

基于移动小波树的电力监控异常数据自动识别算法研究

夏景, 梁薇, 吴珠瑛

(国网上海市电力公司松江供电公司 上海 201699)

摘要: 为保证电力系统正常运行, 以防出现窃电、漏电等异常用电情况, 进行电力监控异常数据自动识别具有重要的现实意义。针对基于数据挖掘、机器学习以及统计的3种数据异常识别算法无法应对不同长度的数据, 导致识别准确性、效率较低的问题, 提出一种基于移动小波树的电力监控异常数据自动识别算法。该算法分为两部分内容: 先是对采集到的用电数据进行处理, 包括数据脱敏、数据填补、颠簸去除等步骤, 提高数据质量, 后构建移动小波树结构, 实现异常数据自动识别。结果表明: 与基于数据挖掘、机器学习以及统计的三种数据异常识别算法相比, 所提算法的准确性和效率均较高, 解决了传统算法存在的问题。

关键词: 移动小波树; 电力监控; 异常数据; 识别算法

中图分类号: TN274

文献标识码: A

文章编号: 1674-6236(2020)18-0148-05

DOI: 10.14022/j.issn1674-6236.2020.18.033

Research on automatic recognition algorithm of abnormal data in power monitoring based on mobile wavelet tree

XIA Jing, LIANG Wei, WU Zhu-ying

(State Grid Shanghai Songjiang Electric Power Supply Company, Shanghai 201699, China)

Abstract: In order to ensure the normal operation of the power system, in order to prevent the occurrence of electricity theft, leakage and other abnormal power use, it is of great practical significance to automatically identify the abnormal data of power monitoring. In view of the problem that the three anomaly recognition algorithms based on data mining, machine learning and statistics can't deal with different length of data, resulting in low accuracy and efficiency of recognition, an automatic anomaly recognition algorithm for power monitoring based on mobile wavelet tree is proposed. The algorithm is divided into two parts: first, the collected power consumption data is processed, including data desensitization, data filling, bump removal and other steps to improve the data quality, and then the mobile wavelet tree structure is constructed to realize the automatic recognition of abnormal data. The results show that compared with the three algorithms based on data mining, machine learning and statistics, the proposed algorithm has higher accuracy and efficiency, and solves the problems of the traditional algorithm.

Key words: mobile wavelet tree; power monitoring; abnormal data; recognition algorithm

为满足生产、生活的需要, 对能源的需求量逐年增加, 尤其是作为基础能源之一的电能, 更是呈现爆

炸式增长。在此背景下, 对异常用电监测也提出了更高的要求。据调查, 每年因为用电异常产生的经济损失已经达到了上千万元。不止如此, 近年来随

收稿日期: 2019-12-04 稿件编号: 201912031

基金项目: 国网上海松江供电公司 2019 年实体营业厅转型升级项目(B2093519000Q)

作者简介: 夏景(1984—), 男, 浙江杭州人, 工程师。研究方向: 电力市场营销。

-148-

着高科技的发展,非法用电的手段更“高级”、方式更隐蔽,造成的影响更是以往的几倍,甚至几十倍,产生的破坏力更是十分巨大,不仅造成了相关电力设施的损坏,还有可能引发火灾,威胁到电力的正常供给和电力系统的稳定运行^[1]。由此可见,进行有效的电力监控异常数据自动识别对于维护电网安全具有重要的实际意义。

对于电力监控异常数据自动识别,目前常用的算法有3种:基于数据挖掘的数据异常识别算法、基于机器学习的数据异常识别算法以及基于统计的数据异常识别算法。以上这3种算法的识别原理均是通过建立固定模型,然后将检测出来的结果与设定的阈值进行对比来完成异常识别^[2]。在数据量一定的情况下,这些算法尚能完成识别,但是在通常情况下,数据的长度都是不同的,且数据具有极高的流动性,历史数据具有的可依赖性极低,因此传统算法获得的数据判断阈值存在较大误差,导致传统算法存在严重误报和漏报情况。

针对上述情况,文中提出一种基于移动小波树的电力监控异常数据自动识别算法。该算法基本原理是通过构建移动小波树结构,使得任意长度的异常数据都能被某个时间窗所包含,从而实现快速异常识别。但是由于受到颠簸数据的干扰,所以检测效率受到一定影响,也存在误报的可能,所以在本文算法中,对颠簸数据处理进行了重点研究,以期提高检测精度和效率。最后为检验本文识别算法的有效性,对其识别精度和识别效率进行对比分析。结果表明:与基于数据挖掘、机器学习以及统计的3种数据异常识别算法相比,本文基于移动小波树的算法进行电力监控异常数据自动识别,具有更高的识别精度与识别效率,保证了电力系统运行的安全性,降低了用电异常的风险^[3]。

1 电力监控异常数据自动识别算法研究

用电计量是电力发展中的基础性业务,直接关系到电力企业的经济利益和居民的正常用电,因此十分有必要进行用电异常数据监控,以防出现窃电、漏电等异常用电情况,保证电力系统安全稳定运行^[4]。为此,本文基于移动小波树研究一种新的电力监控异常数据自动识别算法。算法研究流程如下:

第一部分:数据处理。对采集到的电力监控数据进行处理,包括数据脱敏、数据填补、颠簸去除等

步骤,提高数据质量,降低后续异常识别的难度。

第二部分:实现电力监控异常数据的自动识别。首先构建移动小波树结构,然后利用该结构进行异常电力监控异常数据自动识别。该算法能够解决传统异常识别算法无法应对不同长度数据的问题,提高异常识别算法的应用范围^[5]。

1.1 数据处理

在分析电力系统监控数据之前,需要对其进行处理,提高数据质量。处理步骤包括数据脱敏、数据填补、颠簸去除^[6]。

1) 数据脱敏

随着科学技术的发展,智能电网已经取代传统电网,被广泛应用,因此用户用电数据通过与电网公司形成的即时连接网络实现实时参数采样。然而,频繁的数据交换在满足数据管理的同时,也产生了严重的信息泄露现象,因此为保护用户电力数据安全,一般会进行加密处理,而加密过后的电力数据并不能进行异常识别,所以需要进行脱敏。数据脱敏,又称数据漂白、数据去隐私化或数据变形^[7],就是将加密后的数据进行还原处理,电力数据的具体脱敏流程如图1所示。

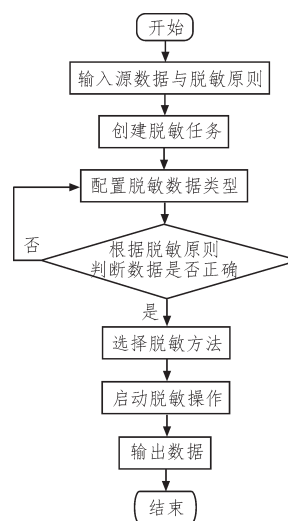


图1 数据脱敏流程

2) 数据填补

数据采集和脱敏过程均会在一定程度上损害数据,破坏数据的完整性,因此还需要进行缺失数据填补^[8]。数据填补流程如图2所示。

3) 颠簸去除

颠簸数据是影响异常识别精度的主要因素,因此需要进行去除。颠簸数据是指数据集中的某些数

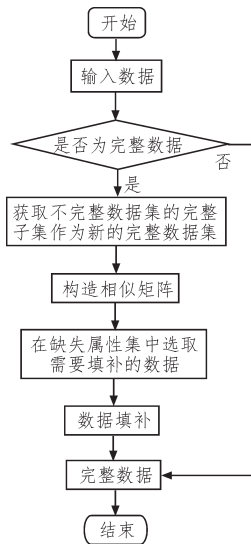


图2 数据填补流程

据存在不稳定情况,但并不足以成为异常的数据。但是,这些存在不稳定情况的数据,极易被识别算法识别为异常数据,从而增加算法的错误率^[9]。而本文研究的基于移动小波树的电力监控异常数据自动识别算法,并不能很好抵御这些颠簸数据的干扰,所以在本章节利用一种比率阈值 a 进行颠簸数据处理。颠簸数据处理的原理如下:假设输入的数据为 $y_i(i=1,2,\dots,m)$,然后选取该数据前后的两个数据,并比较前后两个相同长度窗口的聚集^[10]。如果它们之间的比值超过一定的比率 a ,则认为该数据为异常数据。当检测窗口长度为 b 时,则通过下述表达式进行颠簸数据检测:

$$y_{b+1} + \dots + y_{2b} > a(y_1 + \dots + y_b) \quad (1)$$

公式中, $2b < m$ 。

下面通过一个具体例子进行具体分析。在假定一个数据长度为6,窗口长度为1,数据值分为10、20、19、20.5、19、20,属于颠簸数据,因为

$$y_4 < ay_3 \quad (2)$$

并且

$$y_6 < ay_5 \quad (3)$$

但这个数据点却造成窗口长度2和3出现异常,因为:

$$y_3 + y_4 > a(y_1 + y_2) \quad (4)$$

并且

$$y_4 + y_5 + y_6 > a(y_1 + y_2 + y_3) \quad (5)$$

公式中,阈值比率 a 取值为1.1。

总而言之,若数据只是出现问题,而没有出现异

常,就应该尽量避免颠簸数据造成的误报现象^[11]。

1.2 移动小波树实现异常数据识别

异常检测是数据挖掘领域研究的重点工作,其定义原理如下:

设 $X=(x_1, x_2, \dots, x_n)$ 代表一个数据的时间序列集合; $S=(s_1, s_2, \dots, s_m)$ 代表一个给定窗口大小的集合; F 代表一个聚集函数; $f(s_j), j=1, 2, \dots, m$ 代表对应不同大小窗口的阈值^[12]。识别异常数据就是在数据集中找出的满足如下条件的子序列。

$$\forall i \in (1, 2, \dots, n), \forall j \in (1, 2, \dots, m) \quad (6)$$

$$F(x(i, \dots, i + s_j - 1)) \geq f(s_j) \quad (7)$$

在上述异常识别原理指导下,利用移动小波树进行异常识别,算法基本思路是根据待识别数据异常持续时间的长短,对相应的小波层次进行检测,若某个层次中某个窗口的值超过了事先设定的阈值,则继续搜索其对应的下层窗口,直至找到异常发生的位置^[13]。

1) 移动小波树结构构建

首先从数据的原始时间序列开始,进行第一遍循环,计算每个连续数据的聚集和,完成第一层移动小波树的构建,其他小波分解层次的聚集值由其下层不重叠的数据求和得到,重复该过程,直到最上层结构只包含一个窗口,完成移动小波树结构构建,具体过程如下:

输入:数据序列 $x[1, n]$, $n=2^a$;

输出:迁移小波树 $SWT[1, \dots, a][1 \dots]$ 。

Step1: $b = x // b[]$ 为输入数组;

Step2: $i = 1 \text{ to } a / a = \log_2^n$, 计算各层窗口的聚集; $j = 1 \text{ to } \text{size}(b) - 1 \text{ step } 2 //$, 合并各层相邻的窗口,构造迁移小波树的第 i 层; $SWT[i][j] = b[j] + b[j+1]$; $j = 1 \text{ to } \text{size}(SWT[i]) / 2 // \text{downsample}$ 过程;

Step3: 为下次循环构造输入数组 b , $b[j] = SWT[i][2*j-1]$ 。

2) 实现电力监控异常数据自动识别

在上述构建的移动小波树结构的基础上进行电力监控异常数据自动识别,基本思路如下:首先输入电力监控数据到移动小波数结构第 $i+1$ 层当中,寻找小于设定阈值的子序列,然后计算剩余子序列的时间窗口移动和,再次验证是否超出阈值,从而确定时间窗口,然后从第二层开始,循环识别,直到发现异

常数据为止。

2 算法仿真测试分析

为检测本文所提异常检测算法的有效性,本章节在上述章节的基础上,将基于移动小波树的电力监控异常数据自动识别算法与 storm 平台结合,设计 storm 框架下的识别模型^[14-18]。且在相同条件下,构建基于数据挖掘、机器学习以及统计的3种传统数据异常识别算法模型,作为对比项,进行算法检测准确性和检测效率对比。

2.1 测试数据

本章实验来源于某市5 000位用户2018年一年的用电信息,从中选出2 328位用户的用电信息作为样本数据。由于这2 328位用户的用电信息数据都是正常的,因此为了满足识别需要,需要人工生成异常数据。在这里使用高斯分布生成异常数据样本。表1为测试数据样本的划分情况。

表1 测试数据

类型	正常	异常
类型1	234	24
类型2	345	45
类型3	354	34
类型4	375	10
类型5	875	32

2.2 实验平台配置

实验平台配置如表2所示。

表2 实验平台配置

名称	参数
服务器类型	浪潮 P8000 工作站
语言编程环境	MATLAB 6.5
处理器	Intel Xeon CPU ES-2620 v3 2.40GHz*2
内存	16GB
存储	SAS 300GB*3
网卡	Gigabit Ethernet Card*4
操作系统	Windows 2007

2.3 评价指标

评价指标如下所示。

1) 准确率:

$$DR = \frac{a}{b} \quad (8)$$

$$FR = \frac{c}{d} \quad (9)$$

公式中, a 为识别到的异常记录数; b 为数据集中异常记录总数; c 代表误报为异常的正常记录数;

d 为数据集中正常记录数。

2) 识别效率:通过完成一次异常识别所消耗的时间来进行识别效率评价,消耗时间越短,识别效率越高。

2.4 结果分析

从表3中可以看出,利用本文所提算法进行电力监控异常数据自动识别,与基于数据挖掘、机器学习以及统计的3种传统数据异常识别算法相比, DR 值更高、 FR 更低,所花费的识别时间更短,由此说明本文算法的识别性能更强。

表3 结果分析

项目	移动小波树	数据挖掘	机器学习	统计
DR(%)	94.36	90.25	89.33	90.12
FR(%)	3.66	4.58	4.25	5.04
识别时间(s)	574	580	588	603

3 结束语

综上所述,随着对电力需求的增长,电网公司对用户用电管理也越来越严格。因为异常用电现象的频繁发生,不仅给电网公司带来严重的经济损失,也影响电力用户的正常用电,造成的影响和破坏十分巨大。为此,本文针对当前基于数据挖掘、机器学习以及统计的3种传统数据异常识别算法,存在识别准确性与识别效率低的问题,提出一种基于移动小波树的电力监控异常数据自动识别算法。该算法经测试,解决了传统算法无法应对不同长度数据的问题,提高了异常识别准确性,缩短了算法运行时间,为电力系统安全稳定运行奠定基础。

参考文献:

- [1] 郭全军,邓华,陈晓锦,等.500kV变电站35kV所用变异常情况[J].变压器,2017,54(12):58-61.
- [2] 李永攀,门锐,吴俊阳.基于低秩模型的电力状态数据异常检测[J].计算机工程与应用,2019,55(16):255-258.
- [3] 沈小军,付雪姣,周冲成,等.风电机组风速-功率异常运行数据特征及清洗方法[J].电工技术学报,2018,33(14):3353-3361.
- [4] 刘扬,方圆,孙鸿,等.基于物理-信息模糊推理的智能电网攻击检测方法[J].中国科技论文,2016,11(14):1619-1625.
- [5] 余立苹,李云飞,朱世行.基于高维数据流的异常检测算法[J].计算机工程,2018,44(1):51-55.

- [6] 滕予非,吴杰,张真源,等.基于离群点检测的高压并联电抗器本体电流互感器测量异常故障在线诊断[J].电工技术学报,2019,34(11):2405-2414.
- [7] 严宇平,萧展辉.基于聚类分析的离群检测方法在电网数据质量管理中的应用研究[J].现代电子技术,2016,39(15):137-139.
- [8] 于冰洁,夏战国,王久龙.基于高斯过程模型的异常检测算法[J].计算机工程与设计,2016,37(4):914-920.
- [9] 庄池杰,张斌,胡军,等.基于无监督学习的电力用户异常用电模式检测[J].中国电机工程学报,2016,36(2):379-387.
- [10] 左进,陈泽茂.基于改进K均值聚类的异常检测算法[J].计算机科学,2016,43(8):258-261.
- [11] 袁力,刘奇,孙正文,等.现场环境下设备区域电磁干扰检测与识别方法[J].电波科学学报,2017,32(6):650-656.
- [12] 生志荣,程龙生.基于多分类马田系统的半监督数据异常点检测方法[J].数学的实践与认识,2018,48(10):157-162.
- [13] 刘航,孟庆亚,赵元哲,等.基于双树复小波的高速列车转向故障特征分析[J].控制工程,2018,28(8):1386-1392.
- [14] 张志鹏,李勇,曹一家,等.通信和电网联合仿真的配电网局部异常因子故障辨识算法[J].电力系统自动化,2016,40(17):44-50.
- [15] 王禹程.抵抗Web攻击的异常入侵检测算法[J].电子设计工程,2018,26(24):126-130.
- [16] 李龙飞,丁永生.基于机织电极的电势检测系统建模与仿真[J].系统仿真学报,2018,30(1):171-175.
- [17] 王向红,陈潇一.云计算与电力大数据技术的现状及发展趋势研究[J].陕西电力,2017(6):70-72.
- [18] 张利平,张亚军,王家源,等.电力载波通信技术在能源数据采集系统中的应用[J].工业仪表与自动化装置,2018(5):105-107.

(上接第147页)

- [5] 侯思祖,郭威.基于改进EEMD与混沌振子的配电网故障选线[J].仪器仪表学报,2019,11(4):77-87.
- [6] 杨军胜,王森,王承民,等.多频带自适应技术在电力载波方面的研究[J].电测与仪表,2018,56(8):23-24.
- [7] 杨蓉,曹旺斌,尹成群.基于自适应遗传算法的PLC信道动态子载波分配[J].电力系统保护与控制,2019,47(12):111-116.
- [8] 杨帆,贾辉,刘宝树,等. α 稳定脉冲噪声下宽带电力线通信系统性能分析[J].电子与信息学报,2019,41(6):1374-1380.
- [9] 唐松浩,陈鹏伟,肖湘宁.低压配电网超高次谐波传递特性仿真分析[J].电力电容器与无功补偿,2019,8(4):80-87.
- [10] 张慧,赵雄文,卢文冰,等.基于图论的宽带电力线MIMO信道特性研究[J].中国电机工程学报,2019,39(7):22-29.
- [11] 谢文旺,孙云莲.基于OS-ELM的宽带电力线通信解映射优化算法[J].电测与仪表,2018,56(13):35-39.
- [12] 巫志平,陈寿联,张伟.电力线宽带载波技术的智能路灯系统[J].农村电气化,2017,2(3):36-37.
- [13] 张海龙,刘宣,李然,等.基于TTCN-3的宽带载波通信一致性测试系统设计[J].计算机测量与控制,2018,5(4):44-48.
- [14] 廖华,廖前伟.基于广电网络和电力宽带载波的集抄技术[J].中国科技信息,2018,2(1):63-64.
- [15] 刘安灵,黄振刚,王朝霞.基于改进随机漂移粒子群算法的微电网优化运行[J].电子设计工程,2018,26(6):33-37.
- [16] 杜家兵,陈衍鹏,梁满发.基于分布式实时调度策略的智能电网控制系统的设计与实现[J].电子设计工程,2018,26(7):119-122.