

DOI: 10.19783/j.cnki.pspc.200050

电力设备状态监测大数据发展综述

方静¹, 彭小圣², 刘泰蔚², 陈玉竹², 李文泽², 文劲宇², 熊磊³, 王浩鸣⁴

(1. 国网天津市电力公司电缆分公司, 天津 300170; 2. 强电磁工程与新技术国家重点实验室(华中科技大学), 湖北武汉 430074; 3. 国网湖北省电力公司襄阳供电公司, 湖北 襄阳 441000; 4. 国网天津市电力公司, 天津 300010)

摘要: 论述了大数据在电力设备状态监测上的发展趋势与应用前景。首先分析了状态监测数据的大数据特征, 并从大数据技术、大数据思想方法、大数据算法三个层面论述了大数据对电力设备状态监测的提升点。其次给出了基于大数据的电力设备状态监测系统架构, 并从数据采集、数据去噪、特征提取、模式识别、知识挖掘、数据可视化几个方面论述了大数据与状态监测各个环节的结合点。最后通过一个综合监测系统案例, 分析了大数据在多源异构数据融合、综合分析及诊断、设备故障预测上的应用。大数据在电力设备状态监测上的深入应用, 有利于解决设备状态评价和故障预测的难题, 推动该领域朝着更加智能化的方向发展。

关键词: 大数据; 状态监测; 关键技术; 发展趋势; 应用前景

Development trend and application prospects of big data-based condition monitoring of power apparatus

FANG Jing¹, PENG Xiaosheng², LIU Taiwei², CHEN Yuzhu², LI Wenzhe², WEN Jinyu², XIONG Lei³, WANG Haoming⁴

(1. State Grid Tianjin Electric Power Company Cable Branch, Tianjin 300170, China; 2. State Key Laboratory of Advanced Electromagnetic Engineering and Technology (Huazhong University of Science and Technology), Wuhan 430074, China; 3. Xiangyang Power Supply Company, State Grid Hubei Electric Power Company, Xiangyang 441000, China; 4. State Grid Tianjin Electric Power Company, Tianjin 300010, China)

Abstract: This paper presents the development trend and application prospects of big data-based condition monitoring of power apparatus. Based on the analysis of big data attributes of condition monitoring data, three big data application directions of condition monitoring system are discussed, including big data technologies, theories and algorithms. A universal architecture of a big data-based condition monitoring system is presented. Data acquisition, data denoising, feature extraction, pattern recognition and data mining are discussed in detail to present the relationships of big data and all procedures of a condition monitoring system. An application example of a comprehensive condition monitoring system is presented, including big data-based data fusion, condition evaluation and failure-prediction. The application of big data based condition monitoring has great potential to overcome the challenges of condition evaluation and failure prediction and will contribute to the development of intelligent condition monitoring systems of power apparatus.

This work is supported by National Natural Science Foundation of China (No. 51807072).

Key words: big data; condition monitoring; key technologies; development prospects; application fields

0 引言

电力设备状态监测通过信号传感、数据采集、数据处理等步骤获取设备健康状况相关的特征参

数, 评价设备状态, 预测设备故障, 一方面能充分延长健康状况良好设备的停机检修时间, 提高设备运行经济性, 另一方面能检测出有潜在故障的设备, 及时维修, 提高设备的可靠性, 因此在发电机、变压器、气体绝缘开关、电力电缆等主电力设备中获得了广泛的应用^[1-3]。然而由于电力设备固有的机械、物理、电气内部复杂性和设备故障信号固有的传播衰减、耦合干扰、反射振荡等特性, 电力设备

基金项目: 国家自然科学基金项目资助(51807072); 国网天津市电力公司电力科学研究院项目资助(sgtjtk00pjjs1800092)

状态正确评价和故障准确预测是该领域的固有难题,造成了误报警和漏报警等情况,为电力企业带来损失^[4-5]。大数据为该固有难题的解决带来了新的发展契机。

大数据是指其大小超出典型数据软件抓取、储存、管理和分析范围的数据集合^[6-7]。也有观点认为大数据是指需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产^[6-7]。业界对大数据的定义并未统一,但对大数据特性的描述却大同小异,比较有代表性的为 4V 特性,即规模性(Volume)、高速性(Velocity)、多样性(Variety)、价值性(Value)^[7]。规模性(Volume)是指大数据的数据量非常庞大,计量单位达到 TB 级、PB 级甚至 ZB 级;高速性(Velocity)是指数据洪流实时增长,且增长速度快;多样性(Variety)是指大数据的数据类型众多,不仅包括结构化的表格,而且包括半结构的文本、视频、图像、语音及非结构化的文件;价值性(Value)是指大数据中隐藏着价值,但是价值密度稀疏,必须通过知识挖掘才能获取^[8-9]。近年来部分研究者将大数据技术应用用于电力设备状态监测^[10-15]。文献[10]提出一种基于大数据分析的输变电设备状态数据的异常快速检测方法。文献[11]提出一种基于高维随机矩阵理论的大数据模型,以实现设备关键性能评估和异常状态检测。文献[12]提出一种基于无连接层次编码的电力设备状态监测数据模型,以达到优化 OLAP 分析性能的目的。文献[13]开展了基于 Hadoop 云计算平台进行输变电设备状态监测大数据存储优化和基于 MapReduce 的并行分析处理的研究。文献[14-15]中结合大数据技术及数据挖掘分析方法在设备状态评估中应用的现状,分析了大数据技术在电力设备状态评估中的典型应用场景和应用效果。

电力设备状态监测数据,包含在线监测数据、带电检测数据、预防性试验数据等,具有数据量大、增长迅速、类型众多、价值密度稀疏等特点,是典型的大数据^[10]。体现在以下几个方面:

(1) 监测对象数量庞大。电力系统由发电机、变压器、气体绝缘开关、电力电缆等主电力设备构成,这些设备型号各异、种类繁多、数量庞大,实现对这些主设备较大覆盖范围的监测,对象本身已经达到了较高的数量级^[13]。

(2) 监测参数种类繁多。以变压器在线监测为例,含接地电流、振动、温度、微水、介质损耗、油色谱、局部放电等参数,这些监测参数的监测结果中既包含结构化的数据也包含图片、视频、文本等半结构化和非结构化的数据^[16]。

(3) 监测数据增长迅速。电力设备状态需要不间断的数据采集,获取的原始数据呈指数级增长。以三相高压电缆局部放电在线监测为例,其典型的数据采样率为 100 MS/s,每天的监测数据量与电缆数量、采样时间间隔的关系如表 1 所示。从表 1 可以看出该监测参数一天的原始数据就已达到 TB 级。

表 1 电缆在线监测的原始数据数量

Table 1 Data amount of cable condition monitoring systems

电缆监测数量	监测数据量	
	采集时间间隔 1 min	采集时间间隔 1 s
3	8.45 G	507 G
300	845 G	50 700 G
3 000	8 450 G	507 000 G

(4) 数据分析需要综合考虑更多的信息源。电力设备状态监测数据的分析与评价需要综合考虑设备监测历史数据、设备离线实验数据、设备台账信息、同类型设备故障信息、电网工况信息、设备状态评价标准等数据源,才能对设备状态做出正确评价。

(5) 价值密度稀疏。电力设备状态监测数据实时增长,但能捕捉到故障发生征兆的数据却较少,这是因为一方面电力设备发生故障的概率不高,例如文献[17]中统计的变压器故障概率是 1.8%,另一方面部分故障从产生到设备发生事故的时间较短,因此电力设备在线监测数据大部分是平稳数据,价值密度稀疏^[16]。

综上所述,电力设备状态监测数据与大数据的 4V 特征吻合较好,是典型的大数据。采用大数据技术或大数据思想方法开展电力设备状态监测与数据分析,有适应其固有特点的独到优势,并将有利于解决设备状态评价与故障准确预测的难题,进而避免维修不足与维修过剩带来的经济损失,提高电力设备的运行可靠性,保障供电安全^[18-19]。

1 大数据对电力设备状态监测的必要性

大数据对电力设备状态监测的提升主要体现在大数据关键技术、大数据思想方法和大数据算法三个层面。关键技术层面是指处理大数据的分布式计算技术、内层计算技术、流处理技术、批处理技术、分布式存储技术、非关系型数据库技术等能直接用于海量状态监测数据的计算、存储;思想方法层面是指大数据的全局化思想、并行化思想、分布式思想等能对状态监测的系统架构、处理流程、通信协议等方面产生启发性指导;大数据算法层面是指大数据自动化算法对电力设备状态监测中的模式识别、故障预测等方面的提升^[20]。

1.1 大数据关键技术

大数据关键技术含数据集成管理技术、数据处理技术、数据分析技术、数据展现技术等^[7], 这些技术大多来源于大数据兴起的互联网与信息技术领域, 经过最近几年的发展, 具备较高的成熟度和可移植性, 因此将这些关键技术应用到电力设备状态监测领域将会具有较强的生命力。

大数据处理技术含分布式计算技术、内存计算技术、流处理技术等^[7,9,21]。这些技术能在较短的时间能完成 PB 级、ZB 级海量数据的处理, 可直接用于电力设备状态监测海量原始数据的处理。

大数据存储技术含分布式文件管理技术、非关系型数据库技术等。分布式文件管理技术一个典型的代表就是 HDFS(Hadoop Distributed File System)^[21]。HDFS 是采用流式数据访问模式来存储超大文件的文件系统。非关系型数据库一个典型的代表就是 HBase。HBase 是一个构建在 HDFS 之上的分布式面向列存储的数据库系统。大数据存储技术适合数据量较大的电力设备状态监测原始数据和特征参数的存储^[9]。

1.2 大数据思想方法

大数据发展过程中形成的思想方法对电力设备状态监测具有借鉴价值, 这主要体现在全局化、分布式、并行化、低成本四个方面。

1.2.1 全局化

大数据从整体而不是从个体的角度开展分析是其具有强大生命力的关键因素之一, 因此大数据是全面的数据, 从大样本中发掘大价值是其核心思想, 以解决“数据爆炸、知识匮乏”的短板^[7-8]。文献[22]用盲人摸象的寓言故事阐述了大数据的全局化思想, 形象说明了小样本分析方法如同盲人一样只能认识庞大系统的局部特性而诊断出错误的结论。

电力设备状态监测具有庞大的对象和复杂的电气物理连接, 而现有的状态监测系统大多从单个设备的单一监测参数开展监测与分析, 容易造成误诊断, 因此将大数据的全局化思想应用到电力设备的监测与诊断分析上, 从标准规范统一、系统整体设计、数据同步采集、数据全局分析等角度对现有的监测系统进行重新设计与改进, 将在很大程度上提升电力设备状态监测的可用性和价值性。

1.2.2 分布式

大数据中普遍采用了分布式文件管理系统和分布式数据库系统来开展原始数据和参数的存储, 以应对海量数据的挑战, 典型的案例有 Apache 基金会的 Hadoop, Google 的 GFS(Google File System) 等^[23-25]。GFS 采用了分布式文件系统, 在全球部署

了超过 200 万台服务器, 以应对每天数以亿计的搜索请求和 24PB 级数据增长^[24]。

由于电力设备在地理上的分散布置特性, 因此电力设备状态监测系统本身就具有分布式特性。这与大数据的分布式思想具有一定的相似性, 但是又有本质的区别, 例如 Google 的分布式文件管理系统虽然是分散布置, 但是能联合工作, 在毫秒级的时间内给出综合搜索结果, 而电力设备在线监测系统大多没有联合工作的特性, 例如某变电站的变压器、GIS、电缆、避雷器的监测, 虽然是分布式布置, 但由于数据采集没有做到时间上完全同步采集, 不能开展综合分析诊断的高级功能, 因此提升现有电力设备状态监测系统各子系统之间的耦合关系, 是大数据分布式思想在电力设备状态监测上可能获得应用的新方向。

1.2.3 并行化

MapReduce 编程模式是大数据处理中普遍采用的方法, 其实现原理在于通过 Map 将任务分解成小的模块, 发送给较多的节点并行计算, 再通过 Reduce 将结果汇总, 得到最终结果。其核心思想是“分而治之、移动逻辑、屏蔽底层、处理定制”^[9]。并行化的鼻祖可以追溯到曹冲称象的故事, 通过将无法称量的大象分散为可以计量的石块, 用小工具并行处理, 再将结果汇总, 得到大数据的结果, 就是典型的 MapReduce 思想。

MapReduce 的思想已经在电力设备在线监测中得到使用, 并可以继续发展, 发挥其更为广阔的内涵: 状态监测系统本身包含不同的监测子单元, 例如变压器在线监测系统包含局部放电、接地电流、微水、油色谱、温度等不同的监测单元, 这些单元联合工作完成变压器综合监测就体现了 Reduce 的思想; 在较多台电力设备的在线监测中采用时分复用的思想, 将监测单元的计算资源拆分使用, 即为 Map 的思想; 在线监测数据的数据处理, 采用多线程技术, 也是 Map 思想的应用。

在海量监测原始数据的处理中, 将数据分割成较小的子模块, 发送给子单元处理, 再将结果汇总, 可大大减少计算时间, 例如, 各种电力设备在线监测单元本身就是一个高性能的处理子站, 现在的电力设备综合监测系统数据传输大多是单向的, 即原始数据和指标量从现地子监测单元向总站传输, 在未来的监测系统中, 设计双向数据高速传输通道, 利用 MapReduce 思想将总站的海量数据发送到子站并行处理, 是一个值得研究的方向。

1.2.4 低成本

大数据获得成功的另一个关键因素是采用了低

成本的设计思想, 例如 Google、百度、Facebook 等大型互联网公司都没有采用昂贵的超级计算机而采用了廉价的计算机组成的集群来存储大数据, 才能在互联网大数据中获得成功, 因此低成本是大数据得以生存之本^[9]。

电力设备状态监测普遍存在监测系统成本过高的问题, 部分电力设备状态监测系统的成本达到了主电力设备成本的 10% 甚至更高, 这是电力设备状态监测没有普及的重要原因之一, 因此如何降低传感器成本、现地监测单元成本、数据服务器成本、软件成本、系统维护成本等, 将状态监测系统“平民化”是推广状态监测中的迫切问题。例如将昂贵的工业控制计算机用廉价的单片机和 ASIC (Application Specific Integrated Circuit) 代替, 是降低监测成本的一个有效途径。

1.3 大数据算法

大数据在发展过程中形成了一些优秀算法, 其中与状态监测匹配度较高的一个方向就是大数据预测算法。大数据预测算法一个最典型的应用就是 IBM 的第三代超级计算机 Watson^[25-26]: 2011 年在美国最受欢迎的电视问答类节目《危险边缘》中, IBM 的第三代超级计算机 Watson 采用大数据技术, 打败了人类两位问答冠军。Watson 的技术实现了“让机器像人一样思考”, 在几秒钟的时间内搜索海量的资料, 给出预选的答案, 并通过预测模型 6 次反复预测与评估, 给出最终的答案。IBM 的高管们认为 Watson 是代表了第三个计算机时代的第一台

机器, 并誉其为人工智能的真正到来^[21-22]。IBM 正在为 Watson 申请美国的医疗执照, 在不远的将来即将应用到医疗领域, 为广大患者带来福音。基于海量病例库、药物报告、治愈方案、医疗期刊、病人历史治疗记录、病人基因序列、病人化验报告等数据源的大数据精准医疗, 将会更加准确地预测出病人的病情走势并给出科学合理的治疗方案, 这将带来医疗领域的革命。

电力设备状态监测通过形式各异的传感器获取设备健康状况与医生通过化验测试获取病人病情有很强的相似性, 因此借鉴 Watson 对海量数据的处理与预测方法, 基于海量故障案例库、设备台账信息、设备在线监测信息、设备历史监测记录、同类型号设备故障信息、电网工况信息等数据源, 设计基于大数据的电力设备故障诊断与预测方案, 有望为该领域带来新的变革。

2 基于大数据的电力设备状态监测总体架构

基于大数据的电力设备状态监测系统总体架构如图 1 所示, 主要包括电力设备在线监测系统、大数据数据源、大数据存储、大数据处理、大数据分析、高级应用功能以及大数据展现等模块。其主要思想是以电力设备状态监测相关数据为数据源, 利用大数据存储技术、大数据处理技术以及大数据分析技术实现大数据在电力设备状态监测领域的高级应用, 并通过大数据展现技术对结果进行可视化展示。

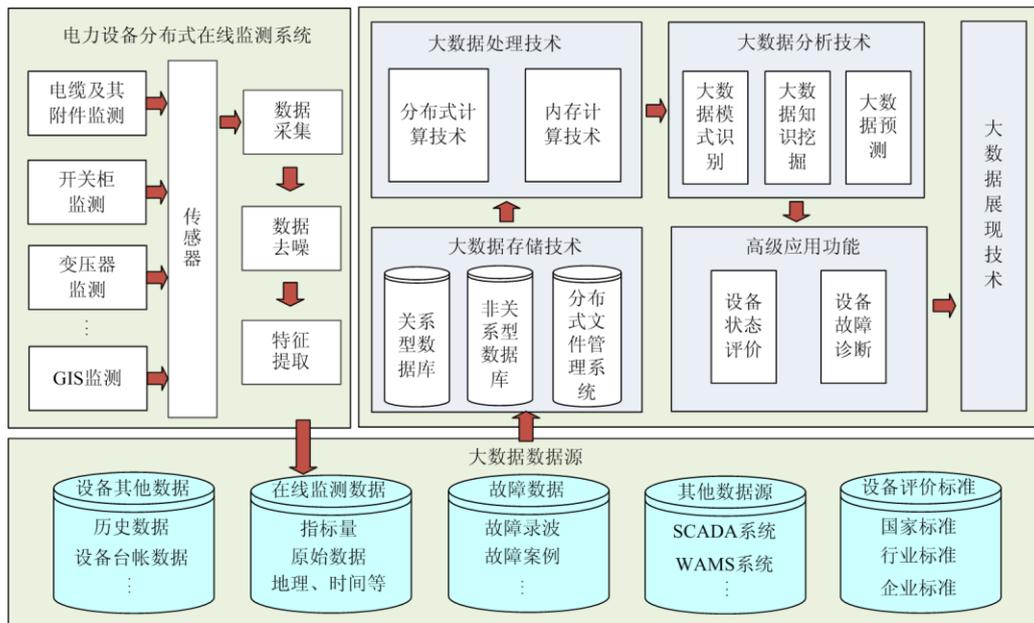


图 1 基于大数据的电力设备状态监测系统总体架构

Fig. 1 Architecture of big data based condition monitoring of power apparatuses

除了设备在线监测数据之外, 状态监测大数据源还包括设备其他数据、故障数据、设备评价标准、其他数据源等。设备的其他数据包含监测历史数据和设备台账等; 故障数据包括故障录波数据、海量故障案例库等; 其他数据源包含 SCADA 系统、WAMS 系统、MIS 系统等; 设备的评价标准含国家标准、行业标准和企业标准等。

状态监测大数据通过大数据存储技术和处理技术进行集成管理和数据处理。适用于状态监测大数据的存储技术既包含传统的关系型数据库, 也包含非关系型数据库和分布式文件。大数据处理技术可采用分布式计算技术、内存计算技术等。存储和计算技术是实现技术手段, 其主要目的是通过大数据分析实现设备状态评价、设备故障预测等高级应用功能。大数据分析技术含大数据知识挖掘、大数据模式识别、大数据预测等。该通用架构还包括大数据的展现模块, 将大数据预测及高级应用功能的结果通过互联网、三维可视化模型、空间信息流等可视化技术向运行管理人员实时展示。该架构中的大数据存储技术和处理技术可采用在互联网和信息技术领域已发展得比较成熟的软硬件方案, 例如 HDFS、HBase 等。而数据采集、数据去噪、特征提取、模式识别、知识挖掘等环节则需要结合状态监测的需求和特点, 提升和完善传统技术。本文将在下节对这些技术进行详细阐述。

3 基于大数据的电力设备状态监测关键技术

电力设备状态监测的基本流程含数据采集、数据去噪、特征提取、模式识别、知识挖掘等环节。大数据技术、思想方法、自动化算法在各个环节都能得到结合与应用, 这一方面是因为大数据技术本身能在海量数据的存储、处理、展现上具有独到的优势, 另一方面是因为大数据高级分析功能对各个环节提出了新的要求。

3.1 面向大数据应用的数据采集技术

由状态监测参数含振动、摆度、温度、湿度、压力、局部放电、电流、电压、微水、色谱、红外、图像、气体含量等机械、物理、化学信号^[12,27], 这些信号通过传感器变送成电压或电流信号, 并通过数据采集单元转换成原始数据。原始数据的采集受传感方式、采样频率、采样时间间隔、触发方式等多种因素的影响。大数据分析对这些因素提出了新的要求, 其发展动向主要体现在以下几个方面:

1) 数据采集的标准化。大数据分析融合更多的数据源进行全面分析, 然而现有的数据采集系统采用的标准各异, 造成了数据融合困难, 无法直接进

行数据分析与知识挖掘。通常需要对这些数据进行预处理, 增大了数据处理的难度, 对于某些不同类型的数据, 甚至无法操作。因此基于大数据的数据采集技术应使用统一的数据采集标准。

2) 数据采集的同步化。大数据分析采用更为全面的数据进行横向和纵向的对比分析, 例如对一个包含变压器、GIS、电力电缆的综合监测系统进行分析, 可以对比同一时刻信号的差异性开展综合分析从而提高诊断的准确性, 但是这要求这三个电力设备的状态监测系统共用同一个 GPS 时钟以实现数据采集的完全同步, 因此数据采集的同步化是电力设备状态监测开展大数据分析的一个重要的前提条件。

3) 数据采集的高可靠性。大数据分析虽然可以开展多源异构数据的融合和异常数据检测^[10], 但是低可靠性的数据采集系统获取的数据, 将加大后期数据处理的难度。现有的电力设备状态监测系统由于传感器异常、信号调理电路故障、数据采集硬件设备可靠性低、数据采集软件缺陷等原因, 存在数据采集异常终止造成数据缺失的情况, 因此从质量监测与采购管理的角度提高数据采集的可靠性是提升状态监测价值的一个重要途径。

4) 数据记录的全面性。大数据分析采用的数据源不仅有设备状态监测数据, 还需综合考虑故障传播耦合、故障发展趋势、设备故障记录、设备维护记录等更加众多的数据源, 因此数据采集单元在存储原始数据与指标量的时候必须同步记录时间信息、设备信息、地理信息、参数设置等信息, 使数据逐级上传后不丢失关键信息, 便于开展大数据分析与知识挖掘。

3.2 面向大数据的数据去噪技术

设备故障信号与噪声和干扰信号混在一起, 大大降低了监测信号的可用性^[27-30]。干扰信号的误判断会产生误报警, 降低运行管理人员对状态监测系统的信赖; 故障信号没有准确提取会产生漏报警, 带来重大损失, 因此数据去噪与干扰抑制是电力设备状态监测领域的一个亟待突破的难题。

数据去噪技术在电力设备状态监测领域已经得到了蓬勃的发展, 国内外的研究者已经将包括 FFT 去噪、小波去噪、自适应去噪、经验模态分解等数学方法引入到设备故障信号的去噪中, 部分方法获得了较好的效果^[28-30]。

基于大数据的状态监测数据去噪的发展动向主要体现在以下几个方面:

1) 分布式与集中式相结合。如果将电力设备状态监测原始数据集成起来开展数据去噪, 将需要昂

贵的计算存储资源和较大的计算代价。大数据的分布式思想可以应用到原始数据去噪, 因为电力设备状态监测系统现地监测单元一般采用具有较高运算速度的工业控制计算机作为运行控制核心, 大量的现地监测单元分布在广泛的地理空间, 本身就是一个分布式计算系统, 因此对监测数据中与故障信号频段相差较大的噪声与干扰信号在现地完成数字去噪, 将大大降低后期集中处理的难度。

2) 大数据干扰抑制技术。状态监测信号的干扰源有白噪声、随机干扰、周期干扰、通信干扰、广播干扰等。部分干扰信号与待提取的故障信号在频段上有较多重叠部分, 例如电力电子技术在电力设备上大规模使用使得可控硅信号成为一个重要的干扰源。通过大数据技术建立典型干扰信号的特征参数库, 在运行控制中枢开发基于海量样本库的干扰信号抑制技术, 将有效识别传统的去噪技术无法剔除的干扰信号。

3) 综合分析判断。位置临近的电力设备具有较强的电气物理连接, 这使得同一故障信号在不同的电力设备状态监测系统中都能检测到, 从而造成误诊, 因此在更高的层面开展综合分析判断, 可以进一步剔除临近电力设备耦合进来的干扰信号, 减少误诊断。

3.3 面向大数据的特征提取技术

特征提取是开展大数据模式识别和知识挖掘的前提。基于大数据的特征提取的发展动向主要体现在以下几个方面:

1) 大数据降维。从不同的时间尺度、描述方法、组合方式来描述故障信号会得到不同的特征, 例如局部放电信号的特征就包括统计特征、分形特征、小波特征、矩特征、波形特征、组合特征、过程特征、序列特征、盒维数特征、信息维数特征等, 多达数十种甚至上百种, 采用粗糙集、主成分分析等方法从数十种特征中优选出核心特征参数, 降低特征参数的维度, 是提高大数据表现力一个重要的手段^[30-32]。

2) 特征抽取。电力设备状态监测信号具有价值密度稀疏的特点, 典型表现为信号平稳时间长、故障过度时间短, 因此通过特征抽取从较长监测时间段内的数据中获取最具代表性的特征参数, 能提高后续模式识别和知识挖掘准确性。

3) 并行化特征提取。海量样本的特征提取需要繁重的计算代价和较长的计算时间, 因此借鉴大数据并行化的思想, 将特征提取并行化, 可以在有限的时间内得到理想的结果, 例如基于 Spark 的特征提取、基于 MapReduce 的特征提取、基于云计算的

特征提取等。

3.4 面向大数据的模式识别技术

大数据识别可以从传统模式识别方法的并行化和大数据模式识别新方法两个方面入手。

1) 传统模式识别的并行化。麦肯锡认为可用于大数据分析的模式识别算法包括关联分析、机器学习、神经网络、时间序列预测模型、遗传算法等^[33-34]。文献^[20]也指出“把传统的机器学习算法运用到大数据环境下一个典型的策略是对现有的学习算法并行化”。

2) 大数据深度学习算法。大数据算法中应用得最成功的一个算法就是大数据深度学习。大数据深度学习结合了大数据的样本优势和深度学习的特征学习优势, 能够刻画更为复杂的非线性关系, 在语音识别、手写字体识别、图像识别等传统模式识别领域获得了巨大的成功。例如 2011 年 Google 的语音识别研究人员采用大数据深度学习技术, 将语音识别错误率降低了 20%~30%, 是语音识别领域 10 多年来最大的突破性进展^[35]。2012 年在 ImageNet 的大规模图像识别挑战赛上, 参赛者采用大数据深度学习技术, 将错误率从 26% 降低到 15%, 引起广泛的关注^[31]。电力设备状态监测中普遍采用模式识别技术, 例如局部放电识别、振动信号识别、红外图像识别、油色谱识别等^[28-31,36]。部分监测信号的模式识别存在泛化能力欠缺、识别精度较低等问题, 因此将大数据深度学习算法应用到状态监测数据模式识别上, 为这些难题的解决带来新的思路。

3.5 面向大数据的数据挖掘技术

特大数据知识挖掘一方面可以将传统的知识挖掘方法在高级计算资源和并行化计算平台下推广应用^[9,20], 另一方面可采用的一个重要方法就是大数据相关关系性分析。所谓“相关性”是指 2 个或 2 个以上变量的取值之间存在某种规律性, 其目的是找出海量数据集里隐藏的关系网, 一般用支持度、可信度、兴趣度等参数反映相关性^[7]。大数据相关性分析在商业领域获得了巨大的成功, 例如, 沃尔玛通过大数据分析天气情况、顾客购买清单发现飓风来临时, 蛋挞和飓风用品具有很强的相关性, 通过调整飓风来临时两者的摆放位置, 大大提高了销量^[37]。大数据这种只注重相关关系而不注重因果关系的数据分析方法带来了科学研究思维方法的重大改变。数据密集型科研被誉为科学研究的第四范式^[7,38]。

大数据相关性分析对电力设备状态监测中的设备状态评价有很好的适用性。对电力设备状态监测而言, 通过特征提取获得的特征量可能有几十种甚

至上百种，如何获取核心特征量评价设备健康状态是该领域的难题，通过大数据样本开展相关性分析，将有很大的可能性推动该科学问题的发展。

3.6 面向大数据的可视化技术

电力设备状态监测数据的可视化已经得到了广泛的应用，例如基于互联网的状态监测可视化、基于三维动画建模的可视化、手机短信报警提醒等。随着信息技术的不断发展，尤其是移动互联网和智能可穿戴设备的逐步应用，电力设备状态监测大数据可视化技术向着更加贴近运行管理人员的方向发展^[10,13,16]。

移动互联网是互联网与移动通信各自独立发展后互相融合的产物^[39]。从终端上定义，移动互联网是指用户使用手机、上网本、笔记本电脑、平板电脑、智能本等移动终端，通过移动网络获取移动通信网络服务和互联网服务^[40-42]。随着移动互联网时代的到来，电力工业生产逐渐引入移动互联设备，促使电力企业的生产经营更加高效。

基于移动互联网的状态监测大数据可视化技术，能进一步缩短运行管理人员与主设备之间的距离，方便用户随时随地获取设备健康状况和发展趋势，缩短预警时间，必将在电力设备状态监测中得到广泛的应用。

4 基于大数据的综合监测与分析系统应用案例

本节通过典型的具有电气和物理连接的变压器、GIS、电缆综合监测系统案例阐述大数据在电力设备状态监测上的应用前景。该综合监测系统的拓扑结构如图2所示，系统中的GIS和变压器通过三相高压电缆连接，高压电缆中间有两个中间接头。图中所示的综合检测系统分为三层：即现地层、站控层和中枢层。

现地层包含三个设备的现地监测单元，实现对16个监测参数的实时数据采集、数据去噪、特征提取。现地层三个监测单元通过同一个GPS时钟实现数据采集的同步。

站控层一方面实现对实时监测数据的集成管理，另一方面通过大数据存储技术实现变电站SCADA系统数据、设备评价标准、设备台账信息、设备监测历史数据的汇总，以此为基础并融合大数据思想方法，实现多源数据融合分析判断、专项参数对比分析判断、设备状态实时精确评价等大数据高级分析功能。

中枢层集成了更多的数据源，一方面存储站控层传输的实时监测数据，另一方面集成了设备海量故障案例库、设备监测历史数据、同类型设备故障记录、干扰信号案例库，以此为基础开展异常数据检测、干扰信号识别、故障模式识别、设备故障预测等大数据高级分析功能。

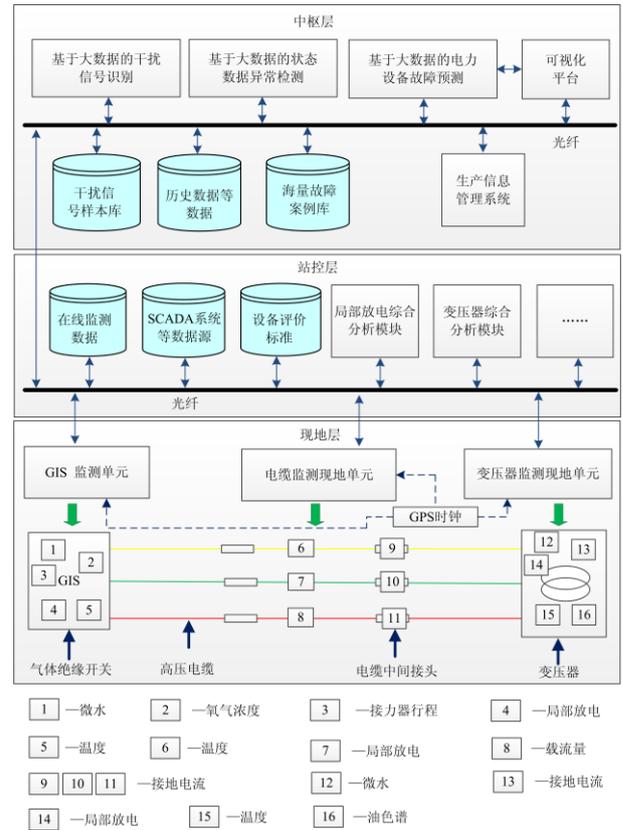


图2 电力设备综合检测系统结构图

Fig. 2 Architecture of a comprehensive condition monitoring system of power apparatuses

4.1 大数据应用点一：基于大数据的多源数据融合分析

对于电力设备的同一监测目标，从单一监测参数判断设备状态，存在误诊断的可能，从而误导检修人员。大数据全局化分析思想可以应用到对单一电力设备的多源数据融合分析上，实现对设备状态的精确评价，提高设备诊断的可靠性。

图2中所述的变压器绝缘状态监测多源数据融合分析流程如图3所示。局部放电监测、油色谱监测、温度监测三个参数都能反映变压器的绝缘状况，通过融合三个传感器实时监测数据，在站控层建立变压器绝缘状况综合分析模块，可实现传感器异常检测、变压器绝缘状况实时精确评价等高级功能。

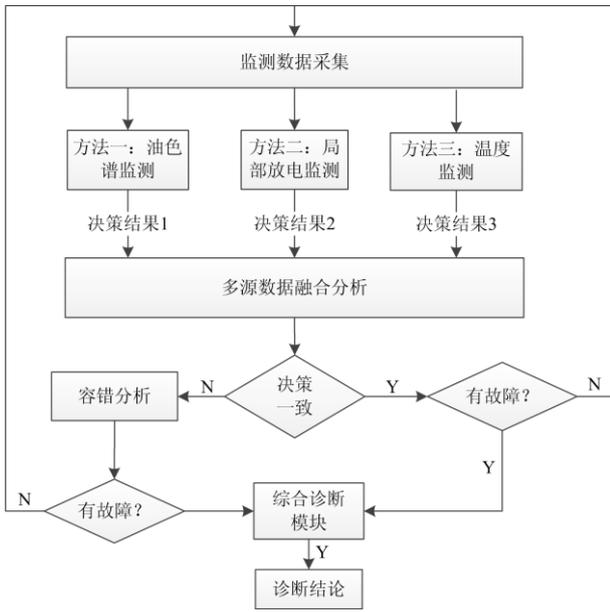


图 3 变压器绝缘状态监测多源数据融合分析流程
Fig. 3 Flowchart of data fusion of insulation condition monitoring of power transformers

图 2 中所述的电缆监测单元含温度、局部放电、载流量、接地电流 4 个监测参数, 以这些监测数据为基础, 通过大数据相关性分析, 可以建立电缆局部放电与温度、载流量、接地电流之间的关系, 实现对电缆局部放电特性的深入分析与评价。

4.2 大数据应用点二: 基于大数据的多目标设备综合分析

通过对多目标电力设备的同一监测参数综合分析与判断能提高电力状态监测诊断结果的可靠性, 减少误诊断。例如图 2 所述的变压器、GIS、电缆都包含局部放电监测模块, 由于局部放电信号沿有电气连接的电力设备传播, 造成一个源头的局部放电信号在三个电力设备的监测单元都能捕捉到, 从而相互成为干扰信号, 造成误诊断的情况。一个典型的多目标电力设备局部放电综合分析与诊断的流程如图 4 所示。该综合诊断分析一个重要的前提就是三个电力设备的局部放电监测数据采集通过同一个 GPS 时钟同步触发。通过分析三个监测单元局部放电信号时域特征和传播特征, 能精确诊断出局部放电信号的源头, 从而向各个子监测单元给出反馈信号, 修正现地预警结果。

4.3 大数据应用点三: 基于大数据的设备故障预测

电力设备故障预测是该领域的难题, 通过类似本文 1.3 节所述的大数据预测方法建立类似于 IBM Watson 的预测系统, 为该难题的突破带来新的解决思路。

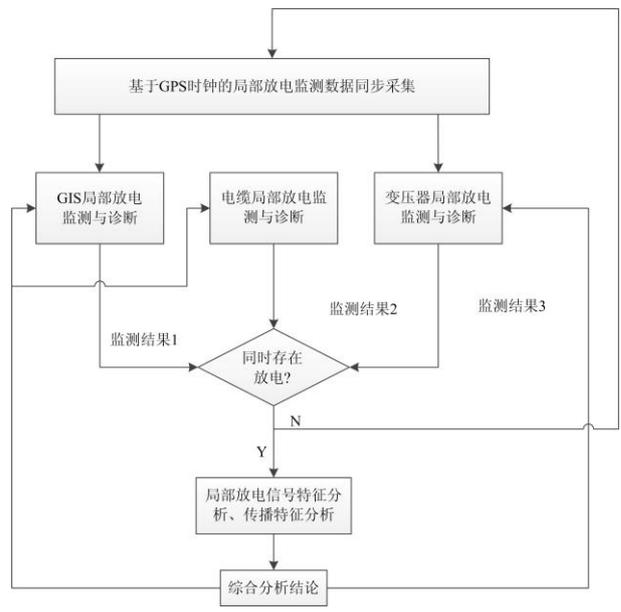


图 4 多目标电力设备局部放电综合分析与诊断流程
Fig. 4 Flowchart of data analysis and diagnostics of partial discharge condition monitoring of multi power apparatuses

一种基于大数据的电力设备故障预测系统原理如图 5 所示。该预测系统包含: 基于 MapReduce 的大数据并行处理模块, 任务管理、任务调度和任务监控模块, 设备故障预测模块, 基于 HDFS 的海量原始数据管理模块, 基于 HBase 的开源非关系型数据库模块, 结果发布与预警反馈模块。

设备在线监测数据(含设备在线监测原始数据、数据采集时间、数据采集地点、设备型号、设备编

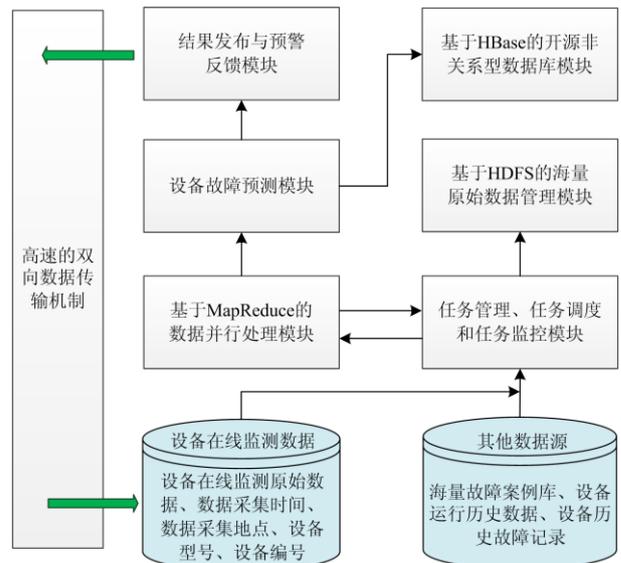


图 5 基于大数据的电力设备故障预测系统诊断流程
Fig. 5 Flowchart of big data based fault prediction of power apparatuses

号)通过高速的双向数据传输机制传输到基于大数据的电力设备预测平台。该平台通过任务调度和任务监控模块分析管理设备在线监测数据和其他数据源的数据(含海量故障案例库、设备运行历史数据、设备历史故障记录等),并将多源数据存储在基于HDFS的海量原始数据管理模块中,同时通过基于MapReduce的数据并行处理模块实现监测数据的实时分析,分析结果送入设备故障预测模块。设备故障预测模块通过大数据预测技术,给出具有较高准确度的诊断结果,并将诊断结果和相应指标量存储在基于HBase的开源非关系型数据库模块,同时将诊断结果送入结果发布与预警反馈模块。结果发布与预警反馈模块一方面通过大数据展示技术,实现对监测结果的多媒体发布,另一方面将通过高速的双向数据传输机制,将诊断结果反馈到现地监测单元,实现对现地预警信息的更新。

5 总结

本文探讨了电力设备状态监测数据的大数据特征并从大数据技术、大数据思想方法、大数据算法三个层面分析了大数据与电力设备状态监测的结合点,给出了一个具有通用性的电力设备状态监测大数据平台架构,对电力企业状态监测系统的搭建具有参考价值。同时本文从数据去噪、特征提取、模式识别、知识挖掘、数据存储、数据可视化深入分析和展望了大数据与电力设备状态监测各个环节的结合点。这些结合点一方面是因为大数据技术、大数据思想方法、大数据算法在状态监测各个环节具有独到优势,另一方面是因为大数据高级分析应用对这些环节提出了新需求。最后结合一个包含变压器、GIS、电缆的综合监测系统,深入分析了电力设备状态监测大数据应用前景。

基于大数据的电力设备状态监测既需要理论支撑又需要大数据工程的检验,而大数据样本的获取是制约其发展的瓶颈。本文虽然从整体理论框架和实现细节探讨了大数据在电力设备状态监测上的应用方向与前景,并给出了分析案例,但是没有与具体大数据工程的结合是本文的不足之处,需要进一步研究和完善。

大数据在电力设备状态监测上的深入应用将提高电力设备状态评价和故障诊断的准确性。可以预见,通过大数据样本、大数据软硬件技术、大数据思想方法、大数据算法建立电力设备状态评价与故障诊断医院,实现对电力设备疑难杂症的科学诊断,将引领行业发展、推动行业变革,促进绿色、自愈、安全、高效的坚强智能电网建设。

参考文献

- [1] PENG Xiaosheng, YANG Fan, WANG Ganjun, et al. A convolutional neural network-based deep learning methodology for recognition of partial discharge patterns from high-voltage cables[J]. IEEE Transactions on Power Delivery, 2019, 34(4): 1460-1469.
- [2] 汪可, 廖瑞金, 吴高林, 等. 采用双向改进模糊2DLDA算法提升多因素影响的局部放电识别可靠性[J]. 电工技术学报, 2014, 29(11): 210-220.
WANG Ke, LIAO Ruijin, WU Gaolin, et al. Improvement of partial discharge recognition reliability considering influence of multi-factors based on two-directional modified fuzzy 2DLDA algorithm[J]. Transactions of China Electrotechnical Society, 2014, 29(11): 210-220.
- [3] 张轶议, 廖瑞金, 杨丽君, 等. 基于云理论的电力变压器绝缘状态评估方法[J]. 电工技术学报, 2012, 27(5): 13-20.
ZHANG Yiyi, LIAO Ruijin, YANG Lijun, et al. An assessment method for insulation condition of power transformer based upon cloud model[J]. Transactions of China Electrotechnical Society, 2012, 27(5): 13-20.
- [4] PENG Xiaosheng, LI Jinshu, WANG Ganjun, et al. Random forest based optimal feature selection for partial discharge pattern recognition in HV cables[J]. IEEE Transactions on Power Delivery, 2019, 34(4): 1715-1724.
- [5] DONG Xian, JING Xiaoping, QIAN Kejun, et al. A smart grid needs smart monitoring[J]. International Journal of Smart Grid and Clean Energy, 2013, 2(3): 322-328.
- [6] 鲁宗相, 徐曼, 乔颖, 等. 风电功率预测的新型互联网运营模式设计[J]. 电网技术, 2016, 40(1): 125-131.
LU Zongxiang, XU Man, QIAO Ying, et al. New internet based operation pattern design of wind power forecasting system[J]. Power System Technology, 2016, 40(1): 125-131.
- [7] 彭小圣, 邓迪元, 程时杰, 等. 面向智能电网应用的电力大数据关键技术[J]. 中国电机工程学报, 2015, 35(3): 503-511.
PENG Xiaosheng, DENG Diyuan, CHENG Shijie, et al. Key technologies of electric power big data and its application prospects in smart grid[J]. Proceedings of the CSEE, 2015, 35(3): 503-511.
- [8] BIRNEY E. The making of ENCODE: lessons for big-data projects[J]. Nature, 2012(489): 49-51.
- [9] HU Zhuangli, HE Tong, ZENG Yihui, et al. Fast image recognition of transmission tower based on big data[J].

- Protection and Control of Modern Power Systems, 2018, 3(2): 149-158. DOI: 10.1186/s41601-018-0088-y.
- [10] 严英杰, 盛戈皞, 陈玉峰, 等. 基于大数据分析的输变电设备状态数据异常检测方法[J]. 中国电机工程学报, 2015, 35(1): 52-59.
YAN Yingjie, SHENG Gehao, CHEN Yufeng, et al. An method for anomaly detection of state information of power equipment based on big data analysis[J]. Proceedings of the CSEE, 2015, 35(1): 52-59.
- [11] 严英杰, 盛戈皞, 王辉, 等. 基于高维随机矩阵大数据分析模型的输变电设备关键性能评估方法[J]. 中国电机工程学报, 2016, 36(2): 435-445.
YAN Yingjie, SHENG Gehao, WANG Hui, et al. The key state assessment method of power transmission equipment using big data analyzing model based on large dimensional random matrix[J]. Proceedings of the CSEE, 2016, 36(2): 435-445.
- [12] 王德文, 周青. 一种电力设备状态监测大数据的分布式联机分析处理方法[J]. 中国电机工程学报, 2016, 36(19): 5111-5121, 5392.
WANG Dewen, ZHOU Qing. A method of distributed on-line analytical processing of status monitoring big data of electric power equipment[J]. Proceedings of the CSEE, 2016, 36(19): 5111-5121, 5392.
- [13] 宋亚奇, 周国亮, 朱永利, 等. 云平台下输变电设备状态监测大数据存储优化与并行处理[J]. 中国电机工程学报, 2015, 35(2): 255-267.
SONG Yaqi, ZHOU Guoliang, ZHU Yongli, et al. Storage optimization and parallel processing of condition monitoring big data of transmission and transforming equipment based on cloud platform[J]. Proceedings of the CSEE, 2015, 35(2): 255-267.
- [14] 廖瑞金, 王有元, 刘航, 等. 输变电设备状态评估方法的研究现状[J]. 高电压技术, 2018, 44(11): 3454-3464.
LIAO Ruijin, WANG Youyuan, LIU Hang, et al. Research status of condition assessment method for power equipment[J]. High Voltage Engineering, 2018, 44(11): 3454-3464.
- [15] 江秀臣, 盛戈皞. 电力设备状态大数据分析的研究和应用[J]. 高电压技术, 2018, 44(4): 1041-1050.
JIANG Xiuchen, SHENG Gehao. Research and application of big data analysis of power equipment condition[J]. High Voltage Engineering, 2018, 44(4): 1041-1050.
- [16] 国家电网公司生产技术部. 电网设备状态检测技术应用典型案例[M]. 北京: 中国电力出版社, 2012.
- [17] 余绍峰, 高胜友. 基于五年统计数据的变压器故障概率计算[C] // 2011年全国输变电设备状态检修技术交流研讨会, 2011, 中国, 珠海.
YU Shaofeng, GAO Shengyou. The probability of failure of the five-year statistical data based on the calculated transformer[C] // National Power Transmission Equipment State Maintenance Technology Seminar in 2011, Zhuhai, China.
- [18] HE Jinghan, LIN Liu, LI Wenli, et al. Development and research on integrated protection system based on redundant information analysis[J]. Protection and Control of Modern Power Systems, 2016, 1(2): 108-120. DOI: 10.1186/s41601-016-0024-y.
- [19] 廖瑞金, 刘捷丰, 杨丽君, 等. 电力变压器油纸绝缘状态评估的频域介电特征参量研究[J]. 电工技术学报, 2015, 30(6): 247-254.
LIAO Ruijin, LIU Jiefeng, YANG Lijun, et al. Investigation on frequency domain dielectric characteristics for condition assessment of transformer oil-paper insulation[J]. Transactions of China Electrotechnical Society, 2015, 30(6): 247-254.
- [20] 何清, 李宁, 罗文娟, 等. 大数据下的机器学习算法综述[J]. 模式识别与人工智能, 2014, 27(4): 327-334.
HE Qing, LI Ning, LUO Wenjuan, et al. Summary of machine learning algorithms under large data[J]. Pattern Recognition and Artificial Intelligence, 2014, 27(4): 327-334.
- [21] WHITE T. Hadoop: the definitive guide[M]. Third Edition. O'Reilly Media, Inc, 2011.
- [22] WU Xindong, ZHU Xingquan, WU Gongqing, et al. Data mining with big data[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(1): 97-107.
- [23] LAM C. Hadoop in action[M]. Manning Publications, 2010.
- [24] 曲朝阳, 朱莉, 张士林. 基于Hadoop的广域测量系统数据处理[J]. 电力系统自动化, 2014, 37(4): 92-97.
QU Zhaoyang, ZHU Li, ZHANG Shilin. Data processing of Hadoop-based wide area measurement system[J]. Automation of Electric Power Systems, 2014, 37(4): 92-97.
- [25] 阿里研究院. 互联网+从IT到DT[M]. 北京: 机械工业出版社, 2015.
- [26] 罗素. 人工智能——一种现代的方法[M]. 北京: 清华大学出版社, 2013.
- [27] HAO Liwei, LEWIN P L. Partial discharge source discrimination using a support vector machine[J]. IEEE

- Transactions on Dielectrics and Electrical Insulation, 2010, 17(1): 189-197.
- [28] TANG Ju, LIU Fan, ZHANG Xiang, et al. Partial discharge recognition based on SF6 decomposition products and support vector machine[J]. IET Science, Measurement & Technology, 2012, 6(4): 198-204.
- [29] SU M, CHIA C, CHEN C, et al. Classification of partial discharge events in GILBS using probabilistic neural networks and the fuzzy c-means clustering approach[J]. International Journal of Electrical Power & Energy Systems, 2014, 61: 173-179.
- [30] PENG Xiaosheng, ZHOU Chengke, HEPBURN D, et al. Application of k-means method to pattern recognition in on-line cable partial discharge monitoring[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2013, 20(3): 754-761.
- [31] 陈仁刚, 杨荣华, 曲文韬, 等. 一起 500 kV 变压器内部放电故障的检测与分析[J]. 电气技术, 2014(12): 106-110.
- CHEN Rengang, YANG Ronghua, QU Wentao, et al. Detection and analysis for accident caused by internal discharge of 500 kV transformer[J]. Electrical Engineering, 2014(12): 106-110.
- [32] 白俊良, 梅华威. 改进相似度的模糊聚类算法在光伏阵列短期功率预测中的应用[J]. 电力系统保护与控制, 2014, 42(6): 84-90.
- BAI Junliang, MEI Huawei. Improved similarity based fuzzy clustering algorithm and its application in the PV array power short-term forecasting[J]. Power System Protection and Control, 2014, 42(6): 84-90.
- [33] McKinsey & Company. Big data: the next frontier for innovation, competition, and productivity[M]. New York: McKinsey Global Institute, 2011: 1-28.
- [34] TIMOTHY H, BEZDEK J, LECKIE C, et al. Fuzzy c-means algorithms for very large data[J]. IEEE Transactions on Fuzzy Systems, 2012, 20(6): 1130-1146.
- [35] 余凯, 贾磊, 陈雨强, 等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展, 2013, 50(9): 1799-1804.
- YU Kai, JIA Lei, CHEN Yuqiang, et al. Deep learning: yesterday, today, and tomorrow[J]. Journal of Computer Research and Development, 2013, 50(9): 1799-1804.
- [36] 王风雷. 电力设备状态监测新技术应用案例精选[M]. 北京: 中国电力出版社, 2009.
- [37] 维克托·迈尔-舍恩伯格. 大数据时代(生活工作与思维的大变革)[M]. 杭州: 浙江人民出版社, 2013.
- [38] 高小明. 基于移动互联网的输变电巡检系统的设计与开发研究[D]. 广州: 华南理工大学, 2013.
- GAO Xiaoming. Design and research of power equipment inspection system based on mobile internet[D]. Guangzhou: South China University of Technology, 2013.
- [39] 黄罡, 刘譞哲, 张颖. 面向云-端融合的移动互联网应用运行平台[J]. 中国科学, 2013, 43(1): 24-44.
- HUANG Gang, LIU Xuanzhe, ZHANG Ying. Cloud-oriented integration platform for mobile internet applications[J]. Science China, 2013, 43(1): 24-44.
- [40] 曾治安, 姚树友, 郑晓玲, 等. 基于移动互联网技术的继电保护设备智能运维管理模式探讨[J]. 电力系统保护与控制, 2019, 47(16): 80-86.
- ZENG Zhi'an, YAO Shuyou, ZHENG Xiaoling, et al. Discussion on intelligent operation and maintenance mode of relay protection equipment based on mobile internet technology[J]. Power System Protection and Control, 2019, 47(16): 80-86.
- [41] 艾精文, 党晓婧, 吕启深, 等. 基于物联网的具有全景功能的全维度设备状态监测系统研究[J]. 电力系统保护与控制, 2019, 47(16): 122-128.
- AI Jingwen, DANG Xiaojing, LÜ Qishen, et al. Research on full dimension equipment status monitoring system with panoramic function[J]. Power System Protection and Control, 2019, 47(16): 122-128.
- [42] 高旭, 刘宏君, 杜丽艳, 等. 基于综合评价识别法的智能变电站虚回路在线状态监测技术研究[J]. 电力系统保护与控制, 2019, 47(3): 182-187.
- GAO Xu, LIU Hongjun, DU Liyan, et al. Study on the online status monitoring technology of intelligent substation virtual circuit based on comprehensive evaluation and recognition method[J]. Power System Protection and Control, 2019, 47(3): 182-187.

收稿日期: 2020-01-13; 修回日期: 2020-04-16

作者简介:

方静(1981—), 男, 高级工程师, 研究方向为电力电缆检修、运维、检测等;

彭小圣(1983—), 男, 通信作者, 博士, 副教授, IEC工作组专家, 研究方向为电力系统大数据理论与应用、电力系统主设备智能监测、局部放电在线监测与模式识别、新能源功率预测等. E-mail: XiaoshengPeng@hust.edu.cn

(编辑 周金梅)