

基于数据中台的电网数据全链路监控方法

萧展辉¹, 甘莹², 柯婷², 梁置铭²

(1. 南方电网数字电网研究院有限公司, 广东 广州 510700;

2. 南方电网数字电网集团有限公司, 广东 广州 510700)

摘要: 电网通过监控全链路电网数据可以发现电网异常, 但是数据监控方法存在杰卡德系数低、监控准确率低的问题。在此背景下, 研究一种基于数据中台的电网数据全链路监控方法。该方法以数据中台三部分功能作为支撑, 实施电网数据全链路采集与预处理, 提取电网数据样本中提取特征, 包括角度离散程度、平均角的正弦和余弦、圆角标准差, 以特征为输入, 利用改进神经网络算法构建异常识别模型, 实现电网数据全链路监控。结果表明: 所研究方法的杰卡德系数更大, 该值均在0.9以上, 数据异常监控准确率在97.5%以上, 证明了所研究监控方法的准确性。

关键词: 数据中台; 电网数据; 数据采集; 特征提取; 改进神经网络算法; 全链路监控

中图分类号: TP393; TP183 文献标识码: A 文章编号: 1003-7241(2025)02-0100-05

Power Network Data Full Link Monitoring Method Based on Data Center

XIAO Zhan-hui¹, GAN Ying², KE Ting², LIANG Zhi-ming²

(1. China Southern Power Grid Digital Power Grid Research Institute Co., Ltd., Guangzhou 510700 China;

2. Digital Grid Group, CSG, Guangzhou 510700 China)

Abstract: The power grid can find the grid anomaly by monitoring the full link power grid data, but the data monitoring method has the problems of low Jackard coefficient and low monitoring accuracy. Under this background, a full link monitoring method of power grid data based on data center is studied. This method is supported by the three functions of the data center, implements the full link collection and preprocessing of the power grid data, extracts the features extracted from the power grid data samples, including the degree of angle dispersion, the sine and cosine of the average angle, and the standard deviation of the rounded corners. With the features as the input, the improved neural network algorithm is used to build an anomaly identification model to realize the full link monitoring of the power grid data. The results show that the jackard coefficients of the methods studied are larger, the value is above 0.9, and the accuracy of data anomaly monitoring is above 97.5%, which proves the accuracy of the monitoring methods studied.

Keywords: data center; grid data; data acquisition; feature extraction; improved neural network algorithm; full link monitoring

0 引言

电网在整个电力系统中起到了重要的连接作用^[1], 将电力运输到各个环节, 实现电力的分配、调度。电网在工作过程中, 会产生大量的电网数据, 这些数据对于电力公司具有非常大的利用价值。通过这些数据可以帮助电力公司制订战略计划, 提高服务质量。电网在长期运行下, 某一链路上难以避免地会出现异常或者故障问题。一旦链路出现问题, 就会导致电网数据出现异常值。异常值的发现对于电力公司安排维护工作至关重要, 若不能及时发现, 应用结果会与实际结果存在较大的偏差, 因此, 电网数据全链路监控具有重要的现实意义。通过监

控能及时发现异常数据, 保证原始数据的完整性和准确性。

目前, 关于数据的监控研究有很多。例如刘文倩等人提出了一种基于数字孪生的监控方法, 首先进行时空数据关联, 获取数据, 然后利用数字孪生网络分析数据, 最后通过图元动态展示功能实时可视化显示数据分析结果^[2]。张华兵等人以微服务架构为基础, 首先利用压缩感知方法采集全链路电力数据并构建测量与感知矩阵, 然后利用小波变换预处理电力数据, 最后从横向与纵向两个维度实现数据监控^[3]。苏国梁等人提出了一种电网海量故障全景数据的远程监控方法, 该方法采用5类监控手段建立保护熵并获取相对保护熵, 将相对保护熵作为依据, 提出两个判据对电网故障全景数据进行判断, 实现数据监控^[4]。但是上述方法均存在杰卡德系数低, 监控效果

*基金项目: 数据运维工具(全链路监控)研发项目(0002200000076144)

收稿日期: 2023-08-31

差的问题。

数据中台是电力行业中的常用的大数据可视化管理系统,在大数据处理与分析中起到了重要作用。基于此,结合前人研究经验,提出一种基于数据中台的电网数据全链路监控方法。通过该方法研究以期提高电网数据质量,方便电网数据管理和应用。

1 电网数据全链路监控研究

1.1 基于数据中台的应用框架

全链路电网数据是电网运行全过程的体现,具有重要的挖掘价值和析价值。然而,电网运行经常发生故障^[5],而电网全链路监控可以通过数据异常的监控发现故障,因此,电网全链路监控对电网具有重要的现实意义。针对这一点,研究基于数据中台的电网数据全链路监控方法。数据中台是一种大数据管理平台,主要由数据湖处理单元、数据分析中台以及应用中心三个部分组成,如图1所示。

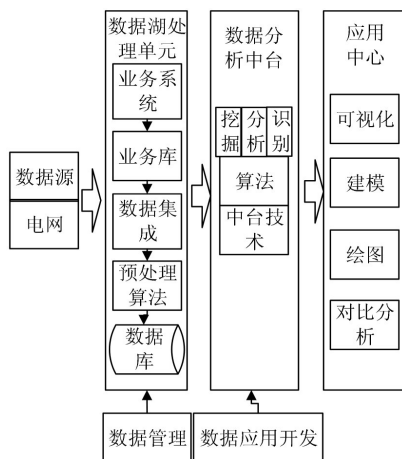


图1 数据中台基本架构

基于数据中台三大组成部分的功能,构建基于数据中台的应用框架,包括电网数据全链路采集与预处理、全链路电网数据特征提取、全链路电网数据可视化监控^[6]。

1.2 电网数据全链路采集与预处理

利用数据处理单元中的各个程序算法对电网数据进行全链路采集与预处理。首先利用非侵入式探针技术旁路截获全链路电网数据,该技术属于旁路连接方式,因此不需要单独设置客户端,提高了数据采集的效率和便利性^[7]。采集过程如下:

步骤1:将旁路服务器接入网关,建立旁路服务器与主服务器之间的通信以及旁路服务器与网关之间的通信。

步骤2:将旁路服务器网卡设置为混杂模式,旁路服务器对数据进行旁路捕获,得到与旁路服务器网卡直接连接的网关发出的所有数据包^[8]。具体采集过程如下:

(1) 设置待采集数据的源和目的MAC地址等信息;

(2) 将网卡设置为混杂模式;

(3) 采集电网数据包;

(4) 分析帧首部、分析IP首部、分析传输层首部、分析应用层数据;

(5) 依据分片等信息重组数据包构成数据段;

(6) 根据应用数据中的偏移量信息,将多个数据段组合并为数据块并存储于旁路服务器。

步骤3:数据实时同步更新。

经过上述过程采集全链路电网数据后,降维和标准化处理数据,具体过程如下所示:

(1) 电网数据降维

降维,即降低数据维度。电网数据一般具有很高的维度,不便于数据分析^[9-10]。为此,需要降维处理电网数据,假设高维空间原始数据集为 $S=\{s_1, s_2, \dots, s_n\}^k$, n 为数据集规模; k 代表数据维度;然后计算 S 中 k 维第 i 个电网数据 s_i 与其他 $n-1$ 个电网数据之间的距离,计算公式如下:

$$D_{ij} = \sqrt{\sum_{t=1}^k (s_{it} - s_{jt})^2}, j=1, 2, \dots, n-1 \quad (1)$$

式中, D_{ij} 代表 k 维第 i 个电网数据 s_i 与第 j 个电网数据 s_j 之间的距离。

根据 D_{ij} 选取排序前 M 个与 s_i 邻近的近邻点,然后计算这 M 个近邻点与 s_i 之间的权重,计算公式如下:

$$w_{ij} = \frac{D_{ij}}{\sum_{j=1}^M D_{ij}} \quad j=1, 2, \dots, M, \text{这里 } j=1, 2, \dots, M \quad (2)$$

式中, w_{ij} 代表 s_i 与邻近点之间的权重。

最后计算低维嵌入空间, $Q=\{q_1, q_2, \dots, q_a, \dots, q_n\}^m$, $a=1, 2, \dots, n$, m 为低维维度。即通过求解下述公式最小值来获得。即:

$$\begin{cases} \min Q = \sum_{i=1}^n \left| q_i - \sum_{j=1}^M w_{ij} q_j \right|^2 \\ s.t. \begin{cases} \sum_{i=1}^n q_i = 0 \\ \frac{\sum_{i=1}^n q_i q_i^T}{n} = 1 \end{cases} \end{cases} \quad (3)$$

利用寻优算法求解满足式(3)的 n 个 q_a , q_a 为 m 维度下的第 a 个电网数据,即完成电网数据降维^[11]。

(2) 电网数据标准化

标准化处理降维数据,公式如下:

$$\hat{q}_i = \frac{q_i - \min q}{\max q - \min q} \quad (4)$$

式中, \hat{q}_i 代表标准化后的第 i 个电网数据; q_i 代表原始第 i 个电网数据。

经过上述过程完成电网数据全链路采集与预处理工作,为下一步研究做准备。

1.3 全链路电网数据特征提取

针对上述处理好的全链路电网数据,基于数据中台的第二部分功能从电网数据样本中提取特征,用于最后的数据监控^[12]。在这里采用的提取方法为圆分布法。提取公式如下:

首先计算电网数据的信息熵,计算公式如下:

$$R_i = -d \sum_{i=1}^n Y(\hat{q}_i) \cdot \ln Y(\hat{q}_i) \quad (5)$$

其中,

$$Y(\hat{q}_i) = \sqrt{\frac{\hat{q}_i}{\sum_{i=1}^n \hat{q}_i}} \quad (6)$$

式中, R_i 代表第*i*个电网数据的信息熵; d 代表常数; $Y(\hat{q}_i)$ 代表电网数据 \hat{q}_i 的占比。

其次计算加权信息熵:

$$H_i = -d \sum_{i=1}^n \frac{\bar{q} \cdot Y(\hat{q}_i) \cdot \ln Y(\hat{q}_i)}{\hat{q}_i} \quad (7)$$

式中, \bar{q} 代表电网数据平均值; H_i 代表第*i*个电网数据的加权信息熵。

最后基于 H 数值,结合圆分布法,对 H_i 进行角度转换,得到角度值 θ_i ,然后提取特征。提取公式如下:

(1) 角度离散程度

$$G = B_x^2 + B_y^2 \quad (8)$$

其中,

$$B_x = \sqrt{\frac{\sum H_i \cdot \cos \theta_i}{\sum H_i}} \quad (9)$$

$$B_y = \sqrt{\frac{\sum H_i \cdot \sin \theta_i}{\sum H_i}} \quad (10)$$

式中, G 代表角度离散程度; B_x 、 B_y 分别代表*x*方向和*y*方向角度偏差。

(2) 平均角正弦和余弦

$$\begin{cases} \sin \bar{\theta} = \frac{B_y}{G} \\ \cos \bar{\theta} = \frac{B_x}{G} \end{cases} \quad (11)$$

式中, $\sin \bar{\theta}$ 、 $\cos \bar{\theta}$ 分别代表平均角 $\bar{\theta}$ 的正弦、余弦。

(3) 圆角标准差

$$F = \sqrt{\frac{-2 \ln G}{\pi}} \quad (12)$$

式中, F 代表圆角的标准差。

经过上述过程,完成全链路电网数据特征提取工作,为异常全链路电网数据识别奠定基础。

1.4 异常全链路电网数据识别

利用数据中台中的第二部分和第三部分功能构建基于改进神经网络算法的异常识别模型,利用该模型可视化监控全链路电网数据。改进神经网络结构图如图2所示。

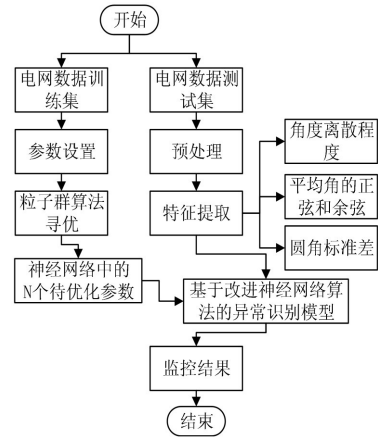


图2 基于改进神经网络算法的异常识别模型

在基于改进神经网络算法的异常识别模型中主要分为两部分,一部分是利用粒子群算法求解神经网络算法的权值和阈值^[13]。以往神经网络算法的权值和阈值是通过反向传播调整的,这种方法缺乏全局性,且存在可能局部最优的特性。面对这种情况,利用粒子群算法代替传统反向传播法,调整权值和阈值。执行粒子群算法,粒子位置和速度更新公式如下:

(1) 速度更新公式:

$$\zeta_i^m = \delta \zeta_i^{m-1} + \beta_1 \gamma_1 (\phi_i^m - x_i^m) + \beta_2 \gamma_2 (\varphi^m - x_i^m) \quad (13)$$

式中, ζ_i^m 、 x_i^m 代表第*i*个粒子的速度和位置; ζ_i^{m-1} 代表前一次迭代的速度; δ 代表惯性权重; m 代表迭代次数; β_1 、 β_2 代表个体和社会学习因子; γ_1 、 γ_2 代表随机函数; ϕ_i^m 代表第*i*个粒子历史最优位置; φ^m 代表所有粒子历史最优位置。

(2) 位置更新公式:

$$x_i^m = x_i^{m-1} + \zeta_i^m \quad (14)$$

式中, x_i^{m-1} 代表前一次迭代的位置。

经过粒子群算法优化神经网络之后,将全链路电网数据特征作为输入,经过隐含层、输出层运算,得出各个数据异常类型对应的数值,即异常监控数值。

$$s_i = \sum_{i=1}^n w_i \psi_i - \varepsilon_i \quad (15)$$

式中, s_i 代表输出层第*i*个神经元的输出,即异常全链路电网数据识别结果; w_i 、 ε_i 代表连接权值和阈值; ψ_i 代表隐含层输出; n 代表输出层神经元数量。

基于输出 s_i ,参照最大值归属原则,得出最终数据异常类型,完成全链路电网数据可视化监控。

2 方法应用测试

2.1 测试环境及参数设置

在测试设计方法性能前,需要构建测试环境,为方法应用测试作准备,在该测试中主要应用服务器、数据库、测试工具等,具体测试环境如表1所示。

表1 测试环境

服务	环境
服务器	腾讯云服务器
数据库	MySQL
系统环境	CentOS 7.0
测试工具	Chrome,Android APP

在完成测试环境构建后,设置测试参数,具体参数设置如表2所示。

表2 测试参数设置

参数名称	参数符号	设置值
数据维度	k	2
个体学习因子	β_1	4.3
社会学习因子	β_2	3.1
近邻点排序号	M	10
数据集规模	n	100

2.2 电网数据样本

完成测试环境构建和参数设置后,利用数据中台采集电网数据,采集的部分数据如图3所示。

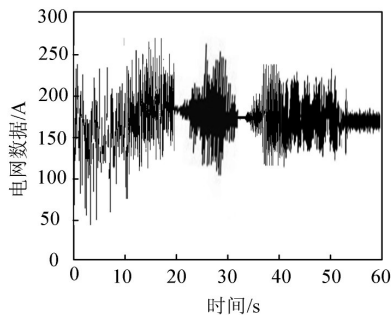


图3 电网数据样本示例

将上述采集的电网示例数据样本进行处理,组成电网数据样本,绘制成表,结果如表3所示。

表3 电网数据样本

数据类型	样本数量/个	
	训练样本	测试样本
正常	80	82
异常1	84	80
异常2	78	80
异常3	89	80
异常4	85	81
异常5	86	92
异常6	88	85

2.3 数据特征提取实现

基于1.3节研究,提取每类电网数据样本的角度离散程度、平均角的正弦和余弦、圆角标准差四个特征值。以电网数据样本集合中每类样本的前3个样本为例,计算出

的数据特征值如表4所示。

表4 数据特征值

样本类型	样本序号	角度离散程度	平均角正弦	平均角余弦	圆角标准差
正常	1	2.54	2.45	10.15	3.22
	2	2.74	1.72	10.45	2.52
	3	3.56	1.20	12.45	1.45
异常类型1	1	5.41	0.22	10.45	8.54
	2	7.12	0.85	15.62	6.23
	3	1.25	0.42	10.54	5.22
异常类型2	1	0.56	0.52	11.23	4.12
	2	1.55	0.33	11.52	2.22
	3	1.85	3.22	16.32	5.65
异常类型3	1	2.52	2.12	14.45	8.51
	2	2.86	2.56	8.55	7.12
	3	3.15	0.85	9.23	5.23
异常类型4	1	4.54	0.21	5.42	6.31
	2	2.87	0.14	12.23	4.21
	3	5.43	3.26	25.52	5.23
异常类型5	1	3.86	5.12	14.52	2.02
	2	2.40	1.45	16.33	0.12
	3	4.55	2.20	15.52	0.52
异常类型6	1	2.44	3.56	10.12	3.12
	2	1.41	5.85	21.20	1.45
	3	0.56	7.45	12.36	2.55

2.4 电网数据异常监控结果

将上述7个类型若干个样本数据混合在一起,应用所提方法、基于数字孪生的方法(文献[2])和基于微服务架构的方法(文献[3])进行电网数据异常监控,以监控准确率为性能指标,监控结果如图4所示。

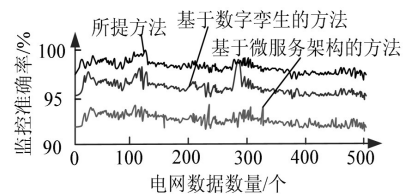


图4 电网数据异常监控数值

从图4中可以看出,所提方法的电网数据异常监控准确率最高,整体准确率曲线在对比方法之上,并且最高准确率达到100%,而最低准确率也达到了97.5%,对比其他方法,基于数字孪生的方法整体在95.5%以上,最高为98.1%,基于微服务架构方法的监控准确率整体在92.0%以上,但是最高值仅为94.5%,三种监控方法相比,所提方法的监控准确率最高,整体高出2.0%,因此,所提方法有效提高电网数据异常监控的准确率。

2.5 监控方法性能对比分析

相同测试条件下,利用基于数字孪生的方法、基于微服务架构的方法、所提方法,得出监控结果。然后计算所研究监控方法和三种对比监控方法应用下的杰卡德系数。该系数越大,监控方法越准确,进一步验证设计方法

的价值,结果如图5所示。

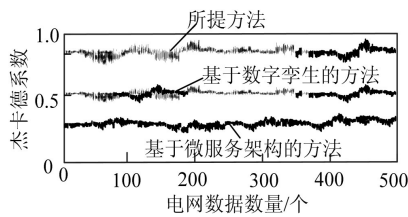


图5 杰卡德系数

从图5中可以看出,与基于数字孪生的方法、基于微服务架构的方法相比,所提方法的杰卡德系数更大,该系数均在0.75以上,而基于数字孪生方法的杰卡德系数在0.50左右,基于微服务架构方法的杰卡德系数在0.30左右,三种方法相比,所提方法提高了0.25和0.45以上,进一步验证了所提方法电网数据全链路监控性能更高,其具备更高的应用价值。

3 结束语

综上所述,电网在电力系统中起到了重要的传输作用,一旦电网出现异常或者故障,电力输送工作就会面临中断。通过监控电网数据,能有效发现电网运行是否处在正常状态。在此背景下,研究一种基于数据中台的电网数据全链路监控方法。该方法通过数据中台的三个部分功能实现从数据采集、特征提取到异常识别,完成数据全链路监控。最后通过大量的样本测试了方法性能,得出的方法的杰卡德系数要大于基于数字孪生的方法、基于微服务架构的方法的杰卡德系数,由此证明了方法的有效性。然而,本研究还需要进行方法效率验证,以进一步完善方法性能对比。这将是下一步研究的重点。

(上接第60页)

频次难以达到要求。自动化机房及办公大楼机房UPS放电,在真实负载放电时,通过所提系统可以在保护电压的情况下完成深度放电测试,为蓄电池的后期维修工作提供科学的参考。

参考文献:

- [1] 樊欣欣,丁晖,陈秀国,等.基于模糊逻辑的变电站蓄电池在线健康状态评估[J].电子器件,2021,44(1):136-140.
- [2] 于港旭,李中伟,金显吉,等.基于光伏和蓄电池的智能小区能量调度优化策略研究[J].自动化仪表,2022,43(7):21-27.
- [3] 张文爽,李键,余文涛,等.基于新一代电源控制器的锂电池在轨管理分析[J].电源技术,2022,46(1):104-107.
- [4] 朱晓荣,韩啸.储能电池与常规机组配合参与一次调频的自适应控制策略[J].电测与仪表,2023,60(9):34-42.
- [5] 魏丽君,章若冰.高精密铅酸蓄电池充放电管理系统的研究[J].仪表技术与传感器,2022(2):57-60,64.

参考文献:

- [1] 李楠,徐新宇,单露.含分布式电源的配电网电能质量扰动源自动定位方法[J].自动化技术与应用,2024,43(12):55-58,79.
- [2] 刘文倩,韩利峰,黄丽,等.基于数字孪生的TMSR-SF0数据监控与可视化方案[J].核技术,2022,45(2):83-90.
- [3] 张华兵,周英耀,徐磊,等.基于微服务架构的电力信息系统全链路监控技术[J].沈阳工业大学学报,2022,44(4):409-414.
- [4] 苏国梁,李国庆,范晨亮,等.电网海量故障全景数据的远程监控方法研究[J].电网与清洁能源,2022,38(3):42-46,54.
- [5] 韦恒,黄超,杨彦,等.基于数据挖掘技术的电网设备参数风险识别方法研究[J].自动化技术与应用,2022,41(2):47-50,97.
- [6] 黄龙达,庄卫金,龙苏岩,等.电力交易平台数据指标监视技术研究与应用[J].电网技术,2021,45(8):3200-3206.
- [7] 林大崮.基于模糊识别的高速公路交通流数据实时监控方法[J].科技通报,2020,36(8):65-69.
- [8] 张晓琴,汪云飞,胡春强.基于改进极限学习机的数据采集与监控系统攻击检测模型[J].南京航空航天大学学报,2021,53(5):708-717.
- [9] 王锋,高欣,贾欣,等.一种基于对数区间隔离森林的电力调度数据异常检测集成算法[J].电网技术,2021,45(12):4818-4827.
- [10] 夏爱明,伍雪冬.基于改进多目标海樽群算法的电力系统优化调度[J].电测与仪表,2023,60(7):77-82.
- [11] 胡文曦,肖先勇,汪颖,等.现代电网电力扰动数据分析与主动应用[J].电力系统自动化,2021,45(4):1-11.
- [12] 李文俊,杨学强,杜家兴.基于数据中台的装备保障数据集成[J].系统工程与电子技术,2020,42(6):1317-1323.
- [13] 王毅,王智微,何新.智能电站数据中台建设与应用[J].中国电力,2021,54(3):61-67,176.

作者简介:萧展辉(1975-),男,硕士,高级工程师,研究方向:计算机技术。

- [6] 原艳南,韩文兴,杨诚,等.基于DSP和蓄电池的控制棒驱动电源电路设计[J].自动化仪表,2021,42(z1):218-221,225.
- [7] 蔡勇超,余勇,曹小冬,等.以实际负载放电的蓄电池远程核容装置设计[J].蓄电池,2021,58(2):92-96.
- [8] 杨忠亮,黄世回,王汝钢.蓄电池在线核容放电关键装置与远程控制策略的研究[J].蓄电池,2020,57(1):5-11.
- [9] 于港旭,李中伟,金显吉,等.基于光伏和蓄电池的智能小区能量调度优化策略研究[J].自动化仪表,2022,43(7):21-27.
- [10] 王瑜瞳,袁平亮,李颖,等.大规模高效液流电池自动储能技术研究[J].自动化技术与应用,2023,42(10):133-136.
- [11] 胡晓工,苏志坚,贾晓亮,等.基于分布式容积信息卡尔曼滤波的锂电池SOC估计[J].自动化技术与应用,2024,43(12):1-7.

作者简介:杨劲松(1979-),男,本科,高级工程师,从事电力系统及自动化、科技创新与成果转化孵化、产业化等工作。