

超/特高压 GIS 设备运维电力大数据数据库的构建研究

张刘杰¹, 刘锦蕙¹, 贾 罡¹, 张 利¹, 张文博¹, 王榕泰²

(1. 国网河南省电力公司超高压公司, 河南 郑州 450007; 2. 武汉理工大学, 湖北 武汉 430070)

摘 要: 气体绝缘组合电器(gas insulated switchgears, GIS)在运行过程会产生大量数据, 且其规模庞大、种类复杂、结构不统一, 呈现多源、异构的特征, 导致数据难以统一存储与利用。文中通过收集并处理大量 GIS 设备数据, 研究 GIS 设备多源异构数据清洗方法, 并开发基于 MySQL 的 GIS 数字孪生电力大数据数据库。

关键词: GIS 设备; 多源异构数据; 数据清洗; 电力大数据库

中图分类号: TM 73 **文献标志码:** A **文章编号:** 1006-348X(2025)04-0029-05

0 引言

GIS 在运行过程产生大量数据, 源头包含设备传感器实时采集数据、仿真数据、外部导入数据等等^[1]。如图 1 所示, 淡绿色代表静态数据, 橙黄色代表动态数据。静态数据包含了电力设备的基本信息、几何结构、材料属性、仿真数据等, 此类数据较为稳定, 随时间的变化离散进行迭代; 动态数据主要是设备运行过程中各类传感器记录的大量实测信息, 此类数据随时间的变化持续迭代, 是实现虚实融合、构建镜像孪生体的关键。

GIS 设备数据有如下特点:

1) 数据规模庞大。由于安装了诸如局放传感器、气压传感器、温度传感器、振动传感器等, 瞬时数据量大。建立数据库需要长年累月的数据积累, 整体数据规模十分庞大。

2) 数据种类复杂。由于需要监测的运行数据种类繁多, 对应传感器监测得到的数据种类也十分复杂。

3) 数据结构不统一。GIS 设备数据采集装置收集各类运行数据, 数据的格式及采集间隔各不相同, 导致数据结构难以统一。

4) 数据采集稳定。由于传感装置对数据进行实时持续的采集, 数据具有随时间变化的特点, 数据主要为插入添加, 基本不涉及删除操作。

构建 GIS 设备电力大数据数据库的数据来源多样, 如何处理多源异构的动静态数据是一大难题。文中通过收集并处理大量 GIS 设备数据, 研究 GIS 设备多源异构数据清洗方法, 并开发出基于 MySQL 的 GIS 数字孪生电力大数据数据库。

1 GIS 数据清洗

GIS 多源异构数据由结构、形式及存储介质不同的多个数据源得来, 一般通过异构分布式数据库存储。融合多源异构数据是建立 GIS 设备电力大数据数据库的必要条件^[2]。

GIS 多源异构数据清洗过程如图 2 所示, 具体流程如下:

1) GIS 设备数据预处理

受工作环境、安装位置影响, GIS 设备数据传感

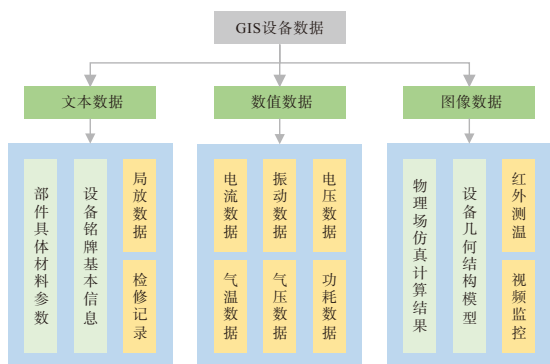


图 1 GIS 设备数据

收稿日期: 2025-02-18

基金项目: 国网河南省电力公司 2022 年科技项目(5217M021N003)。

作者简介: 张刘杰(1995), 男, 硕士, 工程师, 研究方向为电力系统分析。

器进行数据采集时存在噪音污染,并存在数据缺失。通过对初始传感器采集数据进行筛选,剔除存在明显瑕疵的问题数据,可提升数据质量^[3]。

2) GIS 设备数据特征选择

在经过预处理的GIS设备数据中选择特征字段,将其转化为具有特定格式的特征子集。通过选取更优特征子集,实现数据清洗过程中的泛化能力改善,并降低数据过拟合风险^[4]。

3) GIS 设备数据异常检测

数据经过预处理后仍有部分处于离群点、噪声点、偏差点,此时需要对数据进行异常检测,将异常数据进行筛查并单独处理。

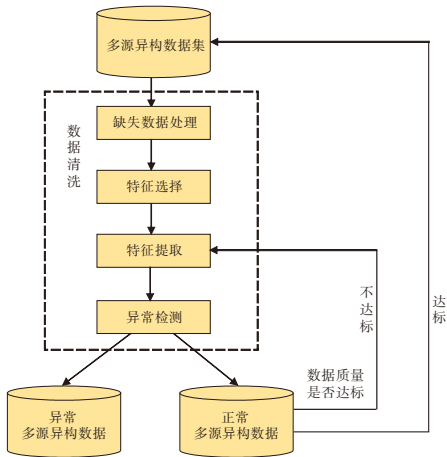


图2 多源异构数据清洗过程

4) GIS 设备数据存储

当GIS设备数据质量达标时,进行数据回流至数据仓进行存贮;如数据质量未达标,则重新从特征选择步骤进行数据清洗,直至数据质量达标。一旦多源异构数据在数据清洗过程中消耗殆尽,则结束存储。

文中通过对动态数据相邻采样记录点时间差和采样间隔来检测是否存在重合和缺失数据。设采样间隔为 λ ,相邻记录点时间差为 Δt ,则:

- a) $\lambda = \Delta t$, 相邻记录点正常采样,不存在缺失和重复数据;
- b) $\lambda > \Delta t$, 相邻记录点采样异常,存在重复数据;
- c) $\lambda < \Delta t$, 相邻记录点采样异常,存在缺失数据,缺失数据个数 $n = \frac{\Delta t}{\lambda - 1}$ 。

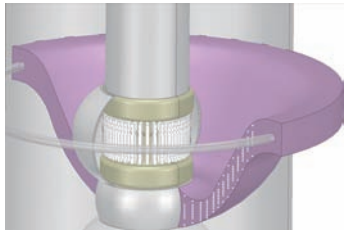
缺失数据的修复存在均值、中位数、众数、固定值填充,前(后)向填充,插值法等。在实际使用中,需要根据数据的种类及特性进行选择。

2 GIS局部放电孪生数据获取

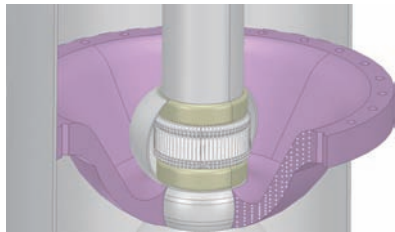
数据根据迭代方式可分为静态数据和动态数据,根据数据获取方式可分为仿真数据和实测数据。GIS局部放电孪生数据属于仿真数据,由孪生计算得来,文中通过孪生仿真计算得出并保存至孪生数据库中。

2.1 孪生仿真模型设置

文中在1100 kV GIS盆式绝缘子上选取了117个可能存在气泡缺陷的位置^[5],在这117个气泡缺陷点位分别进行了4组孪生计算,分别为无缺陷状态下的电场分布计算组、直径1 mm气泡缺陷计算组、直径2 mm气泡缺陷计算组、直径3 mm气泡缺陷计算组,共计468个电场孪生数据;同时,对550 kV GIS盆式绝缘子进行了气泡缺陷的孪生计算,选取了91个可能存在气泡缺陷的位置,同1100 kV盆式绝缘子一同进行了无缺陷、1 mm气泡缺陷、2 mm气泡缺陷、3 mm气泡缺陷的电场计算,共计364个孪生电场数据。如图3所示。



(a) 1100 kV GIS 盆式绝缘子



(b) 550 kV GIS 盆式绝缘子

图3 盆式绝缘子气泡空间位置分布

2.2 GIS孪生数据结果

将1100 kV盆式绝缘子最靠近凸面的13个气泡位置划分为第1组,其上的13个气泡位置作为第2组,以此类推盆式绝缘子中最靠近凹面的13个气泡为第9组。从左至右各个气泡所在位置分别称为位置1、位置2,直至位置13。表1为无缺陷条件下组1至组5的仿真结果,表中数值为缺陷点最大电场值,单位为 10^6 V/m。为提高1100 kV盆式绝缘子气泡缺陷数据表格展示的直观性,文中对仿真结果进行二维图像化展示,如图4所示。

表1 盆式绝缘子无缺陷条件下部分仿真结果

	组1	组2	组3	组4	组5
位1	2.1858	2.6280	2.8385	2.7339	2.9381
位2	2.4304	2.6328	2.8002	2.7169	2.8838
位3	2.6532	2.7281	2.8910	2.9206	2.9932
位4	2.8028	2.8456	3.0024	2.9394	3.0932
位5	2.8268	2.8989	2.9767	2.6848	3.0058
位6	2.6925	2.6939	2.6669	2.3877	2.6386
位7	2.4806	2.4134	2.3604	2.1788	2.3319
位8	2.2654	2.1962	2.1617	2.0538	2.1452
位9	2.0976	2.0602	2.0483	2.0301	2.0435
位10	1.9867	2.0202	2.0348	2.0667	2.0346
位11	2.0979	2.0956	2.0319	1.8946	1.9926
位12	2.2456	1.8267	1.8348	1.8946	1.7739
位13	2.1405	1.8267	1.6982	1.7622	1.6345

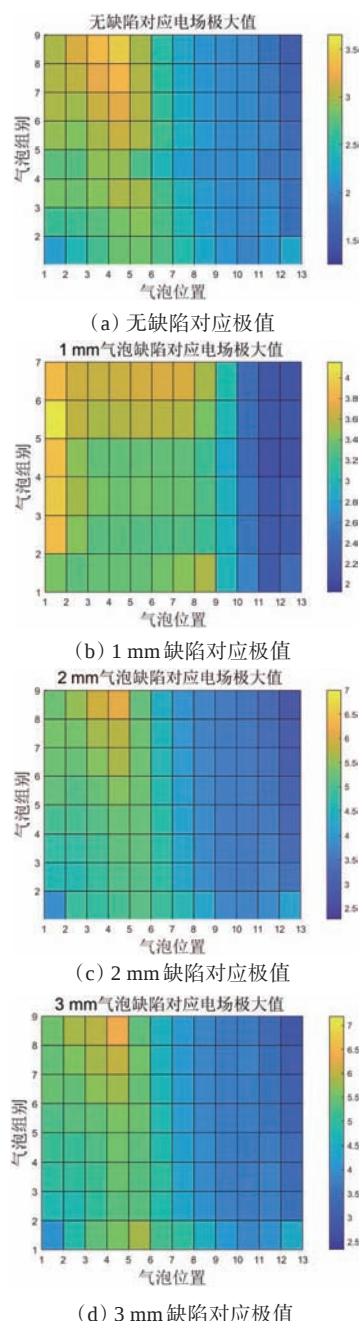


图4 盆式绝缘子气泡缺陷仿真结果图

3 GIS电力大数据构建

GIS电力大数据存储了大量动、静态数据,主要对电力设备的基本信息、几何结构、材料属性、仿真数据以及传感装置采集的实时设备运行参量等进行存储。

3.1 电力大数据库开发

文中基于MySQL对GIS数字孪生电力大数据库进行开发,实现了对GIS设备动、静态数据的快速采集、清洗、融合、存储及调用。

除此以外,GIS电力大数据库还有多种容易被直接存储记录的数据文件,包含运维职工信息、部门信息、设备生产商信息、设备台账信息及检修记录等。通过连接Unity 3D与MySQL数据库,能够实现在三维模型上对数据的立体全方位展示,实现信息的动态读取。通过在Unity 3D项目里创建脚本,应用MySQL Connector/Net能够实现与数据库的连接。

3.2 电力大数据库方案

MVC三层架构的数据库层级清晰,方便系统解耦与扩展^[6],GIS数字孪生电力大数据库采用此种架构进行构建。如图5所示,MVC结构可以分为数据接入层、数据处理层、数据应用层。

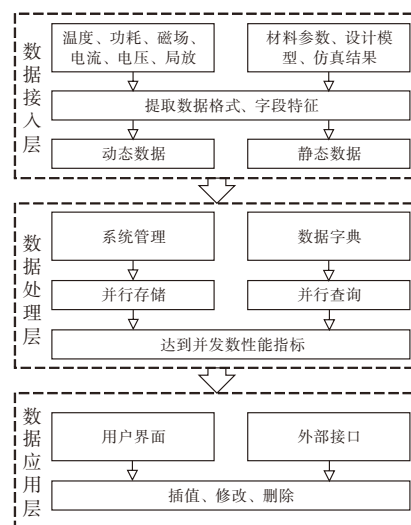


图5 GIS数字孪生电力大数据库架构

在数据库架构的选取方面,选用Browser/Server即浏览器/服务器架构。它由三部分组成:浏览器端、服务器端、中间连接。其优点如下:泛用性强,能在多种运行环境下使用,对环境依赖较小;客户端可以直

接使用浏览器操作,开发维护方面较为便利;业务扩展方便,能通过增加页面升级服务功能。

文中研究的多源异构数据主要包含动态数据和静态数据, GIS 数字孪生电力大数据库的设计方案如下:

1) 数据接入层。该层是数据库的数据底层,通过 GIS 设备上的信号采集装置采集设备实时电压、电流、温度、气压及局放信号等;静态数据包含设备的设计模型、材料参数、仿真数据等。动、静态数据格式及对应数据库的模块设计如图 6 所示。针对动态数据及静态数据的特点,动态数据多需要实时采集,对数据存储的处理速度有一定需求。因此,对动态数据采用热存储,静态数据采用冷存储,以降低数据存储造成的经济压力。

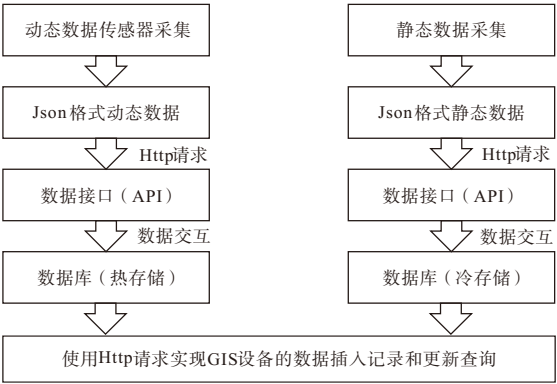


图6 GIS数据库设计方案

采集的动态数据以 JSON 作为储存格式。使用关系数据库 MySQL 对动、静态数据进行接入管理,将 JSON 作为文档序列化的格式,把字段、二进制图像编入索引以完成快速检索。

使用 Java API 的方式实现 GIS 设备与数据库的交互^[7],通过发送一个 HTTP 请求可以插入一个记录,数据库的查询、更新操作也同样可以通过上述请求完成,实现管理人员快速查询、储存多源异构数据的功能。数据将按照表格形式实现分页、分区、分级存储。

2) 数据处理层。该层级负责处理数据,利用多源异构数据的字段特征加速数据的存储,从而提高数据库的存储效率。该层级包含系统管理模块、数据查询模块、并行加载模块,通过多源异构数据库实现了数据库表结构存贮。该存贮方法可以实现数据分页、分块存贮,按照行列索引查询数据内容。

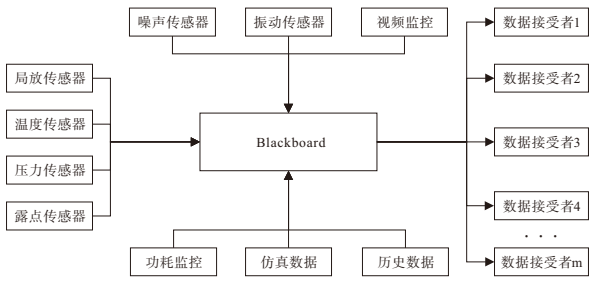


图7 BDP架构

如图 7 所示,数据库采用 BDP (Blackboard Design Pattern) 架构,允许数据的读写同时进行。GIS 设备的数据采集写入与后台人员的信息调用读出能够在时间和空间上完全解耦、互不干扰,满足 GIS 数字孪生电力大数据库的实时性。文中采用的数据库为 Blackboard,此种方案在技术上更容易实现。GIS 数据库在现有的实验条件下,通过 HTTP 协议与设备通讯, GIS 设备的动态数据采集上传至前置机,以转发的形式再由前置机与数据库通讯。

3) 数据应用层。提供方便可靠的应用界面,通过界面实现访问、修改 GIS 设备的多源异构数据,能够利用电力大数据库系统接口完成系统的二次开发^[8-9]。

GIS 数字孪生电力大数据库服务器选择部署在 Linux 系统,其具有网络性能良好、内存管理及调度方法高效、用户管理方便、安全性高、支持多线程访问等优点。GIS 设备数据库管理结构如图 8 所示。

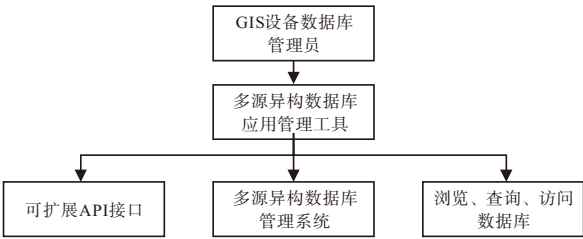


图8 GIS设备数据库管理结构

数据库访问关系图如图 9 所示。数据库前端页面由一个主展示页面和插入页面组成,页面与页面间利用超链接进行跳转。前端界面由首页、温度数据、SF₆ 气压数据、露点数据、功耗数据、噪声数据、振动数据、电流数据、视频监控、仿真数据、历史数据、运维员工数据、部门信息数据、设备厂商数据、设备基本信息数据、设备检修记录数据组成。如图 10 所示,数据库录入 550 kV GIS 盆式绝缘子 3 mm 气泡缺陷仿真数据。

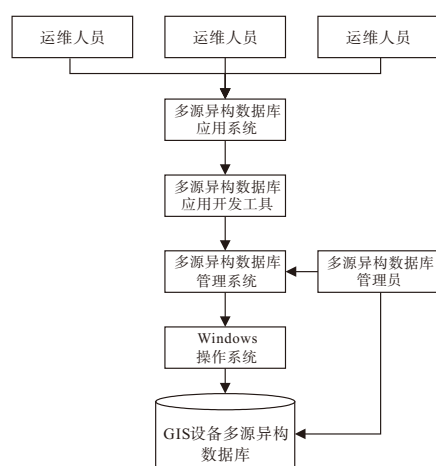


图9 GIS设备数据库访问关系



图10 数据库前端页面展示

3.3 数据库数据安全

GIS 电力大数据库需要考虑数据安全^[10],避免被窃取运行信息,主要包括三个方面。

1) 安全性保障:GIS 电力大数据库的管理操作系统需要保证避免数据库非法入侵访问。

2) 完整性保障:GIS 电力大数据库采取并发控制及封锁技术等方案,保证数据的正确性、完整性,防止错误数据进入数据库,影响数据库的准确性,影响系统安全。

3) 备份与恢复保障:防止硬件故障、软件 Bug、人员操作不当、恶意破坏及病毒造成数据丢失,通过 GIS 电力大数据库的备份与恢复功能将数据进行归档。

4 结语

文中基于 GIS 设备运行数据的特点,收集并处理大量 GIS 设备数据,建立了 GIS 设备电力大数据库,具有以下功能和优势:

1) 通过对 GIS 设备运行数据进行预处理、特征选择、异常检测及数据存储,能够有效地对数据完成清洗。

2) GIS 设备电力大数据库实测监控数据,引入大量局放孪生数据及盆式绝缘子全空间分布缺陷孪生电场数据,有效地丰富了电力大数据库的数据种类多样性及实用性。

3) 使用 MySQL 对 GIS 数字孪生电力大数据库进行开发,实现了对 GIS 设备动、静态数据的快速采集、清洗、融合、存储及调用。

4) 大数据库采用 MVC 架构,其结构较其他结构层次更为清晰,利于系统的解耦合与扩展。

参考文献:

- [1] 蒲天骄,乔骥,韩笑,等.人工智能技术在电力设备运维检修中的应用[J].高电压技术,2020,46(02):369-383.
- [2] 冯薇玺,李清,黄萍.基于数字孪生的多信息源输电网数据融合方法[J].自动化与仪器仪表,2023(09):96-99.
- [3] Y. Liu, T. Dillon, W. Yu, et al. Noise removal in the presence of significant anomalies for industrial iot sensor data in manufacturing[J]. IEEE Internet of Things Journal, 2020, 7 (08): 7084-7096.
- [4] 王罡,刘敬文,李国鹏,等.基于多源异构数据融合的综合管廊电力舱系统保护[J].电力系统保护与控制,2021,49(07):103-109.
- [5] 韩帅,高飞,廖思卓,等.GIS 盆式绝缘子表面缺陷及其诊断方法研究综述[J].绝缘材料,2022,55(02):12-22.
- [6] 赖秋菊,徐政,陈东,等.基于 WEB 服务的高压直流输电基础设计软件的架构[J].中国电力,2018,51(02):75-81.
- [7] 张浚哲.基于强化学习的 JAVA 语言 API 调用序列检索[D]. 哈尔滨:哈尔滨工业大学,2022.
- [8] 李兵,申洪涛,郭荣坤,等.基于数字孪生的计量资产调度立库优化[J].中南民族大学学报(自然科学版),2022,41 (06):720-727.
- [9] Yu. P. Gusev, A. V. Trofimov, V. A. Trofimov. Design Database of Cad System as the Basis of the Digital Twin of the Automation System of Power Plants and Substations[J]. Power Technology and Engineering, 2020, 54(03): 418-423.
- [10] 赵喆,李尚泽,王利军,等.多源数据融合下的电力系统数据传输数字化风险检测技术[J].自动化与仪器仪表,2024 (06):203-207,214.