

输变电设备状态大数据分析应用探讨

王 珣¹,王 馨²,赵 盟¹,许 鹏¹,马 琳³
(1. 国网冀北电力有限公司,北京 100053;
2. 国网冀北电力有限公司电力科学研究院,北京 100045;
3. 北京国网富达科技发展有限公司,北京 100070)

摘 要:随着电网数据急剧增加,传统的数据挖掘和分析方法已经不能适应当前智能电网的要求,而大数据分析为此提供了相应的实现手段。首先简述了大数据技术理论,并从输变电设备状态分析及应用的具体电网业务角度出发,开展数据整合、数据存储、数据计算、数据分析和结果可视化五部分工作;然后针对输变电设备历史缺陷数据和全过程技术监督的问题数据,应用主成分分析法和聚类算法,构建变压器设备缺陷特征分析模型,实现将设备缺陷内容归类特征贴标签,为电网运检人员提供相关决策的数据分析依据;最后介绍了开展输变电设备潜伏性故障关联预测研究的工作展望。

关键词:大数据;电网业务;词云;潜伏性故障

文章编号:2096-4633(2018)01-0001-05 **中图分类号:**TM72 **文献标志码:**B

近年来,电网建设进入了高速发展阶段,电网海量数据急剧增加,仅仅依靠传统的数据库查询检索机制和统计学方法很难获得有效信息,迫切需要能自动地、智能地将待处理的数据转化为有价值信息的技术,即对输变电设备安全、可靠、稳定运行的智能水平提出了更高的要求,而大数据分析为此提供了相应的实现手段。

大数据相关技术是国外率先研究,IBM、SAP 等服务公司都在争相涌入这个行业,技术平台提供者甲骨文、SAS、EMC、Google、亚马逊等都从中盈利^[1]。国内大数据研究多是采用传统的 IT 技术、软硬件工具和数学分析方法,对各种信息进行感知、获取、管理、处理和分析以及结果可视化展示^[2]。目前许多学者主要进行电力生产大数据研究^[3-6],包括大数据技术研究、应用研究等。

1 国内外大数据发展现状

2012 年,美国提出“大数据研究和发展倡议”,积极推进大数据核心技术研究 and 应用。同年,美国麻省理工大学、加州大学圣巴巴拉分校、波特兰州立大学、布朗大学、华盛顿大学和斯坦福大学等 6 所大学建立了以大数据存储、管理、处理、分析和可视化为目标的大数据科学技术中心^[7]。同时,英国牛津大学也成立了首个综合运

用大数据的医药卫生科研中心,通过收集、存储和分析大量医疗信息,探索新药物的研发方向和新治疗手段,为科学家更好地理解人类疾病及其治疗方法提供决策支持^[8]。

2012 年,中国计算机学会和中国通信学会都成立了大数据专家委员会,专门研究大数据分析及应用^[9],旨在推动我国大数据发展。2013 年,中国电机工程学会信息化专委会发布了《中国电力大数据发展白皮书》,将大数据引入到智能电网领域。

同时,中国电力科学研究院等国内电力科研机构开展了大数据技术在智能电网中应用的关键技术研究,并从电力系统的单个环节论述了大数据技术的应用^[10-18]。

2 大数据理论

大数据的定义不是一成不变的,而是呈现多样化发展的状态。广泛通用的定义是 2011 年国际数据中心 IDC 定义的大数据:“大数据技术描述了一个技术和体系的新时代,被设计于从大规模多样化的数据中通过高速捕获、发现和分析技术提取数据的价值”。

这个定义刻画了大数据的 4 个显著特征,即体量(Volume)、多样性(Variety)、价值(Value)和速度

(Velocity)。

- (1) Volume:数据体量大,一般在 TB 级及以上;
- (2) Variety:数据多源异构多样,包括传统的关系数据库存储类型的结构化数据和以文本、图像、视频、音频、e-mail、网页等形式存在的未加工、半结构化或非结构化数据;
- (3) Value:数据价值低,隐藏在海量数据中的有用信息所占比例较小。通过各种分析手段提取有用信息,提高数据质量及其价值;
- (4) Velocity:处理速度快,对海量数据实现近乎实时的分析处理。

大数据价值链可分为数据生成、数据获取、数据存储、数据分析等四个阶段^[4]。

大数据应用流程可以分为数据采集、数据仓库、数据应用服务和数据可视化四个步骤^[7]。

3 输变电设备状态分析具体电网业务的大数据分析方法

从输变电设备状态分析及应用的具体电网业务角度出发,将大数据分析应用分为数据整合、数据存储、数据计算、数据分析、可视化服务五部分内容,如图 1 所示。

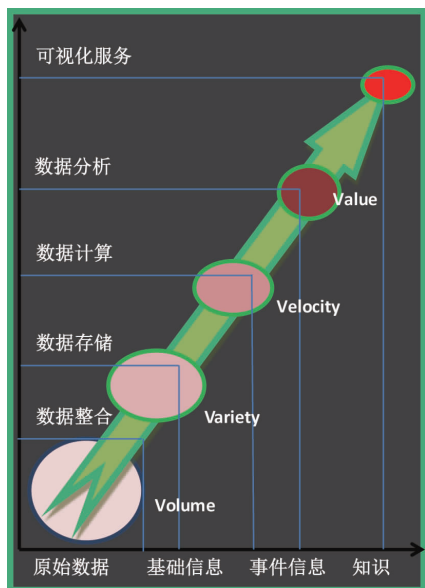


图 1 大数据分析应用示意图

Fig. 1 Application diagram of big data analysis

3.1 数据整合

采用消息队列、数据导入工具、数据抽取工具、数据复制工具等多种技术手段^[19],实现结构化、非结构化、海量历史/准实时、电网空间等数

据接入。

数据源主要包含生产管理系统、状态监测系统、电网 GIS 平台、气象信息系统、雷电监测系统结构化数据和试验报告、现场图片、音/视频文档、设备出厂检验报告、运行报告、故障分析报告、检测报告、日志文件等半结构化及非结构化数据^[20-21]。

3.2 数据存储

对各类接入数据按照统一数据规范进行标准化格式存储,依据应用需求存储在分布式文件系统、分布式数据仓库、非关系型数据库、关系型数据库,实现各类数据的集中存储与统一管理,满足大量、多样化数据的低成本存储需求。

数据存储管理包括基础数据管理、数据质量管理、数据流转监测、数据权限管理和数据运维辅助管理。

基础数据管理:提供对大数据元数据资源的统一管理,保证数据资源的标准化应用,为质量监测、流转监测提供基础数据。

数据质量管理:提供业务明细数据、指标数据等多数据模型的数据质量分析和校核规则,持续提升数据质量。

数据流转监测:实现从数据接入环节、抽取环节、业务处理环节对过程处理进行逐级监测、层层控制。

数据权限管理:提供访问各类数据的授权功能,保证数据的安全及可靠性。

数据运维辅助管理:提供数据日常管理和监控等辅助管理。

3.3 数据计算

提供流计算、内存计算、批量计算、查询计算等分布式计算技术,满足不同时效性的计算需求。

其中,流计算支持实时处理;内存计算支持交互性分析;批量计算支持大批量数据的离线分析;查询计算支持类似 sql 的查询分析技术。

3.4 数据分析

融合传统数据挖掘算法,优化分布式挖掘算法,形成数据挖掘基础算法库。在此基础上,构建支撑输变电设备状态分析业务需求的数据分析算法模型。

对于建模算法选择,主要依赖于所要解决的问题,即输变电设备状态分析的具体业务需求。对于

预测类的问题,有回归和分类预测两种,算法可以选择决策树、逻辑回归、神经网络、机器深度学习应用等;对于描述类的问题,可以选择聚类分析、关联分析、最优化分析等。最后根据所选择的建模算法,通过分析建模工具建立模型。

例如,通过摄像头获取设备检测图像,利用图像/视频分析技术和流计算技术实时分析发现输变电设备故障;利用数据关联分析和主成分分析以及聚类分析技术,关联设备状态信息、电网运行信息、设备故障知识库信息、地理环境信息、气象实况及预报信息等信息,建立设备故障预判模型,实现设备潜伏性故障预判;利用聚类分析和关联分析技术,建立生产管理系统中的设备缺陷信息与物资供应商采购记录、设备使用环境信息、设备监测信息的关联关系,实现对供应商产品质量问题的跟踪和闭环管控。

3.5 可视化服务

基于可视化组件库和设计器,实现可视化配置业务分析场景,提供自助式“DIY”的可视化数据展现。

4 案例

以输变电设备缺陷数据的大数据挖掘词云分析展示为例。

(1)根据现有的电力行业词汇以及从缺陷数据和全过程技术监督的问题数据中寻找行业专用词,并添加到词典中。词典共计 20 084 个中文词及其词性。

(2)利用基于 TF-IDF 和主成分分析法,实现关键词提取,衡量分词的重要性;

(3)应用 R 语言文本分词 jiebaR 算法,构建自适应性最强的混合模型,并通过函数 worker() 建立分词引擎,对缺陷文本和问题文本进行分词,如图 2 所示;

(4)对关键词提取后的词构建词频,对其进行矩阵化,并进行稀疏矩阵的处理,为后续的算法应用作基础;

(5)矩阵化后,运用 k-means 聚类算法,将设备缺陷内容分成冷却器故障、风扇异常、渗油缺陷、呼吸器缺陷、电源故障等,并根据类别特征贴标签;

(6)对每一类别的缺陷进行网省、厂家的统计

分析,并实现可视化展示,如图 3、图 4 所示。

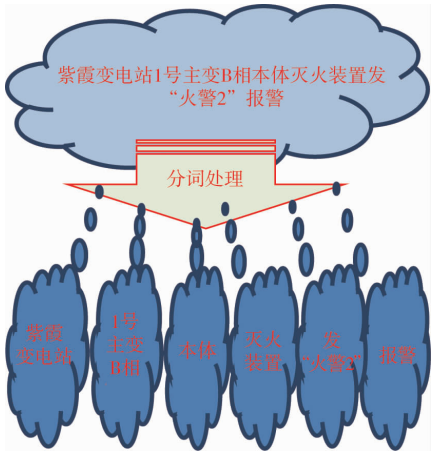


图 2 输变电设备缺陷信息分词示意图
Fig. 2 Defect information of power transmission and transformation equipment

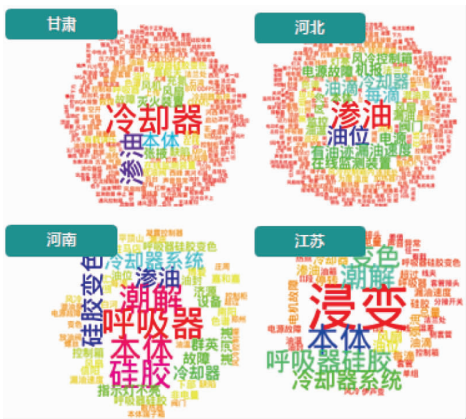


图 3 按网省展示词云示意图
Fig. 3 According to the province display words cloud diagram



图 4 按厂家展示词云示意图
Fig. 4 According to the manufacturer display words cloud diagram

5 总结和展望

文章探索了从输变电设备状态分析及应用的具体业务角度出发,开展电网大数据分析研究方法,并实现了基于输变电设备缺陷数据的大数据挖掘词云分析方法及结果展示。

下一步将综合输变电设备的台账信息、状态监测信息、负荷信息、地理气象信息、历史故障信息和缺陷状况等信息,通过应用朴素贝叶斯算法和 Apriori 关联算法,建立多元线性回归模型,开展输变电设备的潜伏性故障关联预测研究,为开展基于大数据分析的输变电设备状态分析研究进行技术储备。

参考文献:

- [1] 维克托·迈尔·舍恩伯格,肯尼斯·库克耶. 大数据时代[M]. 杭州:浙江人民出版社.
- [2] 李国杰,程学旗. 大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊,2012,27(6):647-657.
LI Guojie, CHENG Xueqi. Research status and scientific hinking of big data[J]. Bulletin of the Chinese Academy of Science, 2012, 27(6):647-657.
- [3] 高志鹏,牛琨,刘杰. 面向大数据的分析技术[J]. 北京邮电大学学报. 2015,03.
GAO Zhipeng, NIU Kun, LIU Jie. Analytics towards big data[J]. Journal of Beijing University of Posts and Telecommunications, 2015, 03.
- [4] 李学龙,龚海刚. 大数据系统综述[J]. 中国科学:技术科学, 2015,45(1):1-44.
LI Xuelong, GONG Haigang. A survey on big data systems[J]. Science China Information Sciences, 2015, 45(1):1-44.
- [5] 方巍,郑玉,徐江. 大数据:概念、技术及应用研究综述[J]. 南京信息工程大学学报:自然科学版,2014,(5):405-419.
FANG Wei, ZHEGN Yu, XU Jiang. Big data: conceptions, key technologies and application[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2014, (5):405-419.
- [6] 宋亚奇,周国亮,朱永利. 智能电网大数据处理技术现状与挑战[J]. 电网技术,2013,37(4):927-935.
SONG Yaqi, ZHOU Guoliang, ZHU Yongli. Present status and challenges of big data processing in smart grid[J]. Power System Technology, 2013, 37(4):927-935.
- [7] 金洁. 基于大数据架构的智能变电站辅助系统规划研究[D]. 华北电力大学,2015.
- [8] 史梦洁,韩笑,程志艳,等. 面向电力需求侧的大数据应用研究分析[J]. 供用电,2014,31(12):20-23.
SHI Mengjie, HAN Xiao, CHENG Zhiyan, et al. Research and analysis on big date of power demand side[J]. Distribution & utilization, 2014, 31(12):20-23.
- [9] Mennis J, Guo D. Spatial data mining and geographic knowledge discovery—An introduction [J]. Computers, Environment and Urban Systems, 2009, 33(6):403-408.
- [10] 刘科研,盛万兴,张东霞,等. 智能配电网大数据应用需求和场景分析研究[J]. 中国电机工程学报,2015,35(2):287-293.
LIU Keyan, SHENG Wanxing, ZHANG Dongxia, et al. Big data application requirements and scenario analysis in smart distribution network[J]. Proceedings of the CSEE, 2015, 35(2):287-293.
- [11] 闫湖,狄方春,袁荣昌,等. 电网智能调度中的大数据及应用场景研究[J]. 电力信息化,2014,10:7-12.
YAN Hu, DI Fangchun, YUAN Rongchang, et al. Research of big data and its application scenarios in grid intelligent dispatching field[J]. Electric Power Information Technology, 2014, 10:7-12.
- [12] 吴军英,刘明硕,常永娟,等. 基于大数据的变压器设备状态及风险分析研究与应用[J]. 电力大数据,2017,20(10):22-30.
WU Junying, LIU Mingsuo, CHANG Yongjuan, et al. Application and research of transformer equipment status and risk analysis based on big data[J]. Power Systems and Big Data, 2017, 20(10):22-30.
- [13] 陈淘,刘利兵. 大数据技术在智能电网中应用[J]. 物联网技术,2016,4:54-57.
CHEN Tao, LIU Libin. Application of big data technology in smart grid[J]. Internet of Things Technologies, 2016, 4:54-57.
- [14] 张东霞,苗新,刘丽萍,等. 智能电网大数据技术发展研究[J]. 中国电机工程学报,2015,35(1):2-11.
ZHANG Dongxia, MIAO Xin, LIU Liping, et al. Research on development strategy for smart grid big data[J]. Proceeding of the CSEE, 2015, 35(1):2-11.
- [15] Xiangyang Wang, Jingwei Chen, Hongying Yang. A new integrated SVM classifiers for relevance feedback content-based image retrieval using EM parameter estimation [J]. Applied Soft Computing, 2010, 11(2):2787-2807.
- [16] Hui Li, Changjiang Li, Xianjun Wu, et al. Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine [J]. Applied Soft Computing, 2014, (19):57-67.
- [17] Xiaobin Tang, Gang Fang. An Algorithm of Mining Spatial Topology Association Rules Based on Complement Set [J]. Energy Procedia, 2011, (11):571-577.
- [18] 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, 25(3):503-511.
- [19] 蔡国保. 电力生产中大数据应用的探讨[J]. 物联网技术, 2015,5(9):55-56.
CAI Guobao. Discussion on big data in power generate [J]. Internet of Things Technologies, 2015, 5(9):55-56.

[20] 秦廷翔. 输配电线路的检修工作探讨[J]. 贵州电力技术, 2016,19(12):54-55.
QIN Tingxiang. The discussion of transmission and distribution lines overhaul work [J]. Guizhou Electric Power Technology, 2016,19(12):54-55.

[21] 姜涛,杨军,荀华. 电力设备状态智能预警系统的设计与应用[J]. 内蒙古电力技术,2017,35(06):45-49.
JIANG Tao, YANG Jun, XUN Hua. Research and application of intelligent alarm system for power equipment status [J]. Inner

Mongolia Electric Power,2017,35(06):45-49.

收稿日期:2017-12-08
作者简介:



王珣(1978),女,博士,高级工程师,主要从事智能电网输电线路及外绝缘领域相关研究。

(本文责任编辑:王 燕)

Discussion on big data analysis and application of power transmission and transformation equipment

WANG Xun¹, WANG Xin², ZHAO Meng¹, XU Peng¹, MA Lin³

- (1. State Grid Jibei Electric Power Co., Ltd., Beijing 100053, China;
2. Research Institute of State Grid Jibei Electric Power Co., Ltd., Beijing 100045, China;
3. State Grid Beijing Fuda Science & Technology Development Co., Ltd., Beijing 100070, China)

Abstract: With the rapid increase of the power grid data, the traditional data mining and analysis methods can not meet the requirement of smart grid, and the big data analysis provides corresponding means to achieve it. The big data technology theory was introduced briefly first, and from the analysis and application of specific power grid business of power transmission and transformation equipment status, the five parts of data integration, data storage, data calculation, data analysis and visualization of results are carried out. Then according to the historical defect data of transmission equipment and problem data of whole technical supervision process, using principal component analysis and clustering algorithm, the defect characteristics analysis model of transformer equipment was built, and realized the equipment defect classification labeling, which providing suggestions to operation and maintenance personnel based on data analysis. Finally the prospects for further work of latent fault prediction of power transmission equipment was discussed.

Key words: big data; power grid business; word cloud; latent fault