

考虑电网动态行为的电力数据异常检测与分析

张 瑞¹ 王婷婷¹ 辛亚峰¹

ZHANG Rui WANG Tingting XIN Yafeng

摘 要

鉴于电网运行环境的复杂性和不确定性,电力数据中常常会包含一些异常值,这些异常值可能会对电网的安全稳定运行造成负面影响。因此,提出了一个考虑电网动态行为的电力数据异常检测与分析方法。在考虑电网动态行为的前提下,通过 SCADA 系统采集了多种类型的电力数据,并对这些数据进行了缺失值处理和标准化处理。随后,应用模糊 C 均值(FCM)算法对经过预处理的多元电力数据进行了聚类分析,目的是检测出其中的电力数据异常值。通过一系列的实验验证,所设计的方法能够显著提高电力数据异常检测的查全率,同时还能有效降低均方根误差。此外,该方法还展现出了良好的加速性能,从而为电网的安全稳定运行提供了有力的技术支持。

关键词

电网动态行为; 电力数据; 数据异常; 异常检测; 检测分析

doi: 10.3969/j.issn.1672-9528.2025.08.042

0 引言

随着我国智能电网建设规模的不断增大,产生的电力数据呈指数级增长。然而,在电力数据的采集、传输和处理过程中,由于设备故障、通信干扰、人为错误等多种因素的影响,电力数据中往往包含大量的异常值。这些异常值不仅破坏了数据的完整性和一致性,还可能对电力系统的分析与决策产生误导,严重影响电网的安全稳定运行。因此,在智能电网环境下,准确检测并剔除电力数据中的异常值,提高数据质量,对于电网故障预测与诊断、负荷预测与优化调度等方面具有重要意义。文献[1]中针对电力数据异常值,设计一种基于改进 K-Means 和 DNN 算法的检测方法,具有较高准确率,该方法在处理大规模电力数据集时,虽然通过并行计算提高了处理效率,但仍然存在计算复杂度高、内存消耗大等问题;文献[2]中针对电力数据异常值,设计一种基于改进 K-均值联合 SVDD 的检测方法,误漏报率较低,由于电力数据通常包含大量的采样点与多维特征,这要求 SVDD 模型具有较高的计算效率和存储能力,但是 SVDD 模型在训练过程中需要计算所有数据点到中心点的距离,导致计算复杂度的急剧增加与内存资源的过度消耗;文献[3]中针对电力数据异常值,设计一种基于伪标签的无监督异常检测方法,检测性能良好,但伪标签的生成更倾向于正常数据,可能导致实际应用中无监督模型对异常数据的检测能力不足。针对传统电力数据异常检测方法的局限性,本文提出了一种考虑电网动态行为的电力数据异常检测与分析方法,为智能电网的安全稳定运行提供更加可靠的技术支持。

1 考虑电网动态行为下获取电力多元数据

电网动态行为,主要指受电力负荷变化、天气条件等因素影响,导致在时间与空间上电力系统运行状态数据不断发生变化。这种变化不仅影响电网安全稳定运行,也对电力数据采集提出了更高要求。因此,在电力数据异常检测中,本文在考虑电网动态行为条件下,通过 SCADA 系统采集多元的电力数据^[4],具体采集对象如表 1 所示。

表 1 电力多元数据提取对象

多元数据类型	提取对象
电力系统运行数据	A、B、C 三相电压
	A、B、C 三相电流
	A、B、C 三相负载功率
源荷数据	电负荷
	热负荷
	冷负荷
	环境温度
	环境湿度
	光照强度
	风速

如表 1 所示,在电网动态行为下,本文考虑环境因素对电力数据的影响,从 SCADA 系统中调取了电力系统运行数据和源荷数据,组成电力多元数据集。但是,在实际电力数据采集,受设备故障、通信中断等因素影响,原始数据中可能存在缺失值,影响后续异常值检测结果的准确性。因此,本文引入拉格朗日插值法进行缺失值填充处理。具体来说,针对原始电力多元数据集,本文将一系列电力参数当成因变量,并将影响电力参数的因素(负荷、环境参数)当成自变量。

1. 国网甘肃省电力公司张掖供电公司 甘肃张掖 734000

然后，对电力多元数据集中缺失数据位置前后的 5 个相邻数据进行提取，组成一个新的数据组，基于该数据组，进行缺失数据的拉格朗日插值填补^[5]：

$$R_n(x) = \sum_{i=0}^n G_i(x) y_i \quad (1)$$

其中：

$$G_i(x) = \frac{(x-x_0) \cdots (x-x_{i-1})(x-x_{i+1}) \cdots (x-x_n)}{(x_i-x_0) \cdots (x_i-x_{i-1})(x_i-x_{i+1}) \cdots (x_i-x_n)} \quad (2)$$

式中： $R_n(x)$ 表示电力多元数据缺失位置的插值结果； $G_i(x)$ 表示拉格朗日基函数值； (x_i, y_i) 表示电力多元数据缺失位置前后提取的数据点，其中 x_i 为数据点位置， y_i 为数据值； n 表示提取数据点的总数。

与此同时，由于原始电力多元数据的量纲和数量级可能存在差异，如果直接进行聚类分析将会造成结果误差较大。因此，本文引入最小-最大法，对缺失值处理后的电力多元数据进行标准化处理，具体表达式^[6]为：

$$X_{\text{new}} = \frac{X_{\text{old}} - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (3)$$

式中： X_{old} 、 X_{new} 分别表示标准化处理前、后的电力多元数据； X_{min} 、 X_{max} 分别表示标准化处理前的电力多元数据的最小、最大值。

根据式（3），可以将原始电力多元数据转换到 [0,1] 区间内，以此消除不同量纲对电力数据异常检测的不良影响。总之，本文通过采集并预处理电力多元数据，获取了高质量的数据集，为后续异常检测提供数据支持。

2 FCM 聚类算法检测电力数据异常值

基于上述电力多元数据，本文引入 FCM 算法进行聚类分析^[7]。在该算法中，应用了模糊划分方式，允许一个电力数据样本点属于多个聚类中心，并通过隶属度来代表样本点属于聚类中心的程度，由于考虑电网动态行为下，电力数据往往具有不确定性，所以 FCM 算法更适合电力数据异常值检测与分析。首先，假设给定的电力多元数据样本集为 $X_i, i = 1, 2, \dots, N$ ，其中 N 为样本总数，对该样本集的聚类中心进行初始化，确定初始聚类中心，聚类中心数量为 M ，最大迭代次数为 D 。然后，在每一次迭代过程中，计算电力数据样本集的模糊隶属度^[8]为：

$$S_{ij} = \begin{cases} \sum_{g=1}^M \left(\frac{L_{ij}}{L_{ig}} \right)^{-2/(\omega-1)}, & L_{ij} > 0 \\ 1, & L_{ij} = 0 \end{cases} \quad (4)$$

式中： S_{ij} 表示第 i 个电力数据样本相对于第 j 个聚类中心的隶属度； L_{ij} 、 L_{ig} 分别表示第 i 个电力数据样本与第 j 个聚类中心、第 g 个聚类中心之间的距离； ω 表示电力数据样本的权重指数。基于上式所求模糊隶属度，本文定义 FCM 算法的目标函数^[9]为：

$$F(X, W) = \sum_{i=1}^M \sum_{j=1}^N S_{ij}^{\omega} L_{ij}^2 \quad (5)$$

式中： $F(X, W)$ 表示电力多元数据样本点到 FCM 聚类中心的加权距离平方和目标函数。

通过最小化式（5）所示目标函数，将原始电力多元数据样本点划分为 M 个组别，完成样本初始聚类划分。划分成功后，进行模糊隶属度和聚类中心^[10]的更新：

$$S_{ij}^{(i+1)} = \sum_{g=1}^M \left(\frac{L_{ij}}{L_{ig}} \right)^{-2/(\omega-1)} \quad (6)$$

$$W_j^{(j+1)} = \frac{\sum_{i=1}^N [S_{ij}^{(i)}]^{\omega} X_i}{\sum_{i=1}^N [S_{ij}^{(i)}]^{\omega}} \quad (7)$$

式中： $S_{ij}^{(i+1)}$ 表示第 i 个电力数据样本相对于第 j 个聚类中心的隶属度更新值； $W_j^{(j+1)}$ 表示第 j 个聚类中心的更新值。

完成电力多元数据样本模糊隶属度矩阵和聚类中心的更新后，按照 FCM 目标函数再次进行样本划分。不断重复上述划分过程，直至达到最大迭代次数 D ，完成电力数据样本的聚类分析。最后，根据各类别电力多元数据的隶属度矩阵来判断样本点的异常性。如果某样本点相对于任何聚类中心的隶属度均较低（小于某个阈值），则认为该样本点为异常点；反之，如果隶属度大于阈值，则认为该样本点为正常点。至此，完成考虑电网动态行为的电力数据异常检测与分析。

3 仿真实验

3.1 仿真设置

本章为了对设计的考虑电网动态行为的电力数据异常检测与分析方法进行实际应用性能校核，将其设置为实验组，并将基于改进 K-means 和 DNN 算法的电力数据异常检测方法和基于改进 K-均值联合 SVDD 的电力数据异常检测方法设置为对照组，展开一个仿真对比实验。首先，按照表 2 所示参数搭建本次仿真实验软硬件环境。

表 2 实验环境参数配置

实验环境	参数	配置
硬件环境	CPU	Intel Xeon Gold 6230
	内存	256 GB DDR4
	GPU	NVIDIA Tesla V100
软件环境	操作系统	Ubuntu 20.04 LTS
	编程语言	Python 3.8
	数据处理库	Pandas 1.2.3 和 NumPy 1.19.5

然后，以我国某电力公司提供的 16 589 条带标签的真实脱敏用户电力数据作为实验数据。数据包含用户 ID、时间戳、电压、电流、功率等多个特征维度，并标记了异常数据。表 3 为实验数据集的节选示例表格。

表 3 实验数据集节选示例

用户 ID	时间戳	电压 /V	电流 /A	功率 /kW	标签
User0001	2024-01-01 00:00	220	10	2.24	正常
User0001	2024-01-02 00:05	230	15	2.76	正常
User0001	2024-02-04 11:08	0	1 000	0	异常
...	...	...	...	...	...
User16 589	2024-12-31 21:55	225	0	2.025	异常

最后，分别应用实验组方法和对照组中两种方法针对上述实验样本进行电力数据异常检测，对比分析检测结果。

3.2 实验指标

在本次实验中，为了综合对比实验组方法和对照组中两种方法对电力数据异常值检测性能的优劣，本文引入查全率 δ_1 、均方根误差、加速比作为实验指标，具体计算公式为：

$$\delta_1 = \frac{N_1}{N_1 + N_2} \times 100\% \quad (8)$$

$$\delta_2 = \sqrt{\frac{1}{N_0} \sum_{i=1}^{N_0} \|Y_i - Y'_i\|^2} \times 100\% \quad (9)$$

$$\delta_3 = \frac{T_0}{T_j} \quad (10)$$

式中： N_1 表示被正确检测出的电力数据异常样本数量； N_2 表示未被检测出的电力数据异常样本数量； N_0 表示所有电力数据样本数； Y_i 、 Y'_i 分别表示第*i*个电力数据异常样本的真实值和检测值； T_0 表示检测方法在单处理器系统上的运行时间； T_j 表示检测方法在第*j*个并行处理器系统上的运行时间。

因此，本文通过上述公式计算的查全率、均方根误差、加速比指标，评价电力数据异常检测方法的全面性、准确性及效率。

3.3 检测结果对比分析

在本次实验中，为避免实验结果的偶然性，将原始电力数据实验样本划分为多个不同组别，在不同数量的实验样本下，分别应用实验组方法和对照组中两种方法进行电力数据异常检测。完成检测后，计算并统计各方法下电力数据异常检测结果的查全率，如图 1 所示。

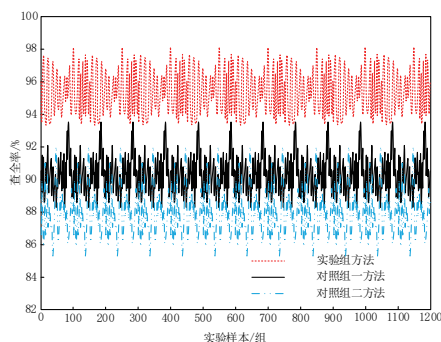


图 1 电力数据异常检测结果的查全率对比

从图 1 中数据可以看出，与对照组中两种方法相比，实验组方法在电力数据异常检测中更具全面性。具体来说，在实验组方法下，电力数据异常检测结果的平均查全率高达 95.98%，较对照组中两种方法分别提升了 5.09%、8.84%。这主要是因为实验组方法考虑了电网动态行为，能够更全面地获取电力多元数据样本，并进行准确检测。

从图 2 中数据可以看出，实验组方法在所有样本数下得到的电力数据异常检测结果的均方根误差平均值仅 0.31%，远低于对照组一方法 1.07% 和对照组二方法的 0.82%，这表明实验组方法在检测电力数据异常样本时具有更高的准确性，所得检测结果更逼近真实值。

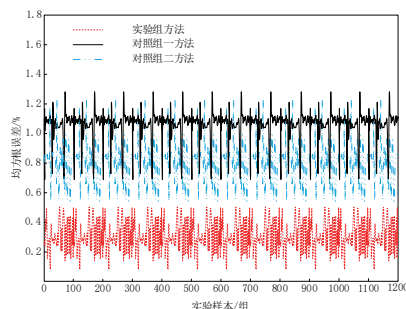


图 2 电力数据异常检测结果的均方根误差对比

从图 3 中数据可以看出，随着电力数据样本的不断增加，实验组方法和对照组中两种方法的加速比均呈上升状态。但是，与对照组中两种方法相比，实验组方法下电力数据异常检测的加速比一直处于一个最高水平上，且增长幅度更大。由此可以说明，实验组方法在并行计算方面表现更出色，可以更有效地利用计算资源，提高电力数据异常检测效率。

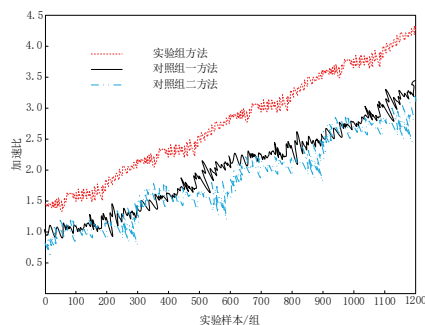


图 3 不同电力数据异常检测方法的加速比对比

综上所述，本文研究的考虑电网动态行为的电力数据异常检测与分析方法是有效且优越的，可以为我国智能电网的安全稳定运行提供更加可靠的技术支持。

4 结语

本项研究工作主要聚焦于传统电力数据异常检测技术在实际应用中所面临的挑战，特别是针对检测精度和检测效率方面的不足。为此，本研究提出并设计了一种新的电力数据异常检测与分析方法，该方法特别考虑了电网的动态行为特性。在研究的初始阶段，对电网的动态行为特性进行了深入

分析,并在此基础上成功获取了包含多种类型信息的电力多元数据集。随后,为了实现对这些数据中潜在异常值的准确检测与分析,引入了模糊C均值(FCM)聚类算法。在仿真实验部分,本研究将所提出的检测方法与现有的几种传统异常检测技术进行了对比分析。实验结果表明,本文所设计的方法不仅能够全面地检测出电力数据中的异常值,而且在检测的准确性方面也表现出了显著的优势。此外,该方法在处理大规模数据集时,也展现出了较高的检测效率。

展望未来,本研究计划继续深入探索电力数据的内在特性以及电网的动态行为规律。通过进一步的研究,期望能够对现有的异常检测方法进行改进和优化,从而提升其泛化能力和实用性。最终,希望本研究的成果能够为我国智能电网的建设和发展提供有力的技术支持,为智能电网的稳定运行和高效管理贡献更多的力量。

参考文献:

[1] 常荣,徐敏.基于改进K-Means和DNN算法的电力数据异常检测[J].南京理工大学学报,2023,47(6):790-796.

[2] 林显奂,胡嘉铭,戴伟力,等.基于改进K-均值联合SVDD的电力数据异常检测方法[J].电力电容器与无功补偿,2023,44(5):99-107.

[3] 林卫伟,白向阳,孔军.基于伪标签的无监督电力数据异常检测框架[J].计算机仿真,2024,41(2):131-136.

[4] 王文博,刘绚,张博,等.基于协议特征的电力工控网络流量

异常行为检测方法[J].电力系统自动化,2023,47(2):137-145.

[5] 王文博,刘绚,林海,等.基于深度学习的电力工控流量应用层报文异常检测[J].电力系统自动化,2023,47(11):69-76.

[6] 徐子东,张镇勇.基于不平衡高维数据特征重要性的电力系统异常检测方法[J].控制工程,2024,31(11):2029-2035.

[7] 薛晓慧,张文,张静,等.基于二次聚类的充电桩执行电价异常检测方法[J].电信科学,2025,41(1):184-190.

[8] 李凯,靳书栋,刘宏志,等.基于IWOA-ELM-AE的电力资产信息管理系统异常数据检测方法[J].沈阳工业大学学报,2024,46(3):255-262.

[9] 王文森,杨晓西,刘阳,等.基于层次聚类分析的变压器油中溶解气体在线监测数据异常检测[J].高压电器,2023,59(1):142-147.

[10] 侯立,王健.改进密度峰值聚类的多源数据异常值检测算法[J].计算机仿真,2024,41(6):565-569.

【作者简介】

张瑞(1978—),女,甘肃张掖人,本科,工程师,研究方向:市场营销。

王婷婷(1991—),女,甘肃张掖人,本科,工程师,研究方向:电气工程及其自动化。

辛亚峰(1992—),男,甘肃武威人,本科,工程师,研究方向:电气工程及其自动化, email: yuxq211@163.com。

(收稿日期:2025-04-15 修回日期:2025-08-12)

(上接第175页)

3 结语

为了提升海量数据处理与分析的效率与准确性,本文开展基于自适应谱聚类海量数据离群点挖掘方法研究,其通过引入信息熵有效减少了海量数据冗余和相关性带来的干扰。利用谱聚类算法提高了海量数据聚类的准确性。实验选用了4个不同的数据集,通过对比离群点挖掘前后的数据分布情况,验证了所设计的数据挖掘方法对数据离群点的挖掘能力。实验结果表明,所设计算法应用后,其检测的离群点占比结果与真实情况完全一致,能够满足预期要求,达到理想的挖掘状态,应用效果较好。

参考文献:

[1] 庄巧萸.异质信息网络快速离群点数据智能挖掘算法[J].黑龙江工业学院学报(综合版),2024,24(10):97-101.

[2] 朱华,乔勇进,董国钢.基于CART决策树的分布式数据离群点检测算法[J].现代电子技术,2024,47(16):157-162.

[3] 周燕,肖莉.基于改进关联聚类算法的网络异常数据挖掘[J].计算机工程与设计,2023,44(1):108-115.

[4] 张忠平,李森,刘伟雄,等.基于快速密度峰值聚类离群因

子的离群点检测算法[J].通信学报,2022,43(10):186-195.

[5] 康耀龙,冯丽露,张景安,等.基于谱聚类的不确定数据集中快速离群点挖掘算法[J].吉林大学学报(工学版),2023,53(4):1181-1186.

[6] 李春燕.基于谱聚类算法的人力资源数据集离群点快速挖掘方法[J].信息与电脑(理论版),2023,35(23):50-52.

[7] 王彩霞,陶健,舒升.基于机器学习的聚类序列离群点数据挖掘算法[J].通化师范学院学报,2024,45(8):28-34.

[8] 杨志强,冯山,尹伊,等.一种多因素融合的高效离群点检测方法[J].山东大学学报(理学版),2024,59(8):118-126.

[9] 何旭,邓安生,葛小龙.基于局部相对密度的离群点检测算法[J].计算机应用与软件,2024,41(12):296-302.

[10] 周玉,夏浩,岳学震,等.基于改进K-means的局部离群点检测方法[J].工程科学与技术,2024,56(4):66-77.

【作者简介】

李婷(1996—),女,山西长治人,硕士,助教,研究方向:机器学习、数据挖掘。

(收稿日期:2025-03-26 修回日期:2025-07-31)