

基于营销大数据的窃电识别综述

邵茹冰

(国网上海松江供电公司, 上海 201699)

摘要 窃电处理是电力营销业务中用电检查工作的重要内容,及时、准确地识别窃电行为对电网的有序发展和安全运行有着重要作用。随着用电信息采集系统的广泛应用和大数据相关算法的引入,为窃电的快速识别和精准定位提供了技术依据。本文针对用电采集大数据的应用,重点分析了高维随机矩阵算法和模糊聚类算法计算的特点,对大数据识别窃电方法进行了总结和展望。

关键词 电力营销;窃电;异常用电;随机矩阵算法;模糊聚类算法

中图分类号: F425

文献标识码: A

文章编号: 2097-3365(2023)12-0064-03

一些不法分子为了自身利益采取各种不法手段盗窃电能,对电力经济的发展和正常的供用电秩序造成了恶劣影响。此外,部分窃电手段不仅对电网的稳定运行带来隐患,甚至可能严重危害人身安全,需要及时查明并采取手段制止窃电行为的发生。及时发现并查处窃电行为不仅能够使供电企业收回经济上的损失,还能够帮助维护整个市场公平有序的竞争环境,营造风清气正的和谐社会氛围。

1 窃电常用手段

窃电常用手段主要有私自改接计量装置和绕越表计用电两种,其目的都是通过减少电能计量装置所计量的电量来减少实际收取的电费,实际上,窃电不仅违反道德要求,更是违反法律的行为。

1.1 私自改接计量装置窃电

1. 电压改接。电压改接窃电是指通过手段对电能表电压线、接线盒、计量熔丝、计量电压互感器等计量装置进行改接,使表计中计量到的电压幅值小于实际用户受电设备接入的电压。一般表现为:将接入表计的电压线路间歇性断开或完全开路;电压计量回路串联电阻分压;安装远程操控装置操控电压回路;故意损坏计量熔丝或计量电压互感器导致电压降低或无法测量等。通过计量电压幅值的减小,来减少计量到的功率,从而减少计量用电量的目的。

2. 电流改接。电流改接窃电是指通过手段对电能表进出电流线、接线盒、计量电流互感器等计量装置进行改接,使表计中计量到的电流幅值小于实际用户负荷产生的电流。一般表现为:断开上述计量装置某处的电流回路;短接上述计量装置某处的电流回路;

故意损坏计量电流互感器或使其失效等。通过计量电流幅值的减小,来减少电能表功率和总用电量。

3. 相序改接。相序改接窃电是指通过手段改变计量装置电压或电流的相位关系使计量的瞬时功率小于实际负荷。一般表现为:改变计量装置中电压或电流的接线相位;利用外部感性或容性电子元器件进行相位转移。

4. 表计内部改接。电能表内部改接窃电是指对电能表内部逻辑元器件的正常运行进行破坏,使电能表本身的计量功能受到影响,增大计量误差甚至计量失灵。一般表现为:直接开盖破坏标记内部元器件;外加磁场等环境影响准确计量;更改表计时钟等。表计内部的改接在日常工作中难以通过直接检查发现,需要提前进行数据分析,并现场借助掌机等专用设备查询电表开盖记录后,再进行开盖检查。

1.2 绕越表计窃电

绕越表计窃电是将用户负荷直接绕过电能表等计量装置接入公共电网中的行为,此方法简单易操作,且拆装快速,故仍然有较多用户尤其是乡村居民在继续私自采用。但相比于计量装置上的窃电,窃电现场更加直观,易于发现也便于窃电的取证,是反窃电工作中查获数量最多的窃电手段。

2 窃电分析的数据来源

随着电子和通信行业技术的飞速发展,各种智能采集终端在电网管理和控制中被大规模地采用,此举不仅提升了营销系统的自动化程度,提升了电力公司的工作效率,同时也为异常用电的检测提供了更加广阔的实现平台,激发了人们围绕这些智能设施对防窃

电进行更加深入的研究^[1]。目前常用的用户数据主要来源于用户的电能信息采集系统,可以实时监测并上传用户的用电电压和电流等关键数据,高压接入的电力用户还可以采集到功率因数、电压电流相序等其他用电信息。营销业务系统中的用户相关数据也为窃电用户的整体行为分析提供了多维度计算的基础。随着用电信息采集装置功能的扩展,以及配电网内监测设备的大量配置,可以为窃电分析提供越来越多的数据源,且随着时间推移逐渐形成体量庞大的海量数据库。

3 窃电识别技术

目前窃电的主要识别技术有直接利用历史窃电信息数据特征关联^[2-3]和根据大数据算法构建数据模型分析计算^[4-5]两大类。直接利用历史窃电信息数据特征关联的技术手段包括:依据用电信息采集数据进行窃电事件关联;基于智能电表数据的灰色关联分析;营销业务系统窃电数据特征提取等。而基于高维随机矩阵、模糊聚类算法等大数据算法的窃电识别定位,因其快速准确的计算能力而更符合智能电网的发展要求。

3.1 基于高维随机矩阵的窃电识别

1. 随机矩阵算法简介。随机矩阵是指随机变量组成的矩阵,高维随机矩阵定义为维数趋于无穷大的随机矩阵,是对复杂系统进行有效分析的计算工具。采用随机矩阵算法对用户用电数据进行分析,将采集的用电历史数据以及当前采集数据、用户基础信息等海量数据信息进行收集整理,构建成为窃电分析的基礎数据库,可以直接运用数据库进行处理和分析计算,避免了电网结构、运行情况等具体物理模型的复杂因素干扰,减小计算难度并且加快计算速度。随着电力系统智能化和数字化的发展,用户用电信息采集系统可以产生海量、多态的数据,符合大数据分布特征,从而构建随机矩阵进行进一步的分析统计。

高维随机矩阵的分布是指矩阵维数趋于无穷时,其特征值的概率密度分布。目前应用较为广泛的高位随机矩阵分布定理包括 M-P 律、半圆律和单环定律。其中, M-P 律描述了随机矩阵的奇异值的渐近收敛特性;半圆率针对 $n \times n$ 的 Hermitian 矩阵的经验谱分布进行描述;单环定律则用于 N 个非 Hermitian 矩阵的分析处理。虽然高维随机矩阵的频谱密度是指维度趋近于无穷大的情况,但在实际应用中,根据现有计算经验和结果,矩阵维度为 10^2 数量级时,已经可以出现较好的收敛结果,为该理论在实际数据处理应用提供了实际操作的可行性。

对于 N 个独立非 Hermitian 矩阵进行相乘并在归一

化后计算其经验谱密度:当非 Hermitian 矩阵的维度无穷大时,特征值复平面上体现为圆环,外圆的半径为 1,内圆的半径随着随机矩阵的行数和列数变化。将配电网数据采集系统收集到的节点电压、电流及功率等电网运行数据,按照时间序列进行储存,生成随机矩阵。由于窃电分析涉及的节点及支路相关的数据量庞大,故直接作为高维随机矩阵进行分析处理。将待测数据输入算法计算后,若新的随机矩阵特征值谱半径平均值处于圆环的环形区域,则待测数据正常,若新的随机矩阵特征值谱半径平均值处于内圆的内侧,则待测数据异常,存在窃电风险,需要结合现场用电情况进行进一步排查。可以推断,当系统正常运行时,随机矩阵的特征值基本处于外圆和内圆之间。若该部分电网数据中包含窃电数据,特征值将向内环内部集中。

2. 算法特点。由于在庞大的电力用户运行数据中,窃电用户的实际数量极少,故数据源中存在大量冗余数据,高维随机矩阵算法可以将这些冗余数据在矩阵特征值计算过程中予以筛选,在复平面上只保留具有代表性特征的数据,既保留了全体数据的整体趋势,也便于对计算结果进行直观的分析 and 统计,对存在窃电情况的配电线路及时进行进一步排查处理。但由于实际异常数据点在计算过程中发生融合,难以直接判断窃电发生的具体位置,适合进行初步感知判断,为现场窃电的查处提供整体方向指导。

3. 数据和处理。数据及其处理包括特征选择和关联随机矩阵的构建。特征选择的优点是在确保选择监测措施的时效性的基础上,确保及时可靠地识别不同类型的窃电,并准确地监测智能网络的状况。联结矩阵网络用于确定节点与网络内区域之间的相对关系。

(1) 合并数据:所有电气特性都必须在一定程度上融合到空间中,以便在一段时间内将这些电子特性结合在一起,检测工作状态,并形成一状态监测矩阵。在扩大高空间和时间维结构的过程中,数据监测矩阵的状况。

(2) 数据分析和可视化:采用滑动窗口法对高维时空状态监测矩阵进行处理,因此,每个电气量之间的相对关系几乎没有改变,空间维度低的数据也统一呈现出高维数据状态。利用随机矩阵理论对化简后的时空状态监测矩阵进行分析检测,实现对电网运行状态的实时判别。

3.2 基于模糊神经算法的窃电识别

3.2.1 模糊聚类算法简介

模糊聚类算法通过对输入数据进行初始类别划分,

并不断迭代更新样本聚类中心点的方式计算最适宜的分类及对应的样本分布。聚类算法通常作为数据提取常用的算法,可以用于处理原始数据作为其他算法的初始数据处理工具,也能单独用作数据分析算法。数据集通过特定学习方式将给定的数据集分成几个类簇,其结果是,同一组中的数据尽可能相似,不同组数据之间的差异是明显的,从而实现了数据分类的目的。

基于划分的聚类算法:给定一个数据集和要划分的类别数,数据集中的所有数据首先进行分类,然后根据每组数据迭代更新其数据中心,将数据根据新的中心重新分类到其所属类别,以便在达到具体条件时提高分类的质量。代表算法有k-means和fuzzy C-means。划分聚类使用简单高效,并且时间和空间复杂度都较低。然而,当数据量较大时,容易发生局部结果最优而停止继续迭代优化的情况,需要根据实际数据情况,预先确定合理的聚类中心,避免局部最优。

模糊聚类算法是数据挖掘模型中的常用算法,在常用的模糊聚类算法中,k均值聚类(k-means)算法与模糊c均值法的使用范围最广,这两种算法通过对比不同分类数据的相似程度,将不同分组的数据中心进行迭代明确,达到数据分析需要的结果。目前,大多数聚类算法是根据数据之间的差异来进行迭代和分类的。然而,可通过考虑数据变量的相关属性进行研究,以获得更多信息。与此同时,可构造被指定对象与其属性变量到同质块最优划分的聚类算法,这种聚类算法被称为块聚类。

根据聚类算法,构建IF-THEN规则,并根据历史窃电数据进行建模,将历史数据进行初步分类并确定每一类数据的聚类中心。

依据模糊C均值聚类方法,求取待测点与聚类中心的距离平方数,即该类别内各样本点与对应中心的偏离程度,再将此类样本的全部距离平方和进行迭代计算。

将待测数据代入历史聚类结果进行计算,能够得出它到每个聚类中心的距离,根据历史聚类的正常或窃电概率来估计该待测数据对应的窃电概率分布。也可以将此数据纳入原始数据库,对隶属度矩阵和聚类中心进行更新,便于后续使用。

3.2.2 算法特点

模糊聚类算法的计算结果提供待测数据窃电的概率并初步筛选出疑似窃电行为的电力用户供进一步排查。优点在于可以直接筛选出窃电可能发生的具体样本点,但与此同时,若要实现对大量用户的同时计算,

在计算速度和结果的输出判别上存在一定的困难,且需要预先给定聚类数目和初始中心,对噪声和离群值比较敏感。

4 窃电识别技术的应用

采用基于大数据算法的窃电识别技术可以对电力用户用电大数据进行初步分析排查,基本确定窃电范围后再进行进一步人工排查分析。某供电公司利用数据分析结果,结合公司营销部日常用户管理工作及反窃电大行动等活动集中进行排摸处理。据统计,仅2023年1月1日至3月8日期间,该公司共查处窃电、违约用电等影响线损的案件99件,完成追补电量41.6万千瓦时,追补电费和违约使用电费共计91.39万元,大大提升了窃电查处效率。由于乡村窃电数量占比大,地理分布广,采用大数据手段大大减少了工作人员的排查工作量。当前,全面推进乡村振兴战略,为乡村振兴提供坚强的能源保障是各供电公司的重要任务之一。加大窃电查处力度、减少线路损耗,保障乡村能源的安全、合理使用,确保计量的准确性是重要一环。结合窃电高损台区治理工作,融入大数据识别技术,对于科学有序推进理论线损、采集设备改造、高损负损治理等达标率提供了高效途径。

5 结语

随着经济社会的发展和国民素质的不断提升,以及对对应窃电查处力度的加大,窃电本身发生的数量已经在逐渐减少。窃电的大数据分析在快速识别和精准定位方面为实际用电检查工作提供了有效的技术手段,大大减轻了窃电用户筛选方面的工作量,提升了工作效率,为配电网智能化发展的经济和技术方面都提供了有力支撑。

参考文献:

- [1] 张晶,刘晓巍,张松涛.基于营销大数据的用电异常事件统计及窃电特征分析[J].供用电,2018,35(06):77-82.
- [2] 史玉良,荣以平,朱伟义.基于用电特征分析的窃电行为识别方法[J].计算机研究与发展,2018,55(08):1599-1608.
- [3] 潘明明,田世明,吴博,等.基于智能电表数据的台区识别与窃电检测方法研究[J].智慧电力,2017,45(12):80-84.
- [4] 薛峪峰,张俊超,马晓琴.基于大数据高维随机矩阵的反窃电识别定位[J].自动化与仪器仪表,2020(11):220-222.
- [5] 郑思达,梁琪琳,彭鑫霞,等.基于模糊聚类的异常用电行为识别研究[J].电测与仪表,2020,57(19):40-44.