

DOI:10.3969/j.issn.1001-4551.2017.06.021

输变电设备多源数据处理系统的功能设计^{*}

龚 霄¹, 邓超志¹, 朱怀金¹, 夏伊乔², 郭创新^{2*}

(1. 贵阳供电局, 贵州 贵阳 550002; 2. 浙江大学 电气工程学院, 浙江 杭州 310027)

摘要:针对输变电设备状态评估数据来源多样、平台繁多以及数据质量参差不齐的问题,通过对输变电设备状态评估数据特点进行研究,提出了一种输变电设备多源数据处理系统的功能设计。设计了可提高设备状态评价、负载能力评估时多源数据质量的包含数据获取、跨平台接入、数据预处理及初步清洗、质量评价模型、规范化转换、数据输出等环节在内的系统功能。将输变电设备多源数据处理系统应用于大数据平台,研究了异构大数据的预处理、分布式存储和高效检索技术;分析了面向输变电大数据分析的核心算法及并行化技术。研究结果表明,运用输变电设备多源数据处理系统搭建的大数据全耦合分析模型框架可应用于电网,为生产运维人员评估输变电设备及电网风险提供数据支持。

关键词:输变电设备;多源数据;处理系统;大数据;全耦合分析

中图分类号:TM732

文献标志码:A

文章编号:1001-4551(2017)06-0659-06

Multi-source data processing system of electric transmission and transformation equipment

GONG Xiao¹, DENG Chao-zhi¹, ZHU Huai-jin¹, XIA Yi-qiao², GUO Chuang-xin²

(1. Guiyang Electric Power Supply Bureau, Guiyang 550002, China;

2. School of Electrical Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract: Aiming at the problems of the data sources and data platform diversity, a multi-source data processing system was proposed. Based on the analysis of electric transmission and transformation equipment condition assessment data, the function including data preliminary processing, cross platform access, preliminary cleaning, quality evaluation model, specification conversion and data output function were designed. In addition, the multi-source data processing system of electric transmission and transformation equipment was applied to big data analysis platform, distribute heterogeneous storage of big data and efficient retrieval technology were researched, and the parallel transmission technology for analysis of big data was focused on. The results indicate that the model framework of big data full coupling analysis can provide data support for the production of power transmission equipment and maintenance personnel.

Key words: electric transmission and transformation equipment; multi-source data; processing system; big data; full coupling analysis

0 引 言

电力系统是一个多维度、动态、高阶非线性的复杂

系统^[1-2],由发电、变电、输电、配电、用电及调度等若干环节构成^[3]。随着电力系统的数字化、信息化、智能化不断发展^[4],电力系统的数​​据呈现爆炸式增长,主要包含设备当前状态数据、系统运行工况数据、设备所

收稿日期:2016-12-03

基金项目:国家高技术研究发展计划(“863 计划”)资助项目(2015AA050204);浙江省自然科学基金资助项目(LZ14E070001)

作者简介:龚霄(1985-),男,贵州贵阳人,主要从事地区电网电力调度管理方面的研究。E-mail: 231610185@qq.com

通信联系人:郭创新,男,教授,博士生导师。E-mail: guochuangxin@zju.edu.cn

处环境信息、历史信息数据等。这些信息一方面呈现出多源异构的特点;另一方面,随着监测点的增多及采样周期的缩短,这些数据的体量呈现高速的增长;同时,输变电设备状态评估数据的信息挖掘也十分欠缺。

近年来,大数据已应用于众多领域^[5-6]。2011 年麦肯锡发布关于大数据的报告^[7]。2013 年《中国电力大数据发展白皮书》发布^[8]。

当前,输变电设备多源状态评估工作面临以下问题与困难:①原有各业务信息系统存有较大差异^[9],输变电设备多源状态评估数据的集成与共享严重不足,传统关系型数据库无法满足海量、多源、异构数据的管理需求;②输变电设备多源数据利用率低,真实价值未能得到充分地挖掘与利用。因此,如何利用新的技术和方法来挖掘海量数据的潜在价值,并指导主网的生产运行具有重要意义。

为此,本研究将在分析输变电设备多源状态评估数据特点的基础上,提出一种输变电设备多源状态评估数据处理系统的功能设计,并在大数据平台上应用。

1 输变电设备状态评估数据特点分析

按照数据的专业属性,输变电设备多源状态评估数据可分为:电网运行数据、故障和保护动作数据、运行分析数据、设备状态监测数据、电能质量数据、视频数据、环境监测数据、作业管理数据 8 大类。具体的数据内容、数据源及数据结构列如表 1 所示。

表 1 输变电设备多源状态评估数据

数据	数据源	数据格式
设备台账、运行日志、运行工况、操作记录、缺陷记录、事故记录、检修记录、测试记录、试验记录。	生产管理系统	结构化
跳闸记录、隐患记录、交叉跨越、监测记录、作业表单、缺陷记录。	输电精益作业管理系统	结构化
巡视记录、避雷器泄露电流及动作次数记录、工器具定检记录;断路器动作次数记录;变压器铁芯及夹件泄漏电流测量记录;SF6 气体压力抄录记录。	变电现场作业数据采集系统	结构化
SCADA 量测数据;在线监测量测数据;保护装置动作记录及录波记录;雷电定位系统数据;覆冰监测数据;微气象监测数据。	运维自动化平台	结构化、非结构化
文档、报表、实验报告	人工管理	非结构化

表 1 所归纳的输变电设备多源状态评估数据具有以下特点:

(1)多源性与复杂性。出自不同的子系统是数据的多源性的重要体现。如事故记录来自生产管理系统;SCADA 量测数据来自运维自动化平台;作业表单则来自输电精益作业管理系统等。另一方面,数据对实时性要求也各不相同,如 SCADA 量测对调度操作影响重大,要求数据实时、准确;而如巡视记录、检修记录等数据对实时性的要求低。因此,对不同数据采取规范化采集管理的工作迫在眉睫。

(2)异构性。输变电设备多源状态评估数据还包含系统日志、表计等半结构化数据,以及视频监控、音频等非结构化数据^[10]。目前上述提到的混合数据类型的半结构化、非结构化数据,大多采用文件直接存储,无法被检索、挖掘、分析并保存在本地系统中。

(3)规模性。省级电网的输变电设备多源状态评估数据每月约有 100 GB 结构化、300 GB 非结构化建设管理数据;每年约 7 亿条/6 T 电网电压、电流等电网运行数据^[11];输变电设备状态信息量大,其中视频数据更达到 PB 级;另外,电网逐年新增海量采集数据、音频数据,电网运行监控将具体信息接入后,省级电网结构化、非结构化和多时空信息达到 TB 级。

综上,输变电设备多源状态评估数据体量为 TB 或 PB 级,关系数据库已缺乏处理能力^[12];同时,当面临视频图像、图片、文档等非结构化数据时,现有关系数据库也无法处理。

2 数据处理系统的功能设计

针对输变电设备状态评估数据多源、量大繁复等问题,笔者设计了输变电设备状态评估数据处理系统。具体的功能设计如下。首先跨平台接入数据并进行含数据标记和数据传输策略的数据预处理;接着对数据进行数据检验、噪声平滑、离群点识别、缺失值填充等步骤在内的数据清洗;并制定质量评价模型分析数据质量;然后对数据格式进行规则规范,完成规范化转制;最后输出数据。具体功能设计流程如图 1 所示。

2.1 数据获取

输变电设备的多源数据都可为状态评估服务。状态评估数据主要来源于信息内网和信息外网相关的业务应用系统,包括了设备当前状态数据、系统运行工况

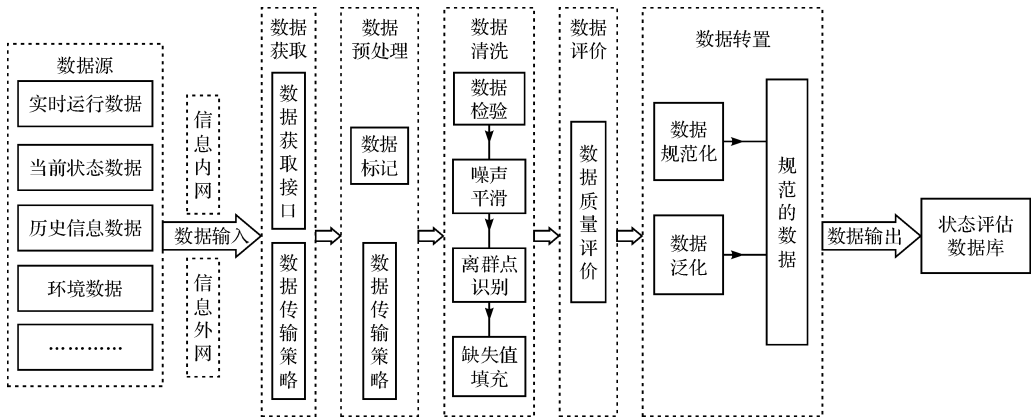


图1 输变电设备多源数据处理系统的功能设计流程

数据、设备所处环境信息、历史信息数据等几部分。其中设备当前状态含产气速率、热点温度、老化情况和风速;系统运行工况数据包含潮流、电压、频率及温度;设备所处环境信息含覆冰、山火、雷电及微气象;历史信息则由事故记录、状态参数、工况记录、环境参数等组成。

数据的获取分信息内、外网两部分^[13],信息内网的数据直接获取并转置,经安全隔离装置可获取、处理信息外网业务系统的数据。数据获取的总体步骤是在企业服务总线的基础上,通过数据接口、数据中心共享、网络隔离下的安全文件传输等方式,配置相关策略,定义相关接口、周期、调用对象和频率等参数,从业务系统中自动选取数据,完成跨平台数据库访问、跨平台大数据库文件高速并发读取、跨平台数据安全传输和同步等关键技术。

2.2 数据跨平台接入

多源异构数据跨平台接入依托输变电多源数据处理系统实现。跨平台输变电多源数据处理系统是集成了相关软件功能的一体化硬件装置,该装置的输入是各业务系统中的实时、历史和环境数据。它以单台或集群的方式部署在信息内网,对于同处信息内网的业务系统(如设备实时运行数据、当前状态数据、历史信息数据等),直接获取由业务系统推送的数据;对于部署在信息外网的业务系统和数据(如设备所处环境信息等),将在隔离系统的保护下进行数据的传输。

输变电多源数据处理系统的主要功能是获取业务数据,经过数据获取、数据初步清洗和数据转换等处理环节,为输变电设备状态评估系统的分析和展示提供数据基础。

2.3 数据预处理及初步清洗

输变电多源数据处理系统经过数据预处理,供设

备评估系统使用,提供日志服务。由于输变电设备状态评估数据来源多样、平台繁多及数据质量参差不齐,不能确保数据的有效性、完整性、一致性等指标,针对获取的数据,装置将对数据进行初步清洗。

基本流程如图 2 所示。

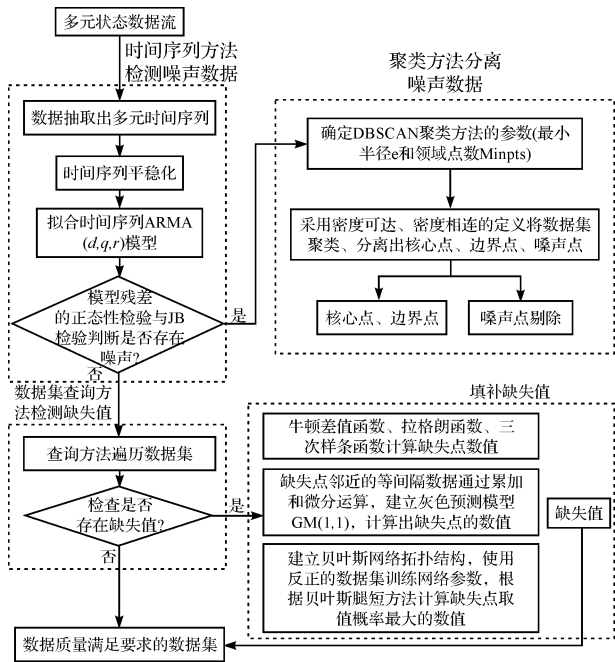


图2 数据初步清洗流程图

2.4 数据质量评价模型

本研究针对数据的数据类型特征以及业务特征总结数据质量评价指标体系,从面向输变电设备状态评估数据分析的应用场景出发,针对数据质量评价指标,结合典型数据质量,构建适合设备状态评估的数据质量评价模型,主要包括评价指标、依赖关系、算法库及关联关系等,初步形成面向输变电设备状态评估数据分析的数据质量评价模型。

数据质量体系和评价模型如图 3 所示。

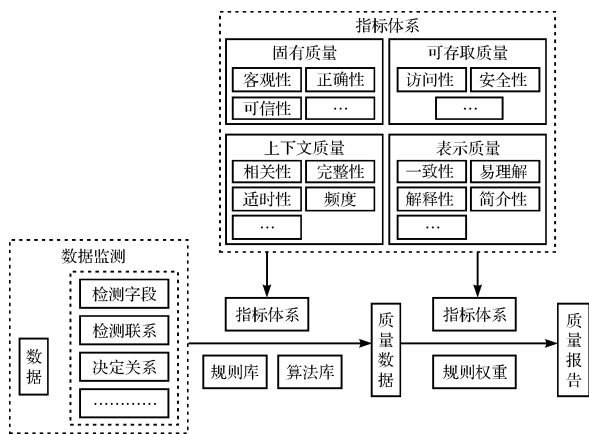


图3 数据质量体系 and 评价模型

2.5 数据规范化转制

数据规范化主要通过数据泛化、规范化、数据属性构造等操作,提高数据或数据流挖掘的性能及精度。数据泛化通过概念分层,将低层次“原始”数据替换为高层次概念,用泛化成用户目标的聚合的、概念层次上的、具有统计意义的元数据替换数据库中的初始数据,使在线分类有优质的数据环境;契合的泛化算法有基于数据立方体的泛化算法及面向属性归纳的泛化算法

等。数据转换是指将数据归并至适合数据挖掘的描述形式,类型有大大表连结、大小表连结、行列转换以及聚集。

输变电多源数据处理系统的规则与规范均采用插件方式,可以灵活部署与扩展。规则与规范可以通过用户定义或业务系统提供,根据不同的应用场景及需求,定义相应的规则与规范。

2.6 数据输出

经清洗、转换好的数据,由装置出口至综合分析系统中,根据数据的不同类型,由通用的数据库访问接口和文件传输接口,写入存储系统,供综合分析系统使用。

3 大数据并行化技术及全耦合分析模型框架

大数据综合分析系统主要作用是提高大数据的存储、检索和预处理服务,支撑数据访问以及各种高性能的数据挖掘分析算法服务。总体架构如图4所示。

下面叙述信息综合分析系统构建的关键技术解决方案。

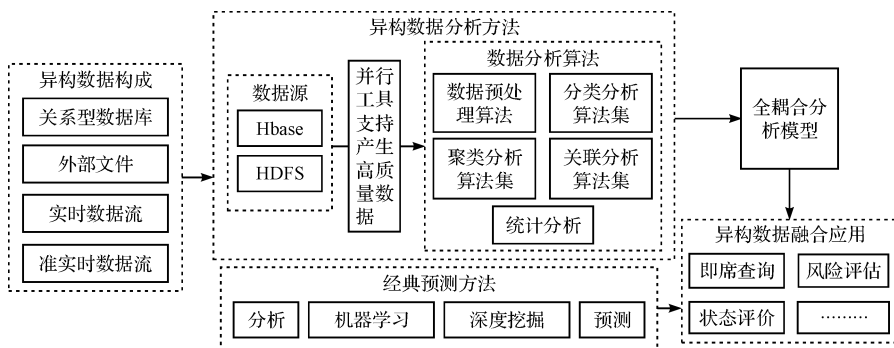


图4 大数据综合分析系统

3.1 异构大数据的预处理、分布式存储和高效检索关键技术

综合分析系统的异构大数据存储采用 hadoop 分布式文件存储,分布式 NoSQL 列数据库、可扩展数据仓库等技术,主要包括分布式文件系统、分布式数据库、分布式协同工作系统、分布式数据仓库、非结构化数据预处理和多级综合索引等功能模块。

分布式文件系统使用 HBase 和 HDFS 文件系统^[14-15]。对于海量的非结构化小文件以及负载多变的结构化数据,使用 HBase 和 key-value 存储^[16];对于较大的单个文件,直接存储在 HDFS 文件系统中。

分布式数据库通过 NoSQL 分布式数据库构建。基于加速组件的主从复制技术,保证多个数据库之间

的数据一致性,实现灾备功能,分担读数据时的压力。分布式协同工作系统基于 YARN 资源管理器,确保分布式系统一致性。分布式数据仓库基于 Hive 数据仓库,根据一致的存储规则将多源数据进行有机结合。

非结构化数据预处理采用非监督式的深度学习方法,针对电力系统中的非结构化数据,如来源于变电站巡检机器人、带电监测、变电站监控系统、输电线路无人机等海量图像数据,实现非人工监督的特征提取和分类。主要方法是构建基于多隐层的机器学习模型和海量训练数据,主动挖掘特征信息。具体实现过程如下:原始图像信息通过卷积神经网络和递归神经网络进行处理,将获得的特征向量输入 SoftMax 分类器进行自主训练和特征学习,实现非结构化数据特征的结

构化转变。

多级综合索引可以提高相似特征数据的检索功能。其中一级索引采用多维 R 树基本结构,依据非传统地理分类规则,将清理后数据对象分成多个字空间每个字空间对应 R 树的一个节点,非叶节点存放叶节点所有子树的最小外接矩形(MBR),叶节点中存放每个空间对象对应的 MBR。二级以下索引采用基于相邻密度(LCF)的对象聚类,根据相对密度的接近程度,实现聚类后的对象划分,克服以简单全局距离标准作为划分依据带来的问题,提升聚类的可调整性、可扩展性和时效性。

3.2 面向输变电设备大数据分析的核心算法及并行化技术

面向输变电设备大数据分析的核心算法包括分类算法、回归算法、聚类算法、关联算法、聚席查询等算法^[17],但已有的实现方法不能满足海量数据的分析需求,继而需要研究大数据应用场景下的算法实现及并行化技术。面向输变电设备大数据分析的核心算法及并行化技术研究步骤如下:

(1)研究核心分析算法的基本流程。找出其中能并行处理的步骤,考虑所处理的数据规模、类型以及用途等因素,提出综合分析系统并行技术框架的选择策略,并进行实现。

(2)针对易于并行化的经典串行算法。结合应用场景选择合适的并行框架;对于难以实现并行化的串行算法,根据分析要求,结合相关应用实际,从并行化角度重新设计并行框架,达到扩大数据规模,提高数据分析效率的目的。

(3)优化所进行的并行算法。使用工具对并行化分析算法的运行过程进行监控,记录计算步骤之间、工作节点之间的资源消耗和数据规模,针对测试过程中的不足,设计优化策略,进而提升算法性能。

3.3 大数据全耦合分析模型框架

全耦合分析模型是在分布式存储和基本核心算法实现基础上,建立的内在关联的设备本征数据分析模型框架,主要包括:特征关联分析、状态特征聚类、状态分类、故障预测和差异化分析等中间层分析。

全耦合分析模型框架的关键技术包括存储模型、分析模型和计算模型 3 种。

存储模型通过基于关系数据库和分布式文件系统的数据仓库,提供了结构化数据与非结构化数据的关联和分领域独立数据的统一视图。

分析模型基于迁移学习的规则耦合,混合计算模式算法实现的封装,模型管理接口和可视化分析接口的开发。针对特征关联分析、状态特征聚类、状态分类、故障预测和差异化分析等不同模型,根据输变电设备状态评估中的要求,灵活选择映射耦合,将一种分析模型作为迁移学习的源领域,另一种分析模型作为目标领域,将源领域中获得的学习信息分析映射到目标领域,在目标领域中以更高的起点获得更加准确的学习分析规则,最终形成闭环形态的全耦合分析模型。

计算模型基于映射-规约的计算模式,对上层分析模型提供统一的计算框架,实现可扩展至 PB 级数据的计算能力。

4 结束语

笔者在研究输变电设备状态评估数据特点的基础上,针对数据多源异构、海量繁杂的问题,开发了输变电设备状态评估数据处理系统,对各环节的功能进行逐项设计。基于大数据平台,研究基于大数据分析的核心算法、并行化技术,搭建大数据全耦合分析模型框架,使输变电设备多源状态评估数据处理系统能在大数据环境下得以应用。

下一步的工作将围绕系统的实现、高级应用功能的算法研发展开,以期为输变电设备的状态评价、负载能力评估提供多角度的数据支持。

参考文献(References):

- [1] 卢 强,梅生伟. 面向 21 世纪的电力系统重大基础研究[J]. 自然科学进展,2000,10(10):870-876.
- [2] 卢嘉栋. 智能电网中电力设备及其技术发展[J]. 机电工程技术,2016,45(12):62-65.
- [3] 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.

(下转第 673 页)

本文引用格式:

龚 霄,邓超志,朱怀金,等. 输变电设备多源数据处理系统的功能设计[J]. 机电工程,2017,34(6):659-663,673.

GONG Xiao, DENG Chao-zhi, ZHU Huai-jin, et al. Multi-source data processing system of electric transmission and transformation equipment[J]. Journal of Mechanical & Electrical Engineering, 2017,34(6):659-663,673.

《机电工程》杂志: <http://www.meem.com.cn>