DOI:10.20033/j.1003-7241.(2025)06-0171-04

基于数字孪生技术的中压配网多源异构数据关联性分析

万 敏，郑剑锋，刘志陆

(广东电网有限责任公司佛山供电局,广东 佛山 528000)

摘 要: 为准确分析中压配网运行数据间的关联性,提出基于数字孪生技术的中压配网多源异构数据关联性分析方法。采集中压配网的台账数据与负荷数据等,基于所采集数据,采用数字孪生技术实施虚实映射处理,构建中压配网孪生模型。采集孪生模型中的多源异构孪生数据,采用关联规则分析法挖掘孪生数据集的最大频繁项集,构建相异度数据结构矩阵,以此为基础确定孪生数据特征量,实现孪生模型中的数据聚类,对比聚类结果与规则库内的相应规则,实现孪生数据关联性分析。实验结果表明,该方法有效构建孪生模型,可准确分析中压配网多源异构数据关联性。

关键词: 数字孪生技术;中压配网;多源异构;数据关联性;关联规则分析;规则库

中图分类号:TP391.9 文献标识码:A 文章编号:1003-7241(2025)06-0171-04

Correlation Analysis of Multisource Heterogeneous Data in Medium Voltage Distribution Network Based on Digital Twin Technology

WAN Min, ZHENG Jian-feng, LIU Zhi-lu

(Foshan Power Supply Bureau of Guangdong Power Grid Co., Ltd., Foshan 528000, China)

Abstract: In order to accurately analyze the correlation between the operation data of the medium voltage distribution network, a method for analyzing the correlation of multi source heterogeneous data in the medium voltage distribution network based on digital twinning technology is proposed. It collects the account data and load data of the medium voltage distribution network, and based on the collected data, implements virtual reality mapping processing using digital twinning technology to build a twin model of the medium voltage distribution network, collects multi-source heterogeneous twin data in the twin model, uses association rule analysis to mine the maximum frequent itemsets of the twin dataset, and constructs a dissimilarity data structure matrix. Based on this, it determines the characteristic amount of the twin data, achieves data clustering in the twin model, compares the clustering results with the corresponding rules in the rule base, and achieves correlation analysis of the twin data. The experimental results show that this method can effectively construct a twin model and accurately analyze the correlation of multi source heterogeneous data in the medium voltage distribution network.

Keywords: digital twin technology; medium voltage distribution network; multi source isomerism; data relevance; association rule analysis; rule base

0 引言

中压配电网是指电压为6 kV-20 kV的农网配电网[1]。中压配网在实际运行过程中会产生多种海量的数据,这些数据具有来源不同、结构不同、时间尺度不同、空间尺度也不同的特征,同时也具有数据量大、增长效率高、类型多、价值高等特性[2],这些特性导致中压配网内的数据分析与应用较为困难。而数据关联性分析能够对多源异构数据进行挖掘分析,提升中压配网运行管理能力。

有学者在研究数据关联性问题过程中,采用向量化方式描述多源异构数据,并将其作为模型输入,利用神经网络挖掘数据特征向量,通过融合特征向量的语义与分布特征,实现数据关联性分析[3]。有学者在研究数据关联性问题过程中,采用大数据技术中的模糊聚类算法提取数据特征,分析对数据产生显著影响的因素,选取主成分分析法分析数据与因素间的相关性[4]。李国栋等[5]在分析数据关联性问题中,采用数据挖掘技术构建数据间的关联关系,并将关联关系分析结果与设定的规则进行对比,实现数据关联性分析。

数字孪生技术是一种新兴技术,其在本质上是一种以数字化形式构建物理实体的虚拟模型[6],该技术利用实体的相关数据,预测其全生命周期过程[7]。本文提出基于数字孪生技术的中压配网多源异构数据关联性分析方法,准确分析数据间的关联性。

*基金项目:广东省科技项目(202100324001)

收稿日期:2023-11-30

1 中压配网多源异构数据关联性分析方法

1.1 中压配网大数据采集

中压配网内包含多个信息源，如中压配网生产管理信息系统、中压配网电能质量监测管理信息系统、中压配网配变负荷监测系统、中压配网用电信息采集系统等。不同信息系统内所获取的数据结构也有所差异，其中既包含结构化数据，也包含半结构化数据。因此在进行中压配网多源异构数据关联性分析前，须采集中压配网内的多源异构数据。表1所示为中压配网内不同结构类型数据的来源。在采集多源中压配网大数据过程中，主要分为信息内网数据采集与信息外网数据采集[8]。设计中压配网多源异构数据采集装置，将其部署在中压配网信息内网中；在采集外网数据过程中，主要依靠安全隔离装置。中压配网大数据采集的整体思路是以电力企业服务总线为基础，通过数据接口与安全文件传输等方式，统一相关接口与周期等参数，自动采集不同来源、不同结构类型的数据。

表1 中压配网内不同结构类型数据的来源

数据类别	数据来源	详细数据
台账数据	中压配网自动化配电系统、中压配网生产管理系统等	中压配网中不同设备元件的相关参数、型号、尺寸、功率与电压电流等参数
实时运行数据	中压配网用电信息采集系统、中压配网生产管理系统以及中压配网电能质量监测系统等	中压配网中的供电、电压等测量数据、不同元件设备的开断情况、检修状态、调度数据、设备维护信息等
负荷数据	中压配网配变负荷监测系统、中压配网负荷控制系统	高级量测数据与SCADA量测数据
环境数据	中压配网气象系统、中压配网地理信息系统	中压配网的风速、温度、湿度与日照强度等数据
用户数据	配网客服系统	用户满意度评价与用户缴费信息等

1.2 基于数字孪生技术的中压配网模型构建

针对所采集的中压配网大数据，采用数字孪生技术进行虚实映射处理，构建中压配网运行状态孪生模型，模型构建具体过程如下：

依照中压配网内不同设备元件的型号、尺寸、功率与电压电流等相关参数构建配网内设备元件的几何模型，在所生成的几何模型内引入设备元件相应的属性信息，通过数据接口将所采集的配网多源异构数据与孪生模型结合在一起。通过孪生模型集成并反映配网运行行为与逻辑，初步融合实际配网设备元件 T_U 与数字孪生配网设备元件 F_U，在此基础上基于“条件—状态—事件”这一逻辑，生成中压配网设备元件运行状态数字孪生模型，该模型可通过式(1)描述：

$$\begin{cases} T_U = I_P + Y_H + R_T + M_K \\ I_P \leftarrow Y_H \triangleright\!\triangleleft R_T \triangleright\!\triangleleft M_K \\ K_E = \left(onP_O, thenB_N\right) \end{cases} \tag{1}$$

式中，I_P和Y_H分别表示数字孪生的中压配网设备元件几何模型集合和物理属性集合；R_T和M_K分别表示数字孪生配网设备元件运行逻辑模型集合和运行动作行为集合；←和$\triangleright\!\triangleleft$分别表示配网设备元件几何模型内运行逻辑与行为的相关性集成和自然连接；K_E、P_O和B_N分别表示需满足的数字孪生配网设备元件运行状态的实时属性集合、匹配的状态模式K_E和同P_O匹配条件下中压配网设备元件的运行行为。

在上述模型内引入所采集的中压配网状态多源异构数据，实现配网设备元件状态多源异构数据的虚实映射，由此构建配网数字孪生模型。采集孪生模型中的多源异构孪生数据，利用聚类方法分析数据间的关联性。

1.3 数据关联性分析

1.3.1 多源异构数据频繁项集挖掘

W和I分别表示中压配网数字孪生模型中的多源异构数据集合和W的项目全集，w_n为I的子集。由此能够得到，w_n具有线程控制标识符，能够描述孪生数据编号。以X表示I的任意子集，即W的项集。如果$w_n \subseteq X$，那么可定义w_n属于X。$P_e(X)$表示W的X出现的频数，通过支持度$P(X)$描述X出现的频率，由此得到：

$$P(X) = \frac{P_e(X)}{|W|} \tag{2}$$

式中，$|W|$表示W内的中压配网数字孪生模型中的多源异构数据量。

在X的$P(X)$大于支持度阈值$P_{min}(X)$的条件下，X为频繁项集；相反，在$P(X)$小于$P_{min}(X)$的条件下，X为非频繁项集。

在生成频繁项集的条件下，需对其实施候选剪枝。详细过程如下：

在分析中压配网多源异构数据关联性过程中，基于专家设定获取$P_{min}(X)$，由此得到支持度下限$C_{min}(X)=P_{min}(X)|W|$。基于W与$P_{min}(X)$对所得频繁项集进行挖掘分析，能够得到所有频繁项集。

若$X \Rightarrow Y$为关联规则，项集X与Y分别为规则的前件与后件，在$X \Rightarrow Y$内，X与Y不相交。并且X与Y均包含大于阈值频繁项集的支持度。基于$P_{min}(X)$与置信度阈值匹配关联规则的标准，则：

$$C_{min}\left(X \Rightarrow Y\right) = \frac{P(X \cup Y)}{P(X)} = \frac{P_e(X \cup Y)}{P_e(X)} \tag{3}$$

式中，$C_{min}(X \Rightarrow Y)$表示最小支持度条件下的关联规则。

$$P(X \Rightarrow Y)=\frac{P_e(X \cup Y)}{q} \tag{4}$$

式中，q表示所得的中压配网数字孪生模型中的多源异构数据条数。

若任意X均可描述m维特征向量$(\chi_{\chi,1},\chi_{\chi,2},\cdots,\chi_{\chi,m})$，那么$I$内的所有子集均可生成幂集格$G(I)$的同构位置格。依照频繁项集的描述，若$X$为频繁项集，那么其中包含的全部子集均为频繁项集，$G(I)$即为基于频繁项集所得的集合。

设定最大频繁项集为G(I)的极大元，基于I的所有子集所生成的$G(I)$与$\{0,1\}^m$，通过类同挖掘的极大元法，完成中压配网数字孪生模型中的多源异构数据最大频繁项集挖掘。

1.3.2 基于相异度数据矩阵的数据聚类

采用机器学习的方法，基于物理或抽象的集合，确定集合的关联性。在基于数据聚类的多源异构数据关联性分析过程中，数据簇的实体情况是通过关联性形式存在的，所以能够获取非关联的不同簇。检测中压配网数字孪生模型中的多源异构数据类簇聚类情况，相同类簇内任意两个数据间的距离与不同类簇内的数据距离相比更小，所以基于聚类的中压配网数字孪生模型中的多源异构数据关联性分析具体描述如下：在给定的中压配网数字孪生模型中的多源异构数据集内，基于数据间的关联水平划分数据集，簇C_i，$C_j \subseteq W$同时满足：

$$\begin{cases} C_i \cup C_j = \varnothing, i \neq j \\ C_i \cup C_j = W \end{cases} \tag{5}$$

式中，β表示相异度矩阵，基于n个对象内有可能存在两个数据对象w_1、w_2间的相异性实现β的存储，可通过$n \times n$型矩阵描述，其中的全部元素$d(w_1,w_2)$就是w_1、w_2间的相异性，描述如下：

$$\beta=\begin{pmatrix} 0 & & & \\ d(2,1) & 0 & & \\ \vdots & \vdots & \ddots & \\ d(n,1) & d(n,2) & \cdots & 0 \end{pmatrix} \tag{6}$$

通常情况下，$d(w_1,w_2)$为非负数，在w_1、w_2较为相似的条件下，$d(w_1,w_2)$值趋近0。$d(w_1,w_2)$同w_1、w_2间差异性之间呈正比例相关，由此能够生成相异度矩阵。基于所构建的相异度结构矩阵，确定配网数字孪生模型中的多源异构数据特征量，实现多源异构数据的聚类，根据聚类结果，对比规则库内的相应规则，实现聚类结果关联性分析。

2 实验结果与分析

本文提出基于数字孪生技术的中压配网多源异构数据关联性分析方法，为验证本文方法在实际关联性分析中的应用效果，选取某中压配网为实验对象，实验对象的主要组成部分如表2所示。

表2 实验对象主要组成

设备	数量/台
发电机	6
变压器	9
电压互感器	6
电流互感器	6
接触器	6
负荷开关	26
隔离开关	28
断路器	6
熔断器	9
母线	28段

2.1 实验对象参数采样误差分析

通过本文方法采集实验对象内的多源异构数据，由于实验对象运行过程受谐波干扰，因此以台账数据内的功率参数为例，分析本文方法在数据采集中的采样误差，结果如图1所示，实验对象运行过程中的功率参数采样误差不受采样时间影响，整体采样误差基本控制在2×10^{-6}～-2×10^{-6}范围内，说明本文方法在采集实验对象运行参数过程中的误差较小，这有利于后续实验对象数字孪生模型构建精度的提升。

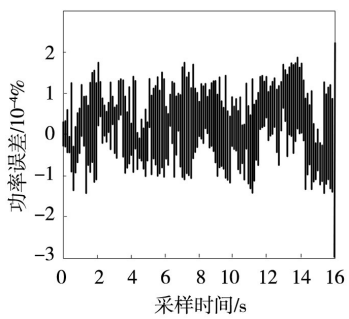


图1 功率参数采样误差分析

2.2 数字孪生模型构建

基于本文方法所采集的实验对象多源异构数据，构建实验对象数字孪生模型，图2所示为实验对象内设备的数字孪生模型构建结果。由图2得到，本文方法基于所采集的实验对象相关数据，通过虚拟映射，能够基本实现1：1的实验对象数字孪生模型构建。

2.3 数据聚类

基于所构建的研究对象数字孪生模型，采集孪生数据，采用本文方法对孪生数据的关联性进行分析，所得结

果如图3所示。由图3可得，采用本文方法能够有效实现不同类别的数据聚类目的。同时在数据类别较多的条件下，本文方法依旧能够获取较高的数据聚类精度，由此说明本文方法具有较高的聚类性能。

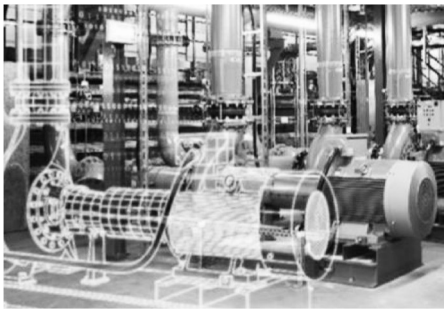


图2 实验对象设备数字孪生模型构建结果

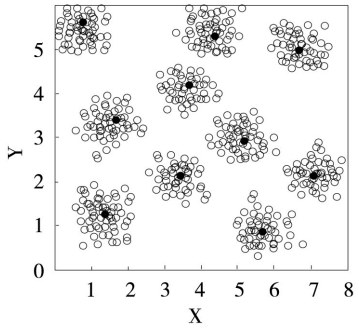


图3 孪生数据聚类结果

2.4 应用分析

基于数据聚类结果，通过数据特征量与规则库的关联性分析。分析表3得到，采用本文方法能够有效依照数据特征量与规则库间关联性的强弱实现数据类型分析，且采用本文方法能够获取准确数据类型分析结果，由此说明本文方法具有较好的应用性。

表3 实验对象数据关联性分析结果

数据类别	数据特征量与规则库关联性强弱	分析结果
发电机	强	发电机
变压器	弱	非变压器
电压互感器	强	电压互感器
电流互感器	强	电流互感器
接触器	强	接触器
负荷开关	弱	非负荷开关
隔离开关	弱	非隔离开关
断路器	强	断路器
熔断器	弱	非熔断器
母线	弱	非母线

3 结束语

本文提出基于数字孪生技术的中压配网多源异构数据关联性分析方法，基于中压配网相关数据构建配网数字孪生模型，并通过基于聚类的关联性分析方法对模型内的多源异构数据进行分析。实验结果显示本文方法能够有效构建中压配网数字孪生模型，准确聚类数据并分析数据类别，具有较好的应用性。

参考文献：

[1] 周慷，温立超，杨震一，等.基于改进数据驱动模型的低压台区拓扑关系自动识别技术研究[J].能源与环保，2022，44(11)：214-219.

[2] 杨翔惟，张洪达，刘霞，等.面向多源异构数据环境的区域水土流失野外调查技术研究与应用[J].干旱区资源与环境，2020，34(10)：139-146.

[3] 刘忠宝，赵文娟.融合语义特征和分布特征的跨媒体关联分析方法研究[J].情报学报，2021，40(5)：471-478.

[4] 白浩，袁智勇，梁朔，等.基于大数据处理的配网运行效率关联性分析[J].电力系统保护与控制，2020，48(6)：61-67.

[5] 李国栋，王峥，贾东梨，等.基于关联分析算法的配电网多级冗余数据校验与挖掘技术研究[J].电测与仪表，2020，57(15)：65-70.

[6] 邹东，冯剑冰.数字孪生技术在城市轨道交通供电系统中的应用场景分析[J].城市轨道交通研究，2021，24(3)：158-162，165.

[7] 李海峰.中美数字孪生研究主题的比较分析——兼论基于结构话题模型的文献主题数据挖掘方法[J].情报杂志，2022，41(1)：156-163.

[8] 杨昊，余芳强，高尚，等.基于数字孪生的建筑运维系统数据融合研究和应用[J].工业建筑，2022，52(10)：204-210，235.

作者简介：万敏(1987—)，男，本科，高级工程师，研究方向：规划、通信。