

基于状态估计的海量多元异构智能电网数据 压缩存储方法

马兴明¹, 董 成², 毛新宇¹, 焦玉新¹, 李 浩²

(1. 国网黑龙江省电力有限公司大庆供电公司, 黑龙江 大庆 163311;

2. 武汉国电武汉电气股份有限公司, 湖北 武汉 430074)

摘 要: 海量多元异构智能电网数据未经处理就进行压缩与存储, 存在压缩误差大、运行时间长的问题, 影响压缩存储效果。因此, 提出基于状态估计的海量多元异构智能电网数据压缩存储方法。融合海量多元异构智能电网数据, 推导出准确的智能电网数据, 通过 Tucker 分解智能电网大数据压缩方法, 压缩海量多元异构智能电网数据。采用可扩展标记语言 (XML) 技术预处理数据, 结合非关系型的数据库技术, 实现海量多元异构智能电网数据的快速存储。试验结果表明, 该方法的线路电阻、电抗动态参数估计准确性高, 数据压缩平均绝对误差、F-范数误差低, 运行时间短, 具有一定的实际应用性能。

关键词: 状态估计; 海量多元异构; 智能电网; 数据压缩; 存储方法; 张量分解

中图分类号: TM711 文献标志码: A 文章编号: 1673-6540(2023)02-0067-06

doi: 10.12177/emca.2022.160

Data Compression and Storage Method of Massive Multivariate Heterogeneous Smart Grid Based on State Estimation

MA Xingming¹, DONG Cheng², MAO Xinyu¹, JIAO Yuxin¹, LI Hao²

(1. State Grid Heilongjiang Electric Power Co., Ltd., Daqing Power Supply Company, Daqing 163311, China;

2. Wuhan Guodian Wuyi Electric Co., Ltd., Wuhan 430074, China)

Abstract: Massive multivariate heterogeneous smart grid data are compressed and stored without processing, which has the problems of large compression error and long running time, affecting the compression and storage effect. Therefore, a data compression and storage method of massive multivariate heterogeneous smart grid based on state estimation is proposed. The massive multivariate heterogeneous smart grid data are integrated to deduce the accurate smart grid data. And the massive multivariate heterogeneous smart grid data are compressed through the Tucker decomposition smart grid big data compression method. Extensible markup language (XML) technology is used to preprocess the data, and combined with the nonrelational database technology to realize the rapid storage of massive multivariate heterogeneous smart grid data. The experimental results show that this method has high accuracy in estimating the dynamic parameters of line resistance and reactance, low average absolute error and F-norm error of data compression, short running time and certain practical application performance.

Key words: state estimation; massive multivariate heterogeneous; smart grid; data compression; storage method; tensor decomposition

收稿日期: 2022-06-06; 收到修改稿日期: 2022-09-09

作者简介: 马兴明(1986—), 男, 硕士, 高级工程师, 研究方向为配网电力设备运维技术。

董 成(1980—), 男, 硕士, 教授级高级工程师, 研究方向为电力系统、大数据、物联网、光技术和电力人工智能。

毛新宇(1985—), 男, 本科, 高级工程师, 研究方向为继电保护二次设备管理。

0 引言

电力是经济发展的基础,我国对智能电网建设力度逐年加大。在智能电网的运行中,会出现海量多元异构智能电网数据^[1-3],其呈现出多样化的特点,会产生数据重复存储现象,导致智能电网系统资源利用率不高^[4-6]。如何处理海量多元异构智能电网数据是智能电网公司亟待解决的难题^[7]。

国内学者针对海量多元异构数据展开了众多研究,如闫会玉等^[8],研究工况数据压缩方法,根据从工业电气设备采集到的多维时间序列数据的变化情况设定死区限值大小,收集设备工况信息并标记工况标签,根据标签序列的变化制定设备运行状态分类规则,实现数据的分类压缩与存储。但是该方法在压缩时未能对海量多元异构数据进行数据处理,海量数据中的坏数据加大了计算难度,增加了压缩误差,影响了最终的压缩存储效果。王鹤等^[9]研究基于分布式压缩感知的数据压缩存储方法,将设计的自适应联合重构算法应用到以分布式压缩感知为边缘算法的云边协同框架中,分析边缘上传的字典原子和测量值,实现电能质量数据的压缩存储。但是该方法在进行储存前没有对数据进行预处理,导致数据压缩存储过程中的冗余信息较多,增加了运行时间。

状态估计是智能电网基础数据处理方法,通过数据分析,预测智能电网设备质量状态,保证智能电网的正常运行。因此,本文提出基于状态估计的海量多元异构智能电网数据压缩存储方法,提高海量多元异构智能电网数据管理质量。

1 方法研究

1.1 基于状态估计的海量多元异构智能电网数据融合

多元异构数据是一种混合型数据,包括结构化数据和非结构化数据。在智能电网的运行中,数据来源较多,数据类型多种多样,且数据量较为庞大,数据融合的难度较大。因此,本文设计的数据压缩储存方法通过测量数据来识别海量多元异构智能电网数据^[10]。状态估计是数据处理的重要方法之一,为提高智能电网数据的分析效率,将采集的海量多元异构智能电网数据进行实时数据

处理,以此对数据状态进行估计,实现海量多元异构智能电网数据融合。

为了推导出准确的智能电网数据,依据状态估计的运算结果进行数据处理与参照物比较,识别海量多元异构智能电网数据。智能电网中状态估计数据与海量多元异构数据的融合流程如图 1 所示。

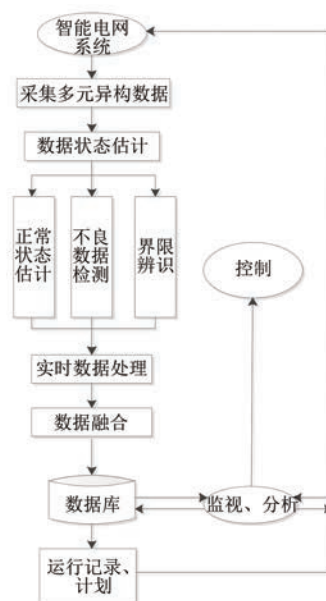


图 1 状态估计数据与海量多元异构数据的融合流程

由图 1 可知,状态估计是智能电网数据处理的核心,利用状态估计实现正常状态估计、不良数据检测和界限辨识,将融合后的多元异构智能电网数据存入数据库^[11],为智能电网的监视、分析和控制提供数据支持。其中状态估计基于加权最小二乘法的状态估计方法实现,流程如图 2 所示。

海量多元异构智能电网数据内存在一些坏数据,若直接采用原始数据实施状态估计会降低准确性。为提升海量多元异构智能电网数据的状态估计质量,必须实施数据处理。坏数据处理包括四个环节,其中利用数据重获来查验智能电网的状态和数据来源,实施相关规则进行坏数据检测,采用决策树理论实施坏数据辨识与修补,经可观测性分析后完成坏数据处理。最后采用分解速度快的加权最小二乘法对处理后的数据实施状态估计。

1.2 基于张量 Tucker 分解的海量多元异构智能电网数据压缩方法

采用张量 Tucker 分解方法对数据融合后的



图2 状态估计流程

海量多元异构智能电网数据实施压缩。首先分析海量多元异构智能电网数据,通过张量定义表明, $B_s \in E^{J_N \times J_t \times J_m}$ 表示数据张量的三个模展开矩阵,用下式描述:

$$\begin{cases} B_{(1)} \in E^{J_N \times (J_t \times J_m)} \\ B_{(2)} \in E^{J_t \times (J_N \times J_m)} \\ B_{(3)} \in E^{J_m \times (J_N \times J_t)} \end{cases} \quad (1)$$

式中: $B_{(1)}$ 为依据站点阶的模展开矩阵; J_N 为海量多元异构智能电网数据的站点; J_t 为海量多元异构智能电网数据的采集时间; J_m 为海量多元异构智能电网数据的量测量; $B_{(2)}$ 为依据时间阶的模展开矩阵; $B_{(3)}$ 为依据量测量阶的模展开矩阵。

通过变换方法分解张量的模展开矩阵,用下式描述:

$$\begin{cases} B_{(1)} = D_{J_N \times J_N} Y_{J_N \times J_t \times J_m} W_{J_t \times J_t}^T \\ B_{(2)} = D_{J_t \times J_t} Y_{J_t \times J_N \times J_m} W_{J_m \times J_m}^T \\ B_{(3)} = D_{J_m \times J_m} Y_{J_m \times J_N \times J_t} W_{J_N \times J_N}^T \end{cases} \quad (2)$$

式中: D 、 W 分别为左、右奇异值矩阵; Y 为三个奇异值矩阵。

为获取三个模展开矩阵的近似奇异值分解公式,需选取奇异值 τ_k 之后的值,得到下式:

$$\begin{cases} BE_{(1)} = DE_{J_N \times P} YE_{P \times P} WE_{J_t \times J_t}^T \\ BE_{(2)} = DE_{J_t \times Q} YE_{Q \times Q} WE_{J_N \times J_N}^T \\ BE_{(3)} = DE_{J_m \times E} YE_{E \times E} WE_{J_N \times J_N}^T \end{cases} \quad (3)$$

式中: DE 、 WE 分别为选取的近似左、右奇异值矩阵; YE 为第 P 、 Q 、 E 个奇异值中选取的近似奇异值对角阵; BE 为奇异值分解近似阵。

通过 Tucker 分解处理 n 阶张量 B_n ,得到下式:

$$B_n = G \times_1 D_1 \times_2 D_2 \times \cdots \times_n D_n \quad (4)$$

$$H = B_n \times_1 D_1^T \times_2 D_2^T \times \cdots \times_n D_n^T \quad (5)$$

式中: G 为压缩过程中的约束参数; D_1 、 D_2 、 $\cdots$ 、 D_n 分别为对应阶的因子矩阵; H 为核心张量。

依据式(5)对海量多元异构智能电网数据张量 B_Y 实施 Tucker 分解,得出压缩后的核心张量,如下所示:

$$H_Y = B_Y \times_1 D_{J_N \times P}^T \times_2 D_{J_t \times Q}^T \times_3 D_{J_m \times E}^T \quad (6)$$

通过上式说明的正交基和原始张量,计算出核心张量 $H_Y \in E^{P \times Q \times E}$ 。所以原数据张量维度降到 P 、 Q 、 E 。

通过选取的近似左奇异值矩阵和核心张量去除海量多元异构智能电网数据张量 B_Y 的冗余数据得到 $\hat{B}_Y$,计算公式如下:

$$\hat{B}_Y = B_Y \times_1 D_{J_N \times P} \times_2 D_{J_t \times Q} \times_3 D_{J_m \times E} \quad (7)$$

在海量多元异构智能电网数据中,非结构化的彩色图片、彩色视频和灰度视频数据^[12,13]的张量模型阶数不同。

(1) 彩色图片数据。五阶核心张量用 H_P 描述,其中 $H_P \in E^{P_3 \times Q_3 \times \cdots \times D_3}$;用 B_P 表示彩色图片,计算出重构数据张量:

$$\hat{B}_P = H_P \times_1 D_{J_N \times P_3} \times_2 D_{J_t \times Q_3} \times \cdots \times_5 D_{J_c \times D_3} \quad (8)$$

式中: J_c 为图像颜色空间。

(2) 彩色视频数据。压缩后六阶核心张量用 H_{CW} 描述,其中 $H_{CW} \in E^{P_2 \times Q_2 \times \cdots \times D_2}$;用 B_{CW} 表示彩色视频,彩色视频的重构数据张量,用下式描述:

$$\hat{B}_{CW} = H_{CW} \times_1 D_{J_N \times P_2} \times_2 D_{J_t \times Q_2} \times \cdots \times_6 D_{J_f \times D_2} \quad (9)$$

式中: J_f 为图像帧。

(3) 灰度视频数据。压缩后五阶核心张量用 H_{DCW} 描述,其中 $H_{DCW} \in E^{P_1 \times Q_1 \times \cdots \times T_1}$;用 B_{DCW} 表示

彩色视频,灰度视频的重构数据张量,用下式描述:

$$\hat{\mathbf{B}}_{\text{DCW}} = \mathbf{H}_{\text{DCW}} \times_1 \mathbf{D}_{J_N \times P_1} \times_2 \mathbf{D}_{J_t \times Q_1} \times \cdots \times_5 \mathbf{D}_{J_l \times T_1} \quad (10)$$

采用张量 Tucker 分解方法对数据融合后的海量多元异构智能电网数据实施压缩,流程如图 3 所示。



图3 张量 Tucker 分解的数据压缩流程

2 数据存储过程

在海量多元异构智能电网数据压缩后的存储过程中,通过采用可扩展标记语言(XML)技术实施数据预处理,将处理后的数据,利用非关系型的数据库(NoSQL)技术,实现海量多元异构智能电网数据的快速存储。

2.1 数据预处理

因压缩后的海量多元异构智能电网数据具有多样性,在进行数据储存之前,需采用 XML 技术标记数据,定义数据类型,对全部基础数据实施数据预处理,处理流程如图 4 所示。

由图 4 可知,海量多元异构智能电网数据包括非结构化数据与结构化数据,针对不同的数据类型采用不同的数据处理流程,将数据以 XML 格式数据表存储至海量多元异构智能电网的各节点中,实现数据预处理。

2.2 数据快速存储优化策略

通过 NoSQL,对预处理后的海量多元异构智能电网数据实施快速存储。实际数据在 NoSQL 上分配海量多元异构智能电网数据存储优化的存储策略如下。

(1) 按照内外网原则对海量多元异构智能电

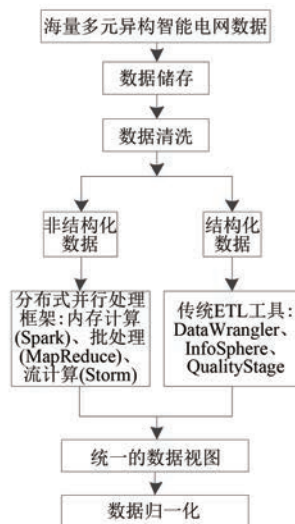


图4 海量多元异构智能电网数据预处理流程

网数据主节点实施冗余计算。

(2) 分别储存智能电网的内外网海量多元异构数据,在对应单位的 IT 资源内存储智能电网的外网数据;智能电网内部 IT 资源只存储智能电网的内网数据。

(3) 按照不同对象分类存储智能电网外部海量多元异构数据;按照不同属性分类存储智能电网内部海量多元异构数据。

(4) 最小路径分配存储策略的出发点是 NoSQL 内的主节点,通过 XML 数据表调用元数据,依据元数据为原始数据配置相对容量的存储空间^[14],假如 XML 数据表内剩余存储空间较小,无法满足需求,依据相邻最近原则,通过续存方式分配所缺的原始数据存储空间^[15]。

(5) 寻找空闲 IT 资源里智能电网节点路径,依据 XML 存储优先分配数据,实现快速存储。

3 试验分析

以某地智能电网 25 个配电站的真实监测数据集作为试验数据,其中包括结构化数据和非结构化数据,验证本文提出的基于状态估计的海量多元异构智能电网数据压缩存储方法的实际应用效果。

在智能电网稳定运行状态下,以某一条输电线路为例,随机选取 6 小时内的数据,并选取两种试验对比方法,分别为文献[8]面向电气设备的工况识别和数据压缩方法、文献[9]基于分布式

压缩感知的配电网电能质量数据压缩存储方法。

3.1 不同方法估计线路电阻、电抗动态参数对比

由于海量多元异构智能电网数据来源较多,电网运行过程中的数据量较为庞大,某时段的电阻与电抗的数据是具有一定波动性的,因此计算电阻与电抗的动态参数能够进一步降低海量多元异构智能电网数据压缩误差。动态参数估计值与实际值越相近,说明该方法准确性越高,越能保证后期海量多元异构智能电网数据压缩效果。分析三种方法估计得到的线路电阻、电抗动态参数,结果如图5和图6所示。

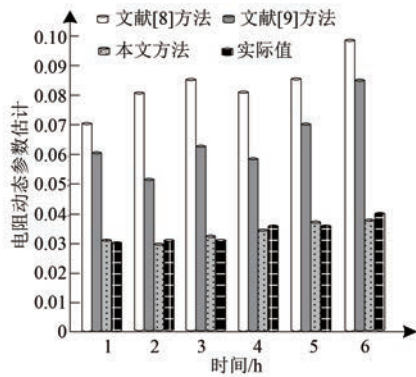


图5 电阻动态参数估计

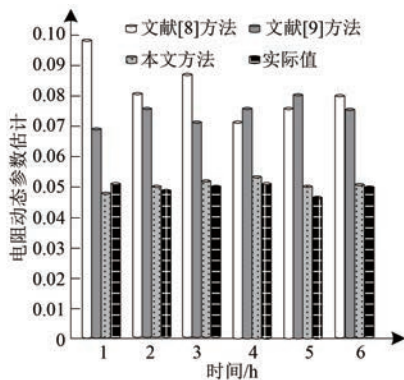


图6 电抗动态参数估计

由图5和图6可知,本文方法估计的线路电阻、电抗动态参数分别为0.032、0.530,结果明显优于其他两种方法,且与实际值非常接近,说明本文方法动态参数的估计准确性较高,能够保证后期海量多元异构智能电网数据的压缩效果。

3.2 不同方法对异常状态估计的结果精度对比

海量多元异构智能电网数据存在较多的变化。由于数据量较大,在数据压缩过程中服务器易出现过载情况。很多伺服驱动器的制动电阻因

过载保护方法不当,会产生电阻突变,从而导致动态参数出现突然变化。因此,需要在进行数据压缩的过程中,对电网异常状态进行估计,准确预测动态参数变化趋势,对电网数据异常进行及时预警,实现智能电网数据的压缩存储。测试三种方法对异常状态下电阻动态参数的估计情况,其结果如图7所示。

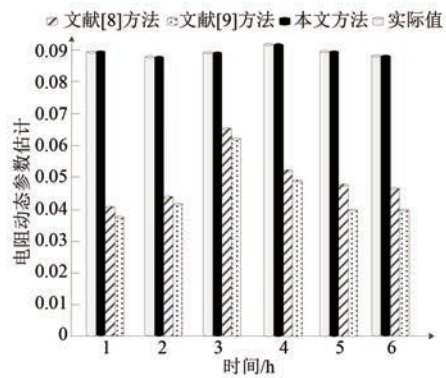


图7 异常状态下电阻动态参数估计

由图7可知,文献[8]方法与文献[9]方法并没有识别出异常状态下电阻动态参数的变化,与实际值差距较大,影响数据压缩效果。而本文方法采用张量 Tucker 分解对数据进行融合,估计的动态电阻与实际值基本一致,可识别出参数的异常变化,识别精度较高,能够实现异常状态下对参数突变的及时预警,实现智能电网数据的压缩存储。

3.3 不同方法的平均绝对误差和 F-范数误差比

海量多元异构智能电网数据类型较多,因此在压缩过程中,很容易由于数据的复杂而增加压缩误差,影响压缩效果。因此计算三种方法的平均绝对误差(MAE)和F-范数误差比(FER),二者越低,说明该方法的压缩误差越低,压缩效果越好。选取验证本文方法数据压缩效果的信息损失水平指标如下。

(1) 其第 h 个量测量的 MAE 如下:

$$MAE(h) = \frac{1}{J_t \times J_N} \sum_j \sum_i |B(i,j,h) - \hat{B}(i,j,h)| \quad (11)$$

式中: $B(i,j,h)$ 为 (i,j,h) 阶的张量; $\hat{B}(i,j,h)$ 为 $B(i,j,h)$ 张量的估计值。

(2) 总体损失水平的 FER 如下所示:

$$FER = \frac{|B - \hat{B}|_F}{|B|_F} \quad (12)$$

式中: B 为张量; $\hat{B}$ 为张量的估计值; F 为范数。

分析三种方法的压缩效果,结果如图 8 所示。

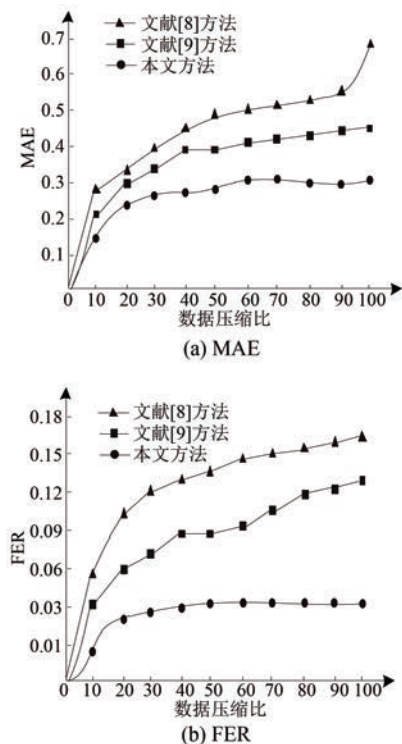


图 8 数据压缩误差结果

由图 8 可知,本文方法所产生的压缩误差处于平稳并缓慢上升态势,尤其在低压缩比下,本文方法的误差明显低于其他两种方法。本文方法在不同数据压缩比下的压缩 MAE、FER 分别为 0.240、0.025,均比其他两种方法的误差低,证明了本文方法的压缩误差低,压缩效果较好。

3.4 不同方法的运行时间对比

在并行计算模式下,测试三种方法的运行时间,结果如图 9 所示。由图 9 可知,本文方法的平均运行时间为 213 ms,分别比其他两种方法的平均运行时间快 353、257 ms,明显少于其他两种方法,说明本文方法的运行效率高,可实现海量多元异构智能电网数据的快速压缩存储。

4 结 语

为了提高海量多元异构智能电网数据压缩存储效率,本文提出了基于状态估计的海量多元异

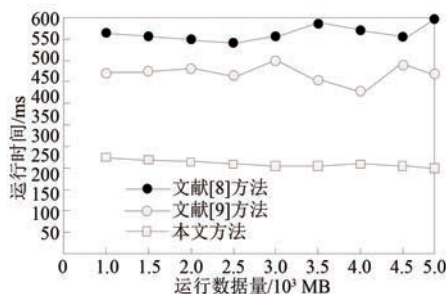


图 9 不同方法运行时间

构智能电网数据压缩存储方法。采用基于加权最小二乘法的状态估计实施数据融合,结合张量 Tucker 分解数据压缩方法和 NoSQL 技术实现海量多元异构智能电网数据的压缩存储。试验结果表明,本文方法的精度高,误差小,运行时间低至 213 ms,运行效率高。

【参考文献】

- [1] 吴争荣,俞小勇,董旭柱,等. 基于状态估计的配电网实时态势感知与评估[J]. 电力系统及其自动化学报,2018, 30(3): 140.
- [2] 李雪,李雯婷,杜大军,等. 拒绝服务攻击下基于 UKF 的智能电网动态状态估计研究[J]. 自动化学报,2019,45(1): 120.
- [3] 沈晨晖,汪伟,张晓良,等. 基于 SR-CKF 的电液伺服系统状态估计和故障诊断[J]. 中国测试,2018, 44(1): 101.
- [4] 盛万兴,方恒福,沈玉兰,等. 考虑量测延时基于 3 种数据融合的配网状态估计[J]. 电力系统及其自动化学报,2019,31(12): 108.
- [5] 王鹏,陈蕾,陈艳波,等. 一种适应于配电网的交叉逼近状态估计方法[J]. 电力自动化设备,2019,39(12): 102.
- [6] 赵会群,李春良. 基于密度划分的数据存储方法与技术[J]. 计算机工程与设计,2020,41(9): 2482.
- [7] 赵洪山,马利波. 基于张量 Tucker 分解的智能配电网大数据压缩[J]. 中国电机工程学报,2019,39(16): 4744.
- [8] 闫会玉,宋宏,郑泽宇,等. 面向电气设备的工况识别和数据压缩方法[J]. 计算机应用,2018, 38(S2): 148.
- [9] 王鹤,李石强,于华楠,等. 基于分布式压缩感知和边缘计算的配电网电能质量数据压缩存储方法[J]. 电工技术学报,2020,35(21): 4553.

(下转第 81 页)