

分类号:
学号: 20222108010

密级: 公开
单位代码: 10759

石河子大学

硕士学位论文



基于云计算的电力负荷预测与异常检测方法 研究与实现

学位申请人	杨一坤
指导教师	陈文斌 高工 李志刚 教授
申请学位类别	专业硕士
专业名称	电子信息
研究领域	计算机技术
所在学院	信息科学与技术学院

中国·新疆·石河子

2025年5月

分类号:
学号: 20222108010

密级: 公开
单位代码: 10759

石河子大学

硕士学位论文



基于云计算的电力负荷预测与异常检测方法 研究与实现

学位申请人	杨一坤
指导教师	陈文斌 高工 李志刚 教授
申请学位类别	专业硕士
专业名称	电子信息
研究领域	计算机技术
所在学院	信息科学与技术学院

中国·新疆·石河子
2025年5月

**Research and Implementation of Electric Power Load Forecasting and
Anomaly Detection Methods Based on Cloud Computing**

A Dissertation Submitted to

Shihezi University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Engineering

By

Yang Yi-Kun

(Computer Technology)

Dissertation Supervisor:

Sn Engr Chen Wen-bin

Prof. Li Zhi-gang

May, 2025

石河子大学学位论文独创性声明及使用授权声明

学位论文独创性声明

本人所提交的学位论文是在我导师的指导下进行的研究工作及取得的研究成果。据我所知，除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示谢意。

研究生签名：

杨一坤

时间： 2025 年 5 月 27 日

使用授权声明

本人完全了解石河子大学有关保留、使用学位论文的规定，学校有权保留学位论文并向国家主管部门或指定机构送交论文的电子版和纸质版。有权将学位论文在学校图书馆保存并允许被查阅。有权自行或许可他人将学位论文编入有关数据库提供检索服务。有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

研究生签名：

杨一坤

时间： 2025 年 5 月 27 日

导师签名：

李志刚

时间： 2025 年 5 月 27 日

摘要

深入分析与挖掘电力大数据对构建新型电力系统及保障电网运行的高效、安全和稳定至关重要。尤其在电力负荷预测与异常检测这两个关键领域，其研究进展直接关系到电力系统的优化运行和故障预防能力。本文聚焦于电力负荷预测与异常检测方法的研究，并取得了以下主要研究成果：

(1) 本文提出了一种结合随机采样重组策略的混合卷积-多层感知器网络的用户用电量预测方法。针对非平稳且非线性的用户电力数据，传统的时间步逐次输入方法难以有效挖掘深层次的时间序列特征。为此，本研究引入了随机采样重组策略，将原始时间序列随机分割为多个片段并重新组合，从而增强了数据的多样性和时间序列特征的可提取性。在特征提取阶段，首先利用卷积操作从重组后的时间序列数据中捕获细致的局部特征，这有助于识别复杂的模式。随后，采用多层感知器来捕捉整个时间序列的全局特征。通过上述特征提取与融合步骤，最终实现了对用户用电量的精确预测。最后，进行了详尽的实验验证，实验结果表明，本文所提方法在预测精度上表现出显著优势，在 4 种预测长度上 MAE 分别提升 1.69%、1.28%、3.58%和 4.59%。

(2) 本文解决了用电数据集中类别不平衡的问题，并提出了一种改进的混合卷积-多层感知器网络方法用于电力负荷异常检测。首先，针对用电数据中正常用户与异常用户之间显著不平衡的现象，提出了欠采样与过采样结合的混合重采样方法，以优化样本分布并提升模型训练效果。随后，采用改进的混合卷积-多层感知器网络对处理后的数据进行重建。在此基础上，设计了一个异常分数模块，将重建结果转化为具体的异常分数，用以量化评估用户的用电异常行为。最后，进行了详尽的实验验证，实验结果表明，本文所提方法 AUC 值为 0.9867。

(3) 为了使电力公司工作人员直观的了解用户的用电情况，本文设计并实现了一个基于云计算的负荷管理平台，该平台集成了数据管理、任务调度以及数据可视化等核心功能模块。平台的数据管理模块采用 Doris 对电力负荷数据进行分布式数据存储和处理，确保了大规模电力负荷数据的高效管理和一致性维护。任务调度模块则采用 DolphinScheduler 数据同步和调度工具，其支持多种复杂的任务并行执行。数据可视化模块采用 Superset 将复杂的数据分析结果转化为易于理解的视觉信息，帮助决策者快速识别关键趋势和异常情况。综上所述，本文所构建的基于云计算的负荷管理平台不仅显著增强了电力公司对用户用电情况的理解与掌控能力，也为电力系统的智能化管理和优化运行提供了强有力的技术支撑。

关键词：预测；异常检测；云计算；电力负荷

Abstract

In-depth analysis and mining of power big data is crucial to the construction of new power systems and ensure the efficient, safe and stable operation of power grids. Especially in the two key fields of power load prediction and anomaly detection, the research progress is directly related to the optimal operation and fault prevention ability of power system. This thesis concentrates on the research of power load forecasting and anomaly detection methodologies, achieving the following principal research outcomes:

(1) This thesis proposes a hybrid convolution-multilayer perceptron network combined with random sampling recombination strategy to predict user electricity consumption. For the non-stationary and nonlinear user power data, the traditional time-step successive input method is difficult to effectively mine the deep time series characteristics. To this end, this thesis introduces a random sampling recombination strategy, which randomly splits the original time series into multiple segments and recombines them, thus enhancing the diversity of data and the extractability of time series features. In the feature extraction stage, the convolution operation is first used to capture the fine local features from the reorganized time series data, which helps to identify complex patterns. Subsequently, a multi-layer perceptron is employed to capture the global features of the entire time series. Through the above feature extraction and fusion steps, the accurate prediction of user electricity consumption is finally realized. The experimental results show that the proposed method shows significant advantages in prediction accuracy, and the MAE of the four prediction lengths is increased by 1.69%, 1.28%, 3.58% and 4.59%, respectively.

(2) This thesis solves the problem of class imbalance in the electricity data set, and proposes an improved hybrid convolution-multilayer perceptron network method for power load anomaly detection. Firstly, in view of the significant imbalance between normal users and abnormal users in the electricity data, a hybrid resampling method combining undersampling and oversampling is proposed to optimize the sample distribution and improve the training effect of the model. Subsequently, the improved hybrid convolution-multilayer perceptron network is used to reconstruct the processed data. On this basis, an abnormal score module is designed to convert the reconstruction results into specific abnormal scores, which can be used to quantitatively evaluate the abnormal behavior of users. Finally, detailed experimental verification is carried out. The experimental results show that the AUC value of the proposed method is 0.9867.

(3) To facilitate the intuitive understanding of user electricity consumption by power company staff, this thesis designs and implements a cloud-based load management platform that integrates core functional modules including data management, task scheduling, and data visualization. Specifically, the data management module employs Doris for distributed storage and processing of power load data, ensuring efficient management and consistency maintenance of large-scale datasets. The task scheduling module utilizes DolphinScheduler, an advanced data synchronization and orchestration tool, to support the parallel

execution of complex tasks, thereby optimizing resource allocation and workflow automation. For data visualization, Apache Superset is employed to transform intricate data analysis outcomes into readily comprehensible visual representations, such as charts and dashboards, enabling decision-makers to swiftly identify critical trends and anomalies. In summary, the cloud-based load management platform developed in this thesis significantly enhances the power company's capability to understand and control user electricity consumption. Furthermore, it provides robust technical support for the intelligent management and optimized operation of the power system. This thesis lays a solid foundation for innovative applications within the power industry under the context of big data environments, offering both theoretical insights and practical advancements.

Key words: forecasting; anomaly detection; cloud computing; electric power load

目录

摘要.....	I
Abstract.....	II
第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 电力负荷预测研究现状.....	2
1.2.2 电力负荷异常检测研究现状.....	4
1.2.3 云计算在电力系统中的研究现状.....	6
1.3 研究内容与技术路线.....	6
1.3.1 研究内容.....	6
1.3.2 技术路线.....	7
1.4 本文组织结构.....	8
第 2 章 相关理论及技术.....	9
2.1 电力负荷预测及异常检测相关概念.....	9
2.1.1 电力负荷特性.....	9
2.1.2 电力负荷预测的分类.....	10
2.1.3 电力负荷数据异常机理.....	11
2.2 预测相关方法.....	12
2.2.1 传统统计模型.....	12
2.2.2 基于深度学习的预测方法.....	13
2.3 异常检测相关方法.....	15
2.3.1 基于距离的异常检测方法.....	16
2.3.2 基于重建的异常检测方法.....	16
2.3.3 基于预测的异常检测方法.....	17
2.4 云计算相关技术.....	18
2.4.1 虚拟化技术.....	18
2.4.2 分布式存储.....	19
2.5 本章小结.....	21
第 3 章 基于混合卷积-多层感知器网络的负荷预测方法研究.....	22

3.1 引言.....	22
3.2 负荷预测模型构建.....	23
3.2.1 序列分解.....	23
3.2.2 随机采样重组.....	24
3.2.3 卷积-多层感知器混合结构.....	25
3.3 数据集介绍及预处理.....	27
3.3.1 数据集介绍.....	27
3.3.2 数据集预处理.....	28
3.4 实验设置与结果分析.....	29
3.4.1 实验设置.....	29
3.4.2 评价指标.....	29
3.4.3 结果分析.....	30
3.4.4 消融实验.....	34
3.5 本章小结.....	35
第4章 基于数据重采样的卷积-多层感知器网络的负荷异常检测方法研究.....	36
4.1 基于重采样的数据集平衡方法研究.....	36
4.1.1 数据不平衡问题.....	36
4.1.2 过采样方法.....	37
4.1.3 欠采样方法.....	39
4.1.4 混合采样方法.....	40
4.1.5 重采样数据质量分析.....	40
4.2 负荷异常检测模型构建.....	42
4.2.1 模型概述.....	42
4.2.2 异常检测模块.....	43
4.3 实验设置与结果分析.....	44
4.3.1 实验设置.....	44
4.3.2 评价指标.....	44
4.3.3 结果分析.....	45
4.4 本章小结.....	49
第5章 基于云计算的负荷管理平台设计与实现.....	50
5.1 云计算实验平台环境.....	50
5.2 需求分析.....	50
5.2.1 功能性需求.....	50
5.2.2 非功能性需求.....	51

5.3 平台技术架构.....	52
5.4 平台开发实现.....	53
5.4.1 数据存储模块实现.....	53
5.4.2 任务调度模块.....	54
5.4.3 可视化模块.....	55
5.5 平台展示.....	56
5.5.1 整体信息展示.....	56
5.5.2 预测结果展示.....	57
5.5.3 异常检测结果展示.....	58
5.6 本章小结.....	59
第 6 章 结论与展望.....	60
6.1 结论.....	60
6.2 未来展望.....	61
参考文献.....	62
致谢.....	68
作者简介.....	69

第1章 绪论

1.1 研究背景及意义

电力系统是现代社会的基础设施，我国目前发电装机总容量、远距离输电能力、电网规模、全社会用电量等指标均稳居世界第一。截至2024年10月底，全国发电装机容量约31.9亿千瓦，其中，太阳能发电装机容量约7.9亿千瓦，风电装机容量约4.9亿千瓦，风力光伏发电合计装机（12.8亿千瓦）已超过煤电装机（11.7亿千瓦）^[1]。随着大规模新能源电力接入电网，其结构形态、运行控制方式以及规划建设与管理发生根本性变革，形成了以新能源电力生产、传输、消费为主体的新一代电力系统，即新型电力系统。随着大数据、云计算、人工智能技术等计算机技术在传统配电网中的日益广泛应用，电力行业也正在经历深度转型，分布式智能电网逐步走向成熟，未来将形成“分布式”与“大电网”并驾齐驱的局面，新型电力系统的运行模式也正朝向源网荷储互动的方向转变，不同类型的电气设备的大规模采用以及可再生能源发电快速集成的新挑战给智能电网系统带来了新的挑战。

新型电力系统整合了信息通信技术（Information and Communication Technology, ICT）、控制、传感和测量、新型储能、自动化、分布式发电、可再生能源集成等传统电网不具备的智能技术领域。广泛部署的信息和通信基础设施充当这些智能技术的骨干，使源网荷储之间的双向数据流成为可能。通过实施先进的计量基础设施（Advanced Metering Infrastructure, AMI）技术，将实现双向电力和信息通信^[2]，为新型电网提供近乎实时监测、显示消费者能源消耗数据的能力，并支持动态电价谈判和负载控制等多种应用，进一步衍生出对电力负荷的精准预测、配网韧性增强、电力的灵活调配等新的需求。

随着我国市场经济体制的不断完善，电力工业正逐步向市场化转型，这对负荷预测的准确性、实时性和可靠性提出了更高的要求。新型电力系统的建设推动了电力数据的管理分析和终端智能交互的发展。新型电力系统需要在随机波动的负荷需求与发电端之间实现能量的供需平衡，因此负荷预测作为智能电网电力数据分析的核心问题之一，对于提高电力系统的可靠性和稳定性具有重要作用。随着电力市场改革的持续深化与结构性演进，一系列创新技术和新兴市场参与者不断被引入并深度整合到电力系统的运作框架中。这些创新包括需求响应机制、共享经济模式、负荷聚合服务以及虚拟电厂等概念^[3]。这一进程显著增强了电力负荷特性的复杂性和动态性，使其呈现出前所未有的多维

性和变异性。与此同时，电力需求侧管理的深化进程持续加速，促使需求侧资源的潜力得到更为充分的挖掘与利用，进而推动了源网荷储协同互动与深度融合。通过负荷预测，可以实现对电力需求的精准预测，保障源荷双侧的电力电量平衡，使电力需求响应更加灵活精准，提高供需匹配效率。

在新型电力系统中，数据作为基础性资源，其质量和准确性对系统分析及决策结果具有至关重要的影响。异常数据的出现可能会显著干扰系统的正常运行。输配电层面的电网电能损失包括技术损失(Technical Losses, TL)和非技术损失(Non-Technical Losses, NTL)^[4]。技术损失是不可避免的，其发生在输配电所需的设备中而自然发生的损耗。而非技术损失则由多种因素引起，包括未计费电力、设备故障、计费错误、基础设施质量低下以及非法用电行为，这些因素会导致输配电公司及合法消费者的成本增加。随着AMI系统的不断建设与完善，其提供的大量用电数据为数据挖掘技术在电力异常检测中的应用提供了广阔的空间。AMI系统不仅能够实现高精度的数据采集和实时监控，还为电力系统的智能化管理提供了坚实基础。面对AMI系统带来的如此庞大的数据量，电力工作人员在监测与精准识别异常数据方面面临着巨大的挑战。因此，有效地识别和处理这些异常行为对于增强电网的安全性和运营效率至关重要。

1.2 国内外研究现状

1.2.1 电力负荷预测研究现状

电力负荷预测作为时间序列预测的一个重要应用场景，其核心在于通过对历史负荷数据的分析，挖掘并预测未来负荷的变化趋势。多年来，国内外学者在电力负荷预测方法上进行了广泛研究，这些方法大致可以分为传统模型预测方法和机器学习预测方法两大类。传统模型预测方法包括灰色理论预测，自回归差分移动平均，指数平滑预测等方法。机器学习预测方法则包括循环神经网络、长短时记忆网络等方法^[5]。

在传统模型预测方面，吴迪等人^[6]提出一种二次指数平滑多目标组合模型 ES-BiLSTM-LSTM 来预测短期电力负荷。郭崇等人^[7]改进了基本灰色预测模型并结合需求侧响应的约束条件，最终在预测精度上显著提高，且具有较高的效率和加速比。王洁等人^[8]提出将灰色模型结合滤波法、相关分析等方法，考虑到基准负荷的周期性和增长性特点，采用了灰色模型 GM(1,1)，并融入了滤波法、相关分析等策略，构建了日峰降温电力负荷与人体舒适度指标之间的分段回归预测模型。但是，传统模型预测在面对日益复杂且庞大的新型电力系统生产的电力数据时逐渐不再合适，随着云计算、人工智能算法在电力负荷预测领域的应用，可以解决传统模型在处理大规模复杂数据时建模效果不佳的问题。

随着深度学习在时间序列预测方面的不断发展,依据循环神经网络(Recurrent Neural Networks, RNN)框架的方法^[9, 10],通过采纳马尔可夫假设来构建连续数据点间的动态关系模型,在针对时间序列数据的预测任务中,展现出了具有潜力的积极预测效果。邓斌等人^[11]设计了一种创新的 RNN 架构,该网络通过引入膜积分器,对传统 RNN 的神经元状态方程进行了优化改进,旨在克服传统 RNN 中存在的自我修正能力薄弱和易于陷入局部最优解的固有缺陷。实验结果显示,该模型在中长期电力负荷预测任务中展现出了卓越的表现。但是 RNN 由于反馈信号要么消失要么爆炸,只能回顾大约 10 个时间步。长短期记忆神经网络(Long Short-Term Memory, LSTM)则解决了这个问题^[12]。陈振宇等人^[13]介绍了一种融合 LSTM 和极端梯度提升树(XGBoost)的组合预测模型对电力负荷数据进行预测,实验结果验证了该组合模型在超短期电力负荷预测任务中展现出的高精度特性。卷积神经网络(Convolutional Neural Networks, CNN)是为图像数据集设计的,其核心优势在于能够有效提取在空间维度上具有平移不变性的局部特征关系^[14]。L. Li 等人^[15]使用 CNN 对输入的电力数据进行精确的聚类,随后使用另一个具有三个隐藏层的神经网络来预测电力负荷,其中考虑各种外部影响因素,例如温度、湿度、风速等,在准确性和效率方面都表现良好。为了使 CNN 适应时间序列数据集, Bai S 等人^[16]提出了时序卷积网络(Temporal Convolutional Networks, TCN),该架构通过使用膨胀卷积层和残差连接来实现长程依赖性的建模。T. Cheng 等人^[17]在微电网负荷预测任务上以 TCN 为基础模型,考虑到注意机制在根据特征重要性自适应赋权方面的优势,建立了 TCN-ATTENTION 短期负荷预测模型并具有较高的预测精度。最初因在自然语言处理(Natural Language Processing, NLP)领域取得突破而声名鹊起的 Transformer 模型^[18],现已广泛应用于时间序列分析。通过其独特的自注意力机制,Transformer 模型能够有效地捕捉长程依赖关系,展现出卓越的性能和广泛的适用性。随着时间序列分析领域的不断发展,一系列基于 Transformer 架构的改进模型应运而生,包括 Informer^[19]、Autoformer^[20]、FEDformer^[21]、Crossformer^[22]、iTransformer^[23]与 PatchTST^[24]等。这些模型不仅致力于相对减轻计算负担,还显著提升了对复杂时序模式的捕捉能力,特别是在处理长程依赖关系方面表现尤为突出。

综上所述,电力负荷预测作为时间序列分析的重要应用,经历了从传统模型到现代机器学习方法的演变。传统方法如灰色理论、ARIMA 和指数平滑等,在处理相对简单和小规模的数据集时表现良好,但面对复杂和大规模数据时逐渐显现出局限性。近年来,随着深度学习技术的发展,特别是 RNN 及其变体 LSTM、TCN 以及 Transformer 架构的引入,显著提升了预测精度和效率。这些先进模型通过捕捉长时间依赖关系和局部特征,在短期和中长期负荷预测上表现出色。在此背景下,本文探索使用随机采样重组策略挖掘序列中更深层信息、使用混合卷积-多层感知器网络捕捉序列的依赖关系和局部特征,同时降低计算成本,为电网公司提供更适用电网实际的电力负荷预测方法。

1.2.2 电力负荷异常检测研究现状

在电力负荷异常检测任务上，国内外学者做了多年的研究工作。在电力负荷异常检测任务上，国内外学者做了多年的研究工作。根据检测技术的差异性，本文分别从基于硬件和基于机器学习两个方面详细对现有电力负荷异常检测方法进行剖析。

(1) 基于硬件的电力负荷异常检测方法

基于硬件的电力负荷异常检测方法通过部署具有多功能事件记录性能的计量设备和特定功能的附属设备，结合地理信息系统，可以有效实现对电力负荷异常的检测与定位。Y. Zhou 等人^[25]通过在数据中心部署数字保护继电器 (Digital Protective Relays, DPR) 来实现多租户数据中心电力盗窃问题，DPR 是电力系统中用于故障检测和事件记录的设备。

(2) 基于数据驱动的电力负荷异常检测方法

基于数据驱动的电力负荷异常检测方法可以根据是否依赖标注数据分为两大类：有监督学习方法和无监督学习方法。其中，有监督学习方法包括分类与回归方法，无监督学习方法包括聚类等。

① 有监督学习方法

基于支持向量机 (Support Vector Machines, SVM) 的方法目的在于构建一个决策边界 (普遍形式为超平面或超球体)，以确保多数正常样本被包含于边界内部，而异常样本则位于边界外部^[26]。J. Nagi 等人^[27]提出了一种融合遗传算法与支持向量机的电力异常分析框架。其中，遗传算法通过随机性与预填充基因组的综合应用，显著提升了收敛性能，并为支持向量机超参数的全局优化提供了有力支持。I. Parvez 等人^[28]则将电压互感器的连续数据作为小波变换、滤波器的输入信号，随后将滤波后的特征数据集输入至单类支持向量机 (One-Class Support Vector Machine, OCSVM) 所训练的模型中。若存在干扰，则该模型能够检测并标记为异常状态。

人工神经网络在电力异常检测中被广泛应用，J. Fang 等人^[29]采用了 LSTM 和修改后的 CNN 来进行历史数据预测，并实验发现对其实际智能电网管理有效。M. Priyadarsini 等人^[30]将反馈回路引入智能电网异常预防的研究工作，专注于检测虚假数据注入攻击，同时降低传感器收集数据的噪声，使用 CNN 估计状态，并创建一种反馈机制，用于使用物联网设备的行为预测异常。对于需要处理大量数据的情况，X. Liu 等人^[31]提出了一种基于预测的异常检测的方法，并与三个基准方法进行了比较，在集群环境下采用独特的 Lambda 架构来高效处理海量数据，确保了对大规模数据集的快速响应和高精度预测。R. Jaiswal 等人^[32]提出了一种分层分布式雾计算架构，用于部署经过机器学习训练的异常检测模型。雾计算架构通过将计算资源分布在网络边缘，减少了数据传输延迟，并提高了系统的可扩展性和鲁棒性。

② 无监督学习方法

无监督学习方法无需标签信息，通过密度、距离、近邻度等方法筛选出与多数用户相异的用户为疑似异常用电用户。

M. M. Breunig 等人^[33]提出了一种通过对比分析数据集中各数据点与其邻近数据点的局部密度差异，以识别异常点的方法，即局部离群因子 (Local Outlier Factor, LOF)。Y. Peng 等人^[34]将 K-means 聚类算法与 LOF 算法相结合，利用 LOF 算法计算 K-means 聚类所选出的候选异常值的异常程度，并设计了一个针对 AMI 系统应用场景的异常检测框架。然而，对于大规模数据集，LOF 算法面临着较高的计算成本和较长的执行时间的问题。针对这一挑战，殷锋等人^[35]提出了一种基于混合查询剪枝策略的球树模型来优化 LOF 算法。通过重新设计数据对象的存储结构及查询、搜索机制，该方法显著提升了算法处理大数据集的能力。

聚类是典型的无监督学习方法，主要可以分为两类：硬聚类与软聚类。在硬聚类中，各目标样本属于同属于某一聚类，包括 K-means^[36]、DBSCAN^[37]等方法。在软聚类中，各目标样本在一定程度上可属于多个聚类，常用的方法有模糊 C 均值方法。Y. Yining 等人^[38]结合差异算法和 K-means 聚类来识别电力异常。DBSCAN 不需要事先指定聚类的数量，并且能够识别出噪声点。黄书俊等人^[39]通过结合 Copula 分布函数分析居民用户的水电用量之间的相关性，并利用 DBSCAN 算法对台区用户的距离相关系数进行聚类，将水、电使用数据距离相关系数曲线与其他用户差异较大的识别为异常用户。模糊 C 均值 (Fuzzy C-means, FCM) 是一种软聚类算法，FCM 允许数据点同时属于多个聚类，这种模糊处理方式使得 FCM 在处理边界不明确或重叠的聚类时更加灵活和有效^[40]。郑思达等人^[41]针对小型用户群体，提出了一种基于 FCM 算法的异常用电检测方案，实现对异常用电行为的有效识别及精确定位其发生的具体时间段。H. Liu 等人^[42]使用遗传模拟退火算法改进了 FCM 聚类算法，通过因子分析，对用户功耗特性进行了尺寸规格化，提高了模型检测的效率。

综上所述，现有研究主要分为基于硬件和基于数据驱动的两大类方法。基于硬件的方法通过部署多功能计量设备和特定功能的附属设备，结合地理信息系统，实现对异常的精确定位，但无法应对高科技手段窃电。无监督学习方法通常依赖于数据的分布特性来识别异常点，但在实际应用中，正常用户的用电模式也可能出现较大波动，导致误报率较高；仅依靠用电数据进行训练的有监督学习方法可能无法捕捉到所有类型的异常行为，因为窃电行为可能是多维度的，涉及时间、空间、用户行为等多种因素。在此背景下，本文探索利用数据重采样技术，使用电数据中异常用户与正常用户的数量比例适合与检测模型，利用混合卷积-多层感知器网络对序列进行重建，利用异常检测模块对重建样本和原始样本进行计算识别出异常用户，为电网公司提供更适用电网实际的电力负荷异常检测方法。

1.2.3 云计算在电力系统中的研究现状

新型电力系统所生成的数据展现出显著的特性，具体包括数据体量庞大、数据类型多元化、数据变化速率高（即动态性显著）以及数据价值密度相对较低（因大量数据记录的是正常运营状态，而具有实际应用价值的数据占比较低）。这些特征使得新型电力系统产生的数据高度契合基于云计算架构的大数据存储与处理技术的需求与应用场景。云计算技术凭借其强大的并行处理能力和可靠的存储机制，成为理想选择。相较于传统集群系统，云计算不仅提升了系统的处理效率和可靠性，还大幅降低了运维成本和复杂度。因此，云计算技术更加适宜作为新型电力系统的运作平台。

当前的相关研究主要聚焦于系统架构的规划、实现策略的探索以及未来发展趋势的展望等领域。冯桂玲等人^[43]采用 Map-Reduce 架构，部署于 Hadoop 平台，利用分布式计算方式进行电网负荷的精准建模和预测分析。李磊等人^[44]依托 Spark 云计算大数据处理平台，设计了一种创新的电力负荷预测方法，但未涉及电力数据的分布式存储和查询等需求。王均^[45]则着重对电力大数据领域中的高速存储及高效检索关键技术展开了深入的探索与研究。针对云计算技术在电力行业中的具体应用与实践，苏寅生及其团队^[46]针对南方电网总调中心在传统集中式架构中所面临的诸多挑战，依据地域特性和实际需求，创造性地设计了一种兼顾高性能与安全性的分布式解决方案，该方案以当前互联网技术前沿的 Kubernetes 集群管理系统和 Docker 容器化部署技术为基石，构建了高效、可靠、绿色的共享计算平台。

综上所述，新型电力系统产生的数据具有体量大、类型多、动态性强及价值密度低的特点，云计算技术因其强大的并行处理和可靠存储能力成为理想解决方案。上述研究展示了云计算在提升电力系统性能和安全性方面的巨大潜力。本文探索利用 Docker 容器化部署分布式 Doris 数据库管理电力负荷数据，利用分布式易扩展的可视化工作任务调度平台 DolphinScheduler 负责数据处理任务调度，为电网公司提供基于云计算的新型电力系统负荷管理平台的构建途径。

1.3 研究内容与技术路线

1.3.1 研究内容

本文主要研究内容如下：

(1) 电力负荷预测方法研究。通过深入分析负荷序列的特性，设计并实现了一种结合随机采样重组策略的混合卷积-多层感知器网络模型。该方法首先通过对原始时间序列数据进行随机分割和重组，以增强数据的多样性和时序特征的可提取性。随后，利用