洁后所提升的发电量收益之间的平衡来实现清洁最大化收益。

在光伏电站运维中，系统的预警和报警都扮演着更为重要的角色。预警更多关注的是潜在的故障或异常情况，采取预防性维护措施，协助运维人员及时发现问题并且迅速处理，从而有效降低设备故障出现的概率。报警是对当前即将或已经发生的重要问题的警示，其强调的是紧急情况下的快速响应，以减少任何潜在的损失或危险。随着光伏电站装机量的逐年增加和规模的不断扩大，导致光伏电站运维成本也不断提升。及时的预警模块能为整个系统起到防患于未然的作用，减少经济效益的损失。常见的预警可以通过组串电流、组串电压、离散率和满发小时数等指标进行及时预警。大部分电站都会使用阈值法进行预警，比如满发小时偏差大于某阈值属于日满发时长偏低，离散率大于某阈值属于组串电流异常等。在如今深度学习快速发展的背景下，产业希望通过光伏组串现有数据进行智能化预测对应指标参数来进行分析预警。

本文基于上述问题深入改进，设计并实现了一个功能齐全的光伏远程运维系统，以便于系统运维人员可以对光伏设备进行实时监测、远程控制、数据分析等操作，并在此基础

上增加具备动态清洁和维护预测的功能。其中动态调整的清洁决策实现清洁收益最大化，光伏板的电流预警又能实现维护预测能力。

1.2 国内外研究现状

1.2.1 光伏电站运维发展现状

目前国内光伏装机规模正在迅速发展，但光伏电站运维水平仍存在发展滞后的问题。光伏电站在技术管理方面存在未定期开展无人机巡检等问题，在节能方面存在未定期开展系统效率、功率衰减检测计算等问题，在发电单元方面存在积灰故障等问题^[6]，导致发电效率的降低和安全事故的发生使得电站质量和安全无法保证，因此更需要从技术层面解决运维的瓶颈问题。

受益于物联网、大数据、云平台等数字技术的迅速发展，光伏电站的运维开始趋于智能化和标准化。物联网技术利用各种传感网络实现不同传感设备之间的互联，使得设备之间具有感知互联、数据上传、远程下发等功能。在这个大背景下出现了一系列智能软件系统和硬件设备，如光伏电站清扫机器人、智能巡检无人机系统、远程抄表等等。这些智能设备和系统能够有效地降低人力成本，提高清洁和巡检效率，为光伏企业带来了新的运维方式和管理手段。同时针对故障预警、故障报警、清洁决策等一系列问题，企业利用数据采集和大数据分析等方法对现有数据进行实时分析预测，从而即时发现设备潜在隐患并且提升发电效率。而云平台和微服务的快速发展提高了光伏运维后台的稳定性。其中各个服务以独立的形式部署，互不影响且可扩展性强从而能够适应不同业务需求，多线程和缓存技术又能提高系统并发能力和响应效率。电站的一系列实时数据基于消息推送技术经过采集、分析、存储，并且结合天气情况对相关参数进行分析，从而实现实时监测电站数据，及时提醒相关人员运维等功能^[7]。

1.2.2 光伏电站运维系统研究现状

由于国外发达国家在光伏产业发展较早，因此对电站运维的关注度较高，20 世纪末便对光伏监控系统开展研究工作。美国国家可再生能源实验室（NREL）针对光伏并网电站展开分析^[8]，并对电站各个设备的实时数据和环境数据进行研究，从而量化评估电站的发电效率和设备损耗。1995 年，国际能源署（IEA）在开展的 PVPS Task-2 研究项目中^[9]，对世界各地不同气候条件下的电站数据进行分析并且提出了多个表征光伏系统工作特性的参数和概念。而全世界最大的光伏电站运维公司 MeteoControl 在光伏运维领域中独占鳌头^[10]，其拥有超过 40 多年的电站维护经验，并且成功研发了适用于不同装机容量需求的战略解决方案及设备产品。该公司研发的运维模式已经在欧洲地区达到较为成熟的阶段，并且已将大数据等技术运用在各种运维产品上。

而国内在光伏运维系统的研究方面起步较晚。在 2010 年，上海淘科公司（TAOKE）成功研发出中国首个光伏监测平台^[11]，为用户提供网页方式远程监控电站运行信息的服务。

用户能够通过登录网址实时查看报表数据、故障报警等信息，很大程度地提高了运维效率。2014 年远景能源推出了“阿波罗光伏云平台”^[12]，其采集的数据除了逆变器检测的数据外，还会对电表、气象站、光伏组件、汇流箱等设备进行实时数据的采集，从而为后期决策分析和运营管理提供数据支持。2018 年 5 月，特变电工提出 TB-eCloud 智慧能源管理平台^[13]。该运维平台能够全方位实时采集电站的海量数据，推出了知识库、远程专家诊断、故障预警、智能处理等功能，从而解决了由于设备数量庞大、信息收集不全面等问题引发的系统决策缺乏理论依据，技术创新迭代缓慢等问题。

但是随着近些年来电站数量和光伏装机量成倍增长，光伏电站运维系统越来越难满足阈值决策、趋势预警、量化分析等深层次需求。因此，在基础远程监控的基础上，光伏电站更要利用时代发展的产物，整合物联网、微服务、深度学习等前沿技术到光伏运维系统中，这也是本文的研究重点。

1.2.3 光伏板清洁决策研究现状

随着光伏产业的快速发展，高效的电站运维至关重要，而积灰导致光伏组件的发电效率受到影响就是运维痛点之一。光伏板上大量的积灰会显著降低光伏组件的发电效率。这是因为积灰会造成多方面的影响，其中包括遮挡阳光从而导致光伏组件表面玻璃的透射率降低，同时积灰也会减慢光伏组件的散热速率，另外积灰严重时更会产生热斑效应等问题。研究发现，当积灰厚度达到 $3\text{mg}/\text{cm}^2$ 时，阳光的透射率将会减少 50%^[14]。并且 Said^[15]等指出，覆盖在组件表面的积灰会很大程度影响光伏发电功率和发电量。因此定期清洁积灰是提高组件发电效率的重要策略之一。目前光伏组件的清洁决策即清洁周期的确定方法一般可以分为两种：样本法和周期法。

采用样本法进行光伏组件或阵列的实验选取特定的一组光伏组串或一部分光伏阵列作为样本。此外需要每日进行清洁以确保光伏板保持整洁且不会积灰影响，从而起到实验对照的作用。张朝辉^[16]等分别将单块组件和整个电站作为实验对象，测试不同积灰程度下组件的发电功率和整个电站的清洁收益，最后结合实际情况预测电站的清洁周期。但是这种方法需要定期指定运维人员抽取样本的发电数据导致运维人员的工作量较大，并不能利用大量历史的基础数据进行分析。

周期法一般选取发电设备运行稳定的光伏方阵，选取发电设备运行稳定的光伏阵列去分析由于积灰损失的发电量和发电功率，当超过一定的清洁成本时发出清洁的推送提醒，从而去确认整个清洁周期。周期法的具体实现又可以分为以下两种：基于计算机视觉的方法和机器学习方法。

基于计算机视觉的深度学习方法一般是利用摄像头识别相关图像进行分析的方法。张姝伟^[17]提出一种基于图像亮度的积灰检测方法，其发现光伏组件表面亮度随着不同程度积灰的变化而变化。文献^[18]根据图像中像素的空间关系提取有效特征，并且根据人工神经网络（Artificial Neural Network, ANN）划分成多个污染等级。文献^[19]提出了一种基于计算

机视觉的光伏表面土壤和粉尘检测方法，为光伏板提供更有效的清洁技术。但是计算机视觉这类方法需要引入摄像头等设备从而会增加投入成本，对普通的分布式光伏电站推广性不强。

而深度学习方法是针对光伏组件现有历史数据和环境数据进行分析。张博^[20]等提出基于最小二乘支持向量回归（Least Squares Support Vector Regression, LSSVR）的清洁时间优化方法，通过构建发电功率积灰衰减模型对不同清洁时间的经济效益评估进而动态确定清洁周期。但是其构建发电功率积灰衰减模型需要额外的监测装置，并且其对峰值小时数的处理仅依靠于辐照度进行计算，没有考虑到其他环境因素的影响。陈金鑫^[21]等以相关环境气象因素为输入变量，采用改进的粒子群优化算法（Particle Swarm Optimization, PSO）对最小二乘向量机的参数进行寻优，并且在考虑降雨的前提下搭建积灰预测模型，从而得到由积灰造成的光伏输出功率下降程度，但其输入变量也需要用实验方式额外测量积灰量。Herinrich^[22]等基于现有数据采用逻辑回归(Logistic Regression)、支持向量机(Support Vector Machine, SVM)、人工神经网络和随机森林(Random Forest)机器学习算法对清洗决策进行判断，减少辐照度和其他因素对构建模型的影响。

综上所述，可见国内外在光伏板清洁决策领域使用了很多方法，但是很多方法都是通过样本法和添加额外装置或传感器考虑积灰因素，这样只会忽略了对现有基础数据的深入分析，反而会增加设备成本。所以现今研究更应注重基础数据的价值挖掘和智能决策的支持，以实现更高效、智能的光伏运维管理。

1.2.4 光伏板故障预警研究现状

故障预警作为整个光伏运维系统的关键部分，也是国内外研究的焦点话题。整个光伏系统中涉及到多个设备的故障异常，比如光伏组件、逆变器、汇流箱等等。本文主要研究光伏组件的异常故障预警。目前比较常用的光伏组件异常故障预警方法很多，姑且可以分为基于物理特性方法和基于数据驱动的方法。

物理特性方法的原理主要是基于光电转换模型利用光伏系统的 I-V 特性分析功率损耗^[23]。功率损耗是通过计算理想功率与实际功率之差来确定的。通过获取理想单二极管 PV 模型的光生电流、暗饱和电流、PV 的串联和分流电阻以及二极管的理想因数等参数来评估理想功率。然而，这些参数的确定是基于光伏阵列中所有组件具有相同特性的假设，而在实际条件下，每个光伏组件具有不同的特性。此外，该方法采用迭代技术，耗时长，可能存在收敛问题，不同初始化的结果可能会有所不同。

基于数据驱动的方法，是指利用光伏电站的实时数据和历史数据，并运用数据统计分析和机器学习的技术，通过建立模型和规则来对光伏组件的异常情况进行预测和分析。光伏组件预警研究更倾向于使用该种方法，通过深入分析一些光伏组件的重要指标来实施预警处理。基于数据驱动的方法一般分为统计方法和机器学习两种方法。

统计方法也可以用来识别光伏系统的状态。在文献^[24]中，该方法采用最小二乘法求得光伏系统的功率、电压等理想参数与实际参数之间的残差。将遮阳和太阳辐照度信息输入到模糊推理系统中，从而预测确定光伏组件的状态。文献^[25]采用 t 检验对光伏系统的测量变量进行显著性检验，并计算理想功率和实际电压的比值。然后将该比率与从训练算法获得的阈值进行比较从而确定异常情况。文献^[26]通过分析光伏系统变量的统计分布，定义了光伏系统正常运行极限的边界，当系统出现新的变量超出定义的边界时，光伏系统被归为异常。

而目前工业界较为流行使用机器学习来进行光伏组件异常故障预警。文献^[27]中提出了一种结合支持向量机（Support Vector Machine, SVM）模型和 SVR 模型的方式进行异常检测，其中回归模型用来估算不同辐照度下的期望发电量并且将其作为输入参数，再使用 SVM 模型充分考虑到预测发电量与实际发电量之间的差异从而识别光伏组件的异常状态。

综上所述，无论是统计方法还是机器学习，大部分方法都是对实际值和预测值作比较从而来判断异常趋势和异常情况来起到预警的作用。不过对于预测的准确性和异常阈值的确认更是需要着重研究和解决的重点。

1.3 论文主要研究内容以及各章节安排

本文针对目前光伏运维系统存在的问题深入研究，并结合智能网关、云平台、微服务等前沿技术，设计并搭建一个功能丰富的分布式光伏运维系统，包含实时采集、下发命令和远程升级，光伏板清洁决策的判断以及光伏板故障预警等多项功能。

以下是主要的研究概述和工作情况：

第一章：绪论。本章首先详细介绍光伏运维平台及其关键技术的研究背景和意义。然后对光伏远程运维系统、光伏电站清洁决策技术和光伏组件故障预警技术的现状进行调查，并根据需求分析确认具体功能模块的实现思路。

第二章：分布式光伏电站的远程运维系统设计。本章首先进行系统总体设计和业务需求分析，详细介绍整个光伏运维系统整体架构，并且对各个组成部分进行设计实现，包含通信协议，智能网关，数据存储，后端微服务架构和应用层等，最后进行效果展示。

第三章：基于发电功率数据挖掘的光伏电站清洁决策方案。本章首先选择电站历史多次清洁周期内最佳清洁效果的功率数据，并且将这些视为不受积灰影响的功率数据。再结合同一时间采集上来的气象数据进行相关性分析得到环境输入量，最后使用改进的 GS 优化 SVR 的关键参数 c 和 g 并建立理想无积灰情况下功率模型，通过每日的实际功率和实际发电量可以计算出发电量损耗值并和清洁成本对比，从而确定电站的动态清洁周期。

第四章：基于组串电流数据挖掘的光伏组件故障预警方案。本章基于 CNN-BiLSTM 预测模型对光伏组件的时序电流数据进行预测，从输入数据进行空间与时间特征的提取，进而提高预测模型的准确性。并且采用 LSTM、CNN-LSTM、CNN-BiLSTM 模型在晴天、

阴天和雨天三种不同的天气类型下进行预测误差的对比。最后引入偏差率指标并结合滑动窗口等方法建立预警规则，实现光伏组件的故障预警。

第五章：总结与展望。本章会对光伏电站运维系统的整体设计和关键结果进行总结。并且分析设计的不足和对相关方案进行展望。

第 2 章 分布式光伏电站的远程运维系统设计

针对上文分析的传统光伏电站所存在的问题，本文设计并实现了一套基于物联网技术的光伏电站远程运维系统。首先将自主研发的智能网关（通信模块）与以光伏逆变器为核心的多个设备端进行实时数据交互，设备端收集到的数据随后通过无线通信技术上传到设备云平台进解析处理，将光伏设备关键参数和环境参数数据传输给 Web 端交互并展示，同时可以通过微信小程序和发送短信等方式向运维人员发起故障报警、故障预警和清洁提醒等信息，进而以便于实现实时监控和远程操作的功能。

2.1 系统总体设计

2.1.1 需求分析

根据当前传统运维系统中存在的缺陷，对本文设计的运维系统，主要需求有以下部分：

（1）实时监测，通过智能网关可以将光伏设备采集到的电流功率等数据以及环境数据上传至云平台分析处理。同时能查询一个电站安装的逆变器数量、逆变器设备的运行状况以及实时上传的预警数据。

（2）系统运维，逆变器检测到的故障和异常信息进行实时报警以及相关人员进行定期发起巡检计划和任务，通过短信或微信小程序提醒管理人员，并且提供设备的报修服务。

（3）远程操作，远程升级和命令下发服务可以让运维人员在后台通过 HTTP 请求的方式远程控制智能网关的版本升级和网关上下线等命令操作。

（4）数据可视化分析，电流、功率、发电量和辐照度等数据可以通过波形图方式展现，历史信息可以从年月日等维度进行图表展示。部署的电站均能在地图上进行标注。

（5）提供多种可交互的产品，通过 Web 端、微信小程序和 APP 等方式让用户能够便捷进行设备数据的实时监测和对各类设备的远程操作。

除上述需求分析之外，后文需要着重研究的有：

（1）清洁决策的判断是电站运维的一大痛点，需要针对数据库保存的历史清洁记录数据结合机器学习和数据分析技术合理根据发电量损耗进行清洁决策。

（2）光伏组件的故障预警也是重难点问题，预警能在真正故障发生前起到预测作用，也需要结合光伏组件的历史电流数据和环境数据对短期电流进行预测并计算实际偏差值，并设立阈值进行预警处理。

2.1.2 总体架构

整个系统整体框架整体可以由设备端、智能网关、云平台、应用端共同组成。物联网体系架构通常由感知层、网络层和应用层构成。其中本系统中的设备端接入智能网关合为

感知层。云平台与网关则依靠网络层进行通信。云平台和应用端合为应用层。云平台从具体业务功能方面划分为设备云和业务云。系统架构图如图 2.1 所示：

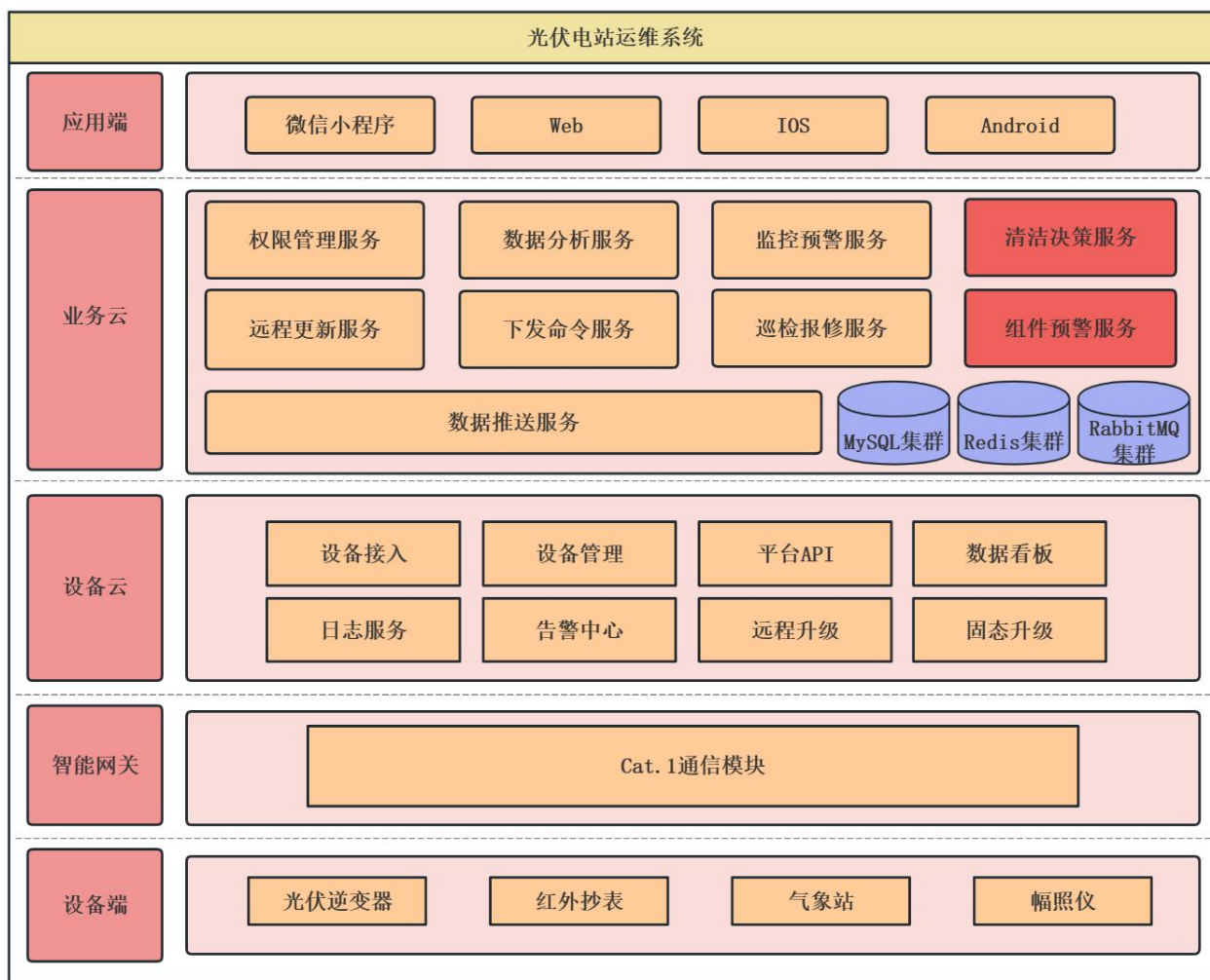


图 2.1 系统框架图

本系统的设备端主要包含光伏逆变器、红外抄表、气象站和辐照仪等。光伏逆变器用来采集实时的光伏板组串电流、光伏发电量、输出功率等重要参数，并且能实时上传报警故障信息。红外抄表用来采集电表有功总功率、电能数据等数据。光伏气象站用来采集实时的天气温度、湿度、气压、风速等环境因素。辐照仪用来采集实时的辐照度。

智能网关是数据传输的桥梁。网关连接设备端的各种设备，其采用 RS485 总线并结合 Modbus 通信协议来实时传输设备端采集到的数据。网关向上对接设备云平台，设备云使用移动公司开发的 OneNET 云平台。两者之间采用 MQTT 通信协议进行数据传输，最后将设备端具体的实时信息保存在设备云平台。同时也能将业务云传过来的下发指令进行解析，解析出设备云上具体的命令发送给下挂的设备端从而达到远程控制的目的。

业务云部分是系统的核心软件模块，它是一个基于微服务技术搭建的后端业务平台。其中最主要的是数据推送服务和数据分析服务。数据推送服务会将设备实时采集的数据放入消息队列中异步处理；数据分析服务中不同的消费者会将对应的数据存储至缓存和数据

库。运维人员可以调用业务云所提供的接口通过设备云对智能网关进行远程升级等命令操作。清洁决策服务通过搭建理想功率模型计算出当日的理想发电量得到当日电量损耗，最后计算总损耗从而判断当日是否清洁。电流预警服务是通过历史电流数据为基础进行预测短期电流的偏差率从而提前进行预警。电流、功率、离散率和满发小时数偏差等实时数据可以在 Web 段、微信小程序和 APP 等应用端通过图表方式进行实时展示，运维人员可以根据网关上传的故障预警信息和具体巡检流程记录等对设备进行远程的实时监控。

2.2 网关设计

光伏电站远程运维系统的主要目标在于实现光伏装备数字化、网络化、在线化的高效运维管理，使设备本身能够良好运转及发挥应有作用。物联网网关是整个光伏系统中的核心部分，网关向下连接逆变器、红外抄表、气象站和辐照仪等设备，向上连接设备云平台，在二者之间充当数据采集和传输的媒介作用，同时又是云端发送控制命令的返回路径。

2.2.1 网关的通信方式与协议

(1) 网关与下挂设备

Modbus-RS485 是传统常用的工业物联网组网方式。其主要是采取 Modbus^[28]的通信协议使用 RS485 总线进行连接通信。RS485 是一中常见的工业总线通信串口，可以支持多点、双向通信，通信距离远且抗干扰能力强^[29]。Modbus 作为一种串行通信协议，由施耐德电气公司于 1979 年为 PLC 通信而研发，其使用简单，易于实现和维护。并支持多种物理层接口，并且适用于各种类型的设备。

Modbus 协议主要通过主从方式的实现通信。系统设计中智能网关归类为主机设备，逆变器、红外抄表、气象站等则被归类为从机设备。本系统使用 Modbus RTU 模式来以不同的格式进行消息传输。Modbus RTU 格式下的帧结构如图 2.2 所示。



图 2.2 Modbus RTU 格式下的帧结构

在上述图中，地址表示下挂设备地址，字长为 1 个字节。功能码表示主机发送的功能指令码，其字长同为 1 个字节。数据码的内容会根据功能码而定，其中包含了从机执行动作或者从设备获取的返回数据，这些数据包括了参考地址或具体数值等信息。校验码中 RTU 采用 CRC16 校验，范围由地址码开始到数据区结束。

(2) 网关与设备云平台

本系统通信协议采用轻量级的 MQTT 协议。MQTT (Message Queuing Telemetry Transport, 消息队列遥测传输协议) [30] 是一种基于发布/订阅模式的通讯协议, 其旨在实现设备之间的低带宽、高延迟或不可靠网络环境下的高效通信。目前该协议在物联网、移动设备等领域广泛应用。在本系统中, 通信模块采集到的数据传输给设备云的过程中, MQTT 连接是建立在 TCP 连接的基础之上的, TCP 提供可靠的数据连接。

MQTT 协议的数据包格式简单规范, 其由固定头部、可变头部和有效载荷三部分构成。固定头部包含了一些控制信息, 包含消息类型和标志位。固定包中的数据包类型决定了数据包是否具有可变头以及可变头的具体信息。有效载荷包含了实际的数据, 其格式也取决于消息类型。MQTT 数据包格式如图 2.3 所示。

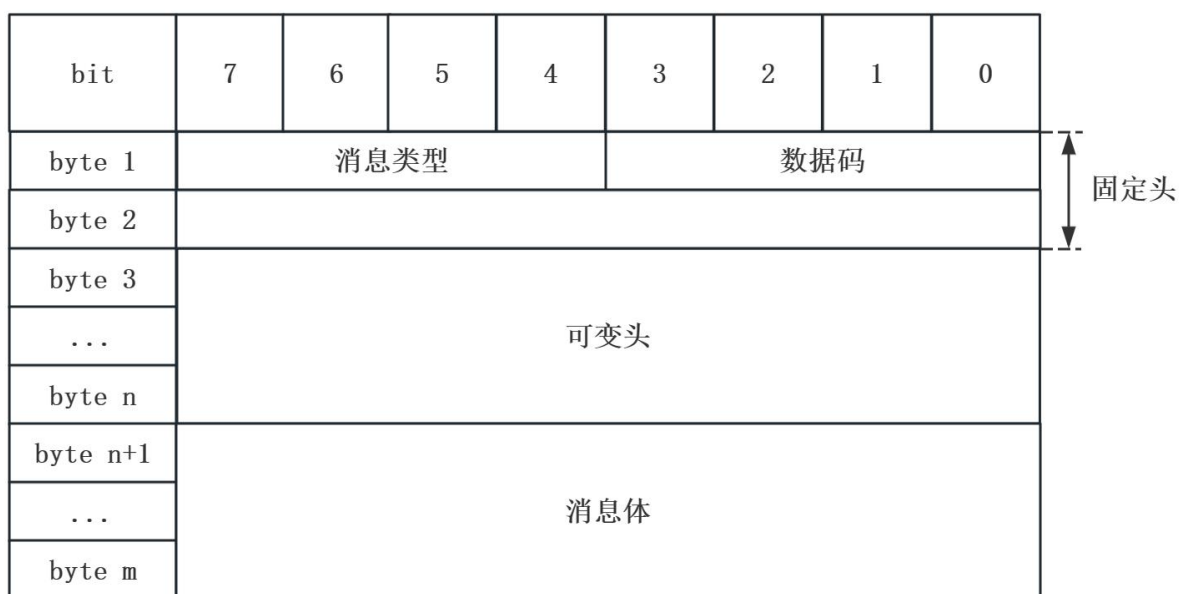


图 2.3 MQTT 数据包格式

2.2.2 网关硬件设计

智能网关硬件部分主要由 RS485 模块, 外部 FLASH 模块, Cat.1 模组, 电源模块, 复位模块等模块构成, 如图 2.4 所示。

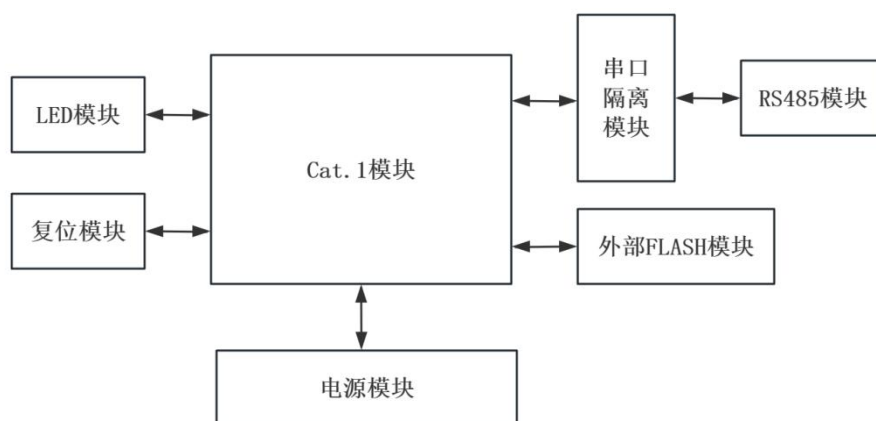


图 2.4 网关硬件设计框图

(1) 外部 FLASH 模块：本系统功能复杂，代码量已超过单片机自带 ROM 存储容量，因此通过该模块增加程序存储量。并且 FLASH 模块会在设备离线情况下进行本地存储。

(2) RS485 模块：该模块主要与光伏逆变器、红外抄表设备、气象站和辐照仪等设备相连，模块主要起到数据采集后的多点通信和长距离传输的作用。

(3) 电源模块：具有防反接功能，当接入电源正负极相反时，装置可自动保护或不予以反应。电源供电采用直流 12-24V 宽幅输入电源，长期工作范围为直流 $24V \pm 10\%$ 电压。

(4) 复位模块：该模块具体使用 SP706 芯片，其主要用于在系统出现异常时重置微处理器，有助于系统从错误状态中恢复，并确保其继续正常运行。

(5) LED 指示灯模块：模块中主要由三个 LED 指示灯构成。第一个是系统指示灯，亮红灯。当设备接通电源时就会长亮。第二个是网络指示灯，亮绿灯。刚接入电源时候会长亮 3 秒后熄灭，然后具体根据闪烁频率动作得到具体信息：闪烁间隔 200ms 表示检测不到 SIM 卡；闪烁间隔 500ms 表示检测到 SIM 卡没有接入到网络；闪烁间隔 1000ms 表示接入网络成功，未连接平台；长亮表示联网成功，可以与云平台建立通信。第三个是通信指示灯，亮黄灯。快速闪烁说明正在与下挂设备进行通信。

(6) Cat.1 通信模块：该模块具体使用 Air724UG 芯片模块，Cat.1 的全称是 LTEUE-Category1。Cat.1 技术能提供更高的数据传输速率，并且能通过具有更低成本^[31]和功耗的 IoT 设备连接到 LTE 网络。由于基站无需升级网络和硬件，其覆盖范围广并且提高了数据传输的可靠性和稳定性。以下总结对比了如表 2.1 所示常用三种无线通信方式。

表 2.1 常用物联网无线网络通信方式比较

参数	LoRaWAN	NB-IoT	Cat.1
数据速率	0.3-50kbps	<100kbps	5.2Mbps-10.3Mbps
覆盖范围	市区 2km-5km，郊区 15km	市区 1km，郊区 10km	范围较小
传输距离	50-100m	>10km	>5km
网络延时	1s-2s	1.4s-10s	<100ms
功耗	一般	高	低
使用场景	适合大规模部署	适合大规模部署	适合小规模部署

虽然 Cat.1 的覆盖范围可能相对于 LoRaWAN 和 NB-IoT 较小，但本系统更注重低延迟和低功耗的特性，Cat.1 的数据传输速度远远快于另外两种。而且 Cat.1 和同系列的 Cat.4 都是 LTE 制式的只不过等级不同，并且成本较低，所以最后考虑 Cat.1 作为通信模块更合适。

2.2.3 网关软件设计

网关软件部分主要由数据上报、命令下发、远程升级三大模块构成。其网关与云平台的数据处理流程如图 2.5 所示。

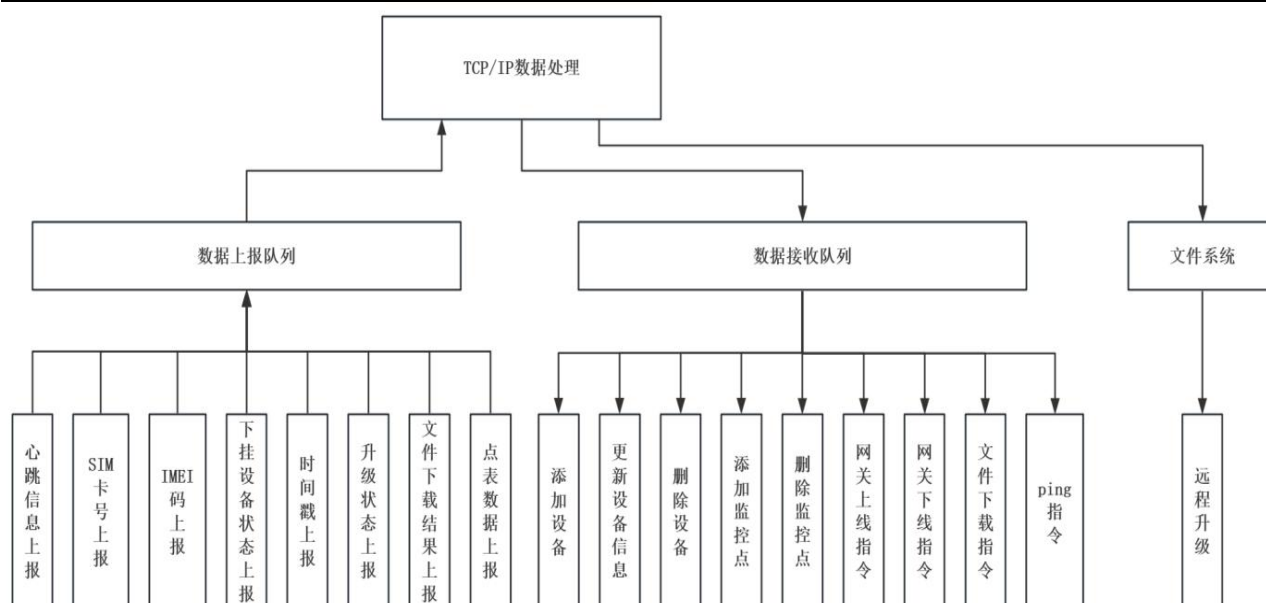


图 2.5 网关与云平台的数据处理流程设计图

下发命令：下发命令是后台运维人员将数据封装成 JSON 格式并发送远程请求给云平台，云平台再通过 MQTT 协议将对应的指令序列化发送给网关设备，当网关设备接收到后便将结果根据 MQTT 协议反序列化解析，从而达到对通信板进行相关指令操作的目的。相关指令具体包含控制设备运行，网关注册、校时功能、添加监控点、修改设备信息、掉电检测等等。

实时上传：本系统中涉及到的逆变器、红外抄表、气象站等设备将监控点数据通过 RS485 总线传输至智能网关，智能网关会把获取的具体数据按一定分类顺序存放在点位标准记录表中，并在 MCU 中进行监控点遍历操作。具体而言，通信板 MCU 通过全局事件轮询的机制在特定时间对点表信息中下挂的各个设备的监控点进行遍历操作，从而实现设备监控数据的实时获取。如图 2.6 所示，点表信息文件中按照设备地址和监控点地址进行了分类，方便后续的数据处理和管理。随后又实时把监控数据、故障数据和异常数据等信息通过 Cat.1 模块上传至设备云平台。通信模块上报的数据信息包含 SIM 卡号，设备唯一标识 IMEI 编码，设备版本号等基本信息。这些信息首先也像上文提到的进行 JSON 格式的封装，再在 TCP 连接的基础之上按照 MQTT 协议需要的传输格式进行在封装，由此将实时信息传输到设备云。

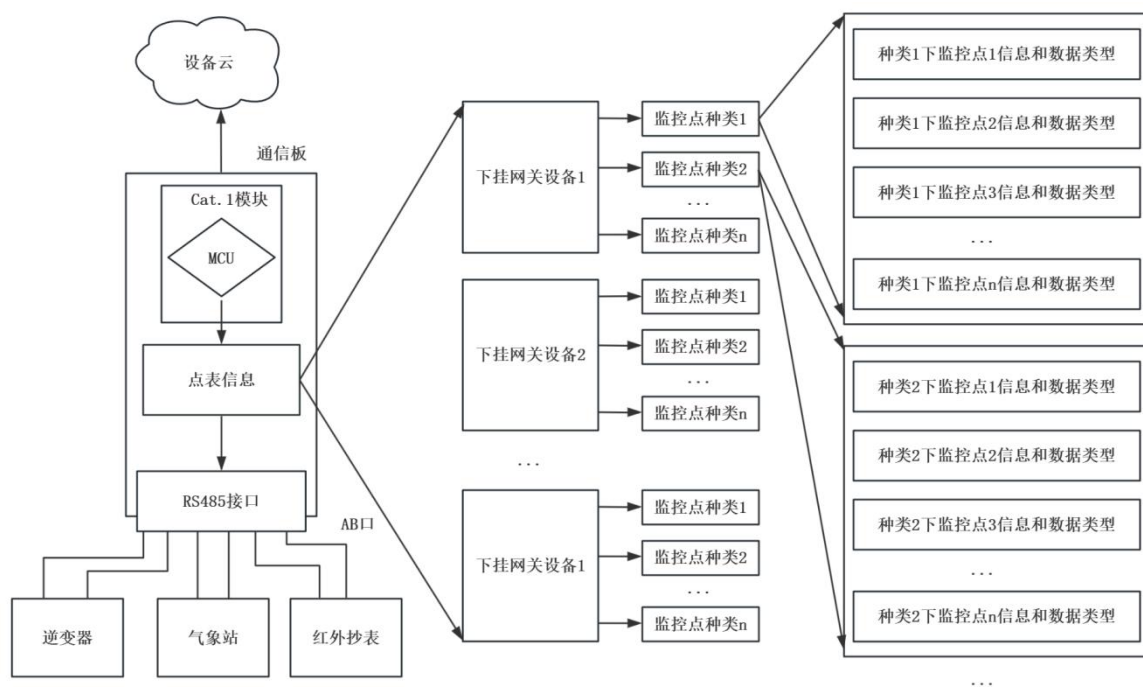


图 2.6 数据上报示意图

远程升级：该模块主要分为嵌入式部分和后端部分，智能网关采用轻量级 TCP/IP 库 LWIP 来发起 HTTP 请求。网关通过 HTTP 协议与云服务器进行连接并通过服务器后台提供的文件下载接口进行远程下载。当用户开始执行远程升级任务时，可以通过业务云平台指定升级版本并且通过后端对应接口对文件进行 JSON 序列化传输给 OneNET 平台，平台通过 MQTT 协议向智能网关传输实时信息，在通信板获得升级文件的信息后，MCU 会实时下载对应文件，并采用直接内存访问的途径将下载的文件存储到 FLASH 中去。下载完成后以文件的形式存放到 FLASH 中并且修改 FLASH 的更新标志位。当系统再次开机时，可以通过判断该 FLASH 标志位是否改变来决定是否已经将之间的旧版本代码进行覆盖，如果已覆盖的话说明完成远程升级。

2.3 后端架构

2.3.1 后端组件选型

（1）数据库

一个可靠稳定的运维后台通常需要一个具备高可用、高性能、高扩展的持久层。由于光伏设备需要实时监控和响应，数据库更需具备高可用性从而确保系统的持续可用。海量的历史报表数据统计需要数据库具备高性能的处理能力，从而能够快速存储和检索。对应的光伏设备数量和数据量的增长也是不可避免的趋势。

综合而言，本系统采用多节点部署数据库集群并根据数据类型的特点选择合适的存储方式，从而可以满足光伏电站平台对于高可靠性、高可用性和性能优越性的需求，为大型软件应用提供坚实的持久层基础。这样的架构不仅满足当前需求，还具备灵活适应未来扩展和变化的能力。

目前市场上主流的关系型数据库中有 MySQL、Oracle、SQL Server 和 PostgreSQL。本文对数据库的几个主要特点进行比较,从而发现尽管 SQL Server 和 Oracle 在稳定性和并发性上表现出色,但是两者使用成本较高并且维护难度较大,鉴于 PostgreSQL 在 Windows 环境下稳定性较高,但在非 Windows 环境中可能有一些限制。综合考虑最终确定以 MySQL 作为关系型数据库。表 2.2 对四种主流关系型数据库的特点进行比较。

表 2.2 主流关系型数据库比较

数据库	稳定性	并发性	维护难度	成本	跨平台性
MySQL	高	高	低	低	高
Oracle	高	最高	高	高	高
SQL Server	高	高	低	高	低
PostgreSQL	较高	高	低	低	高

MySQL 功能强大,首先 MySQL 提供了 SQL (Structured Query Language, 结构化查询语言) 语言的支持,从而实现数据库的创建、修改、查询、删除等操作。并且其支持事务、视图、存储过程和触发器等。事务可以确保数据库在处理大量并发操作时保持数据的一致性和可靠性。其次 MySQL 通过各种索引大大提高检索效率,通过各种锁管理并发访问,提高系统的性能和数据的一致性^[32]。

本系统采用了 MHA (Master High Availability) ^[33] 的 MySQL 高可用方案。它由 MHA Manager 和 MHA Node 两部分构成。每一个 MHA Node 都运行在一个 MySQL 节点上, MHA Manager 会定期检查集群服务器中的所有 master 节点。当 master 节点发生故障后,它就会主动地把所有具有最新数据的 slave 节点更新到 master 节点,最后再把剩余所有的 slave 节点都指向最新的 master 节点。图 2.7 展现了 MHA-MySQL 分布式集群主要架构。

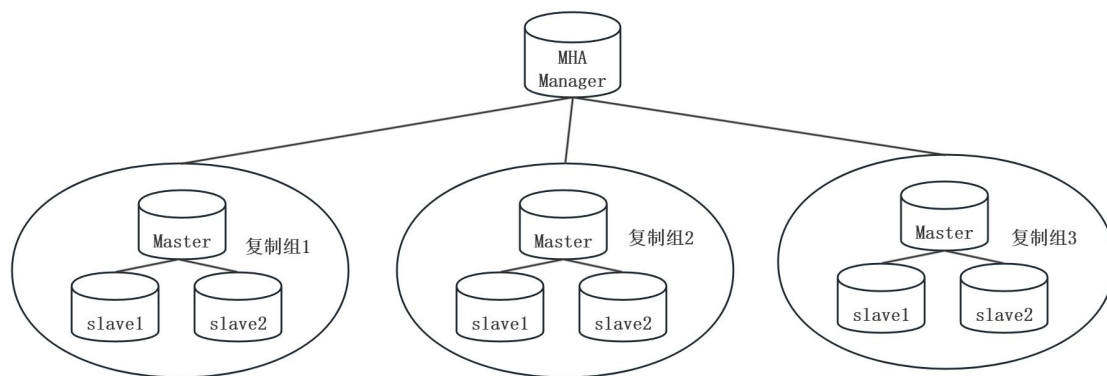


图 2.7 MHA-MySQL 分布式集群架构图

(2) 消息队列

当设备云接收到网关设备实时推送上来的信息后,需要将对应的监控点信息请求给后端进行解析处理。然而大量数据直接进行同步请求可能会存在丢失数据的情况。因此可以通过引入消息队列,将数据的生产者(网关设备推送信息)和消费者(后端系统处理请求)解耦,从而避免直接的同步请求以确保数据的可靠传输和处理。

常见的消息队列包括 RabbitMQ、Apache Kafka、RocketMQ 等。这些消息队列系统提供了可靠的消息传递机制和高效的异步通信能力，能够在分布式系统中进行消息的生产、消费和传递，以实现解耦、削峰填谷、异步通信、消息缓冲等功能。下表将所提到的三种消息队列从吞吐量、稳定性等方面进行对比，可以看出虽然 Kafka 可以实现高吞吐量的消息传递，但是相对于其它 MQ 会存在一定时延。RocketMQ 是阿里巴巴开发的消息中间件，但是相对于其它 MQ 社区支持相对较弱。RabbitMQ 虽然吞吐量低但是能满足本系统下的场景应用需求，并且其通过消息确认机制和持久化可以保证实时推送消息的可靠传递^[34]。表 2.3 对三种主流消息队列的特点进行比较。

表 2.3 主流消息队列比较

消息队列	吞吐量	稳定性	顺序消费	消息重试	死信队列	社区支持
RabbitMQ	1w	高	支持	支持	支持	高
Kafka	10w	高	支持	不支持	不支持	高
PostgreSQL	10w	高	支持	支持	支持	一般

本系统采用 RabbitMQ 的镜像集群模式部署，普通集群模式不能实现系统的高可用，盘节点崩溃可能会导致数据丢失。而镜像集群具有设定数据同步的功能，其既可以选择同步至全部节点，也可以选择同步至一定数量的特定节点，从而确保系统的高可用性。

（3）缓存

光伏远程运维系统可能需要频繁地读取和计算各种监控数据、历史数据、设备状态等信息，需要使用缓存可以加速数据的读取和响应速度从而降低数据库的访问压力。并且巡检模块的定时任务是多机部署，需要分布式锁解决系统中的多进程并发访问资源问题。本系统针对缓存中间件采用 Redis（Remote Dictionary Server）数据库，其为了更快速的读写访问则将所有数据存储在内存中。并且支持多种存储数据类型，例如字符串、链表、集合和哈希等类型。同时为了确保数据的可靠性和安全性，Redis 也提供了持久化机制。

Redis 主要提供了哨兵集群和分片集群两种高可用解决方案，本系统使用哨兵模式来确保系统的可靠性和高可用性。哨兵模式是在主从模式的基础上增加了哨兵进程来实现高可用性。如图 2.8 所示，每个哨兵都是一个独立的节点进程，它能够同时监控多个 Redis 服务器的运行状态。哨兵节点通过心跳机制来检测服务状态，同时每秒发送 ping 命令给集群中的各个实例。如果某个实例在预定时间内未能正常响应，哨兵则认为该实例主观下线。这种情况下此时发现的哨兵会作为 leader 哨兵对其他哨兵发起投票，如果剩余哨兵中多达半数的哨兵认为该实例已经宕机，就需要挑选一个新的节点取而代之。然后会依次对所有从节点的优先级、复制程度和节点 ID 进行三轮考察，挑选出新的主节点并告知其他节点。然后哨兵 leader 会将原主节点下的各个从节点指向现在新的主节点并且告知客户端主从节点的变化信息。最后监视旧主节点当旧主节点醒来后变成变成从节点。哨兵节点间是通过 Redis 的发布/订阅机制发现并相互通信的，并且通过 Raft 一致性算法保证故障时节点之间的和一致性选举，从而可以起到高可用的目的^[35]。

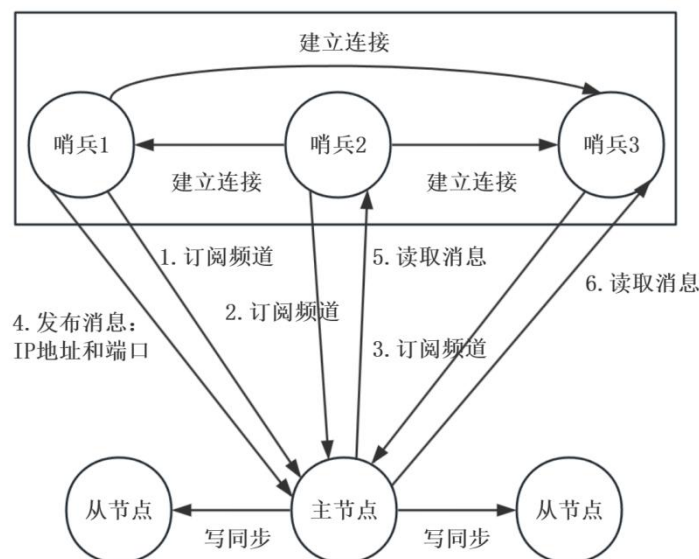


图 2.8 哨兵集群架构图

(4) 注册中心

对于服务注册与发现中心选取，常见的如下表有 Zookeeper、Nacos 和 Eureka。虽然 Eureka 是 Spring Cloud 推荐的服务注册中心组件，但 Eureka 2.0 的开源工作已经停止故不去考虑。Nacos 是由阿里巴巴开源的一款服务注册与发现组件，它支持动态服务发现、服务元数据和配置管理^[36]。Nacos 支持 AP（Availability&Partition tolerance）和 CP（Consistency&Partition tolerance）两种一致性模式，确保在不同的场景下提供合适的服务注册与发现能力。它具备高可用性、更优的性能和较强的伸缩性，适用于构建更加灵活、可扩展的服务架构。而 Zookeeper 集群仅仅支持 CP 一种一致性模式，其可能出现在 leader 节点宕机的时候进行 leader 的选举并且选举期间服务集群瘫痪的问题，所以不能保证服务的可用性。在光伏物联网场景中突发事件或设备故障可能对能源产出造成重大影响。高可用性的系统能够更快地响应应急情况，实现快速故障恢复。本文采用 Nacos 作为服务发现与注册中心。表 2.2 对三种主流注册中心进行比较。

表 2.4 主流注册中心比较

注册中心	语言支持	一致性协议	多数据中心	K8S 集成	社区活跃度
Nacos	JAVA	CP/AP	支持	支持	高
Eureka	JAVA	AP	支持	不支持	低
Zookeeper	C	CP	不支持	不支持	中

Docker 是一个使用 Go 语言编写的开源应用容器引擎，其提供的容器化部署方式以其可移植性、轻量化等优势，能够快速把应用程序的依赖包封装到独立的容器上。其跨平台性能够在不同操作系统上开发和测试容器化的应用。Docker 的沙箱机制实现了容器间资源的隔离，确保各个容器间可以相互独立运行，有效避免了应用间的资源和依赖冲突问题。本系统中提到的 MySQL、Redis、RabbitMQ、Nacos 等中间件服务的镜像都可以通过

Docker 镜像仓库轻松拉取，并且部署到阿里云 ECS 服务器上，从而更好地利用服务器资源，提高硬件利用率，并且能进行横向扩展并处理更多的请求负载。其提供了丰富的管理工具和容器镜像服务，使得应用程序的部署和迁移简便化。最后，通过编写 DockerFile 脚本进一步提高了业务模块的迭代和部署效率。

2.3.2 云平台

2.3.2.1 设备云

本系统采用中国移动公司提供的 OneNET 云平台来对光伏设备的数据进行存储和管理。OneNET 云平台从物联网产业特点出发，旨在简化物联网项目的实施过程，通过提供安全可靠的数据传输存储和多维度的数据分析工具，确保物联网应用的稳定性和可扩展性，从而实现设备的快速接入、数据的实时收集处理以及应用快速开发和上线^[37]。

OneNET 提供了丰富的 API 和 SDK，其中包含设备增删改查、数据点上传、设备命令管理，规则引擎创建等，支持多种行业以及主流标准协议的设备接入，例如 LWM2M、Modbus、MQTT、HTTP 等。同时 OneNET 采用分布式结构和异地双活等多重数据保障机制，从而能为光伏电站设备提供安全的数据存储服务。

本系统中智能网关下挂多台设备，其将采集到的数据通过 MQTT 协议上传到 OneNET 云平台，设备云平台接收数据后对数据流进行进制解析等操作，而后通过全局接口将数据批量推送至业务云平台，光伏运维人员可以通过平台所提供的各种接口完成实时监测、全程管理等操作。设备云架构图如图 2.9 所示。

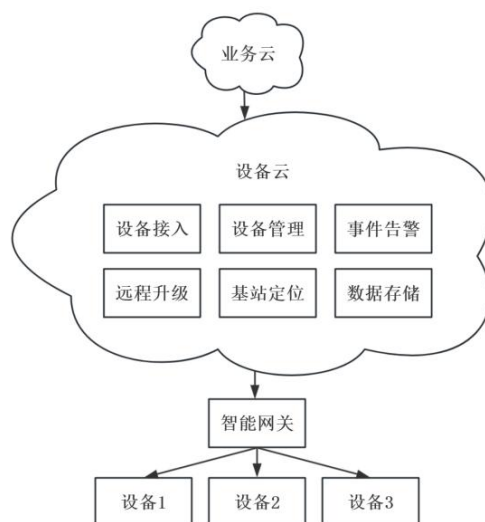


图 2.9 设备云架构图

2.3.2.2 业务云

业务云主要由负载均衡、云服务器、业务组件三方面构成。

首先系统采用 Nginx 负载均衡，当用户请求后端接口时会通过负载均衡进行分发请求起到分流的作用。Nginx 服务器具有高性能、可扩展性强、轻量级、稳定可靠和灵活配置

等优势^[38]。云服务器使用阿里 ECS 云服务器，业务端在其上运行 Tomcat 服务，通过 SpringBoot 和 SpringCloud 框架来实现微服务。SpringBoot^[39]具有简化配置、内嵌式容器、自动化构建和快速启动等优点，极大地提高了 Java 应用程序的开发效率，使开发人员能够更专注于业务逻辑。而 SpringCloud 利用 SpringBoot 的开发便利性简化系统的开发步骤，其专门用微服务开发而设计，提供了服务发现、客户端负载均衡、断路器、服务路由等服务。业务组件中核心部分包含数据库、缓存、消息队列等。为了处理访问量激增对 MySQL 数据库造成的过大压力，业务模块采用 Redis 缓存的方式解决数据库访问速度过慢的问题。消息推送模块则通过 RabbitMQ 进行异步处理来实现设备数据上传时的流量削峰。三种组件均通过集群方式部署来确保系统的稳定性和安全性。整个业务云的架构如图 2.10 所示。

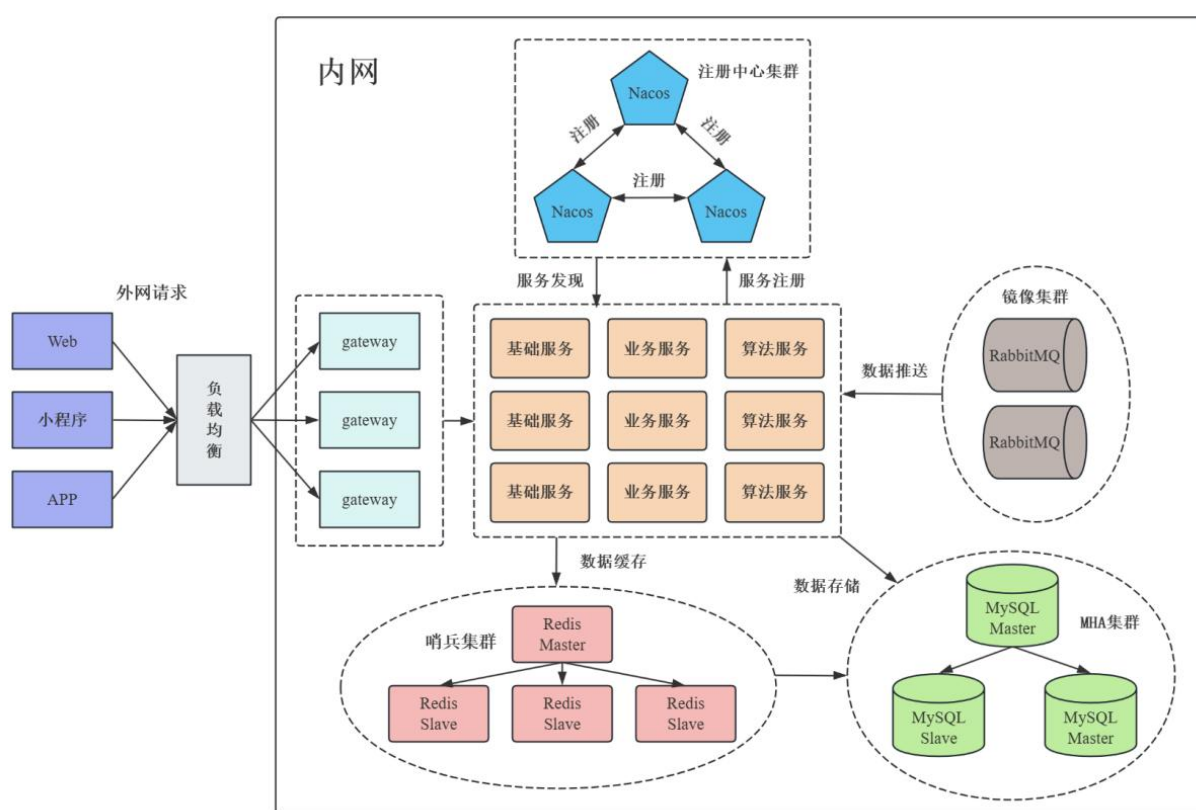


图 2.10 业务云架构图

2.4 服务设计

2.4.1 基础服务设计

2.4.1.1 实时推送服务

实时推送服务是整个数据来源的核心，主要是处理设备云推送过来的大量实时信息。为了防止数据丢失并且减小系统的耦合性从而提升性能，经上文对比本系统采用 RabbitMQ 消息队列进行消息异步处理。

RabbitMQ 主要由生产者、消费者、队列、交换机等几部分构成。首先生产者在发送消息的同时需要带上交换机名称。当交换机接收到信息时，其会根据路由规则针对指定的路由键进行消息转发。最终消息会发送到匹配的消息队列，订阅了该队列的消费者会通过监听器对队列中的消息进行实时消费。

智能网关在实际使用时通过连接多个逆变器，网关监控由逆变器 SN（即逆变器唯一标识）加上对应监控点（监控点约定为 6 个字符）构成，从而区分不同逆变器同一个监控点的数据。RabbitMQ 为了保证数据的持久性，可以设置交换机、队列和消费者信息的持久化机制，开启持久化后默认内存中存储消息确保在 MQ 中消息不丢失。本系统在 RabbitMQ 的基础上进行了改进。首先实时数据传输至设备云平台后会定时请求业务云的后端接口，异步线程池的多个空余线程将接收到的 JSON 格式数据并行发送到消息队列的生产者。生产者处会先将上报上来的逆变器唯一标识拼接上同一时间的戳为键值存入到 Redis 里再发送，从而方便发送失败时重新发送。此时消息队列在确认到达交换机和到达队列处均有自身的 ACK 应答机制。

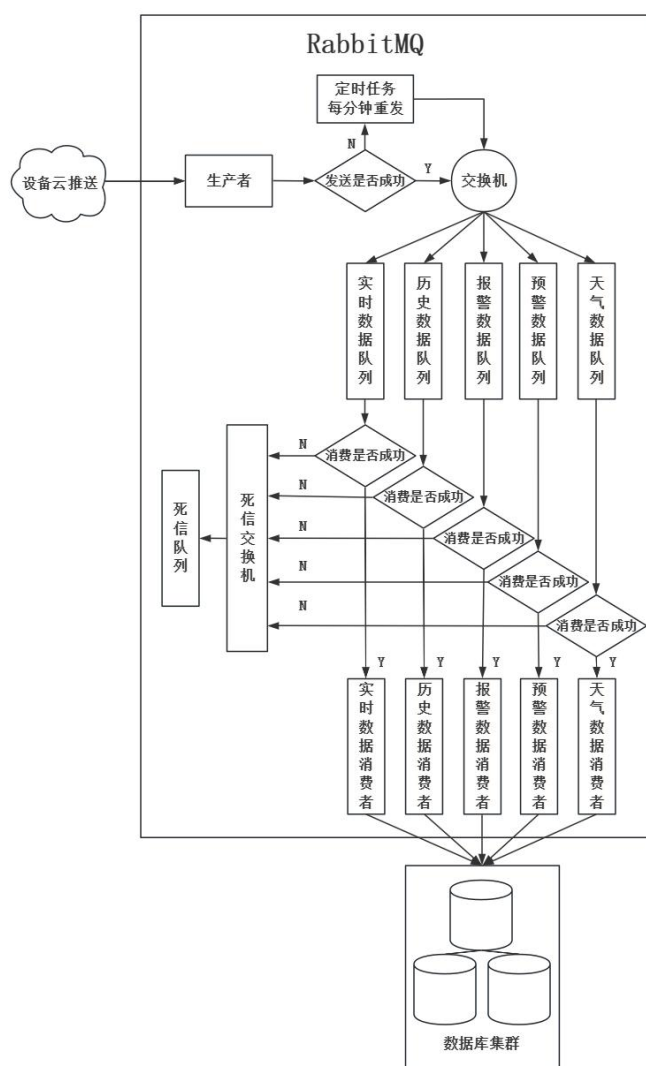


图 2.11 实时推送框架图

本次设计在消息成功到达交换机时会把之前生产者存入的键值从缓存中移除，否则就需要定时任务重发失败消息。而如果消息从交换机到队列失败，则重新把消息存入缓存并且需要定时任务重发失败消息。超时重复机制主要是应对在数据发送时所遇到延迟中断、网络堵塞等问题，导致服务器无法及时地向客户端反馈信息。为了避免客户端长时间地等待响应，本系统采用了超时重发机制来保证消息能够顺利到达服务器。系统利用 RabbitMQ 的默认三次重发机制，当重发次数达到三次仍未收到回复时，即认定设备已经与服务器断开连接，便停止重发。此外，系统还设置了定时任务，每分钟对发送失败的消息进行重发，具体流程如图 2.11 所示。

2.4.1.2 远程升级及优化

远程升级服务主要针对智能网关设备里的通信板进行程序版本升级。该服务的实现同下发命令针对设备云为基础进行操作。远程升级的基本操作流程如图 2.12 所示，

不同版本的网关设备有不同的更新包，首先系统会通过定时任务获取设备云上的最新设备版本信息并且存储至数据库。当发起远程升级任务时，先针对平台不存在的待升级设备和升级包进行选择和添加，升级包以二进制文件的方式存储至数据库。然后查询出与远程升级任务相关的所有设备，挑选指定的升级包进行批量升级。最后确认更新后，远程升级任务会以异步的形式在后台启动。主任务开启后会将数据库的任务状态更改成进行中，并且在数据中添加对应的任务记录，设置对应的创建任务时间和任务 ID。等待最后主任务结束后，数据库的任务状态修改成为结束状态。

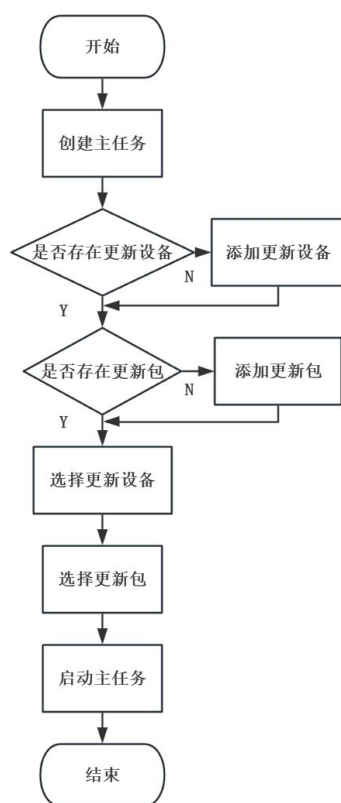


图 2.12 远程升级基本流程图

本次远程升级在原有主任务的基础上进行优化。由于本系统采用的设备云平台中提供的命令下发接口存在一定限制——对下挂设备进行升级时有最多仅支持 50 台设备。而考虑到逆变器设备数量的逐渐增加，单台设备的升级处理需要频繁调用大量的 HTTP 请求。为了达到更高的效率，系统需要通过并发编程的方式对程序进行优化。

本次设计采用基于分支合并思想的 Fork-Join 框架进行并行执行任务，其主要思想是通过递归分割的方式将大任务划分成若干个小任务分别进行计算，最后将小任务的计算结果合并汇总得到最终的计算结果。该框架主要采用工作窃取算法实现任务的调度和执行，其基本流程图如图 2.13 所示。其核心思想是所在队列中的线程可以向其他队列进行窃取任务操作，从而减少线程因为任务空闲而造成的资源浪费。此外，每个子任务均有其单独的处理队列。如果有的线程先完成了自己队列里的任务，就窃取其他队列中的任务进行执行。其优点在于其能够自动化地实现任务分割和调度，提高了任务的并行处理能力，从而使得程序在多核处理器上具有更好的性能表现，更能达到毫秒级统计的效果。

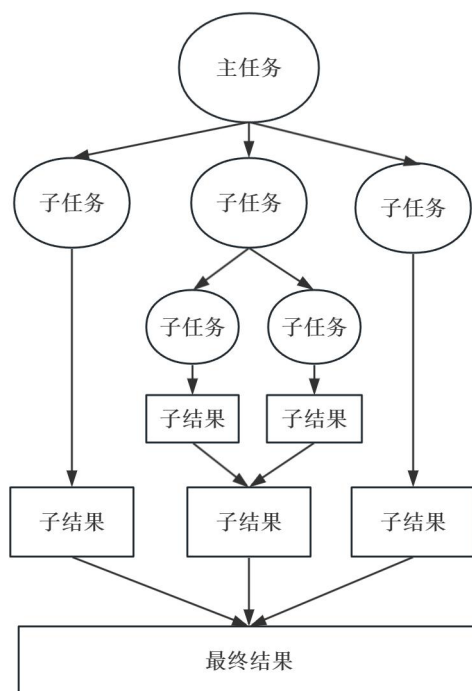


图 2.13 Fork-Join 基本框架图

2.4.2 业务服务设计

本小节将主要通过接口需求设计、数据库设计和逻辑设计几个方面对光伏远程运维系统的后端服务进行设计。

2.4.2.1 接口需求设计

系统接口需求设计主要是对业务逻辑进行拆分，需要按照电站设备、人员管理、数据分析等角度出发考虑系统的完整性，并且涉及到系统内部各个模块间的通信和数据交换。如表 2.4 所示按照需求划分成以下模块接口。

表 2.5 光伏远程运维系统及其功能

模块名称	模块功能	模块 URL
逆变器模块	逆变器相关信息和操作	/inverter
电站模块	电站相关信息和操作	/station
网关模块	网关设备相关操作	/gateway
人员管理模块	运维人员的操作权限管理	/manage
摄像头模块	提供实时视频监控功能	/camera
巡检模块	定期巡检操作	/camera
清洁模块	根据日损失发电量进行清洁决策	/clean
报警模块	对逆变器检测到的故障报警	/alarm
预警模块	对组串电流和满发小时数异常进行预警	/warn
报表模块	生成逆变器、电站、气象站历史记录报表	/report
报修模块	对指定故障进行报修	/repair

整个业务系统分为六大管理模块，分别是综合管理、报表管理、设备监控、定期巡检、清洁决策、预警报警。

(1) 综合管理模块：该模块是对电站信息、逆变器信息、网关信息以及运维人员信息等相关信息进行数据存储。其中包含了角色的权限管理和人员档案管理，并提供账号的相关注册。由于存在多个电站所以业主为指定的人员分配可使用管理的电站。电站信息和逆变器信息的相关管理在新增录入时需要建立。

(2) 报表管理模块：该模块是对网关设备上传上来的数据进行历史记录存放和数据分析。主要分为逆变器、电站、气象站报表，逆变器表报具体记录离散率、满发小时偏差、三相电流、组串电流和输出功率等信息。电站报表具体记录。气象站报表具体记录辐照度、温度、湿度、风向等环境信息。报表可分别根据年月日的层级和列表图标的层级进行展示，并且可以对报表进行导出 Excel 文件。

(3) 设备监控模块：该模块可以对设备的运行状态起到实时监测的作用。网关设备对光伏现场的逆变器电流、功率、发电量等数据的进行实时采集和上报。同时运维人员可以通过 APP、小程序、Web 端等向云平台进行 HTTP 请求，进而实现对光伏逆变器等下挂设备的远程控制，例如逆变器电流电压参数进行设置、网关自身的上下线和版本升级等。另外设备监控支持在线的视频监控，通过嵌入了萤石云 SDK 的方式便可用移动流量和 WIFI 实时上传和展示光伏电站的高清视频摄像。

(4) 定期巡检模块：运维人员可以为指定电站生成巡检计划，计划根据循环创建一次性或者周期性的巡检任务，计划可以指定巡检哪些组件、哪些检查项，设置任务的循环周期、第一次执行时间、任务结束时间、任务提醒时间、责任人、联系电话、额外提醒人等信息。后端定时任务会在任务提醒时间前调用短信和微信公众号接口进行推送，提醒责任人和额外提醒人开启巡检任务，最后根据巡检人对各项指标评价生成对应的巡检记录。

(5) 清洁决策模块：该模块每日会计算各个区域的理想无积灰功率，根据采集上来的实际功率和实际发电量计算出日损耗发电量，并且根据清洁成本判断电站是否需要清洁。

(6) 预警报警模块：报警和预警模块区别于网关定时上传信息，一有异常和故障信息立即上报。预警模块大致分为两类预警：日满发时长偏低和组串电流异常。报警模块根据逆变器厂家提供的具体故障说明上报指定故障信息。

在设备云平台中，用户还可以对数据流进行监控数据点和全局阈值的设定，当某个数据点超过预设阈值时，设备云平台会自动启动全局推送并向业务云平台进行数据传输。运维人员能调用后端提供的 API 进行短信分发工作，同时可以通过 WebSocket 将预警和报警接口与 Web 前端页面实现链接，从而达到实时监测异常的效果。报修模块可以让厂家根据对应的故障信息提交报修订单指派运维人员进行维保工作，提交后的报修申请可以在报修进度中查看具体进度情况。

2.4.2.2 数据库设计

上文提到本系统 MySQL 数据库采用 MHA 集群方式部署。具体地，本系统主要由三个数据库集群构成，分别是基础数据库集群、业务数据库集群和算法数据库集群。基础数据库集群存放有关系统和用户的基本信息。业务数据集群是整个光伏单站运维系统的核心，存放了众多模块的业务信息并且对历史报表信息进行分库分表存储。算法数据库集群用于存储需要预测的服务的相关数据。以下具体说明各个数据库集群的具体设计。

基础数据库集群存放有关系统和用户的基本信息，例如日志记录，用户信息、对应的角色权限分配、远程升级的配置信息和单点登录配置信息等等静态参数设置，相对于具体业务的耦合性不高。核心表及其功能如表 2.6 所示。

表 2.6 基础数据库核心表及其表功能

表名称	表功能
base_api	API 网关路由表
base_sms	设置短信服务
base_oss	设置单点登录
base_role	设置不同的身份权限
base_menu	设置菜单和按键关系
base_role_menu	设置不同身份权限的菜单分配
base_user	管理所有平台已注册的用户
base_dept	管理运维平台下用户所有的部门
base_user_dept	设置对应部门下的用户
base_datasource	数据源配置表
base_log_api	接口日志表
base_log_error	错误日志表

业务系统的存储承担着核心的功能。业务数据库集群不但存放有关光伏电站、逆变器设备、智能网关、监控点等基本信息以及其他模块基本数据存储，还需要对海量的历史报表信息进行分库分表。本系统以智能网关设备为维度进行建表。每次增加一个网关就动态地新建一张表。每个数据库存储一定量的表，达到上限的话就动态新增数据库。核心表及其功能如表 2.7 所示。

表 2.7 业务数据库核心表及其表功能

表名称	表功能
tb_gateway	网关信息表
tb_mapping	监控点配置表
tb_station	电站信息表
tb_inverter	逆变器信息表
tb_repair	设备报修表
tb_camera	设置不同身份权限的菜单分配
tb_inspect_item	巡检项目信息表
tb_inspect_plan	设置具体巡检计划
tb_inspect_task	设置对应部门下的用户
tb_inspect_record	数据源配置表
tb_alarm	接口日志表
tb_warn	错误日志表
tl_inverter_x	逆变器报表，x 代表具体逆变器设备 imei
tl_weather_x	天气报表，x 代表具体逆变器设备 imei

算法分析服务数据库集群主要存储理想清洁时间段的环境数据、辐照度数据以及预警需要分析组串电流数据，这些数据主要针对后两章理想清洁功率预测和组串电流预测。为了避免与基础数据和业务数据发生数据耦合，因此独立出来存储。核心表及其功能如表 2.8 所示。

表 2.8 算法数据库核心表及其表功能

表名称	表功能
algo_current_weather	存储最新的天气数据
algo_power	存储每次清洁周期后的实际输出功率数据
algo_current	存储组串电流数据
algo_result_power	存储算法预测的理想清洁功率数据
algo_result_current	存储算法预测的组串电流数据

2.5 应用层设计

2.5.1 Web 前端设计

光伏运维系统的监控使用主要集中在 Web 前端。现如今 Web 前端技术的发展已经达到了相当成熟的阶段，拥有多样化和丰富的生态开发环境。其中，React、Vue 等技术是最受欢迎的框架，同时第三方库的种类也相当多样，功能也相当全面^[40]。本系统使用了轻量级的 Vue 框架，其由 CSS、HTML 和 JavaScript 等组成，在物联网场景下具有更高灵活性和扩展性。Vue 框架中提供的 npm 开发管理工具具有模块化开发、依赖管理、脚手架工具、环境配置、插件扩展和版本管理等功能，其提供了一整套便捷、高效的开发工具和流程，从而有效地减少开发时间。图 2.14 展示了 Web 客户端的结构设计。

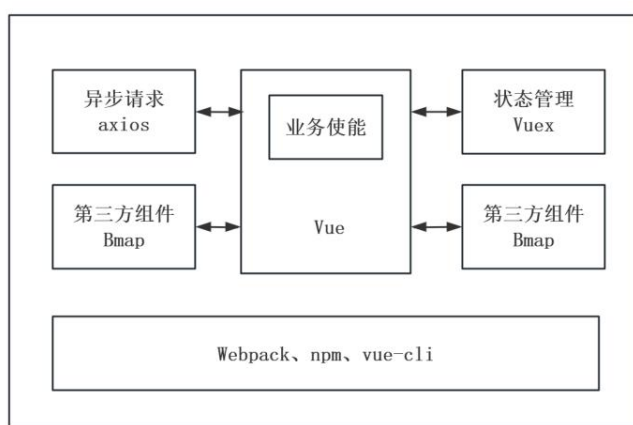


图 2.14 Web 端整体框架图

2.5.2 小程序设计

为了更好地提高光伏运维人员的工作效率和管理水平，本系统在小程序开发上采用了类 Web+原生组件的形式来提高开发效率，实现了和前端功能一样的控制展示界面，为运维人员提供了良好的用户体验和操作感受。图 2.15 展示了小程序的通信模型。

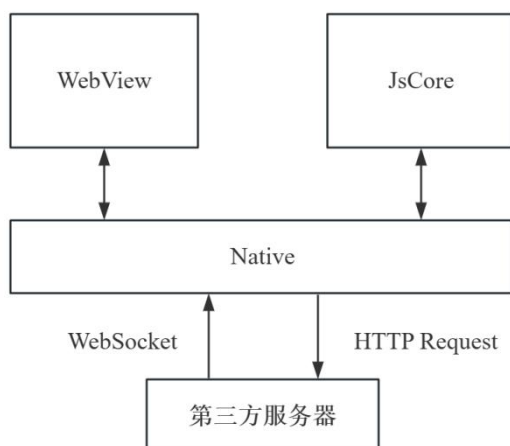


图 2.15 小程序通信模型图

2.5.3 系统展示

分布式光伏电站运维系统中不同模块的硬件实物图和 Web 端、小程序的页面展示如图 2.16 至图 2.29 所示。



图 2.16 屋顶式分布式光伏电站



图 2.17 通信模块电路板硬件图



图 2.18 红外抄表智能网关



图 2.19 光伏逆变器智能网关



图 2.20 地图总览



图 2.21 电站概况



图 2.22 逆变器在线监测

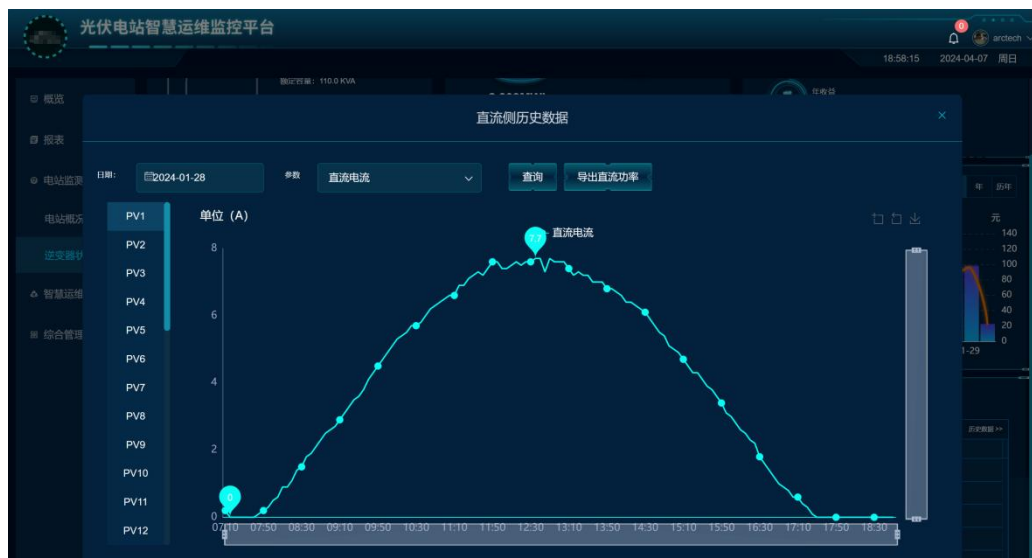


图 2.23 历史电流数据



图 2.24 离散率预警



图 2.25 故障报警



图 2.26 登录首页



图 2.27 电站概况



图 2.28 智能预警



图 2.29 各类通知

2.6 本章小结

本章阐述了光伏电站远程运维系统的设计和实现过程。首先对业务需求和系统设计进行深入剖析，其中重点介绍了智能网关软硬件设计和具体工作流程、网关通信方式与对应协议以及后端整体架构设计和各个模块的具体功能。最后对系统主要的网关硬件和前端界面进行实物展示。

第3章 基于发电功率数据挖掘的光伏电站清洁决策方案

光伏组件上的灰尘累积会减小其透光率并导致电池发电效率的降低，进而影响电站的发电量和经济效益。而当前国内外对光伏板清洁决策的研究多数都是通过样本法和添加传感器来考虑积灰带来的影响，反而会大大增加实验的设备成本花费。上文第2章设计的光伏电站远程运维系统已经实现了数据采集和实时监控等获取光伏数据的功能，从而为后续一系列光伏指标的数据分析提供了基本条件。因此本章在该系统的基础上，主要提出一种基于改进 GS-SVR 算法的光伏电站清洁决策方案，进而从电站实际出发通过历史发电功率数据来量化光伏板积灰程度对光伏发电的实时影响。

3.1 积灰对发电量的影响

光伏板上的积灰通常指沉积在光伏组件表面的灰尘、污垢、沙粒以及其他悬浮物质，其长期累积附着在光伏组件玻璃板表面，从而导致组件发电效率的大幅下降。组件表面的积灰主要来自于自然因素和人为因素^[41]。自然因素主要包括土壤和岩石，它们受到风化作用影响而分解成微小的颗粒，并且通过空气流动被输送到大气中。人为因素引起的灰尘不限于工业排放物、花粉、树叶碎片、鸟类排泄物等，它们主要通过风力、雨水溅射或空气中的悬浮传播而沉积在光伏板上。覆盖在光伏板表面的灰尘会影响组件对入射太阳光的吸收和散射并且减少光伏组件的有效接收面积。而光伏板的发电量直接依赖于光电效应的发生，灰尘会降低光伏板内部的硅电池吸收光子能量的效率，进而造成光伏输出功率和光伏发电量的下降^[42]。图 3.1 展示了光线通过灰尘颗粒后传播过程。

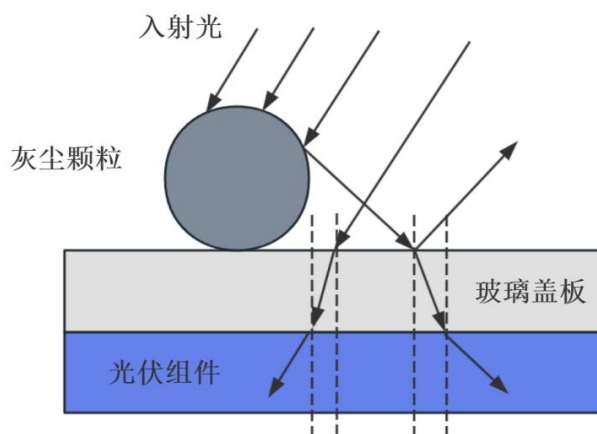


图 3.1 光线通过灰尘颗粒后传播示意图

相关研究表明，光伏组件的性能下降不仅会由其表面灰尘累积造成，当其表面出现大量灰尘累积时，光伏板会局部升温进而出现光伏热斑现象^[43]。热斑效应更会影响光伏板的发电效率，高温条件下会导致电池产生的电压降低，局部区域的过热更会损坏电池，使得光伏板使用寿命缩短，并造成一定的安全隐患。因此更需要制定一个可靠合理的清洁方案。

3.2 清洁决策方案分析

第 1 章已经分析了光伏板清洁决策方案的研究现状与优缺点。目前主流方法是基于数据驱动的方法。这类方法通常利用气象数据和光伏阵列输出数据等数据，利用机器学习算法分析并学习这些参量之间的内在关系，从而实现对未来数据预测或分类。文献^[44]采用神经网络建立回归模型来预测不同辐照度下灰尘粒径对光伏组件功率损耗的影响程度。文献^[45]利用果蝇优化算法的 Elman 神经网络算法对短期时间内的发电量进行预测，通过预测发电量来决定清洗时间。文献^[46]分析了温度及暴露时间对光伏组件积灰率的影响，分别建立了多元线性回归模型及神经网络模型以评估每日积灰速率，基于积灰速率模型给出了光伏清洗建议。黄山峰^[47]等采用一种对光伏组件无积灰理论功率和发电量的估算方法，并通过辐照度与输出功率拟合曲线得到理想发电量并与实际发电量计算出发电量损失值。陈璐^[48]等直接量化在电站光伏组件清洁完成前后 3 天内光伏板表面较为干净并且没有积灰覆盖。但两位都仅考虑辐照度和功率的线性关系，功率的预测精度有很大的提升空间。

本章在分析前人研究成果的基础上，考虑各个算法对光伏数据中非线性特性的预测准确性，采用气象因素结合光伏阵列输出功率的数据驱动方法建立光伏功率预测模型，提供对光伏电站清洁决策的标准化判断。

3.3 光伏功率与气象因素的关系

3.3.1 辐照度对功率的影响

太阳辐照度是指太阳辐射能量在单位面积上的接收量。其强弱受天气状况、海拔高度、纬度位置和日照时间等因素影响。太阳光照能够直接影响光伏电站的输出功率。如图 3.2 为江苏某 1490KW 光伏电站 2023 年 9 月 6 日晴天的辐照度和功率的曲线图。首先大致观察发现光伏输出功率和辐照度的变化趋势几乎一致，且接近正午时输出功率和辐照度达到顶峰。由此可以得出结论，光伏输出功率明显收到辐照度强度的影响，总体呈正向相关性。

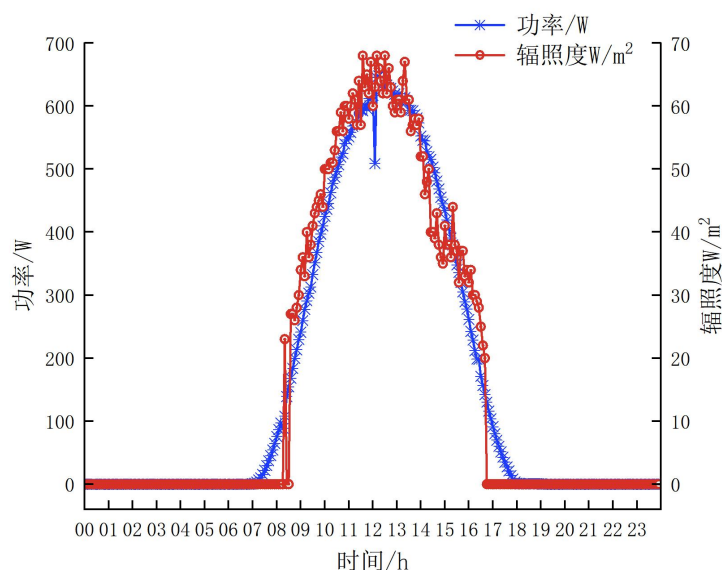


图 3.2 光伏输出功率和辐照度的关系

3.3.2 温度对功率的影响

环境温度的改变会影响光伏组件的性能，其一定程度上会改变组件温度，从而影响光伏出力。随着温度的升高，光伏组件中电池的光电转换效率会降低。而温度与辐照度等气象因素具有互相影响作用的关系。晴天天气情况下中午气温较高，太阳辐照度也很强。如图 3.3 为江苏某 1490KW 光伏电站 2023 年 9 月 6 日晴天的温度和输出功率的曲线图。由图可知，随着温度升高发电效率也会升高，但是当温度达到一定的程度时，其对发电功率的上升速率产生了明显的降低或削弱作用，甚至导致功率的下降。特别是在温度达到最高点时，发电功率明显下降。因此温度对光伏发电功率也具有显著的影响。

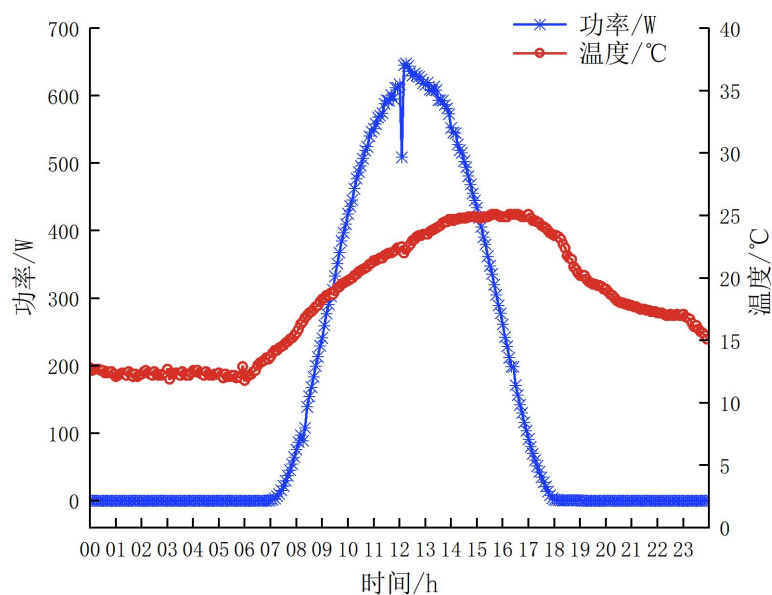


图 3.3 光伏输出功率和温度的关系

3.3.3 湿度对功率的影响

湿度较高时空气中会产生水蒸气，太阳光穿过水蒸气时会发生折射现象，从而能太阳管穿透到空气后到达组件表面的辐射度减弱。另一方面湿度会阻碍光伏板表面的散热能力，从而影响光伏组件的热传导效应。两方面都会直接或者间接影响光伏发电。如图 3.4 为江苏某 1490KW 光伏电站 2023 年 9 月 6 日晴天的湿度和输出功率的曲线图，当湿度达到最低点时，发电功率明显上升。所以湿度可被视为影响光伏发电的重要因素之一。

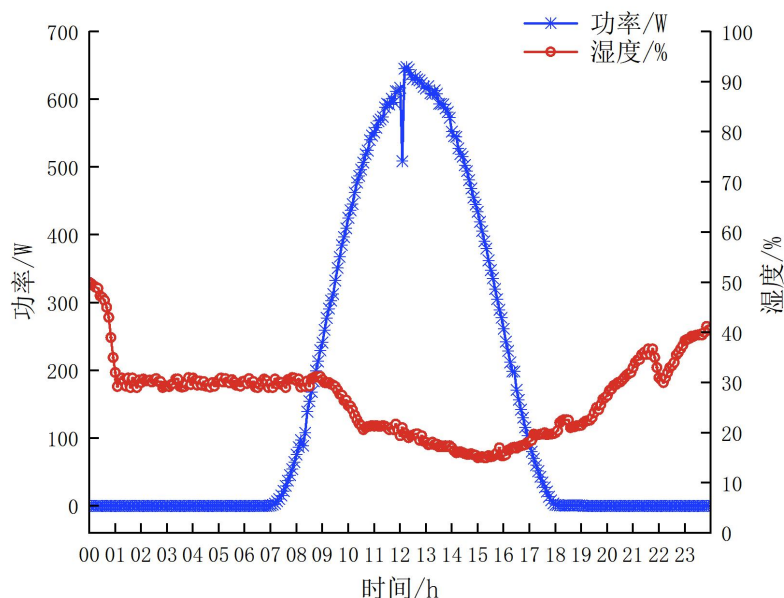


图 3.4 光伏输出功率和湿度的关系

3.3.4 风速对功率的影响

风速对光伏发电功率的影响主要表现在其会影响光伏组件的组件温度和积灰程度。虽然光伏系统一般侧重于太阳能转换，但是风速对光伏功率也有一定的影响。主要影响表现在两个方面：首先风速的增加可以帮助光伏组件散热，从而有助于降低光伏组件的温度。在高温条件下，光伏组件的输出功率通常会下降，而风速的增加有助于降温，从而提高了光伏组件的工作效率。其次适当的风速也有助于减轻光伏板表面的灰尘或杂物积累，保持光伏板表面的清洁，有助于提高光伏板的光吸收效率。如图 3.5 为江苏某 1490KW 光伏电站 2023 年 9 月 6 日晴天的温度和输出功率的曲线图。由图可知，风速表现出较大的不确定性和随机性，并且与光伏发电功率之间并没有显著的关系。

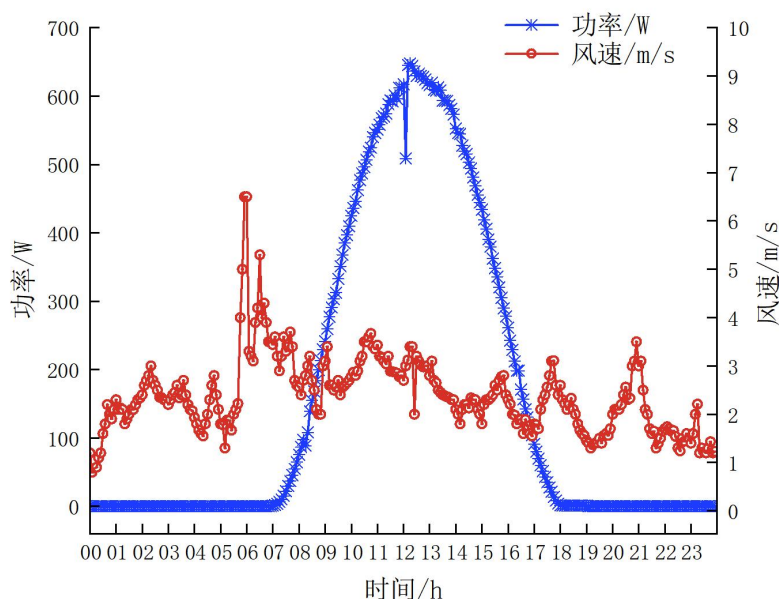


图 3.5 光伏输出功率和风速的关系

3.3.5 其他因素对功率的影响

除了上述分析的辐照度、温度、湿度和风速外，压强、降雨量和风向等因素也会对发电功率造成一定影响。压强对发电功率的影响一般，其会间接通过影响空气流动从而阻碍组件散热。降雨量与光伏功率的影响较复杂，强降雨时候会影响组件的透光率。而当风速一定时，风向会影响光伏板的表面温度。

3.4 数据集定义与分析

3.4.1 相关性分析

根据上文可以初步判断各种气象因素会对光伏输出功率造成大小不一的影响。本章采用皮尔逊相关因素进行相关性分析^[49]，其能描述 2 个变量的相关性程度，以此量化气象因素对功率的影响程度并剔除杂乱因素对实验的干扰。皮尔逊相关系数计算公式如式(3.1)所示。

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3.1)$$

上式中 r 为皮尔逊相关系数， n 为样本长度， X_i 和 Y_i 为实际样本值， $\bar{X}$ 和 $\bar{Y}$ 为实际样本的均值， r 的绝对值越接近 1，表示相关性越高。

以下对选取的环境因素进行相关性热力图分析。

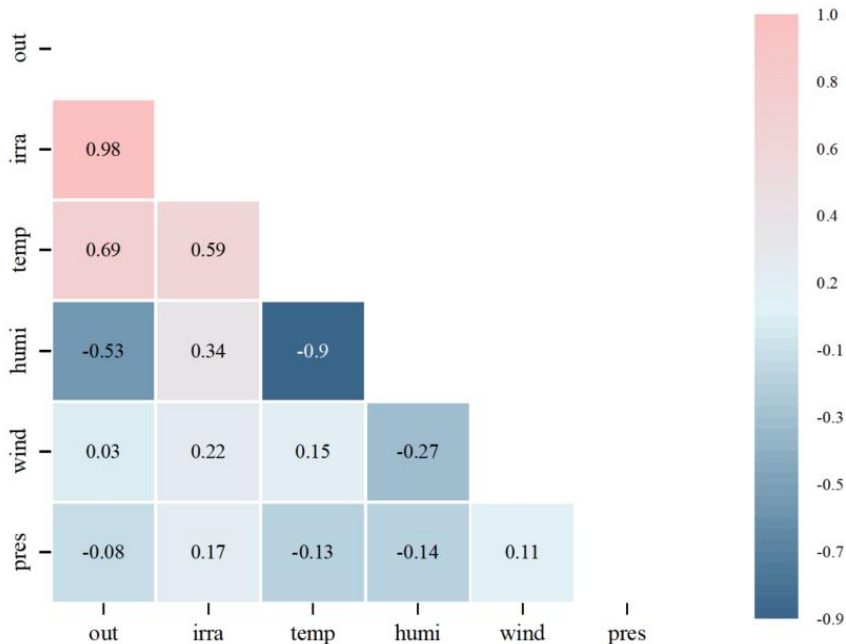


图 3.6 环境因素与输出功率的相关性热力图

本章设置相关系数阈值为 0.5。由以上图 3.6 皮尔逊相关分析结果可以看出，辐照度和输出功率呈极强相关，温度和输出功率呈强相关，湿度和输出功率呈中等相关。综上所述，数据集选取辐照度、温度和湿度这三个环境特征。

3.4.2 数据集定义

经过对光伏运维平台记录的历史数据分析后发现，历史多次电站清洁后光伏组件玻璃盖板上的积灰被充分冲刷，光伏组件的发电效率较好，处于最佳运行状态。而可以将这一段时间称为光伏组件清洁后理想发电周期，记为 T 。本章将光伏组件清洁成后的 3 天定为该理想发电周期。系统将历史电站清洁周期内最佳清洁效果的功率数据和气象数据存储至算法分析数据库，这些功率数据被视为不受积灰影响的功率数据。最终将 n 组数据保存至算法分析数据库，每组数据由某时刻 i 下的辐照度值 I_i 、温度值 H_i 、湿度值 T_i 和理想无积灰输出功率 E_i 组成，从而作为本章实验的数据集。

3.4.3 数据集筛选

根据上文可知组件的输出功率值通过光伏逆变器进行实时检测，对应数据通过智能网关上传至设备云平台。由于系统选取电站历史功率数据进行模型训练，所以对于数据集的筛选着重放在故障筛选部分。数据集只采用设备处于正常运行状态下的数据，从而来排除设备故障带来的影响，进而保证模型的可靠性。

首先后端数据处理服务中会对异常数据进行过滤操作，期间后端已对历史电站的日发电量数据分库分表。由于网络延时、逆变器故障、传感器故障等问题会对数据造成污染，所有采集到的发电量数据会结合历史当日逆变器、组串上报状态和历史发电量数据进行判断。排除光伏组串设备状态为离线、待机、故障以及网络延迟等时间段的数据，且相应参数的误差在可接受范围内，则认为数据有效。

3.4.4 数据归一化

本文的数据集包含光伏历史功率数据和相应环境数据。在使用 SVR 算法进行数据分析过程中，不同特征可能具有不同的度量单位和取值范围，直接使用这些特征进行建模会导致具有较大数值范围的特征可能对模型的训练产生更大的影响，而具有较小数值范围的特征可能被忽略。例如，某天辐照度差值可能达到几百，但是温度差值仅仅为 10°C ，这就导致在 SVR 回归的过程中有较大概率会以辐照度这个特征为准。因此必须在这些变量输入模型前对数据进行归一化处理，从而使得不同特征输入呈现同样的度量尺度。

常见的归一化处理有两种方式^[50]：

（1）最值归一化

最值归一化方法主要将数据缩放到特定的范围内，可以将不同特征的取值范围缩放到统一的区间 $[0,1]$ 之间。

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3.2)$$

式中 x^* 是经过归一化处理的数值， $x_{\max}$ 表示特征向量 x 中的最大值， $x_{\min}$ 表示最小值。

（2）均值方差归一化

如式(3.3)所示, 通过均值方差归一化能将数据集的均值调整为 0, 标准差调整为 1, 从而使得数据集的分布更加均衡, 减少了特征值之间的差异性。

$$x^* = \frac{x - \mu}{\sigma} \quad (3.3)$$

式中 σ 为样本数据的标准差, μ 为样本数据的平均值。

针对以上所述的两种归一化方法, 均值方差归一化更适合处理自身特征值大致服从高斯分布的数据集, 而最值归一化适合处理分布有明显边界的数据集。本章采用的数据集中包含的温度、湿度、辐照度等环境因素均存在明显的上下界, 同时数据集中通常会出现最值不确定的问题, 这可能会导致归一化结果不准确, 因此相比之下最值归一化更适用于本章功率预测的场景。因此本章采取最值归一化对数据集进行预处理操作。

3.5 支持向量机理论

3.5.1 支持向量机

支持向量机 (Support Vector Machines, SVM) 属于二分类模型之一^[51,52], 其最初主要用来解决线性可分的问题。如图 3.7 所示, 其包含空心点和实心点两类点, 这两种点都是线性可分的并且分别代表两种不同的样本类型。其中 H 是分类线, H1 和 H2 线上的训练样本点则称为支持向量, H1、H2 两条线之间的垂直距离则被称为分类间隔。

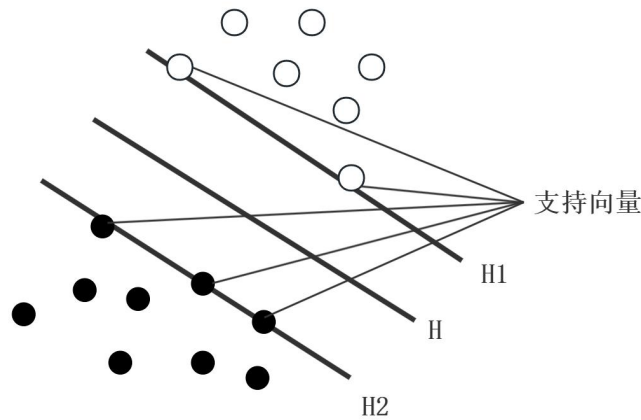


图 3.7 最优分类线示意图

在构建光伏功率预测模型的过程中, 由于光伏发电的特性具有明显的非线性关系, 因此需要采用非线性 SVM 能更好地模拟和捕捉数据中的非线性特征, 提高预测的准确性和稳定性。对于高维数据集而言, 线性 SVM 可能无法准确地区分不同类别的数据点, 从而导致分类错误和较低的预测准确率。而 Vapnik 等人针对非线性问题提出了使用核技巧 (kernel trick) 的方法, 其借助核函数通过非线性映射将原始数据投影到一个更高维的空间, 从而能够有效地处理非线性问题。

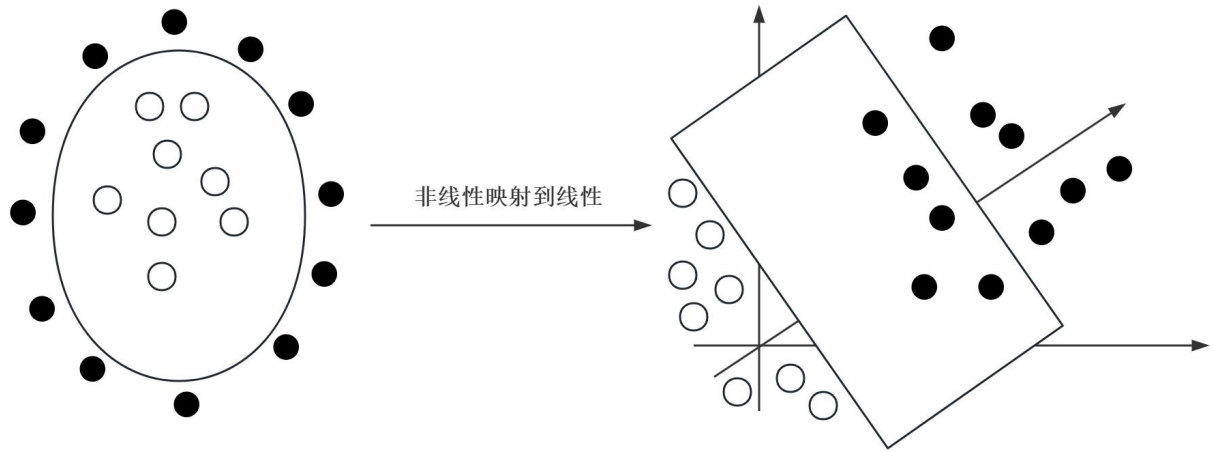


图 3.8 高维空间构造线性划分原理图

3.5.2 核函数

核函数的主要作用是将输入数据从原始低维特征空间映射到一个更高维度的特征空间，以便在新的特征空间中更好地进行分类或回归。这种映射能够将非线性问题转换为线性可分或近似线性可分的问题，从而提高 SVM 的性能和灵活性。

在模型的构建过程中，核函数起到了至关重要的作用。而选择适当的核函数对于 SVM 的性能和预测能力起到了很大的帮助。而核函数的选择依赖于数据的特征和问题的性质。选择过程中对应核函数需要与数据特征的非线性关系相匹配，另外某些核函数的计算复杂度较高可能导致训练时间较长。在大规模数据集上训练时，应选择计算效率较高的核函数。而在有噪声或离群点存在的情况下，应选择具有较好鲁棒性的核函数，以确保模型的泛化能力。常用核函数如表 3.1 所示：

表 3.1 常用核函数

序号	核函数种类	公式	参数
1	线性核函数	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$	
2	多项式核函数	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^d$	$d \geq 1$ 为多项式的个数
3	高斯径向基核函数	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\ \mathbf{x}_i - \mathbf{x}_j\ ^2}{2\sigma^2}\right)$	$\sigma > 0$ 为核的作用宽度
4	Sigmoid 核函数	$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\beta \mathbf{x}_i^T \mathbf{x}_j + \theta)$	$\tanh$ 为双曲正切函数， $\beta > 0, \theta < 0$

一般来说，高斯径向基核函数是最常用的核函数之一，在处理高维数据时其相较于其他几种核函数的表现更为出色，而 Sigmoid 核函数更适合处理低维数据。另外多项式核函数的收敛性更好一点。由于上文提到本章光伏预测涉及多个环境因素及具有高维数据的特点，因此本次研究采用高斯径向基核函数。

3.5.3 支持向量机回归

支持向量回归（Support Vector Regression, SVR）模型是一种基于支持向量机的回归算法^[53,54]，其目标是最小化模型预测值与真实值之间的差异，即最小化预测值与真实值之间的误差平方和，同时还要最大化间隔，即确保超平面和样本点之间的垂直距离总方差最小，从而提高模型的泛化能力。

当 SVM 应用于回归拟合分析时，其核心思想不再是通过最大化类别间的间隔来构建超平面，而是通过找到能够最大化训练样本点与超平面之间间隔的超平面来实现，其目标在于找到一个超平面使得大多数样本点都位于间隔内，并尽量减小样本点与超平面之间的偏差。为了利用 SVM 解决回归拟合方面的问题，Vapnik 等人在 SVM 分类的基础上引入了 ε -不敏感损失函数。其作用主要是在模型中定义一个容忍区间，使得模型对于目标变量的预测在一定范围内不受惩罚。具体而言，不敏感损失函数 ε -将目标变量的真实值与模型预测值之间的绝对差转化为一个非线性损失函数，该损失函数在容忍区间内为零，并在容忍区间外逐渐增加。主要公式如式(3.4)和式(3.5)所示。

$$L(y, f(x, \omega)) = L(|y - f(x, \omega)|) \quad (3.4)$$

$$L(f(x, \omega), y, \varepsilon) = \begin{cases} 0, & |y - f(x, \omega)| \leq \varepsilon \\ |y - f(x, \omega)| - \varepsilon, & |y - f(x, \omega)| > \varepsilon \end{cases} \quad (3.5)$$

其中， $f(x, \omega)$ 为回归预测函数返回的预测值， y 是对应的真实值。如式(3.5)所示，即表示 $f(x, \omega)$ 与 y 之间的差别小于等于 ε ，则损失等于0。

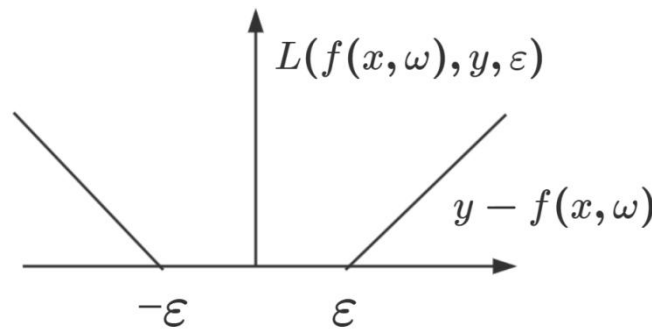


图 3.9 ε 线性不敏感损失函数

设包含 n 个训练样本的训练集为 $\{(x_i, y_i), i = 1, 2, \dots, n\}$ ，其中 y_i 为输出值， x_i 为第 i 个训练样本的输入列向量， $x_i = [x_i^1, x_i^2, \dots, x_i^d]^T$ ， $x_i, y_i \in \mathbb{R}$ 。

可考虑用以下回归函数估计样本数据：

$$f(x) = y = \omega \cdot x + b \quad (3.6)$$

其中 $\omega \in \mathbb{R}^n$ ， $x \in \mathbb{R}^n$ ， $b \in \mathbb{R}$ 。可以由此得出回归问题如式(3.7)所示：

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 \\ \text{s. t.} \begin{cases} y_i - (\omega \cdot x_i) - b \leq \varepsilon \\ -y_i + (\omega \cdot x_i) + b \leq \varepsilon \end{cases} \quad i = 1, \dots, n \end{cases} \quad (3.7)$$

为了处理非线性可分或存在噪声的情况，通过引入松弛变量 ξ_i 和 ξ_i^* 将模型优化问题转化为带有约束条件的凸二次规划问题，从而允许非完美分类。通过调整松弛变量的权重，可以控制模型对误分类样本的容忍程度。而引入惩罚因子 C 可以平衡模型的复杂度和训练数据的拟合程度。因此添加松弛变量和惩罚因子的SVR问题如式(3.8)所示。

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s. t.} \begin{cases} y_i - (\omega \cdot x_i) - b \leq \varepsilon + \xi_i \\ -y_i + (\omega \cdot x_i) + b \leq \varepsilon + \xi_i \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \quad i = 1, \dots, n \end{cases} \quad (3.8)$$

较小的惩罚因子 C 会更加容忍误分类样本，可能导致模型复杂度较高但对噪声更具鲁棒性；而较大的惩罚因子 C 会更加强调模型的简洁性，可能导致更严格的分类或回归边界，但对噪声更敏感。而处理优化问题的同时，可以通过引入拉格朗日乘子将原始优化问题转化为一个只含有等式约束的无约束问题，具体公式如式(3.9)所示：

$$\begin{aligned} L(\omega, b, a, a^*, \xi_i, \xi_i^*) = & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i (\xi_i + \varepsilon - y_i + \omega \cdot x_i + b) \\ & - \sum_{i=1}^n \alpha_i^* (\xi_i^* + \varepsilon + y_i - \omega \cdot x_i - b) - \sum_{i=1}^n (\eta_i \xi_i - \eta_i^* \xi_i^*) \end{aligned} \quad (3.9)$$

上式中 $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ 均表示拉格朗日因子，对上述的参数 $\omega, b, \xi_i, \xi_i^*$ 的偏导都应为0，即：

$$\begin{aligned} \frac{\partial L}{\partial b} = 0 & \Rightarrow \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \frac{\partial L}{\partial \omega} = 0 & \Rightarrow \omega = \sum_{i=1}^n (\alpha_i x_i - \alpha_i^* x_i) \\ \frac{\partial L}{\partial \xi_i} = 0 & \Rightarrow C - \alpha_i - \eta_i = 0 \end{aligned} \quad (3.10)$$

将式(3.10)化简，转换成对偶形式为式(3.11)：

$$\begin{cases} \max \left[-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) (x_i \cdot x_j) - \varepsilon \sum_{i=1}^n (\alpha_i - \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i^* - \alpha_i) \right] \\ \text{s. t.} \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases} \end{cases} \quad (3.11)$$

解方程得最优解 $\alpha^{(*)} = (\alpha_1, \alpha_1^*, \dots, \alpha_m, \alpha_m^*)^T$ ，由此可以计算出 ω 和 b ，最后得到回归函数如式(3.12)所示：

$$f(x) = \sum_{i=1}^n (\alpha_i^* - \alpha_i) (x \cdot x_i) + b \quad (3.12)$$

其中当 $\alpha_i^* - \alpha_i$ 不等于零时其对应的数据被定义为支持向量。由于上文已经选择核函数将数据映射到高维特征空间使其线性可分，然后再进行线性回归分析。此时核函数 $K(x_i, x) = (\phi(x_i) \cdot \phi(x))$ 。回归函数的表达式如式(3.13)所示：

$$\begin{aligned} f(x) &= \sum_{i=1}^n (\alpha_i - \alpha_i^*) (\phi(x_i) \cdot \phi(x)) + b \\ &= \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i \cdot x) + b \end{aligned} \quad (3.13)$$

3.5.4 支持向量机回归模型

SVR 的预测模型如图 3.10 所示，其函数结构本身具有类似神经网络的形式， $(\alpha_i - \alpha_i^*)$ 为网络权重，其中每一个中间节点都与一个支持向量相对应，从而实现输入和输出之间的映射关系。与神经网络架构的每一步都是由经验得到参数值不同，支持向量值都是由算法自动产生。

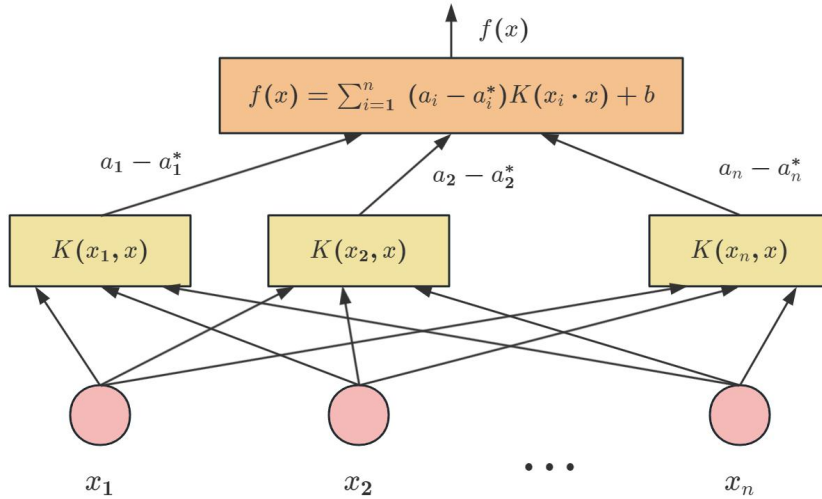


图 3.10 SVR 的神经网络结构

尽管 SVR 模型能够处理非线性数据的问题并且构建过程相对简单，但在某些情况下可能会被困在局部最优解中，影响 SVR 性能的重要参数是惩罚因子和核函数，因此需要优化这两个参数来防止出现过拟合的和欠拟合等问题。传统参数选择过程计算复杂度较高，会大大增加训练时间。并且由于通常基于经验和实践选取最优参数从而缺乏明确的理论指导，因此无法凭借数学证明或理论原则来确定最佳的参数取值。近几年来，进化算法在优化超参数选择方面得到了更广泛的应用^[55]。倪萍等^[56]采用蚁群参数优化算法进行超参数的

优化，刘景旭等^[57]采用了启发式搜索算法进行优化，尽管优化所需的时间有所减少，但模型训练周期仍然相对较长，并且更容易陷入局部最优解。采用遗传算法、粒子群算法等进行参数寻优，效率较高但也可能是局部最优解。而 GS 使用穷举的方式尝试模型各种组合以确保全面搜索，但是计算时间会随参数数量增加而指数级别增长。本研究采用改进的 GS 来减少计算时间并且进行预测结果。

3.6 基于改进 GS 的 SVR 理想无积灰功率预测算法

3.6.1 K 折交叉验证

GS 常常还结合 K 折交叉验证（K-fold cross validation, CV）来调节参数以找到最优的参数组合。该验证法交叉 K-1 份合并为训练集进行分析，在训练集上训练模型并且在测试集上对模型进行评估得到评估指标，最后对 K 次评估结果取平均值作为最终评估结果，从而有效筛选出最优参数组合。如图 3.11 所示，其基本思想是将数据集的所有数据平均划分为 K 份，取第 K-1 份作为训练集，剩下 1 份作为测试集。

交叉验证法评估结果的稳定性和保真性在很大程度上取决于 K 值的大小。当 K=1 的情况下，此时所有数据都被用于训练，因而会出现过拟合导致低偏差、高方差的情况。而随着 K 值的不断升高，单一模型评估时的方差逐渐加大而偏差减小。但从总体模型角度来看，反而是偏差升高了而方差降低了。K 的取值一般为 5、10、20 等，具体取决于数据集大小和计算资源的限制。较大的 K 值可以减小估计误差，但会增加计算的时间复杂度。本章采用了 10 折交叉验证方法，首先将原始数据集划分为 10 个部分，随机选取其中 9 个部分作为训练集，剩下一个部分作为测试集。

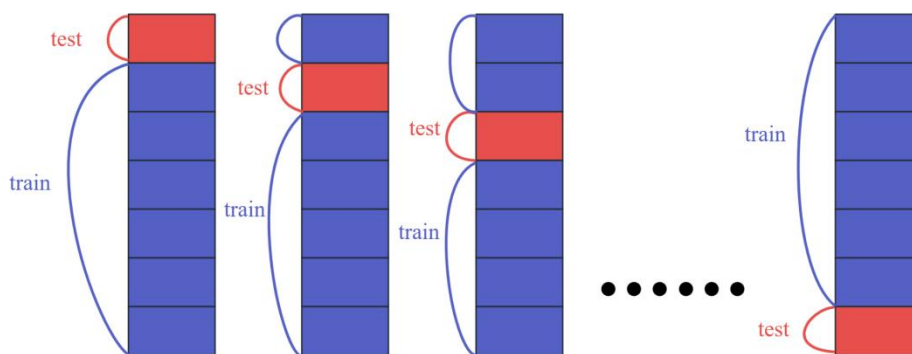


图 3.11 K 折交叉验证示意图

3.6.2 网格搜索法及其改进

网格搜索（Grid search, GS）^[58]是一种参数寻优算法，经常可以用来当作处理非线性问题的优化方法。该方法在满足特定约束条件的前提下，进一步计算网格点的目标函数值，并最终挑选出能让系统性能达到最佳状态的参数组作为最优选择。主要流程如图 3.12 所示。其算法基本思想是：在变量的区域内按照规定的步距划分进行划分，排列组合所有参数的可能取值，得到全部的参数组合形成“网格点”。

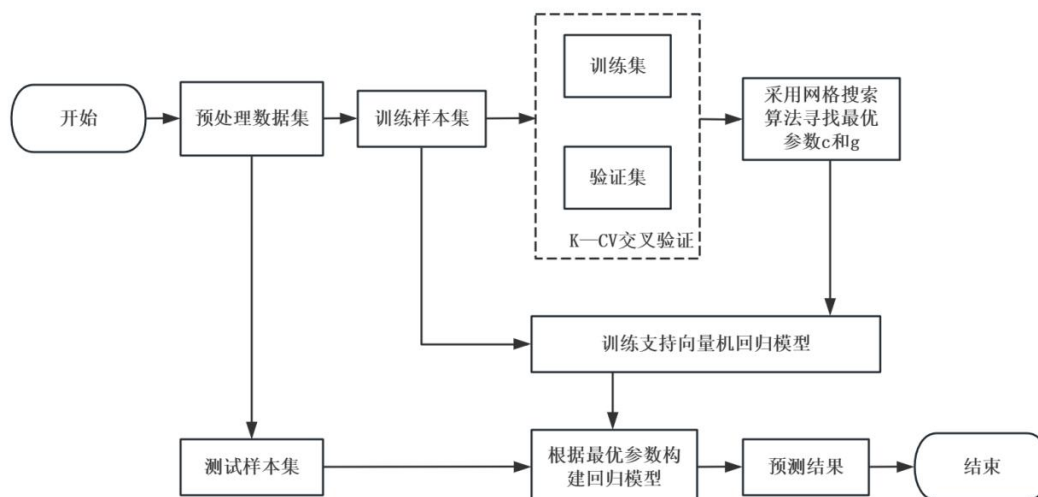


图 3.12 普通的网格搜索法流程图

由于 GS 算法是一种暴力枚举算法，会通过遍历每一组参数组合进行试算。因此该方法也具有一些不足之处：第一，当参数过多时网格搜索法的计算复杂度随着参数数量的增加而指数级增加，其中可能会花费一定的计算成本在不必要的参数组合上。第二，较小的步长相应地需要在更细粒度的参数取值上进行搜索，这会增加搜索的计算量。另外可能导致过度拟合的风险增加，会使模型过于敏感地适应每个具体的参数取值，从而过度拟合训练数据。第三，网格搜索法中可行域的单向逼近容易陷入局部最优解。当参数组合微调后，搜索过程可能会局限于某个局部区域，而无法跳出该区域继续搜索其他可能更优的区域。因此，使用网格搜索法的关键在于合理设置搜索步长的大小。

为了优化算法并且获取最佳参数组合，GS 先通过设置合适的搜索范围和搜索间隔并对所有参数组合依次遍历，再结合 K 折交叉验证法评估性能指标从而调整最优参数。但是在实际应用中较大的搜索范围和较小的搜索间隔都会导致参数寻优过程中的时间复杂度上升，进而降低了模型训练效率。因此本文采用空间二分和变更步长的方法对 GS 进行改进，从而提高算法效率。以下图 3.13 展示了其完整的改进流程。

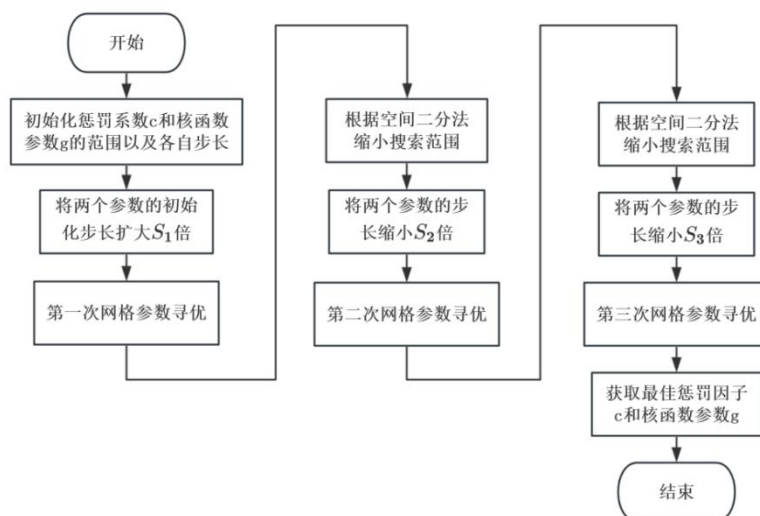


图 3.13 改进的网格搜索法流程图

(1) 首先对惩罚系数 c 和核参数 g 的搜寻范围以及各自的步长进行初始化，然后将两个参数初始化的搜索步长扩大 S_1 倍，并且向不同方向生成网格进行第一次参数寻优。

(2) 采用图 3.14 中的空间二分法将 c 与 g 的搜寻范围缩小，同时将搜索步长缩小 S_2 倍，然后进行第二次网格参数寻优。

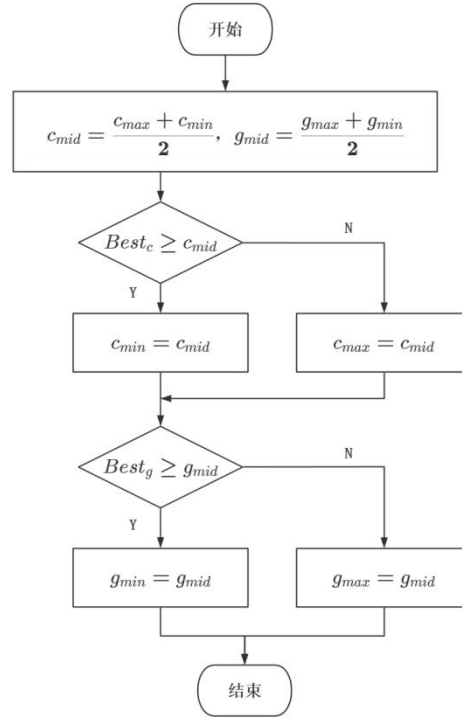


图 3.14 基于二分查找的空间缩小方法

(3) 采用空间二分法再一次缩小了 c 和 g 的搜索范围，同时将搜索步长缩小 S_3 倍，然后进行第三次参数寻优。其中 $S_1=S_2*S_3$ 。

(4) 将步骤 (3) 中获取的最优参数组合代入并且构建理想无积灰功率模型。

(5) 使用构建的理想无积灰功率模型对测试集完成预测。

通过执行以上优化操作，使得搜索过程中采用二分法对搜索范围进行处理，并且缩放搜索法的步长，从而显著提高参数寻优的效率。

3.6.3 基于改进 GS 的 SVR 理想无积灰功率预测算法具体实现

采用上文介绍的改进的 GS 寻找 SVR 的最优参数组合，主要框图如 3.15 所示，构建改进的 GS-SVR 理想无积灰功率模型的步骤如下：

- (1) 预处理样本数据集，并且对缺失或者异常的数据进行相应的修正和填补。
- (2) 对 SVR 参数的惩罚参数 c 、核函数参数 g 的搜索范围及各自的步长进行初始化。
- (3) 针对每个参数组利用 K 折交叉验证法评价每个参数组的平均预测准确率。
- (4) 根据图 3.13 使用改进的 GS 算法变更步长缩减搜索空间进行参数寻优。
- (5) 根据最优 $Best_c$ 和 $Best_g$ 的参数组合构建理想无积灰功率预测模型。
- (6) 利用建立好的模型对理想无积灰功率进行预测。

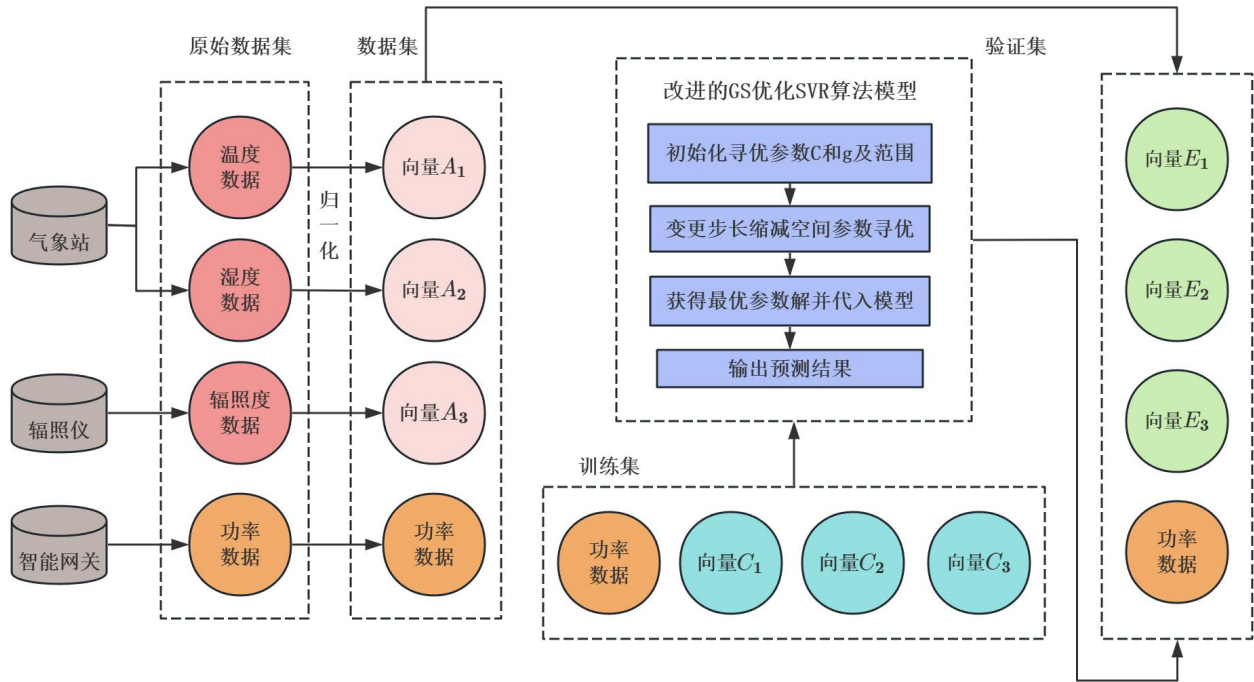


图 3.15 基于改进的 GS-SVR 预测理想无积灰功率框图

3.7 实验测试与结果分析

3.7.1 数据预处理

本章以江苏某 1490KW 光伏电站为例，通过系统提供的 2021 年 9 月 2 日到 2023 年 11 月 13 日共 10 次历史清洁数据，整个电站由 12 台逆变器去实时监测各自区域光伏阵列的数据，本次实验选取其中 1 台逆变器获取到的区域光伏组件的功率数据和对应时刻的温度、辐照度数据进行实验。上文提到每次清洁后最佳清洁效果周期 T 定为 3 天，故一共选取该台逆变器上传的 26 天天气情况较好的数据。

光伏理想功率预测的数据集中包含的属性为温度、辐照度以及对应时刻的光伏输出功率。考虑到夜晚和凌晨基本没有太阳辐照度，即光伏发电基本不出力或者出力很小。因此，从智能网关采集到的数据中选取从白天 7 点到晚上 7 点的数据，由于网关以每 5 分钟的频率进行数据采集，即一天共 144 个采样点。同时对数据进行一系列的预处理步骤，包含补全缺失数据，去除奇异点、统一时间等，具体等于 3688 组。

3.7.2 模型评估指标

上文提到在光伏功率预测中会受到辐照度、温度、湿度等多个因素的影响，预测结果往往存在一定的误差和不确定性。因此需要采用合适的评估指标来衡量预测结果与实际观测值之间的差异并且评估不同模型的预测精度。常用的误差评定方法包括以下几种：

(1) 平均绝对误差 (MAE)，是用于衡量预测结果与实际观测值之间差异的指标。它计算了预测值与真实值之间的绝对差值的平均值。较小的 MAE 值表示预测结果与实际

观测值之间的差异较小，即预测的准确性较高；而较大的 MAE 值表示预测结果与实际观测值之间的差异较大，即预测的准确性较低。

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (3.14)$$

(2) 均方根误差 (RMSE)，是一种用于衡量预测模型在连续性数据上的预测精度的指标。它衡量了预测值与之间的均方根差异，表示预测值与真实值之间的平均偏差程度，是回归任务中常用的性能评估指标之一。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (3.15)$$

(3) 复判定系数 (R^2)，评估的是拟合回归线周围数据点的散布情况。对于同一数据集，较高的 R 平方值表示观测数据与拟合值之间的差异较小。当 R^2 越接近 1 时，表示相关的方程式参考价值越高；相反，越接近 0 时，表示参考价值越低。

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \quad (3.16)$$

(4) 算法性能，在对比误差的同时也要考虑不同模型的实际运行性能，模型训练所需的时间能够直观反映模型的运行性能。

3.7.3 实验仿真和分析

(1) 准备训练样本

根据以上相关性分析，训练样本包含温度、湿度、辐照度和输出功率四个特征值，并且在每次清洁后自动存储到算法分析表中。

(2) 参数初始化设置

如表 3.2 所示，对改进 GS-SVR 的参数进行初始化。

表 3.2 预测模型参数配置

模型	参数类型	参数设置
改进的 GS-SVR	惩罚参数 c	$[2^{-8}, 2^8]$
	核参数 g	$[2^{-8}, 2^8]$
	c 的步长	0.2
	g 的步长	0.2
	步长放缩参数 T	16
	步长放缩参数 F	4
	步长放缩参数 S	4

(3) 模型训练

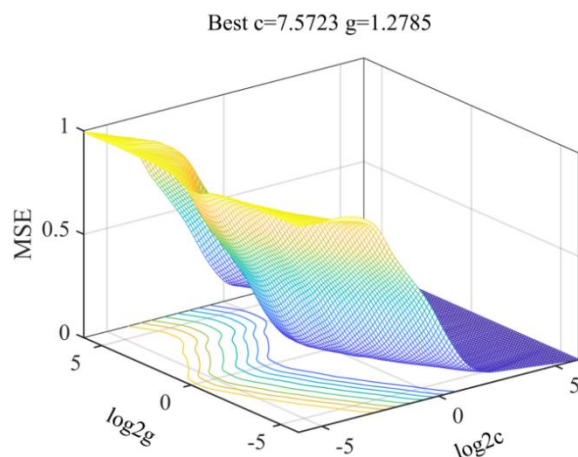


图 3.16 改进的 GS-SVR 模型参数选择 3D 视图

仿真实验均在 PyCharm2022.1.1 和 Python3.8 的环境下进行。以网格搜索为切入点进行最优解计算时，在参数空间内，在搜索过程中同时计算全部网格交叉点，直至将最优解找出来。如图 3.16 所示，当 SVR 核函数的参数 c ， g 不断变化的过程中，光伏输出功率预测的准确率也在不断变化，最后采用改进的 GS 寻优得到 SVR 的最优参数组合 c 为 7.5723， g 为 1.2785。

(4) 模型验证

以上述最优参数建立理想无积灰功率模型，选择了其中一组逆变器的功率样本来进行模型的验证，其识别的结果如图 3.21 展示。为了展示改进后的 GS-SVR 模型所带来的积极影响，本研究将其与标准 GS-SVR 和未经优化的 SVR 做了详细比较。在这个过程中，标准的 GS-SVR 寻参阶段采用了标准 GS 算法来进行参数查找。未经优化的 SVR 保持不变，图 3.17、图 3.18 和图 3.19 分别展示了最终的识别成果。

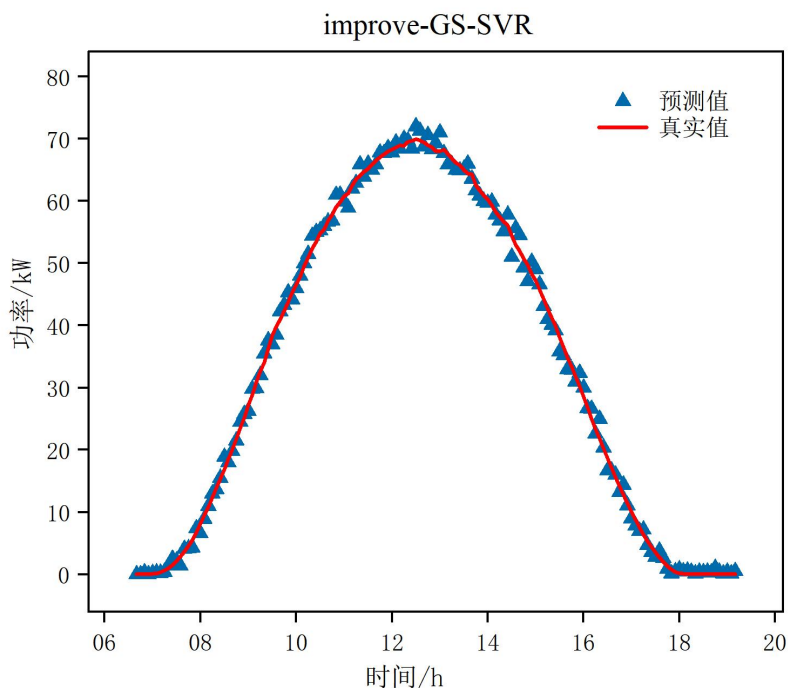


图 3.17 改进的 GS-SVR 结果

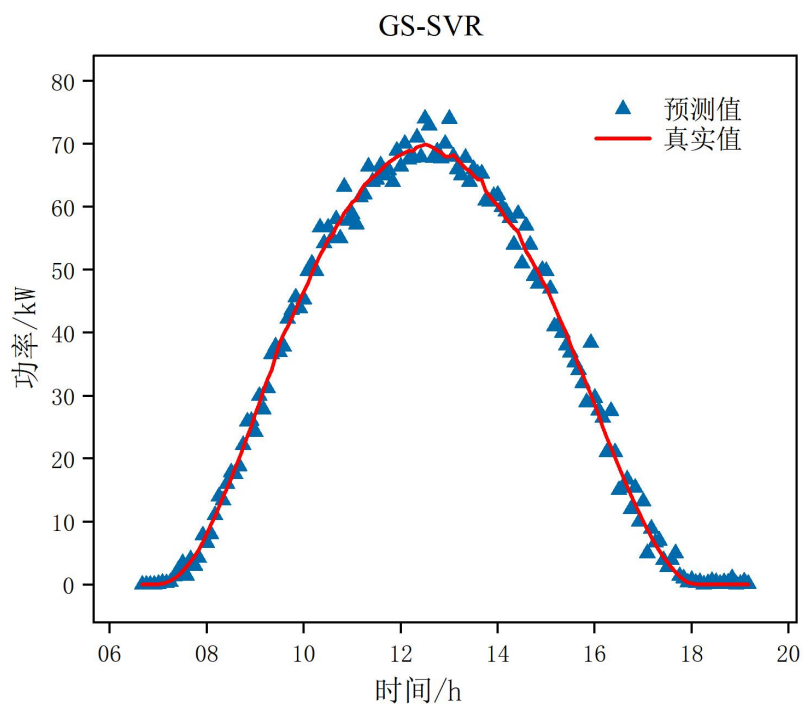


图 3.18 GS-SVR 结果

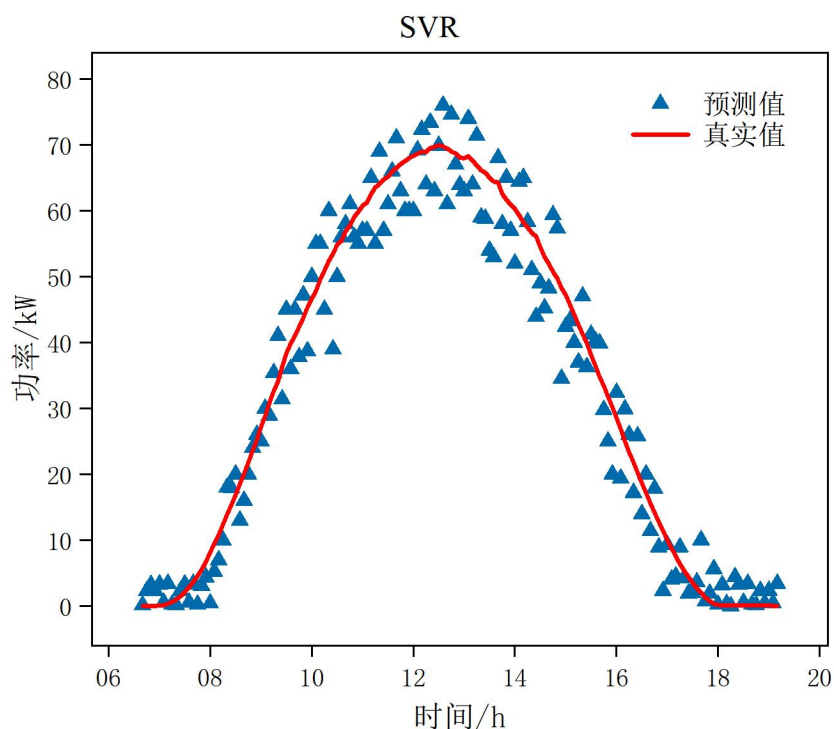


图 3.19 SVR 结果

根据以上各个模型和真实功率点线图可以看出，普通的 SVR 相较于真实值曲线偏差较大。而改进的 GS-SVR 和 GS-SVR 预测效果较好，经过改进的 GS-SVR 模型的拟合曲线和实际曲线重复程度较高，即预测结果更加接近真实数值。

(5) 模型对比

表 3.3 对这三种模型的评价指标进行对比。

表 3.3 模型评估结果对比

模型	评价指标			
	MAE	RMSE	R ²	样本训练时间/s
SVR	2.072	3.618	0.8233	235
GS-SVR	1.306	1.932	0.8824	768
改进的 GS-SVR	0.944	1.213	0.9107	359

从表中数据可以看出基于改进的 GS-SVR 模型预测效果最好。对于 MAE 指标，改进的 GS 模型较于 SVR、普通的 GS-SVR 模型，误差分别减少了 54.4%、27.7%。对于 RMSE 指标，改进的 GS 模型较于 SVR、普通的 GS-SVR 模型，误差分别减少了 66.5%、37.2%。对于 R² 分数，分别提高了 10.6%、3.2%，优化效果比较明显。改进 GS 对于预测结果有小幅提升，但是对算法训练时间大大缩减，所以总的来说该模型相比之下最为优秀。

3.8 清洁决策制定与实例验证

3.8.1 无积灰理想发电量的计算

当得到无积灰理想功率模型后，由于清洁结束后的 3 天的功率数据作为模型训练的数据集，因此以清洁结束后的第 4 天开始作为起始计算时间，从该天开始计算当日的无积灰理想发电量。首先，从运维系统中实时获取太阳辐照度数据、温度数据和湿度数据，计算得到此时区域光伏组件的理想光伏组件输出功率，并通过智能网关实时采集该光伏组件的实际输出功率。然后在夜晚对一天的实际输出功率曲线和理想输出功率曲线分别进行积分运算，计算得到的每日实际发电量 Q'_{real} 和理想发电量 Q'_{ideal} ，则该区域光伏阵列每天的发电能力 L_p 可表示为：

$$L_p = \frac{Q'_{real}}{Q'_{ideal}} \times 100\% \quad (3.17)$$

发电能力指标 L_p 可以反映每日积灰对光伏组件输出功率和发电量的影响程度。这样即便光伏电站的气象站和辐照仪存在偶尔测量不精确的情况，也可以经过前期数据筛选提高精度，不会对积灰对光伏发电的影响程度的分析结果造成影响。根据光伏远程运维平台采集的逆变器数据得到光伏阵列实际日发电量 Q_{real} ，但该数据与实际输出功率通过积分运算获得的实际发电量 Q'_{real} 有所不同。由于系统损耗和设备老化问题存在一定误差，不可以直接采用积分运算得到的实际发电量 Q'_{real} 。因此通过式(3.20)计算该区域组件无积灰理想发电量 Q_{ideal} ，即：

$$Q_{ideal} = \frac{Q_{real}}{L_p} \quad (3.18)$$

进而可得到日积灰损失发电量 Q_{loss} ，即：

$$Q_{loss} = Q_{ideal} - Q_{real} \quad (3.19)$$

3.8.2 确认清洁时间

设 E 为从上一次清洁时间开始累计整个电站光伏组件清洁后产生的经济效益，计算如下所示：假设第 1 天到第 n 天的某个逆变器监测的区域光伏组件日光伏发电量损失值为 $\{Q_{1j}, Q_{2j}, \dots, Q_{n-1j}, Q_{nj}\}$ ，并且假设有总共有 m 个逆变器来监测各自区域光伏组件数据（上文提到的电站共计 12 个逆变器），即每个区域的损耗值积分可以得出整个电站的日损失发电量，再结合当前电价计算一个周期内损失的经济效益即清洁效益，则有如下公式：

$$E = A \sum_{i=1}^n \sum_{j=1}^m Q_{ij} \quad (3.20)$$

式中 A 为当前电价，具体电价调整定期更新至数据库。

总电站的日损失发电量的计算流程如图 3.20 所示。

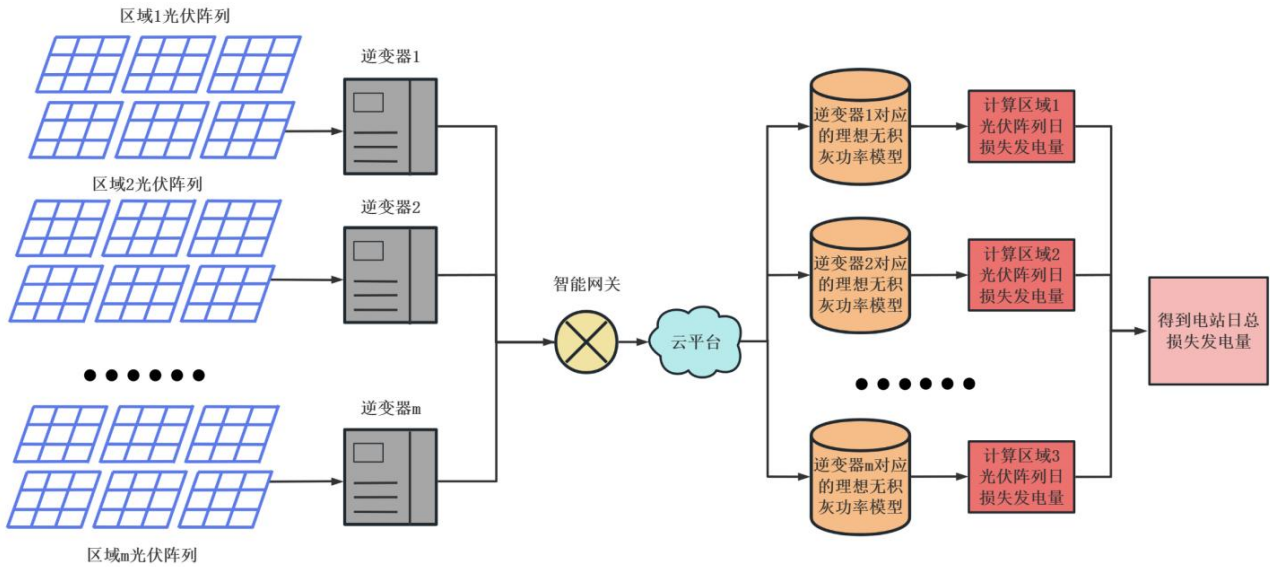


图 3.20 电站日损失发电量计算示意图

设 M 为一清洁电站总成本，其由每千瓦光伏组件的清洁成本 C 和光伏电站的总装机容量 I 得到。有如下公式(3.21)所示：

$$M = CI \quad (3.21)$$

当 E 大于 M 的时候且未来 3 天没有下雨，网关下发清洁建议，否则的话不下建议。清洁决策判断如图 3.21 所示。

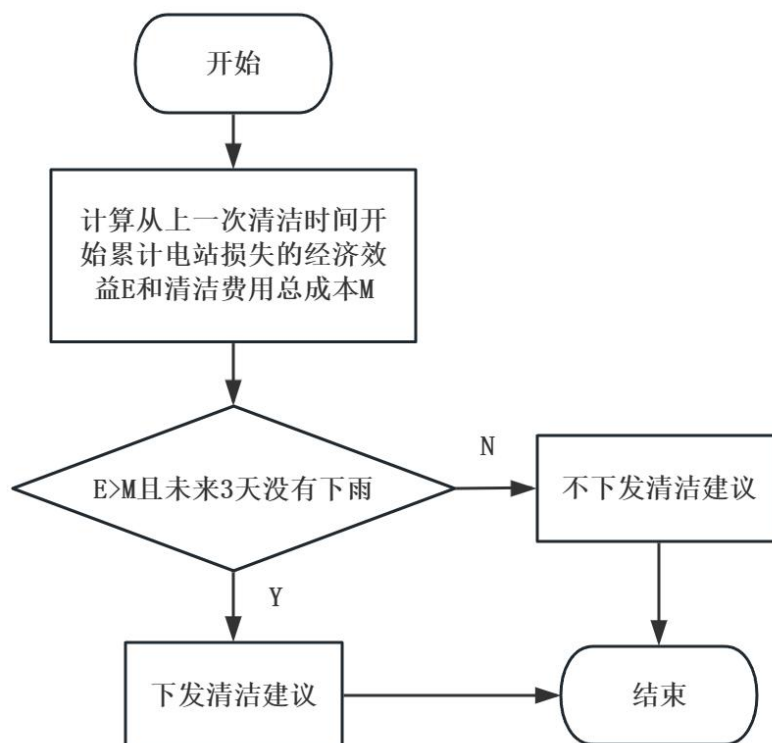


图 3.21 清洁判断流程图

最后清洁决策判断需要结合未来天气预报信息，将未来当地天气预报信息作为判断是否需要对光伏组件进行清洁的另一判断条件。在后端服务中可通过请求当地气象实况天气接口并且每天更新 3 次来获取稳定、可靠的当地天气数据来源。

虽然雨水可以冲刷掉光伏组串上的部分积灰，但是降雨量的大小对光伏板积灰的冲刷程度有很大影响并且很难进行评估，所以本方案忽略雨水冲刷积灰的影响，仅仅结合未来天气因素和日发电量损耗收益来动态判断是否下发清洁提醒。如果当下清洁收益大于清洁成本，但由于清洁时或清洁后下雨会让清洁收益大打折扣，因此接下来是雨天则不下发清洁提醒并做延期处理，这也避免了人为经验判断或者固定的清洁方法会影响到清洁收益最大化的最佳决策判断。所以最终清洁方案为：如果计算的发电量累计损失金额大于清洁的人工成本且未来 3 天没有下雨则下发指令，通知有关运维人员或光伏清洁机器人等自动化设备执行清洁任务。该清洁收益对应的清洁周期为最优清洁周期。

3.8.3 实例验证

以江苏的一个 1490.00kW 的分布式光伏电站为研究对象，该电站采用的是自主使用余额的上网方式，其所在地区的度电收入为 0.38 元，而每千瓦的清理费用是 2 元。该电站从 2021 年 9 月至今进行了总共 10 次清洁工作，上文已通过这 10 次记录对该电站所有区域构建各自区域的理想无积灰功率模型。现对该电站 2023 年的一次电站清洁进行下一次时刻清洁时间进行量化决策。该次清洁完成时间为 2023 年 11 月 10 日。上文提到清洁后前三天的组件处于无积灰理想状态，所以从 2023 年 11 月 14 日开始计算每日的无积灰理想发电量和其对应的积灰损失发电量，具体如表 3.4 所示。

表 3.4 积灰损耗发电量统计

日期	实际发电量 (KWh)	无积灰理想发电量 (KWh)	日积灰损失发电量 (KWh)
2023-11-14	3069.2	3172.6	103.4
2023-11-15	3112.1	3206.4	104.3
2023-11-16	2836.5	2960.0	123.5
2023-11-17	2661.8	2784.1	122.3
2023-11-18	2610.8	2718.2	107.4
2023-11-19	2891.8	3003.2	111.4
2023-11-20	2945.7	3100.1	154.4
2023-11-21	2880.2	3032.6	152.4
2023-11-22	2529.5	2701.9	172.4
2023-11-23	328.4	408.8	80.4
2023-11-24	583.9	673.2	89.3
2023-11-25	1230.8	1420.1	189.3
2023-11-26	312.3	389.7	77.4
2023-11-27	212.4	301.8	89.4
2023-11-28	726.6	859.0	132.4
2023-11-29	3171.4	3404.6	233.2
2023-11-30	2059.5	2312.7	253.2
2023-12-1	889.6	1019.8	130.2
2023-12-2	2878.5	3166.9	288.4
2023-12-3	3060.2	3296.7	236.5
2023-12-4	3115.7	3348.9	233.2
2023-12-4	3282.0	3569.2	287.2
2023-12-6	2848.6	3099.0	250.4
2023-12-7	3108.6	3305.8	197.2
2023-12-8	3053	3395.3	342.3
2023-12-9	2140.4	2384.9	244.5
2023-12-10	2405.8	2629.2	223.4
2023-12-11	2270.6	2643.8	373.2
2023-12-12	629.6	760.8	131.2
2023-12-13	2368.2	2634.4	266.2
2023-12-14	2132.1	2507.3	375.2
2023-12-15	2325.2	2513.4	188.2
2023-12-16	833.8	1057	223.2

日期	实际发电量 (KWh)	无积灰理想发电量 (KWh)	日积灰损失发电量 (KWh)
2023-12-17	2444.9	2760.2	315.3
2023-12-18	2523.9	2846.3	322.4
2023-12-19	2170.9	2469.3	298.4
2023-12-20	2134.8	2424.1	289.3
2023-12-21	2300.5	2643.8	343.3

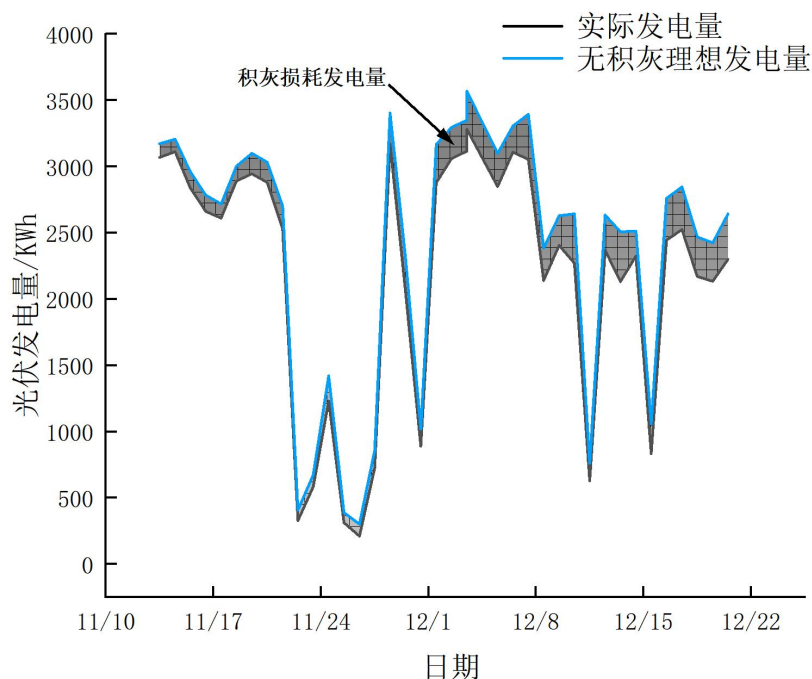


图 3.22 积灰损失发电量损失示意图

根据图 3.22 可以观察到随着时间推移，每天的日积灰损失的发电量逐渐增加，雨天情况下实际发电量和无积灰理想发电量差值较小，天气状况良好的情况下日损耗发电量增加的趋势更为明显。而到 2023 年 12 月 21 日为止，该电站距离上一次清洁所累积的总积灰损失发电量已经超过由清洁成本、装机容量和度电收益换算而成的电量（7842.11kWh），即清洁经济效益超过了清洁成本，此时光伏运维后台实时发出清洁建议。

3.9 本章小结

本章设计了一种光伏组件的清洁决策方案，实现了对光伏组件智能周期性清洁管理。首先介绍了积灰带给光伏发电的损失，同时研究了各种气象因素对光伏输出功率的具体影响，并且通过相关性分析选择合适的环境因素作为特征输入。然后对 SVM 的基本原理进行简要阐述，提出采用 GS 优化 SVR 最优参数组合的搜索过程，并且在此基础上针对 GS 算法训练时间长的缺点进行变步长和空间二分缩减的优化，建立理想无积灰功率预测模型。经实验测试对比，本方案相较于未优化的 GS-SVR、SVR 有较大的精度提升，并将判别系数 R^2 提高到 91.07%，并且大大缩短训练时间。通过每日采集上来的实际功率和实际发电

量计算得到日积灰损失发电量，再结合清洁成本动态地为光伏组件清洁提供合理的决策依据。本章最后通过对历史清洁周期的预测实验分析，得出该方案能合理动态地规划光伏组件的清洁判断，并能为提高清洁经济效益提供有力的技术支持。

第 4 章 基于组串电流数据挖掘的光伏组件故障预警方案

当前光伏运维系统的功能研究更集中在故障报警等模块，故障预警功能仍处探索阶段尚未完善。本章着重聚焦光伏电站的预警功能，主要提出了一种基于 CNN-BiLSTM 的模型对光伏组串电流进行预测，并针对不同天气情况分别建模，与 LSTM、CNN-BiLSTM 两种预测模型进行实验对比。最后把预测结果与光伏组串实时上报的实际值进行对比计算出电流的偏差率指标，并结合滑动窗口思想建立报警阈值机制，从而在故障报警真正发生之前进行预警检修。

4.1 预警方法选型

目前常见的光伏组件故障预警指标有组串电流离散率、日满发小时偏差、组串电流偏离值等。而较多的异常判断更针对电流值的异常判断。常见的故障预警方法有基于离散率的故障预警模型和基于偏差率的故障预警模型等。

离散率是衡量光伏组件运行性能的重要指标，其中离散率模型可以用于评估光伏电站发电性能的一致性。其本质就是通过组串电流的标准差和平均值计算反映此时该逆变器所监控的区域光伏阵列中多个组串的整体发电状况。离散率越低，说明各路组串发电性能越集中，一致性越好，发电情况越稳定。离散率是通过每个时刻离散率的加权平均值来进行反映的，多个时刻点的离散率加权平均计算得出一天的离散率，其可分为 4 个等级：组串电流离散率取值范围在 0~5% 以内，说明支路电流运行稳定；组串电流离散率取值范围在 5%~10% 以内，说明支路电流运行情况良好；组串电流离散率取值范围在 10%~20% 以内，说明支路电流运行情况有待提高；组串电流离散率取值大于 20%，说明支路电流运行情况较差，需要运维人员即时处理^[59]。

而基于偏差率的故障预警模型更针对光伏电站系统中组件设备在一段时间内的运行参数，通过参数变化趋势来表现该时间段内组件的发电能力，在组件设备故障报警之前起到及时预防潜在风险的作用，从而能够保证组件设备稳定运行，提高电站发电量。Abd^[60]等提出了一种阵列故障检测程序，将实际电气参数与预期电气参数进行比较，通过详细的 PV 计算精确的参考数据，以此实现光伏阵列的故障预警。文献^[61]提出了一种基于 ANN 算法预测交流发电量，然后将来自光伏系统的实时趋势数据与模型的预测值进行比较，并分析残差向量以检测异常并生成每日预测性维护警报。宋海明^[62]等使用非线性状态估计、小波变换等理论构建出设备的状态模型，根据计算出的实际数据与预测数据之间的相似度判断设备状态，从而完成设备故障预警。Choi^[63]等使用随机森林算法与过去的功率数据去预测未来一段时间内的功率数据，当预测模型的精度大于某个 MAPE（平均绝对百分比误差）值时，会检测预测数据中存在的异常值。而常见的偏差率算法关键在于设置异常阈值，并且对组件设备的各个性能指标偏差与阈值进行比较从而得到相应结果。假如该组件设备某

个支路中参数误差小于一定阈值范围，则判定组件设备在该时间段的运行性能良好，反之则出现异常需要发出预警信息。同时阈值门限也包含数值门限和比例门限^[64]。

本章主要设计从偏差率的角度出发对光伏组串电流数据故障预警方案进行设计和实现。

4.2 预测模型选型

光伏组串电流的预测和第三章的输出功率的预测类似，两个参数都具有非线性非平稳的特点，在光伏发电过程中会受到温度、湿度、辐照度等环境因素所影响。因此要求所构建的电流预测模型也需要同第三章采用的 SVR 算法一样具有擅长处理非线性和非平稳特征问题的能力。

电流异常的预测和识别是两个不同却密切相关的电站管理任务。两者主要区别在于所依据的数据和目标。电流异常识别主要基于当前监测数据进行判断从而关注当前状态。电流异常预测是基于历史数据和时序特征来进行未来趋势的预测，其主要关注未来的状态。

基于以上两点要求可以得出，构建光伏组串电流预测模型需要兼顾过去和现在的电流数据以全面了解电流变化的变化趋势，同时针对光伏组串发电能力受气象环境影响的特点，也不能忽视环境因素对预测结果的影响。表 4.1 对几种时间序列算法的优缺点进行对比。

表 4.1 时间序列算法优缺点对比

算法名称	优点	缺点
ARMA	在一个统一框架下适用于所有时间序列。	模型形成需要基于知识的经验判定 ^[65] 且要求原始时序数据稳定。
ARIMA	模型简单，只需要内生变量而不需要借助其他外生变量。	要求时序数据稳定或差分后稳定，并且本质上只能捕捉线性关系
BP	学习速率固定，非线性映射能力强，有着柔性的网络结构。	算法收敛速度慢，结构选择不一，应用实例与网络规模存在矛盾 ^[66]
RNN	对具有序列特征的数据非常有效	只能记忆最近的序列，记忆性差 ^[67]
LSTM	具有长序列依赖处理能力，设计结构可以防止梯度消失问题	不利于并行化，效果较差 ^[68]

表中五种时间序列模型中的 ARMA（Autoregressive Moving Average Model）和 ARIMA（Autoregressive Integrated Moving Average Model）算法属于基于统计的预测模型，对平稳线性的时序数据预测效果更好，但是在复杂的时序应用中会有很大的局限性。由于光伏发电会受到气象因素的影响从而呈现非平稳非线性的特性，所以该两种模型不适合光伏组串电流的预测建模。而后三种基于神经网络的预测模型中 BP 神经网络（Backpropagation Neural Network）虽具有自学习能力，但是算法收敛速度较慢，可能会陷入局部最优并且网络层数有限。LSTM 模型是 RNN（Recurrent Neural Network）的一种特殊形式，它将 RNN 中的每个隐藏单元换成了具有记忆功能的细胞，从而避免出现梯度

消失和梯度爆炸的情况。因此在光伏组串电流故障预警模型中，选取 LSTM 为基础的时间序列模型来对光伏组串电流进行预测。

4.3 CNN-BiLSTM 模型

4.3.1 LSTM 模型

LSTM 是一种特殊的 RNN，其是专门设计用于处理序列数据的模型。相较于传统 RNN 在结构中具有多次反向传播的特性，LSTM 模型通过引入门控机制来更好地捕捉和管理序列中的长期依赖关系^[69]，缓解梯度消失或梯度爆炸等问题，改变了传统 RNN 依靠串行计算的特性，从而提升处理序列数据时的并行处理能力和训练效率。如图 4.1 为 LSTM 基本结构示意图。

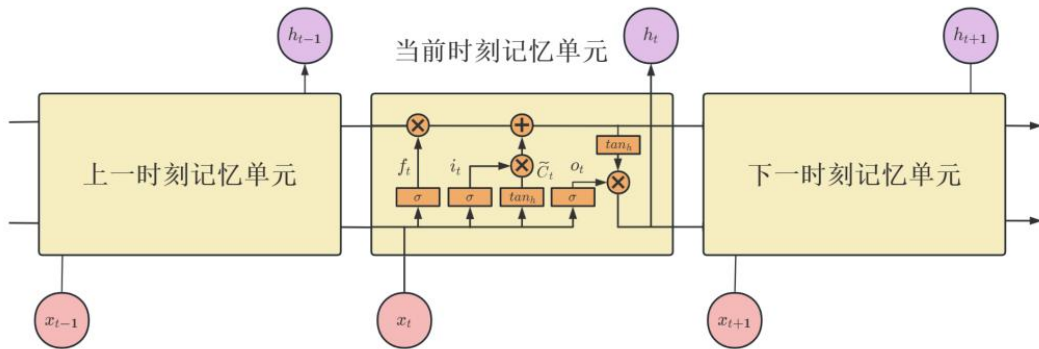


图 4.1 LSTM 基本结构示意图

LSTM 模型包含一个或多个 LSTM 单元，在每个时间步上接收输入和前一时刻的隐藏状态作为输入。每个单元由四个关键的门控单元组成：遗忘门、输入门、更新门和输出门。

(1) 遗忘门：其决定了上一层网络中单元状态将哪些不重要信息遗弃。其利用 h_{t-1} 和 x_t 之间的信息输出一个 0 到 1 的向量，向量的数值决定了 c_{t-1} 中信息是否留存，0 表示将信息丢弃，1 表示将信息保留。遗忘门的结构和计算如图 4.2 所示，式(4.1-2)所示。

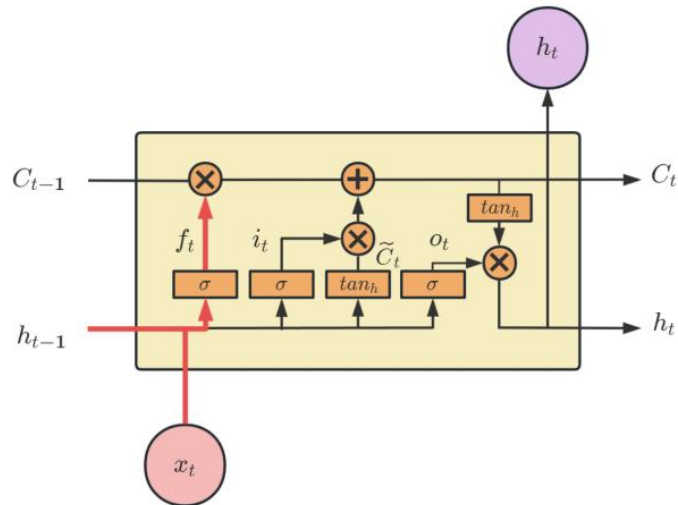


图 4.2 LSTM 的遗忘门

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4.1)$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (4.2)$$

其中 W_f 为权重矩阵， b_f 为偏置向量，遗忘门参数 f_t 由上一时刻输出 h_{t-1} 和当前时刻输入 x_t 决定， σ 为 sigmoid 激活函数。

(2) 输入门：其负责处理短期记忆，LSTM 的输入门包含了 sigmoid 层和 tanh 层两个重要组成部分，前者用于决定新信息添加的数量，后者则可根据前一个输出值 h_{t-1} 和当前输入值 x_t 生成一个新的当前状态向量，并将其加入到已有的状态矩阵中。输入门的结构和计算如图 4.3 所示，式(4.3-5)所示。

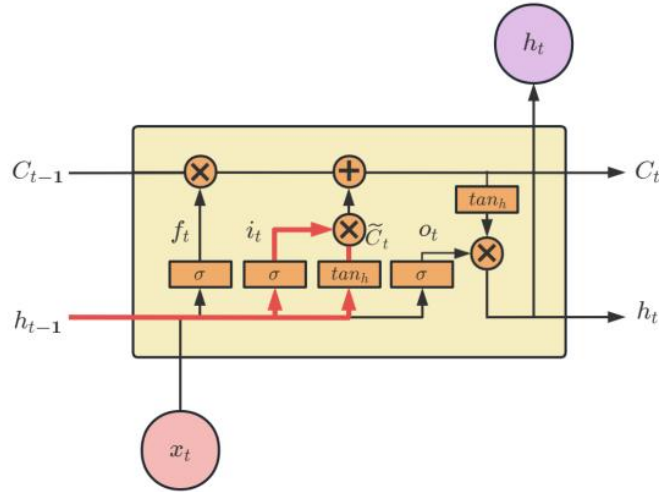


图 4.3 LSTM 的输入门

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4.3)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4.4)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4.5)$$

W_i 和 W_c 是输入门传播到 i_t 和 $\tilde{C}_t$ 时的权重， b_i 和 b_c 是传播到 i_t 和 $\tilde{C}_t$ 时的偏置， $\tanh$ 和 σ 是激活函数， $\tilde{C}_t$ 表示新单元的状态。

(3) 更新门：其主要结合遗忘门和输入门的数据进行输出，具体先遗忘之前决定要遗忘的信息，而后加上新状态的添加值。如图 4.4 所示为更新门的结构。

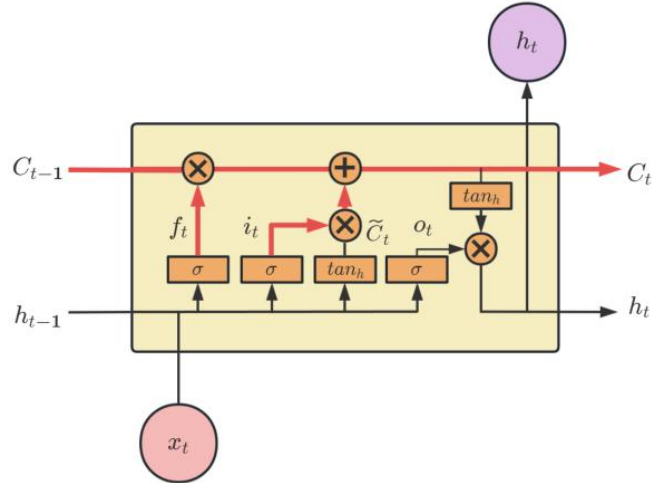


图 4.4 LSTM 的更新门

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \quad (4.6)$$

其中“ $\circ$ ”为 Hadamard 矩阵乘积运算。

(4) 输出门：其主要用于控制当前时刻的隐藏状态（记忆细胞）中的信息如何被传递到下一层或输出层。输出门由一个 sigmoid 激活函数控制，它的输入是当前时刻的隐藏状态和当前时刻的输入。该门输出的值范围在 0 到 1 之间，其中 0 表示完全阻止，1 表示完全通过。输出门的结构和计算如图 4.5 所示，式(4.7-8)所示。

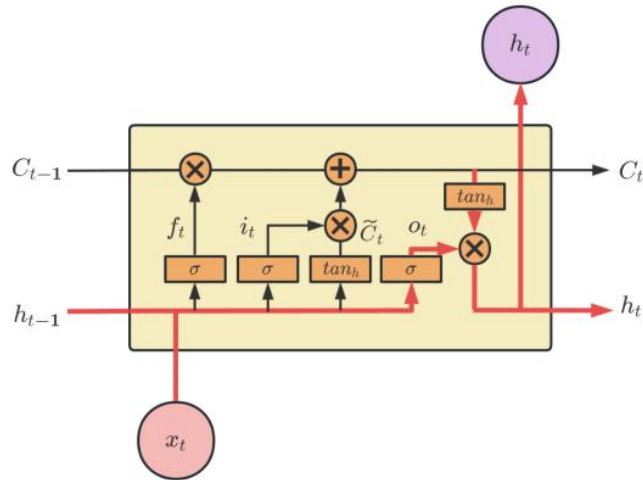


图 4.5 LSTM 的输出门

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4.7)$$

$$h_t = o_t \circ \tanh(C_t) \quad (4.8)$$

其中 W_o 为权重矩阵， b_o 为偏置向量，输出门参数 o_t 由上一时刻输出 h_{t-1} 和当前时刻输入 x_t 决定。当前时刻输出 h_t ，受当前时刻状态向量 C_t 影响。

4.3.2 双向长短时记忆神经网络原理

双向长短时记忆网络（Bi-directional Long Short-Term Memory, BiLSTM）^[70]是一种扩展的 LSTM，传统 LSTM 只能捕捉到当前时刻之前的上下文信息，而 BiLSTM 则通过同时

考虑历史上下文和未来上下文，能够更全面地理解和建模序列数据。BiLSTM 通过在时间步之间引入一个正向 LSTM 和一个逆向 LSTM 来实现双向建模，一个从前向后对输入序列中提取长期依赖特征，一个从后向前对输入序列中提取长期依赖特征^[71-73]，其具有向过去和未来进行递归和反馈的能力。训练阶段中正向和逆向 LSTM 分别处理输入序列并将它们的隐藏状态拼接在一起，从而作为输入传递给后续层或用于任务的输出。预测阶段中正向和逆向 LSTM 分别处理输入序列并将它们的隐藏状态拼接在一起，从而进行各种预测任务。

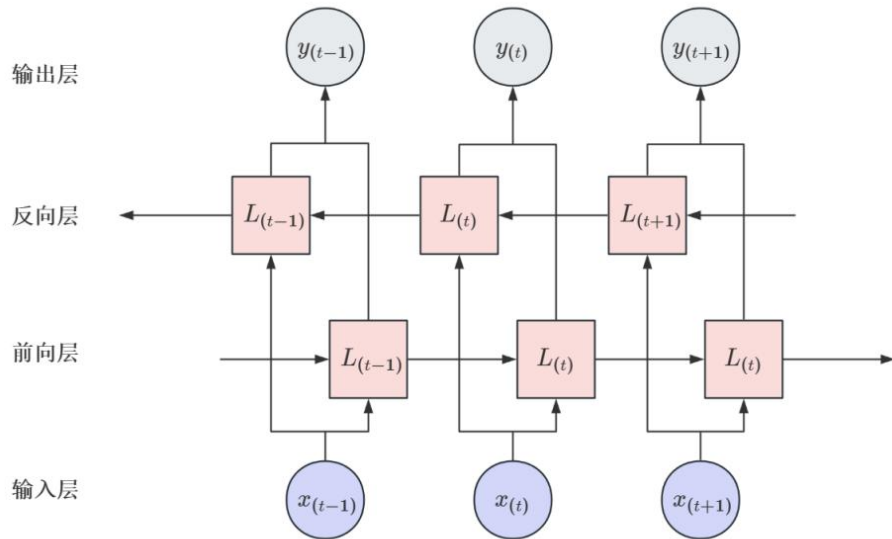


图 4.6 BiLSTM 网络结构图

如图 4.6 所示为 BiLSTM 网络结构图，其中 x 和 y 表示输入和输出， L 表示隐藏层的神经元状态。模型在计算过程中包含前向层和反向层的并行运算，两个层分别处理输入序列的正向和反向信息，处理完成后两个层的输出会在每个时间步中合并，从而为模型提供了完整的上下文信息。而后合并信息可以被用来进行各种下游任务，如分类、回归或序列生成。对于分类任务，通常会在这个特征表示上应用一个或多个全连接层，以学习并提取更高级别的抽象特征，并最终输出分类结果。

总的来说，针对时序类问题，BiLSTM 对于处理同时包含两个时间步长的数据信息具有较强的能力。因此可以通过充分挖掘数据之间时序上的联系以达到更好预测数据的效果。

4.3.3 卷积神经网络原理

卷积神经网络 (Convolutional Neural Networks, CNN)^[74]属于前馈神经网络，与传统的全连接神经网络不同，CNN 通过在输入数据上应用卷积操作来提取局部特征，并通过训练过程自动学习这些卷积操作的参数。该网络凭借其优秀的非线性拟合能力和特征提取能力，能构有效提取二维图像和高维数据的能力，已经被广泛应用到诸多领域中^[75]，可用于解决一些复杂的预测问题。该模型主要特点是权值共享和局部连接。由于同一个卷积核在不同位置上的权值是相同的，从而可以大大减少模型参数。而卷积层中每个卷积核只与输入数据的一部分进行卷积运算，而不是与整个输入数据进行卷积运算，从而使其具有提取

数据特征、降低模型复杂度^[76]和内存消耗等优势。因此本章结合卷积神经网络搭建光伏组串电流预测模型。其具体流程如图 4.7 所示。

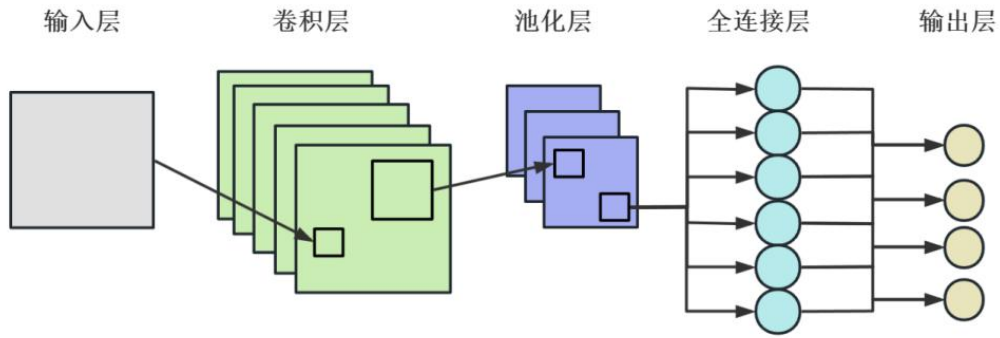


图 4.7 卷积神经网络结构图

(1) 卷积层

卷积层通过卷积操作提取特征，每个卷积层通常包括多个卷积核，每个卷积核对输入数据进行卷积运算，得到一个特征图。同时卷积核自带纠正线性单元（Rectified Linear Unit, ReLU）作为激活函数^[77,78]来完成非线性计算。卷积神经网络中，卷积层内的神经元只能与上一层的一部分神经元相互连接，形成了一种稀疏连接结构。卷积层中权重参数和偏置参数共同塑造了神经元之间的连接关系和网络的输出行为。权重参数决定了输入数据与卷积核之间的线性组合方式。偏置参数则为每个卷积核的输出增加了一个可学习的常数项，允许模型在没有输入或者输入很小时也能产生非零的激活，从而提供了输出的灵活性和调整空间。卷积层性能由卷积核个数与大小、移动步长等共同决定。

$$y_j^l = f \left(\sum_{i=1}^s x_j^{l-1} * W_{ij}^l + b_j^l \right) \quad (4.9)$$

$$ReLU(x) = \max(0, x) \quad (4.10)$$

其中 y_j^l 为第 l 层的第 j 个特征映射输出， f 为激活函数， s 为输入特征映射数量； x_j^{l-1} 为第 $l-1$ 层的第 j 个映射向量， W_{ij}^l 为可训练的卷积核， b_j^l 为偏置^[79]。

(2) 池化层

池化层的主要功能是通过降采样来减少特征图的尺寸，增强模型的鲁棒性和特征提取能力。常用的池化操作包括最大池化和平均池化，它们分别取局部区域中的最大值和平均值作为输出。本章采用最大池化对区域内特征点提取最大值，即提取最明显的特征，该方法适合提取区域内的详细信息。其具体操作如图 4.8 所示。



图 4.8 最大池化操作示意图

（3）全连接层

全连接层主要是通过将特征映射到一个高维特征空间中，再通过 softmax 函数进行分类或回归。在训练过程中，通常采用反向传播算法来更新网络参数，以最小化损失函数。在回归任务中，全连接层的输出结果可以是一个连续值，如位置坐标、物体大小等。通过调整权重矩阵和偏置项，全连接层可以学习到输入特征与回归结果之间的关系。

（4）输出层

输出层的作用是对前面层级中提取到的特征进行最终的分类、回归或其他预测任务。

本章在 BiLSTM 模型的基础上加入 CNN 模型，其主要利用 CNN 的特征提取能力，在构建预测模型过程中从原始的电流数据中提取出有效的特征，从而对数据信息进行深度挖掘。并且通过参数共享和稀疏连接来减少模型的参数数量，提高了模型的效率和泛化能力。

4.4 基于 CNN-BiLSTM 组合光伏组串电流预测模型的实现

图 4.9 为一种基于 CNN-BiLSTM 的光伏组串电流预测流程图。在该组合模型中，首先采用一维卷积层对输入进行特征提取。然后提取后的特征将会输入至池化层，池化层采用最大池化法对区域网格中最大的数据作为输出从而起到特征压缩作用。随后将 CNN 提取的特征通过 Dropout 层按照时间步长传输到单层 BiLSTM 网络中，通过网络中独特的门控单元来提取数据中的时序关系并学习和建立一种双向时间上的拟合关系。最后数据通过 Dropout 层连接全连接层并通过全连接层对加强数据特征，输出模型的预测结果。

CNN-BiLSTM 模型训练流程：

- （1）将数据集进行归一化处理，并按照 8:2 的比例划分为训练集和测试集；
- （2）设置 CNN-BiLSTM 网络模型的初始化参数，设置迭代次数。
- （3）训练 CNN-BiLSTM 网络模型，循环执行以下步骤直到收敛或达到最大迭代次数：首先将训练数据输入 CNN-BiLSTM 进行前向传播，计算预测值。然后计算损失函数 Loss 并且反向传播计算梯度。在根据梯度和学习率更新网络参数。最后检查是否满足收敛条件，如果不满足则继续迭代，否则停止训练。
- （4）最后将测试集输入 CNN-BiLSTM 进行前向传播预测，得到电流预测结果。

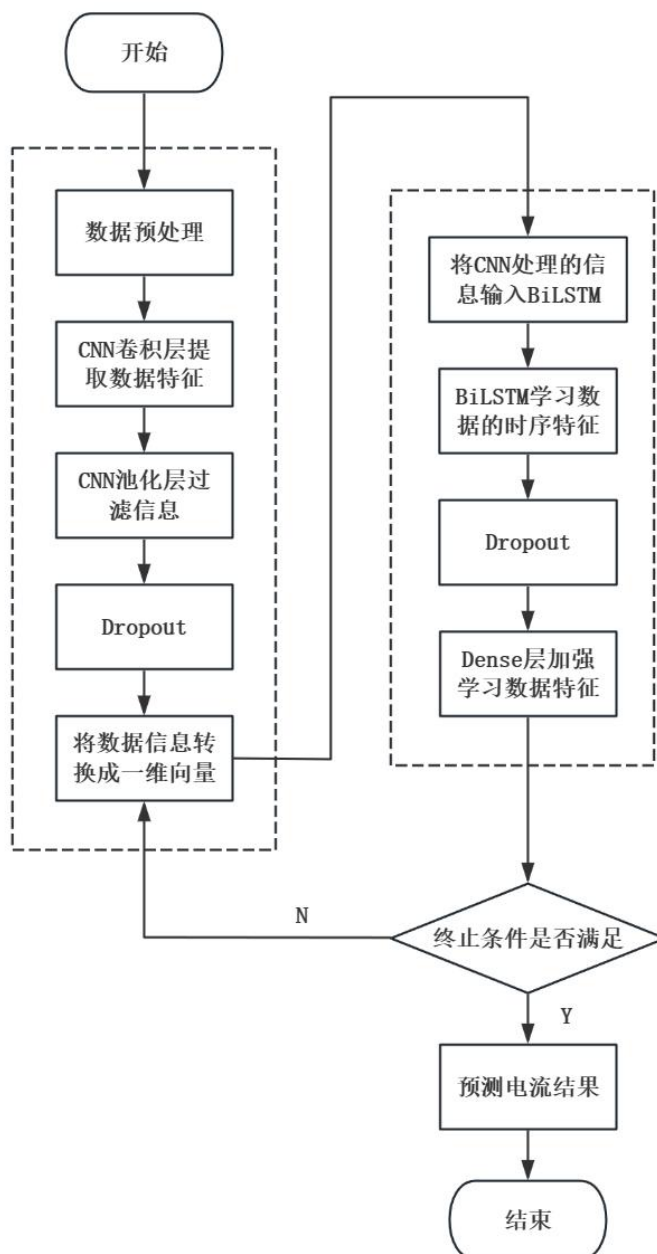


图 4.9 CNN-BiLSTM 预测流程图

4.5 实验测试与结果分析

4.5.1 数据预处理

本章依旧以上文提到的江苏某 1490KW 光伏电站为例，采用的数据是系统提供的 2023 年 1 月 1 日到 2023 年 12 月 31 日的历史数据。光伏组串电流预测的数据集中包含的属性为温度、辐照度、湿度以及对应时刻的光伏组串电流。具体时间范围同第三章一样选取白天 7 点到晚上 7 点的数据，一天共 144 个采样点，并且对数据进行预处理。具体数据一共 22530 组。为了验证 CNN-BiLSTM 模型在不同天气情况下的预测精度，从历史数据中筛选晴天、阴天和雨天的数据集进行分别训练。

4.5.2 模型评估指标

本章选用第三章所提到的平均绝对误差（MAE），均方根误差（RMSE）和复判定系数（ R^2 ）验证预测模型的有效性。

4.5.3 模型初始化

本章基于 CNN-BiLSTM 的光伏组串电流预测模型使用 CNN 对数据进行空间特征信息提取，使用 BiLSTM 提取数据的时间特征信息，集成 CNN 特征提取的优点与 BiLSTM 能够彻底挖掘数据时序特征的优点，使两者互补，实现预测精准度的提升。模型每次输入四个时刻的数据，预测下一时刻的数据。表 4.2 表示模型主要参数的初始化设置。

表 4.2 预测模型参数设置

模型	参数类型	参数设置
CNN-BiLSTM	CNN	1 维卷积
	卷积核	2
	滤波器数	32
	BiLSTM	1 层
	优化器	Adam
	迭代次数	500
	学习率	0.005
	激活函数	Relu
	损失函数	MSE

本章采取 Adam 作为模型的优化器。其能够自动调整学习率，适应不同参数的变化情况，提高模型的收敛速度和稳定性。此外，Adam 也对学习率进行了偏差修正，使得初始阶段的学习率较大从而有利于快速收敛，后续逐渐减小从而有利于精细调整。本章模型采用 Dropout 层来防止过拟合。在模型训练时，每层后面均会添加一个 Dropout 层，对应参数值设置为 0.25。Dropout 层会广泛应用于每个隐藏阶段，并且在每次前向传播和后向传播时，以一定概率随机失活一些节点，并进行参数更新，从而提高模型的泛化能力和稳定性。

4.5.4 实验仿真和分析

为验证本文提出 CNN-BiLSTM 预测算法的有效性，仿真实验同第三章均在 PyCharm2022.1.1 和 Python3.8 的环境下进行，并将 CNN-BiLSTM 与 LSTM^[80]和 CNN-LSTM 预测模型作为实验对照组，并且分别选取 2023 年 9 月 6 日、2023 年 5 月 17 日和 2023 年 7 月 14 日作为晴天、阴天和雨天的预测日，观察预测模型是否可以准确预测不同天气情况下电流运行状态，从而更好验证模型的准确性。预测结果及误差如下图所示。

（1）晴天预测结果分析

如图 4.10 所示为晴天情况三种模型的预测结果。通过观察图中的预测曲线，从可以直观得出三种模型在晴天的条件下预测结果较好，预测曲线和实际值重合程度高。说明三种模型受到晴天的环境因素影响小，晴天的环境特点使得数据预测更具有规律性。其中 CNN-BiLSTM 模型与实际值更为接近，说明该模型表现出更高的预测准确性。

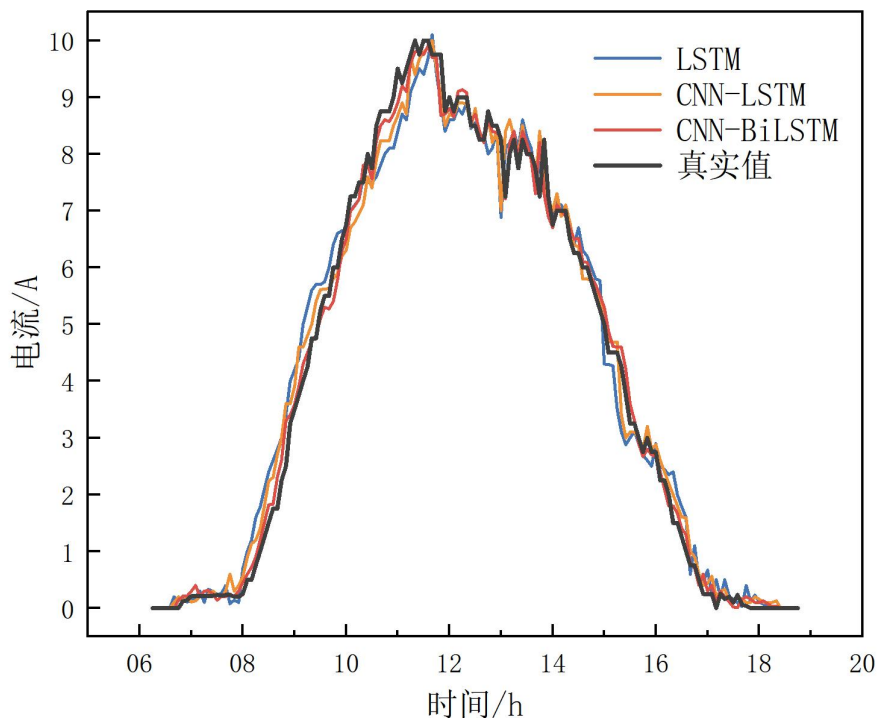


图 4.10 晴天情况三种模型的预测结果（2023/09/06）

表 4.3 晴天情况三种模型的评价指标

预测模型	MAE	RMSE	R ²
LSTM	0.347	0.423	0.9313
CNN-LSTM	0.281	0.363	0.9601
CNN-BiLSTM	0.268	0.331	0.9766

上表 4.3 所示为晴天情况三种模型的评价指标。首先对于 MAE 指标，CNN-BiLSTM 模型具有最小的预测误差，CNN-BiLSTM 相较于 LSTM、CNN-LSTM 分别减少了 22.7% 和 4.6% 的误差。针对 RMSE 指标来说，CNN-BiLSTM 模型的预测误差值也是最小的，误差分别比 LSTM、CNN-LSTM 减少了 21.7% 和 8.8%。最后 CNN-BiLSTM 模型计算得出的指标 R² 相较另外两种模型更大，指标值为 0.9766。因此可以得出 CNN-BiLSTM 在晴天情况下的预测精度更为卓越。

（2）阴天预测结果分析

如图 4.11 所示为阴天情况三种模型的预测结果，仅从网关实时上传的电流真实值曲线可以看出阴天情况相较于晴天情况组串电流受环境影响起伏较大，从普通的 LSTM 模型预测曲线看出单一模型面对该天气情况存在较大误差。而 CNN-BiLSTM 预测模型在阴天情况下相较于另外两种模型更接近组串电流的真实值。

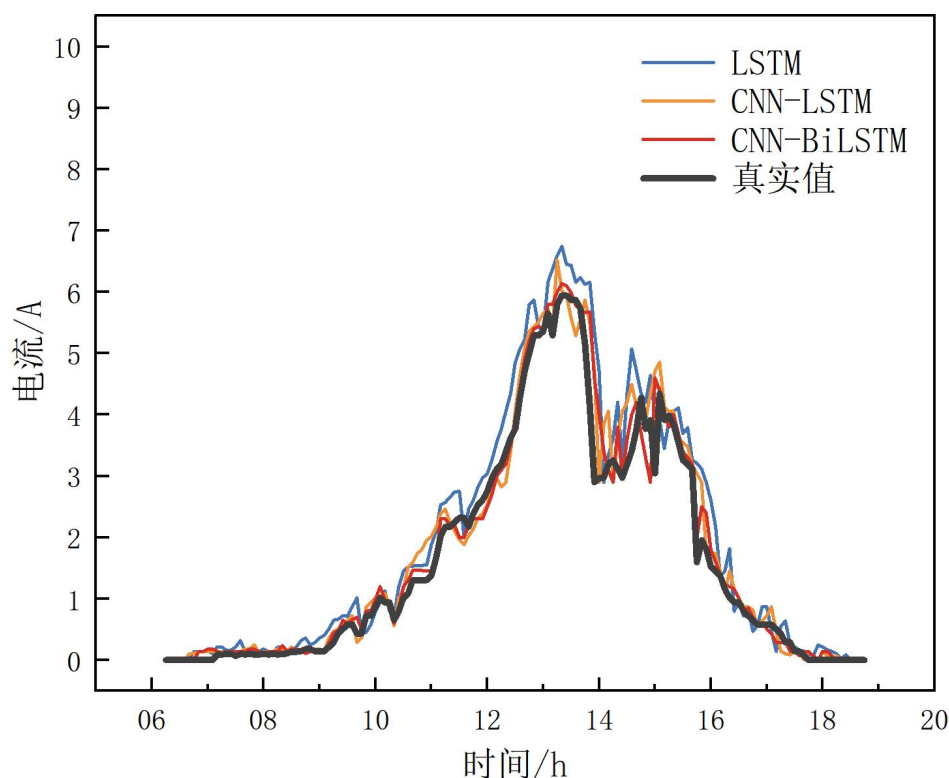


图 4.11 阴天情况三种模型的预测结果（2023/05/17）

表 4.4 阴天情况三种模型的评价指标

预测模型	MAE	RMSE	R ²
LSTM	0.684	0.912	0.8460
CNN-LSTM	0.511	0.721	0.8974
CNN-BiLSTM	0.393	0.475	0.9352

表 4.4 所示为阴天情况三种模型的评价指标。针对 MAE 指标来说，CNN-BiLSTM 预测模型相较于另外两种模型指标值最小，CNN-BiLSTM 和 LSTM、CNN-LSTM 相比分别减少了 42.5%和 23.1%的误差。针对 RMSE 指标来说，误差分别比 LSTM、CNN-LSTM 减少了 47.9%和 34.1%。因此在阴天情况下三种模型中较为突出的还是 CNN-BiLSTM，并且两中组合模型比 LSTM 单一模型具有更小的预测误差。

（3）雨天预测结果分析

如图 4.12 所示为雨天情况三种模型的预测结果，雨天天气下太阳辐射会被云层吸收，整个光照时间会比晴天短。组串电流值较低并且波动更密集，部分波形会受到雨量影响。由下图可以得出雨天情况下，LSTM 和 CNN-LSTM 模型的预测曲线波动较大，而组合模型 CNN-BiLSTM 在预测过程中与真实值较为接近。

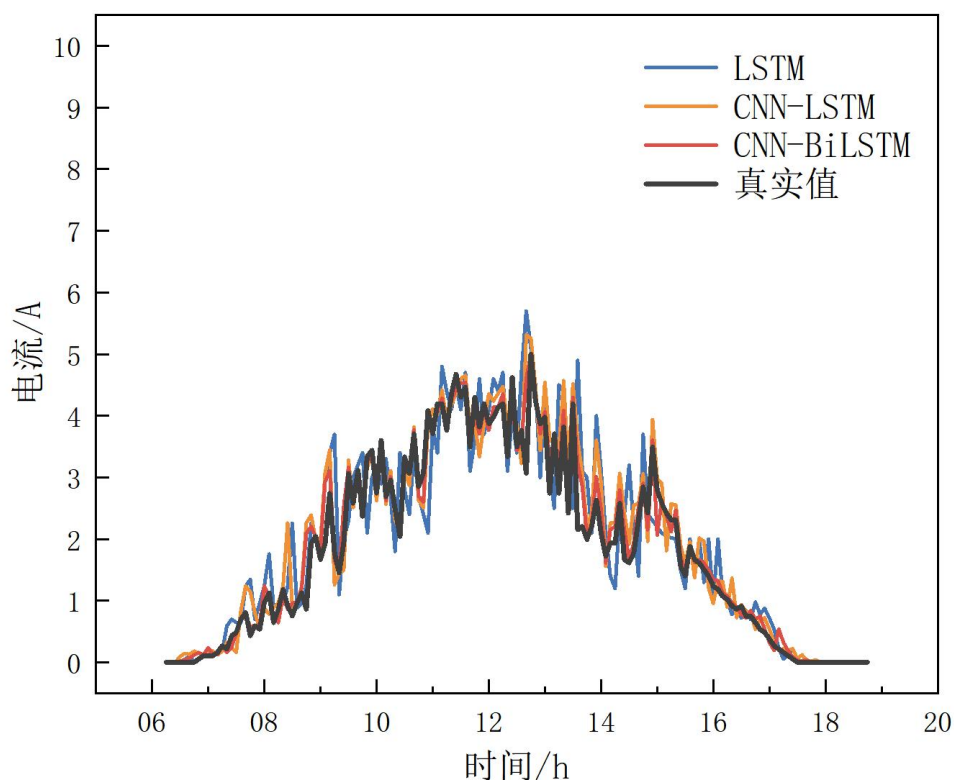


图 4.12 雨天情况三种模型的预测结果 (2023/07/14)

表 4.5 雨天情况三种模型的评价指标

预测模型	MAE	RMSE	R ²
LSTM	0.567	0.730	0.8544
CNN-LSTM	0.442	0.573	0.9152
CNN-BiLSTM	0.361	0.458	0.9403

如表 4.5 所示为雨天情况三种模型的评价指标。通过上表可以看出采用 CNN-BiLSTM 模型所得到的 MAE 和 RMSE 两个指标值相较于另外两种单一模型最小, R² 最大且达到了 0.9403。针对 MAE 指标来说, CNN-BiLSTM 和 LSTM、CNN-LSTM 相比分别减少了 36.3%和 18.3%的误差。针对 RMSE 指标来说, 误差分别比 LSTM、CNN-LSTM 减少了 37.2%和 20.1%。因此可以得出雨天情况下 CNN-BiLSTM 预测模型具有较好的预测精度。

通过以上三种不同天气情况下的预测误差汇总分析可知, 在光伏组串电流预测方面, CNN-BiLSTM 模型展现出了良好的表现, 预测误差较小, 并且具有强大的高维特征提取能力和对时间序列良好的预测能力, 体现光伏系统中组串电流的运行过程中的内在联系与规律。本章对电流异常的判断是结合性能指标偏差与预置的门限参数比较而得出的, 根据上述三种天气情况的预测可发现当前预测模型对不同天气状态预测能力较好, 对不同天气状况所带来的误差影响在可接受范围内可忽略, 所以本章采用的基于 CNN-BiLSTM 的预测模型构建不需要根据天气情况划分模型来进行分别预测。

4.6 故障预警规则制定与实例验证

4.6.1 偏差率

光伏组串电流数据的预测模型是在组件稳定运行的前提下建立的，上文在晴天、阴天和雨天三种不同天气情况下构建的模型预测中可以发现，当组件处于正常工作状态下，智能网关采集上来的电流实际值和模型的预测值误差很小。因此可以通过对两个值的比较来展现出组串电流的异常趋势，从而能对光伏组件发生故障前实时预警。本章提出一种光伏组串电流的偏差率指标 dev 。公示(4.11)如下：

$$dev = \left| \frac{Y_{true} - Y_{model}}{Y_{model}} \right| \quad (4.11)$$

公式中 Y_{pred} 是模型输出的光伏组串电流预测值， Y_{true} 是通过光伏逆变器采集的实际值。

当被预测的组件处于正常运行状态时，假如模型预测能力精确稳定，预测值和实际值的预测误差越小，计算出来的偏差率指标也会很低。而当所预测的组件处于异常运行状态时，其实际值与预测值之间的差异会更加明显，并且偏差率指标会显著上升，此时需要实时上报故障预警。下图 4.13、图 4.14、图 4.15 依次对上文的三种天气情况下的预测值进行偏差率计算和分析。

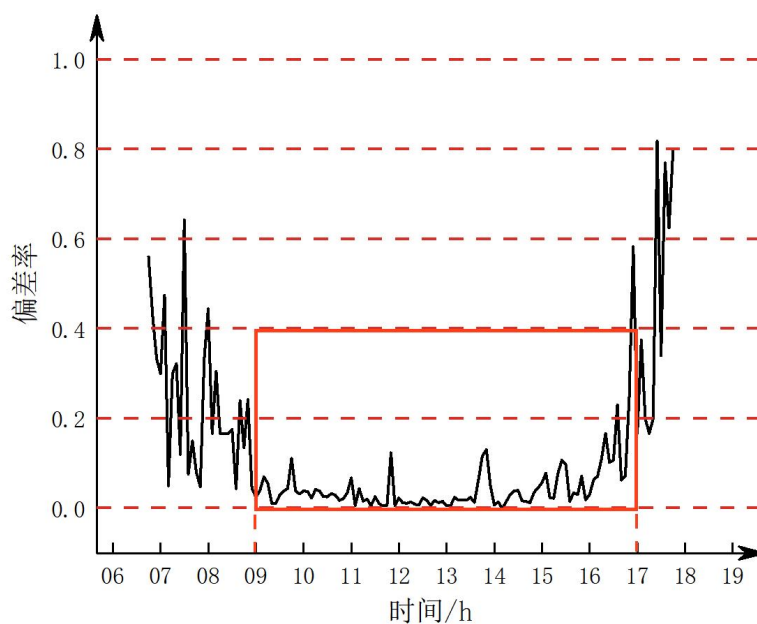


图 4.13 正常组件晴天偏差率

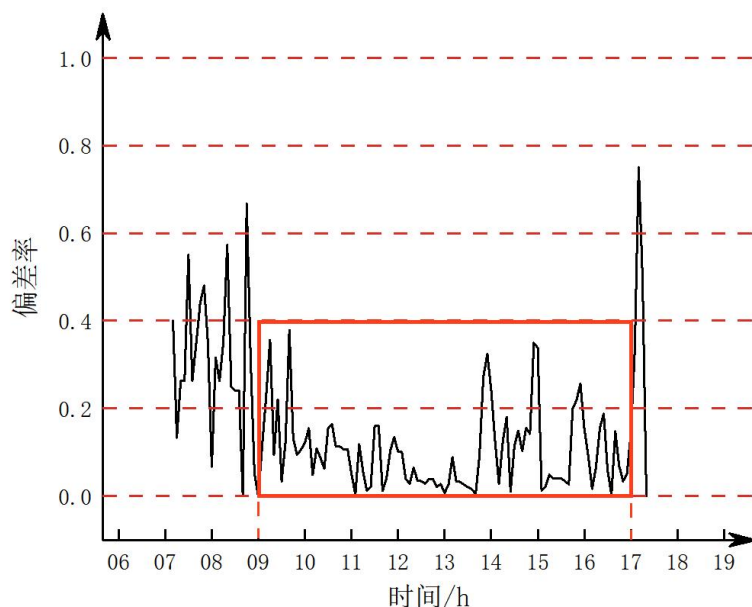


图 4.14 正常组件阴天偏差率

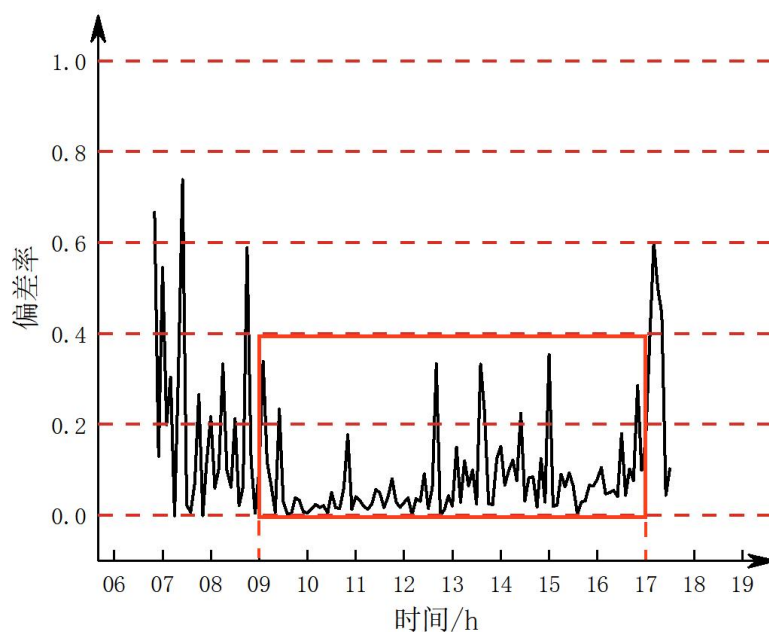


图 4.15 正常组件雨天偏差率

根据三种情况下的偏差率曲线图可以发现，偏差率曲线在早上 9 点前和傍晚 5 点后抖动严重，在这两段时间模型预测误差较大，偏差率甚至会达到 0.6 到 0.8 之间。而偏差率曲线在 9 点到傍晚 5 点之间趋于稳定，偏差率在晴天情况下基本可以稳定在 0.2 以下，在阴天和雨天情况下偶尔会波动到 0.2 到 0.4 之间。这种毛刺现象的发生基本是由于环境因素的波动引起偏差率曲线出现起伏，然而这种波动很快会下降至 0.2 以下并且重新趋于稳定。所以先设定具体预警规则的适用时间范围在每日的早上 9 点到傍晚 5 点之间，这一段时间是光伏组件的主要发电时间段，而其余时间段模型预测效果不好并且发电量少，因此忽略对于光伏组件异常电流的判断。

上文提到传统偏差率算法核心思想是将设备内部不同的性能指标和预先设置的的阈值参数进行比较，根据比较结果做出相应的预警判断。经上图分析设定偏差率等于 0.2 为一级故障预警阈值，等于 0.4 为二级故障预警阈值。若低于阈值则为正常，超出则为异常。图 4.16 为故障预警的阈值示意图。

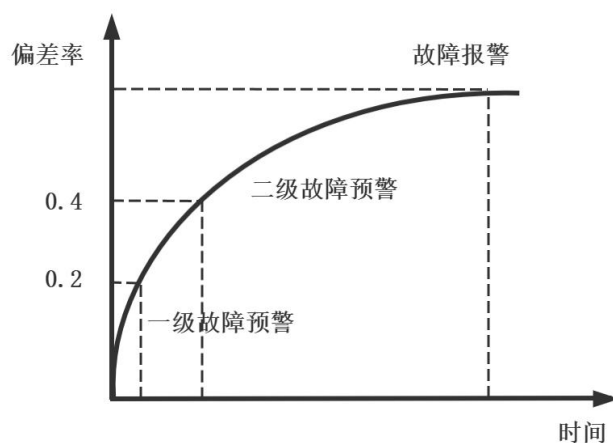


图 4.16 故障预警阈值图

但是如果云平台计算出来的组串电流偏差率一旦超过预警阈值就进行预警推送，就会忽略到上文提到的气象条件带来的毛刺影响，恶劣天气情况下的环境因素很可能会对正常组件的偏差率带来毛刺抖动的影响，甚至会超过设定的二级故障预警阈值，所以显然一超过预警阈值就进行推送的方案不可取，需要更进一步改进。

4.6.2 滑动窗口优化

滑动窗口方法可用于对时间序列数据进行分析。在进行某一段电流分析时，滑动窗口可以将整个时间序列分割成多个固定大小的窗口，并在每个窗口内进行特征提取和分析。这种方法可以有效地捕捉电流数据的局部模式和变化趋势。滑动窗口的优势在于它能够处理连续、时变的电流数据，并充分利用时间上的相关性。

滑动窗口主要由窗口大小和步长这两个属性组成，每次确定窗口长度大小后，会向后滑动一个步长的长度后，再次划分出新的窗口，同一条电流信息可能会出现在多个窗口内。如图 4.17 所示，滑动窗口大小为 3，步长为 1，则长度为 6 的电流序列可以划分为 4 组序列，即 $\{\Delta t_1, \Delta t_2, \Delta t_3, \Delta t_4\}$ 。

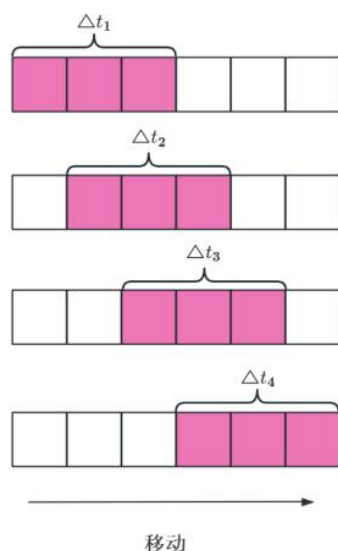


图 4.17 滑动窗口原理图

系统中智能网关每间隔 5 分钟会将光伏逆变器检测到的各个支路的组件的组串电流数据。因此偏差率计算的间隔时间也设为 5 分钟。但上文提到组串电流偏差率不能一超过预警阈值就进行预警推送，所以基于滑动窗口思想进行优化，将包含当前时刻在内的前 1 小时的电流数据作为一组序列，由于每隔 5 分钟计算因此一个窗口有 12 组序列数据。每个支路的组件的组串电流偏差率计算完都会存入到数据库中。

当某一时刻窗口内的 12 组电流序列的偏差率均超过 0.2 时则进行一级故障预警，偏差率均超过 0.4 时则进行二级故障预警。这一优化能极大程度解决因为某一时刻的气象条件引起的毛刺抖动所造成的误警率。由于预警规则的适用时间范围从每日的早上 9 点开始，即真正滑动窗口计算偏差率从每日 10 点开始。

4.6.3 实例验证

为了直观的验证本设计的可靠性，下表统计 2023 年 1 月 1 日到 2023 年 12 月 31 日智能网关所实时监测到的 12 台光伏逆变器对各个支路的所有报警信息，如表 4.6 所示。

表 4.6 故障内容和出现次数统计表

故障内容	出现次数
设备异常网络离线	25
PV 绝缘故障	8
PV 接地故障	36
无电网	21
直接母线不均压	5
组串过流	23
直流过压	51

选取 PV 绝缘故障、PV 接地故障、直接母线不均压、组串过流、直流过压五种与光伏组件有关的 123 次报警信息，筛选各支路中处于正常运行状态的组件作为该支路的参照物，

采用该组件 2023 年 1 至 12 月的组串电流数据作为训练集，构建该支路组串电流预测模型，预测上述选取的各类组件故障至故障前 1 个月的组串电流并与实际电流计算得到偏差率。下图 4.18 和图 4.19 展示光伏阵列第 3 支路电流在 11 月 15 日发生 PV 接地故障报警前的预警偏差率曲线。

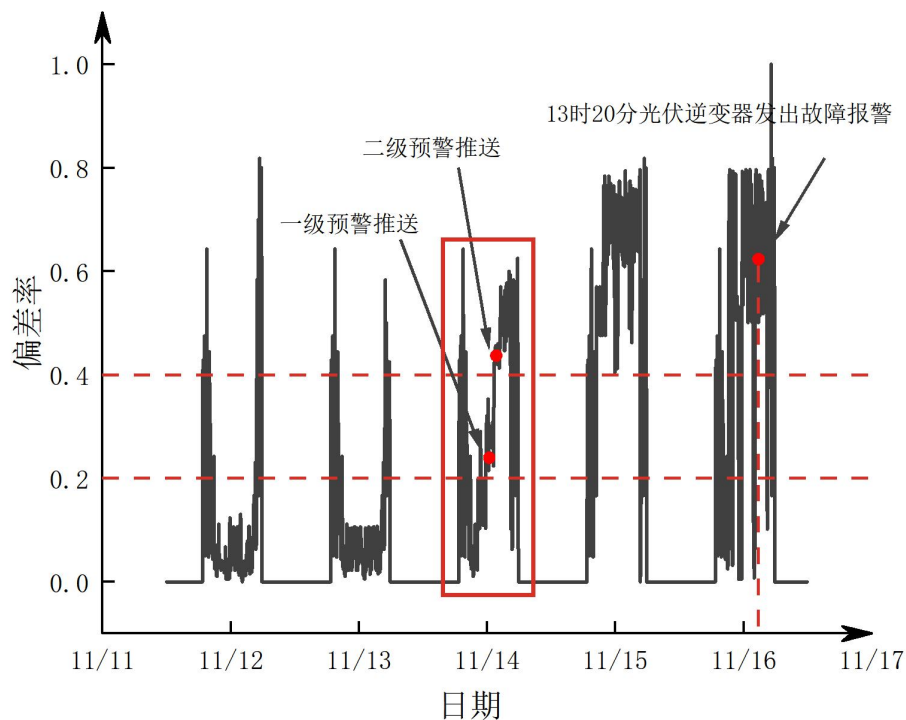


图 4.18 故障周期内预警偏差率曲线

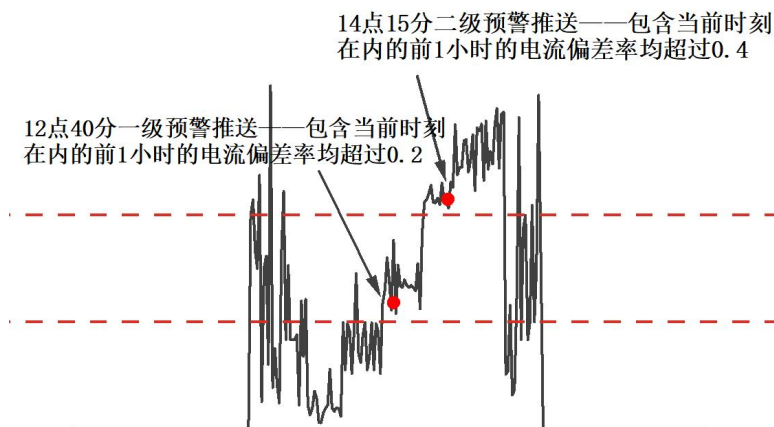


图 4.19 11 月 14 日预警偏差率曲线

通过实例分析，上图显示出逆变器在 11 月 15 日发出正式的报警信息，15 日和 18 日的偏差率基本在 0.6 到 0.8 之间。12 日和 13 日的偏差率除去清晨和傍晚预测的偏差率波动之外，其余的光伏出力的时间段基本在 0 到 0.2 之间。11 月 14 号偏差率抖动上升，在 12 点 40 分时达到了一级推送的标准，每间隔 5 分钟进行预测即 11 点 40 分到 12 点 40 分之间的 12 个预测点偏差率均超过了 0.2。在 14 点 15 分时达到了二级推送的标准，每间隔 5 分钟进行预测即 11 点 15 分到 12 点 40 分之间的 12 个预测点偏差率均超过了 0.4。而后 11

月 15 日和 16 日的偏差率基本在 0.6 到 0.8 之间。光伏逆变器在 11 月 15 日发出正式的报警信息。因此针对光伏阵列第 3 路电流的实验成功预警，证明了该预警方案在实际应用中的实用性。

对上文提到总共与光伏组件有关的 123 次报警信息进行预警检验。具体如表 4.7 所示。

表 4.7 检验预警率

故障内容	出现次数	二级预警次数	二级预警率
PV 绝缘故障	8	6	75%
PV 接地故障	36	34	94.44%
直流母线不均压	5	5	100%
组串过流	23	22	95.65%
直流过压	51	48	94.11%

根据表中数据得出 PV 接地故障、组串过流和直流过压三种故障情况的预警率都超过了 90%，说明这两种故障与组串电流的偏离率有极大关系。PV 绝缘故障和直流母线不均压两种情况由于样本出现次数过少，参考价值不大。因此可以得出在与光伏组串电流有关的故障内容中，PV 接地故障、组串过流和直流过压等故障内容可以通过本章设计的预测组串电流计算偏差率的方案做到提前预警，在实际应用中具有较高的预警率，从而能在组件故障真正发生前起到预测异常趋势的作用。

4.7 本章小结

本章针对光伏组串电流的时序特征和发电过程中受到气象条件表现而产生的不确定性，提出一种基于 CNN-BiLSTM 的组合模型进行电流预测，从空间和时间角度对数据进行特征提取，进而提高组串电流的预测精度。并分别采用 LSTM、CNN-LSTM、CNN-BiLSTM 模型在晴天、阴天和雨天种天气类型下进行组串电流的预测对比。然后构建 CNN-BiLSTM 的电流预测模型，通过比较网关实时上报的实际值与模型分析的预测值从而计算出电流偏差率，通过对偏差率设定阈值结合滑动窗口的方法进行电流异常的预警。最后通过测试实验得到该方法在与光伏组串电流有关的故障内容中能起到良好的预警效果，在报警信息上报之前就能帮助运维人员及时发现电流异常，进行预测性的维护工作。

第 5 章 总结与展望

5.1 总结

本文研究了光伏电站运维系统发展现状和运维过程中遇到的技术问题和需求痛点，设计了一种结合物联网和机器学习的分布式光伏远程运维系统。该系统具有实时推送、远程升级、清洁决策以及故障报警等多种功能。为了解决以上提到的问题，本文提出了一种基于改进 GS 的 SVR 的清洁决策方案，以及提出了一种基于 CNN-BiLSTM 的光伏组件故障预警方案。本文的主要工作如下：

(1) 本文设计并实现了一个分布式光伏电站远程运维系统。硬件部分主要采用以 Air724UG 为内核的通信模块来对光伏逆变器为核心的多个设备端进行实时的数据交互。后端组件根据功能和数据特点选用 MySQL、RabbitMQ、Redis 和 Nacos 等作为后端中间件，并且通过容器化部署保证系统响应速度和容灾能力。业务云平台主要采取微服务架构，设计了远程升级服务、实时推送服务、常规业务服务和数据分析服务等各类服务。远程升级采用 Fork-join 框架进行并行处理优化服务响应时间。实时推送服务在消息队列中间件上进行改进，保证了数据推送的实时性和完整性。后台将光伏设备关键参数和环境参数数据传输给 Web 端交互并展示，同时可以通过微信小程序和发送短信等方式向运维人员发起故障报警、故障预警和清洁提醒等信息，进而以便于实现实时监控和远程操作的功能。最终，实现了一个稳定可靠、功能完备的分布式光伏电站远程运维系统。

(2) 针对现如今大部分分布式光伏电站清洁周期往往基于估计或静态设置，并且无法对组件表面积灰程度量化分析的问题，本文提出一种光伏无积灰理想发电量算法思想并且筛选电站历史清洁周期内的功率数据作为不受积灰影响的功率数据，然后通过实际功率来计算组件发电能力从而分析积灰带来的影响，根据实际发电量来计算日发电量损耗值并与清洁成本对比来动态确认电站清洁周期。本文具体研究了温度、湿度、辐照度等气象因素对光伏输出功率的影响，为后台分析算法提供了理论依据，并且通过相关性分析选择合适的气象因素作为输入。接着构造数据集使用 SVR 算法来训练无积灰功率模型。针对 SVR 模型参数寻优过程中的敏感问题，设计改进的 GS 来进行优化，从而大幅提高模型泛化能力和预测精度，并且减少了预测时间。

(3) 针对现如今光伏预警功能薄弱，更多靠实时报警获取故障信息并且无法预防或减轻潜在的风险等问题，本文以电流偏差率的角度出发，设计并实现了一种基于 CNN-BiLSTM 的组串电流预测模型来进行故障预警。然后对晴天、阴天、雨天三种天气情况进行光伏组串电流的预测，该算法在预测的精度上高于 LSTM、和 CNN-BiLSTM。最后根据偏差率指标和滑动窗口优化等方法建立报警规则，实现光伏组件的故障预警。

5.2 展望

本文分析了光伏运维系统的现有不足，并且从完善光伏清洁和故障预警的角度出发，结合物联网、微服务和机器学习等多种技术，设计了一个功能完善的分布式光伏运维系统。但是由于本人水平有限，论文研究仍然存在一些不足，后续希望可以进行一次改善：

（1）目前的光伏运维系统中由于实际成本的限制，尚未部署后台监控系统去监控各个中间件的运行情况，因此整个系统的稳定性有待提升。后续系统迭代过程中可以部署 Zabbix 等监控环境来保证服务器系统的安全运行。

（2）本文清洁决策在最初提出无积灰理想发电量的概念过程中将最佳清洁效果周期设定为 3 天，该周期设定还可以继续考究，不同地区环境因素不同，因此周期设定应具有差异性，后续需要加强实验证明。而且本文是采用改进的 GS 优化 SVR 算法来建立无积灰功率模型，由于 SVR 算法存在内存占用高，运行时间长等缺点，后续可以考虑不同深度学习算法建立模型进行性能比较。

（3）目前对于光伏运维系统的预警部分，无论是离散率分析还是偏差率分析，很多异常判断实质都是对光伏组串电流值的异常检测。但是光伏逆变器上报的故障信息种类众多，本文设计的故障预警方案能在与光伏组串电流有一定关系的故障内容类别进行提前预警，但是其他故障类型适用程度不高。对此后续可以从其他指标去考虑其他故障信息的预警方案。

致谢

时间总是一晃眼，三年时光转瞬即逝。人生不过由一段段经历构成，而我最后的学生时代也将拉下帷幕，回顾当年考研上岸时的欣喜到现在，研究生三年是我度过最难忘的三年。相较于本科时期的稚嫩，读研带给我的经历更加宝贵。如今我将踏入社会，预示着我度过了人生最青春的年华，纵有万般不舍，但仍然心怀感激。

首先感谢我的导师陈科明教授，陈老师严谨的治学态度和深厚的工程能力深深地影响到我，其在工业物联网领域深耕多年，无私地将自身多年的独到见解传授于学生，拓宽了我们的眼界和格局。老师的教诲让我在思考问题时能逐渐跳出狭隘的学生思维，学习以第三人称视角去锻炼自身的解决问题能力，并且在各个项目中发挥自己的才能创造价值。在本次论文撰写过程中，陈老师也从选题和撰写过程中给了我悉心教导。在此谨向尊敬的陈老师表达我最诚挚的敬意和感激。

其次感谢我的父母，父亲总是不善言辞但是总在关键的时候想方设法为我提供帮助，母亲爱唠叨但是重复不知多少遍的话语从她嘴里说出总能在我最失落的时候给我力量。你们一直是我最大的支持者和鼓励者。养育之恩，无以为报，我将尽我所能努力成为你们的骄傲。在此，祝愿我的父母及家人们能幸福安康、健康长寿。

接着感谢我的同窗朋友。感谢同门叶宗昆、王从舸和雷宇素，三年以来我们互帮互助，互相成就。感谢师兄师姐在秋招过程的帮助，特别感谢徐志杰师兄，工作繁忙之余给予我很大的指点。感谢一路以来身边帮助过我的朋友，在我困惑落寞时开导我，骂醒我。感谢初入职场帮助过我的前辈们，是你们的指导让我不断成长。在此祝大家都有美好的未来。

然后感谢一个女孩，我们走过一段很美好的时光，祝她考公顺利上岸，越来越好。

另外我要感谢不是很厉害但一路走来不容易的自己，感谢硬着头皮上以解决问题态度处理事情的自己。我从未放弃。一直以来我都是一个自卑、纠结和犹豫的人，我总需要他人的开导才能看清点什么，但三年以来我也逐渐从面对困难时一如既往地质疑自己，慢慢变得能在最开始为自己打气且相信问题总能解决。我总有慕强思想，但我也逐渐放下总与他人比较的态度与自己和解，因为回头看去我已经比三年前的自己厉害不少了。如今的我马上踏入社会，希望自己能继续努力，珍惜眼前抓得住的，去见识一切新事物，享受人生不同阶段带给自己的体验和收获。

还有感谢摇滚乐和说唱音乐，许多独特的经历在我累得喘不过气的瞬间给予我力量。

最后感谢论文评阅和答辩委员会的老师，感谢你们给予的耐心以及宝贵的意见。感谢培养教育我的杭州电子科技大学，感谢母给予我锻炼学习的平台。

至此，由衷的祝愿所有人，工作顺利，万事如意！

参考文献

- [1] 丁怡婷.太阳能发电装机突破 6 亿千瓦[N].人民日报,2024-01-31(002).
- [2] 曹林,王静,王宝友等.太阳光伏能源技术及其标准化[J].信息技术与标准化,2011(Z1):30-34.
- [3] 杨洪雷.集中式光伏电站与分布式光伏电站工程建设管理方法的思考[J].上海节能,2023(11):1719-1723.
- [4] 董伯先.浅析光伏电站清洁机器人的应用现状[J].太阳能,2020(03):46-50.
- [5] 侯治国,张光奎,莫鹏辉等.光伏组件自动清洗系统的应用研究[J].太阳能,2023(11):80-87.
- [6] 周亚男.光伏电站运维现状分析[J].太阳能,2024(01):12-19.
- [7] 冯相赛.分布式光伏电站运维现状及发展趋势[J].能源与节能,2018(04):57-59.
- [8] 杨金焕,于化丛,葛亮.太阳能光伏发电应用技术[M].电子工业出版社,2009.
- [9] Masson G, Kaizuka I, Lindahl J, et al. A snapshot of global PV markets-the latest survey results on PV markets and policies from the IEA PVPS Programme in 2017[C]. 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC). IEEE, 2018: 3825-3828.
- [10] Meteocontrol[EB/OL].<https://www.meteocontrol.com/>.
- [11] Taoke 第三方光伏监控数据服务平台.光伏监控平台[EB/OL].[2014-1-2].<https://www.lvsedianli.net/>.
- [12] 钟银燕.用大数据平台促进光伏业理性繁荣[N].中国能源报,2015-04-13(004).
- [13] 李帅.特变电工的泛在新战场[J].能源,2019(07):70-72.
- [14] 曲宏伟,王靖雯.积灰对光伏板输出特性影响理论和试验研究[J].太阳能学报,2018,39(08):2335-2340.
- [15] Said S A M, Hassan G, Walwil H M, et al. The effect of environmental factors and dust accumulation on photovoltaic modules and dust-accumulation mitigation strategies[J]. Renewable and Sustainable Energy Reviews, 2018, 82: 743-760.
- [16] 张朝辉,石祎炜,刘晓豹等.光伏电站中光伏组件的清洗效果分析及清洗周期预测[J].太阳能,2022(12):53-61.
- [17] 张姝伟.基于视觉的光伏板积灰状态监测与性能评估[D].东北电力大学,2019.
- [18] Unluturk M, Kulaksiz A A, Unluturk A. Image processing-based assessment of dust accumulation on photovoltaic modules[C]. 2019 1st Global Power, Energy and Communication Conference (GPECOM). IEEE, 2019: 308-311.

- [19] Abuqaoud K A, Ferrah A. A novel technique for detecting and monitoring dust and soil on solar photovoltaic panel[C]. 2020 Advances in Science and Engineering Technology International Conferences (ASET). IEEE, 2020: 1-6.
- [20] 张博,黄帅,丛伟伦等.基于最小二乘支持向量回归的光伏电站清洗时间动态优化方法[J].太阳能学报,2021,42(09):55-61.
- [21] 陈金鑫,潘国兵,欧阳静等.自然降雨下光伏组件积灰预测方法研究[J].太阳能学报,2021,42(02):431-437.
- [22] Matthias Heinrich, Simon Meunier, Allou Samé, et al, Bernard Multon. Detection of cleaning interventions on photovoltaic modules with machine learning[J]. Applied Energy, 2020: 263.
- [23] Silvestre S, da Silva M A, Chouder A, et al. New procedure for fault detection in grid connected PV systems based on the evaluation of current and voltage indicators[J]. Energy Conversion and Management, 2014, 86: 241-249.
- [24] Andrianajaina T, Sambatra E J R, Andrianirina C B, et al. PV fault detection using the least squares method[C]. 2016 International Conference and Exposition on Electrical and Power Engineering (EPE). IEEE, 2016: 846-851.
- [25] Dhimish M, Holmes V. Fault detection algorithm for grid-connected photovoltaic plants[J]. Solar Energy, 2016, 137: 236-245.
- [26] Platon R, Martel J, Woodruff N, et al. Online fault detection in PV systems[J]. IEEE Transactions on Sustainable Energy, 2015, 6(4): 1200-1207.
- [27] Jufri F H, Oh S, Jung J. Development of Photovoltaic abnormal condition detection system using combined regression and Support Vector Machine[J]. Energy, 2019, 176: 457-467.
- [28] 吴晨红.基于 Modbus 通信协议的信号采集系统[D].合肥工业大学,2021.
- [29] 任瑶瑶,马金平.基于 RS485 的多主机对等通信系统的设计与实现[J].山东大学学报(工学版),2023,53(05):156-164.
- [30] Alshammari H H. The internet of things healthcare monitoring system based on MQTT protocol[J]. Alexandria Engineering Journal, 2023, 69: 275-287.
- [31] Li M, Amin A, Torrea R, et al. Processor based 20Mhz 4×4 Cat-5 LTE MIMO receiver with advanced detectors[C]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 2669-2673.
- [32] 熊群毓.大数据时代 MySQL 数据库的应用分析[J].信息与电脑(理论版),2023,35(14):209-212.
- [33] 丁岚,商毛红,冯振颖.基于 MHA 的 MySQL 高可用方案研究[J].电脑编程技巧与维护,2019(01):59-60+89.
- [34] 肖顺,严碧波.一种基于 RabbitMq 的消息推送系统的设计与实现[J].电子世界,2019(07):204-205.

- [35] Almeida D, Lopes M, Saraiva L, et al. Performance Comparison of Redis, Memcached, MySQL, and PostgreSQL: A Study on Key-Value and Relational Databases[C]. 2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon). IEEE, 2023: 902-907.
- [36] 司琦,李炜毅,蒲鹏.云原生环境基于缓存的配置中心设计与实现[J].软件,2023,44(10):16-20.
- [37] 李民靖,郝东来.基于 OneNET 的家居智能晾衣架系统设计[J].物联网技术,2023,13(11):68-71.
- [38] 金海峰,余傲雪.Nginx 负载均衡器集群架构的实践应用[J].安徽电子信息职业技术学院学报,2023,22(03):1-6.
- [39] Guan J C, Sun Q, Li T, et al. Design and Implementation of Operation and Maintenance Monitoring Platform Based on Data Mining Algorithm[C]. International Conference on Frontier Computing. Singapore: Springer Nature Singapore, 2023: 167-172.
- [40] Serdar Yegulalp. 8 great little Python web frameworks[J]. InfoWorld.com, 2020, 55(24): 28-32.
- [41] 邱建东,李清狮,陈桥等.光伏组件清洗周期研判技术的现状分析与展望[J].科技与创新,2023(15):98-100.
- [42] 岑先富,朱超林.光伏组件积尘对发电量的影响及自动清洁经济效益研究[J].太阳能,2013(21):45-48.
- [43] 林志鸿,曾飞,王哲等.灰尘对光伏电站发电性能的影响[J].科技创新与应用,2022,12(12):90-93.
- [44] Pulipaka S, Mani F, Kumar R. Modeling of soiled PV module with neural networks and regression using particle size composition[J]. Solar Energy, 2016, 123: 116-126.
- [45] W. Han, H. H. Wang, D. U. Wei. Short-Term Photovoltaic Power Forecasting Based on Elman Neural Network with Fruit Fly Optimization Algorithm[J]. Electrical Measurement & Instrumentation, 2014, 2119-2125.
- [46] Hammad B, Al-Abed M, Al-Ghandoor A, et al. Modeling and analysis of dust and temperature effects on photovoltaic systems' performance and optimal cleaning frequency: Jordan case study[J]. Renewable and Sustainable Energy Reviews, 2018, 82: 2218-2234.
- [47] 黄山峰,姜凯,顾志兴等.基于新能源集控平台的光伏组件清洗分析系统的设计[J].太阳能,2023(04):78-83.
- [48] 陈璐,张天文,刘杰斌等.光伏电站组件动态清洗周期的研究[J].中国高新科技,2019(24):6-8.

- [49] Tan M, Liao C, Chen J, et al. A multi-task learning method for multi-energy load forecasting based on synthesis correlation analysis and load participation factor[J]. *Applied Energy*, 2023, 343: 121177.
- [50] Pena J C, Nápoles G, Salgueiro Y. Normalization method for quantitative and qualitative attributes in multiple attribute decision-making problems[J]. *Expert Systems with Applications*, 2022, 198: 116821.
- [51] Vanneschi L, Silva S. Support vector machines[M]. *Lectures on Intelligent Systems*. Cham: Springer International Publishing, 2023: 271-281.
- [52] Wang H, Shao Y. Sparse and robust SVM classifier for large scale classification[J]. *Applied Intelligence*, 2023, 53(16): 19647-19671.
- [53] 许鸣吉, 孙歌, 徐焰等. 基于 PSO-SVR 模型的光伏功率预测研究[J]. *电力与能源*, 2023, 44(02): 150-156.
- [54] Bai Z. Stock Trend Prediction Based on Improved SVR[C]. *International Congress on Communications, Networking, and Information Systems*. Singapore: Springer Nature Singapore, 2023: 148-154.
- [55] Ali Y A, Awwad E M, Al-Razgan M, et al. Hyperparameter search for machine learning algorithms for optimizing the computational complexity[J]. *Processes*, 2023, 11(2): 349.
- [56] Ni L P, Ni Z W, Li F G. SVM Model Selection Based on Ant Colony Algorithm[J]. *Computer Technology and Development*, 2007, 17(9): 95-98.
- [57] Liu J X, Cai H P, Tan Y J. Heuristic algorithm for tuning hyperparameters in support vector regression[J]. *Journal of System Simulation*, 2007, 19(7): 1540-1534.
- [58] Lin H, Chen Y, et al. Prediction of shear strength of rock fractures using support vector regression and grid search optimization[J]. *Materials Today Communications*, 2023, 36: 106780.
- [59] 朱忠雷, 陈晓武, 李淞. 电站光伏组件离散率与热斑研究[J]. *冶金自动化*, 2023, 47(S1): 596-600.
- [60] Abd E, Hossam A, Ahmed E, et al. A new monitoring technique for fault detection and classification in PV systems based on rate of change of voltage-current trajectory[J]. *International Journal of Electrical Power & Energy Systems*, 2021, 133: 107248.
- [61] De Benedetti M, Leonardi F, Messina F, et al. Anomaly detection and predictive maintenance for photovoltaic systems[J]. *Neurocomputing*, 2018, 310: 59-68.
- [62] 宋海明. 基于非线性状态估计的多媒体设备故障预警研究[J]. *佳木斯大学学报(自然科学版)*, 2020, 38(05): 126-129.

- [63] Choi W H, Kim J, Lee J Y. Development of Fault Diagnosis Models Based on Predicting Energy Consumption of a Machine Tool Spindle[J]. *Procedia Manufacturing*, 2020, 51:353-358.
- [64] 谢骊骊,叶青,陈伟等.光伏电站在线监测智能诊断系统设计与实现[J].*太阳能*,2017(01):16-21+27.
- [65] Sansa I, Boussaada Z, Mazigh M, et al. Solar radiation prediction for a winter day using ARMA model[C]. 2020 6th IEEE International Energy Conference (ENERGYCon). IEEE, 2020: 326-330.
- [66] Liu B. Review of swarm intelligence algorithm optimization of BP neural network [J]. *Academic Journal of Computing & Information Science*, 2023, 6(6): 151-155.
- [67] Malhotra R, Singh P. Recent advances in deep learning models: a systematic literature review[J]. *Multimedia Tools and Applications*, 2023, 82(29): 44977-45060.
- [68] Chen H, Chang X. Photovoltaic power prediction of LSTM model based on Pearson feature selection[J]. *Energy Reports*, 2021, 7: 1047-1054.
- [69] Kim D, Kwon D, Park L, et al. Multiscale LSTM-based deep learning for very-short-term photovoltaic power generation forecasting in smart city energy management[J]. *IEEE Systems Journal*, 2020, 15(1): 346-354.
- [70] Ullah F U M, Ullah A, Haq I U, et al. Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks[J]. *IEEE Access*, 2019, 8: 123369-123380.
- [71] Shi H, Miao K, Ren X. Short-term load forecasting based on CNN-BiLSTM with Bayesian optimization and attention mechanism[J]. *Concurrency and Computation: Practice and Experience*, 2023, 35(17): e6676.
- [72] 谢小瑜,周俊煌,张勇军等.基于 W-BiLSTM 的可再生能源超短期发电功率预测方法[J].*电力系统自动化*,2021,45(08):175-184.
- [73] Cai W, Liu B, Wei Z, et al. TARDB-Net: triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification[J]. *Multimedia Tools and Applications*, 2021, 80: 11291-11312.
- [74] Shiri F M, Perumal T, Mustapha N, et al. A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU[J]. *arxiv e-prints*, 2023: arxiv: 2305.17473.
- [75] Sobri S, Koohi-Kamali S, Rahim N A. Solar photovoltaic generation forecasting methods: A review[J]. *Energy conversion and management*, 2018, 156: 459-497.

- [76] Yuan Q, Wang J, Zheng M, et al. Hybrid 1D-CNN and attention-based Bi-GRU neural networks for predicting moisture content of sand gravel using NIR spectroscopy[J]. Construction and Building Materials,2022,350(10):1-15.
- [77] Gupta D, Ramana V V. Solar Power Prediction Based on Recurrent Neural Networks Using LSTM and Dense Layer With ReLU Activation Function[C]. 2023 3rd International Conference on Intelligent Technologies (CONIT). IEEE, 2023: 1-5.
- [78] Kiranyaz S, Avci O, Abdeljaber O, et al. 1D convolutional neural networks and applications: A survey[J]. Mechanical systems and signal processing,2021,151(1):1-21.
- [79] 范志远,杜江.基于相关变分模态分解和 CNN-LSTM 的变压器油中溶解气体体积分数预测[J].高电压技术,2024,50(01):263-273.
- [80] 石小兵.基于数据挖掘的光伏阵列异常检测与预警[D].山东大学,2019.

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] 光伏电站远程运维与智能分析系统开发. 横向课题. 2021.9-2023.7.
- [2] 矿山机械远程运维系统开发. 横向课题. 2022.11-2023.6.

发表的专利：

- [1] 陈科明,王佳敏,孔秀婷等.基于光伏发电量和光照辐射量的光伏组件清洁决策方法及系统[P].浙江省:CN202311827095.X,2024-03-29.