

大规模配电网负荷数据 在线清洗与修复方法研究

刁赢龙, 盛万兴, 刘科研, 何开元, 孟晓丽

(中国电力科学研究院, 北京市海淀区 100192)

Research on Online Cleaning and Repair Methods of Large-Scale Distribution Network Load Data

DIAO Yinglong, SHENG Wanxing, LIU Keyan, HE Kaiyuan, MENG Xiaoli

(China Electric Power Research Institute, Haidian District, Beijing 100192, China)

ABSTRACT: In order to improve data availability in field of distribution network planning and intelligence analysis with reduced data cache cost, effectively analyze large-scale, mixed and inaccurately monitored or collected load data online, and to ensure consistent deviation detection and accurate repair for time series data in each cycle, an online data cleaning and repair method for large-scale distribution network load data is proposed based on analysis of different types of abnormal load causes and distribution features, including abnormal load steam identification method on density and data repair method on collaborative filtering recommendation algorithm. To break through bottlenecks in online data analysis performance for distribution network load, parallel solution on Hadoop platform is given. Verified with actual distribution network operation data, result shows that the proposed algorithm and frame could get effective data preprocessing and yield favorable significance in practice and research.

KEY WORDS: data cleaning; stream data; large-scale distribution network; online cleaning

摘要: 为减少数据缓存成本, 提高负荷数据在配电网规划设计、智能分析等领域的可用性, 充分有效地对大规模、混杂、不精确的监测或采集负荷数据进行在线清洗, 保证每个周期的时序数据得到一致的偏差检测和精确修复, 在分析不同类型异常负荷数据产生原因和分布特点的基础上, 提出一种面向大规模配电网负荷数据的在线清洗与修复方法, 包括基于密度的负荷数据流异常辨识方法和基于协同过滤推荐算法的负荷数据修复方法。为突破配电网负荷大数据在线分析性能瓶颈, 还在 Hadoop 平台上给出相应的分布式并行解决方案, 通过使用实际配电网运行中的负荷数据进行验证, 结果

表明所提算法和框架能够有效预处理配电网负荷数据, 具有实际应用价值。

关键词: 数据清洗; 流数据; 大规模配电网; 在线清洗

DOI: 10.13335/j.1000-3673.pst.2015.11.018

0 引言

配电网负荷数据的有效分析不仅可以满足规划设计、生产调度、智能仿真、负荷预测、电能质量和电能决策工作的需要^[1-3], 还能解决未来配电网面临的精确供能、电力需求侧管理、分布式电源接入、多电源互动以及分散储能等问题。随着传感量测、信息通信、分析决策、自动控制和能源电力等技术与现代配电网的深度融合, 产生了指数量级增长的海量异构、多态、高维的负荷数据, 例如一些东南沿海中、大型城市的市区中压馈线已超过千条水平, 地市供电公司各种系统的监测设备超过 10 万个, 配电自动化系统数据库预留达到 100TB 以上, 年增长量超过 5TB, 用电信息采集系统日采集量近 100 万条, 数据规模达到当前信息学界所关注的“大数据”级别^[4-9], 如何在可承受的时间和存储成本范围内对负荷数据进行清洗和修复, 提高后续分析决策结果的质量和精度, 实现对配电网负荷情况的精确掌握, 充分利用负荷数据指导配电网生产工作是当前配电网面临的挑战^[9-13]。

相较于信息学科提出的互联网、金融、通信等领域大数据而言, 配电网负荷数据具有以下特点: 1) 高并发性, 远方监测终端对象数量庞大且分布广泛, 基于非循环式远动规约同时提交负荷数据, 流量惊人。2) 时序性强, 数据在采集、传输、分

基金项目: 国家电网公司科技项目(EPRIPDKJ [2014]3763 号)。

Project Supported by Science and Technology Project of SGCC(EPRIPDKJ [2014] NO.3763).

析和存储过程中以数据流的形式持续出现，需要以尽可能高效的方式对数据进行即时处理。3) 数据处理代价高昂，负荷数据一般为高精度浮点数且数值敏感性较强，同时配电网分析算法复杂，需要大量硬件资源才能实现负荷数据的传输与处理。所以即使对少量数据进行处理也需要较高的存储和计算成本。

文献[14]针对输变电设备状态评估数据，将设备状态信息等效成各状态量的时间序列，提出一种基于时间序列分析的双循环迭代检验法。文献[15]根据提取的电力日负荷特征曲线的特征，利用支持向量数据描述法对不良数据进行精确定位，最后利用特征曲线对不良数据进行修正。文献[16]基于 SCADA 系统的负荷历史数据，提出一种动态的智能清洗模型。

当前对配电网数据清洗算法的研究着重于算法的具体应用场景和清洗目标，缺少在配电网大数据方向进行具体应用的研究，无法快速清洗在线时序数据或流数据，更无法保证不同周期的数据检测一致性。为提升在线实时负荷数据的质量，从海量时序样本中快速排除干扰点，提高数据抽取和采集的精确度，本文提出了大规模配电网负荷数据在线清洗与修复方法，以典型城市大规模配电网负荷数据为基础，首先分析了配电网负荷数据来源、典型异常数据的原因和特征；然后分别提出了异常负荷数据实时辨识与快速修复算法。本文还具体给出了面向在线负荷数据流的分布式并行清洗框架和修复方案。最后对东南某发达城市的配电网大规模负荷时序数据进行相关验证。实际算例证明，本文提出的方法能够有效辨识配电网负荷数据的无效值与缺失值，同时实现异常数据的精确恢复。

1 异常负荷数据分析

当前配电网负荷数据来源于配电自动化系统、用电信息采集系统、负荷控制系统、用户集中抄表系统、电能质量监测系统和一些地市公司安装的其他检测系统。部分系统对负荷数据的采集情况对比如表 1 所示。受设备、环境和运行状态等因素影响，异常数据特点较为复杂。

表 1 部分配电网负荷数据采集情况对比 Tab.1 Distribution network load data acquisition from different systems				
数据源系统	数据质量	监测对象	采集频率	上传周期
配电自动化系统	较差	中压配网	4 次/h	15 min
用电信息采集系统	较好	配电台区	4 次/h	1 day
电能质量检测系统	一般	配网	—	—
用户集中抄表系统	较好	低压配网	2 次/h	0.5h 或召测

1.1 远方配电终端设备故障及量测误差

安装在高中低压配电网现场的各种远方监测设备、用电信息采集终端和各类通信设备长期暴露于户外，部分设备环境适应能力较差，故障率较高，当出现故障时会产生大量负荷数据漏传。同时受周围环境、天气以及季节变化的影响，配电通信网无法保障终端完成监测信息的实时传输。其次各环节设备的制造工艺和安装缺陷以及调试和使用上的波动都会带来各种各样的误差因素，配电网量测误差至少还包括：电压互感器与电流互感器设备误差、时间不同步导致数据采集的不同时性、三相不平衡及功率因数变化导致的误差。

1.2 配电网运行控制和故障

配电网一次线路故障包括短时短路、大规模停电等产生的冲击负荷或无负荷，虽然能反映实际运行情况，但会影响设计规划和负荷预测结果精度，所以也应该进行清洗；二次方面，电压互感器的一次断线影响遥测信息中电压和功率的正确测量。配电网调度控制时开合断路器产生的励磁涌流也会导致出现异常负荷数据，分布式电源切入切出引发功率逆流，如果没有更换新式电表也会发生一定的量测问题。

1.3 配电通信网通信传输故障

配电网所使用的通信方式在电力系统中较为复杂，信道既包括可靠性很强的光纤以太网方式，也有一些质量较差的配电载波、GPRS 或 3G 无线等通信方式，以及小众的无线扩频、Zigbee 技术，这些信道当遭受过电压、大电流冲击等强电磁干扰时会出现错码断码现象。为保证配电网数据传输速率和实时性，当前大部分通信规约对开放式系统互联(open system interconnect, OSI)参考模型进行了精简(IEC60870-5-101、104 和 DNP3.0)，部分工作层缺少差错恢复、流量控制等功能，这势必导致传输可靠性的下降。另外通信子站或通信接入网也容易发生交换机故障、通信光缆断线，使得大量负荷数据无法集中和转发。

图 1 所示为不同原因下异常负荷数据分布图(假设数据存储分布与监测点在物理位置上分布一致)。图 1(a)(b)和(f)为配电终端监测点出现故障时产生异常数据，分为单点连续性负荷数据异常和抖动性负荷数据异常 2 种。图 1(c)为配电网出现部分馈线故障造成停电产生的异常负荷数据分布，图 1(d)为配电网运行控制对监测点的涌流干扰和配电自动化设备动作造成的短暂零值现象。当配电网通信子站出现故障时容易出现图 1(e)所示多点部分时间

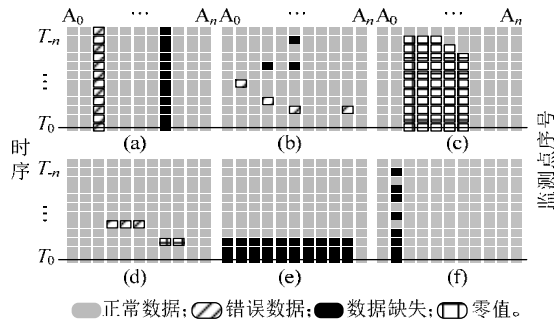


图1 不同原因下异常负荷数据分布
Fig.1 Abnormal load data distribution under different reasons

段数据丢失现象。所以对连续到达主站的负荷数据流,需要进行有针对性的数据清洗工作。图2为典型配电网负荷异常数据,包括配电线路故障、设备故障与通信故障产生的负荷时序数据。

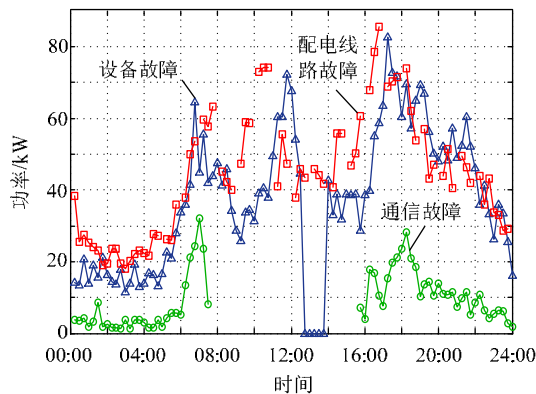


图2 典型配电网负荷异常数据
Fig.2 Typical abnormal data

2 异常负荷数据实时辨识算法

传统对异常数据的检测方法都是通过对比异常点与其他样本点的特征实现数据清洗。而负荷数据异常辨识有别于传统异常数据检测方法主要有以下特征:1)异常点容易聚集,比如零值或因为通信错误产生乱码导致的异常值大多相似。2)面向快速流式数据,需要更加精简的算法进行高效处理。3)考虑异常值分布,配电网线路或通信网络故障会在同一周期产生大量空值数据,需要加入部分历史数据增强样本健壮性。在辨识算法中要考虑以上特征。

本算法的主要思想是计算每个监测点负荷参数局部的异常因子,然后比较所有点的离群程度,判断该程度的依赖对象相对于其本域的孤立情况。具体步骤如下:

设有 n 个负荷数据采集点,每个数据采集点的数据为 m 维, $S(t_0)$ 为当前时刻采集数据样本,不同时刻 $t_k (t_k \in [T, t_0])$ 采集到的负荷数据样本为 $S(t_k)$,为考虑数据时序性,需要合并上一个采集

周期已处理完的“干净”数据进行清洗。

$$Y(S_C(t_{-1})) \cup S(t_0) \xrightarrow{\text{删除冗余点}} S^+(t_0) \quad (1)$$

式中: $S_C(t_{-1})$ 为 t_{-1} 时刻清洗完毕的负荷数据; $Y(S_C(t_{-1}))$ 表示对 $S_C(t_{-1})$ 进行抽样的结果; $S^+(t_0)$ 为合并之后待检测的数据集;删除冗余点是为了防止相同错误点形成较高密度簇影响异常点辨识。

假定 s 为待检测集合 $S^+(t_0)$ 中的点,设置分析目标范围区域半径 $R_{AD}(s)$ 为距离其第 k 远对象的曼哈顿长度 ($d(i, j) = \sum |x_{ik} - x_{jk}|$),对于浮点运算只要计算加减法即可,这就大大提高了运算速度。其实就是求以 s 为中心,能包含 k 个对象的最小范围,设这 k 个对象的集合为 $N_k(s)$,可以看出在分布不均匀的密度较大的区域 $R_{AD}(s)$ 较小,而密度较小的区域 $R_{AD}(s)$ 较大。

以下公式定义了点 s 与点 p 间的可达距离:

$$R_{EA}(s, p) = \max\{R_{AD}(p), d(s, p)\} \quad (2)$$

这样通过式(3)就可以求出集合 $S^+(t_0)$ 内第 i 个节点的异常度 $L_{OF}(s_i)$,通过对其排序设定一定的阈值就可以获得最终的异常点。

$$L_{OF}(s_i) = \frac{\sum_{p \in N_k(s)} \frac{l_{rd}(p)}{l_{rd}(s)}}{|N(s_i)|} \quad (3)$$

其中, $L_{rdk}(s)$ 为局部 $N_k(p)$ 集合中平均可达距离密度的倒数:

$$L_{rdk}(s) = \frac{|N_k(s)|}{\sum_{p \in N_k(s)} R_{EA}(s, p)} \quad (4)$$

从式(2)(3)可以得出结论:如果 s 点的离群度较小,则对于同一类对象的可达距离取 $R_{AD}(s)$ 的可能性较大,局部可达值分布较为平稳;如果点 s 属于异常点,那么局部可达密度方差也较大。若局部离群点因子较大证明该点距离所有簇都相对较远,通过设定检测阈值获取异常点。

3 异常负荷数据快速修复

短时间内配电网负荷数据的变化虽然较为复杂,与时间、天气、电价、突发事件等因素都有关系,但从相邻范围监测点在时序上负荷变化相似性的角度还是可以挖掘一定的规律。本文采用协同过滤推荐算法,以无监督的方式减少错误负荷数据修补的输入依赖,从配电网区域范围内负荷变化横向关联性着手,寻找与当前时刻相似的历史监测点负荷数据集,计算负荷范围推荐度实现异常负荷数据修正。具体步骤见图3。

1) 统计当前时序的异常负荷数据集 $S_W(t_0)$,

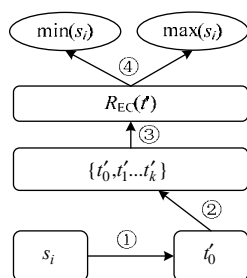


图 3 异常数据修复算法流程

Fig.3 Process of data correction algorithm

循环分析每一个点 s_i ，确定该异常数据的所属负荷点的邻近节点集合，记作 t'_0 ($t'_0 \in t_0$)，通常在配电网负荷数据源系统中负荷数据存储临近的负荷点在实际生产现场地理位置也是邻近的，故很容易得到 t'_0 ，根据负荷数据的这种特点易于对数据进行分区并行修复。

2) 根据 t'_0 读取一定数量的历史数据，形成时间节点集合 $\{t'_0, t'_1, \dots, t'_k\}$ 作为协同过滤算法的输入。

3) 求 t'_0 集合 $\{t'_0, t'_1, \dots, t'_k\}$ 中其他向量的相关系数 r_k ，并进行排序。设置一定阈值得到相似的时序数据集 $R_{EC}(t')$ 。

$$r_k = \frac{1}{k-1} \sum_{X, Y \in t'_k} \left(\frac{X - \bar{X}}{s_X} \right) \left(\frac{Y - \bar{Y}}{s_Y} \right) \quad (5)$$

4) 得到相似时序负荷集合后，就可以进行推荐分析，异常负荷数据上下界估计值为

$$s_i \in [\min_{X \in R_{EC}(t')} (\frac{\bar{X} X_i}{t'_0}), \max_{X \in R_{EC}(t')} (\frac{\bar{X} X_i}{t'_0})] \quad (6)$$

算法的时间复杂度仅为 $O(n \log_2 n) + O(3kn)$ ，由于协同推荐算法的输入为某一时序段部分与异常负荷相邻监测点以及与其对应的前 k 个采

集周期的数据，避免对整体数据集进行修复，提高了异常值的修复速度。同时本算法利用负荷监测点不同时段负荷变化的相关性，无需输入时间、天气、电价等影响因素，既减少了输入数据的规模又使结果更符合实际配电网的运行规律。

4 在线清洗与修复的分布式并行框架

在实际配电网系统中，负荷数据是以数据流的形式持续向主站上传，并发现规模巨大，计算机很难对多个周期数据全貌进行精确分析。负荷数据异常检测和修复的分布式并行框架能够提供针对高并发在线实时的数据清洗，见图 4，配电网负荷数据清洗分布式架构分为数据采集与存储模块、数据准备模块与数据清洗模块。具体功能如下：

1) 数据采集和存储，通过一定的通信规约从远方负荷终端采集或者使用 SQL 语句轮询配电网各负荷数据源系统数据库。对获得的原始数据进行数据解析、差异化处理和时戳添加，采用内存实时数据库对海量并发数据进行缓存，加入流量控制策略优化改善处理延迟，通过与数据清洗模块通信协调限制数据输入带宽；该模块还负责数据归档处理以及向各高级应用分析系统提供数据接口等功能。

2) 数据准备，多个周期负荷数据被读入滑动清洗窗口中，及时对滑动窗口更新已修复完成的“干净”数据。为不失一般性本文在分布式框架中提出了“阶梯型”分布式分片策略，“阶梯型”分片策略通过调节分片斜角 θ 和纵向步长 h ，提高每个集群节点输入样本集内正确数据的覆盖量，如图 5(a)所示，负荷数据样本分为 2 个分片 A 和 B，

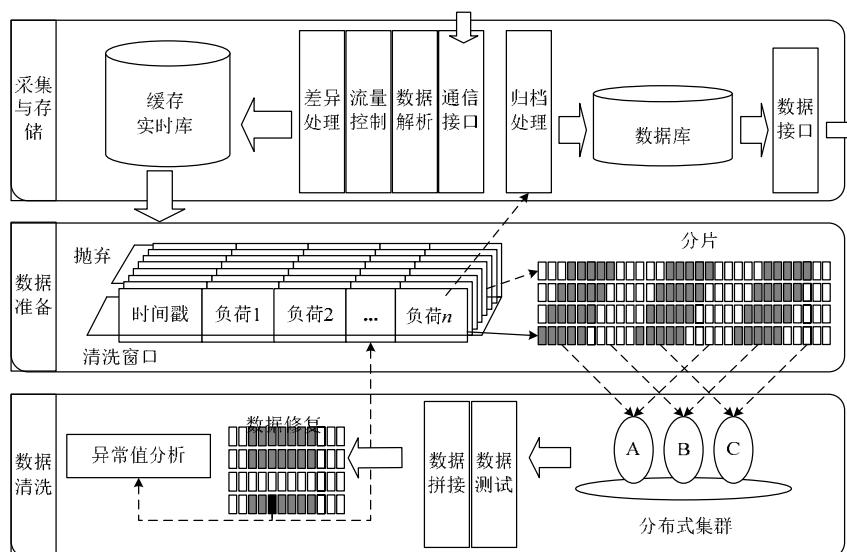


图 4 配电网负荷数据清洗分布式架构

Fig.4 The distributed structure of distribution network data cleaning

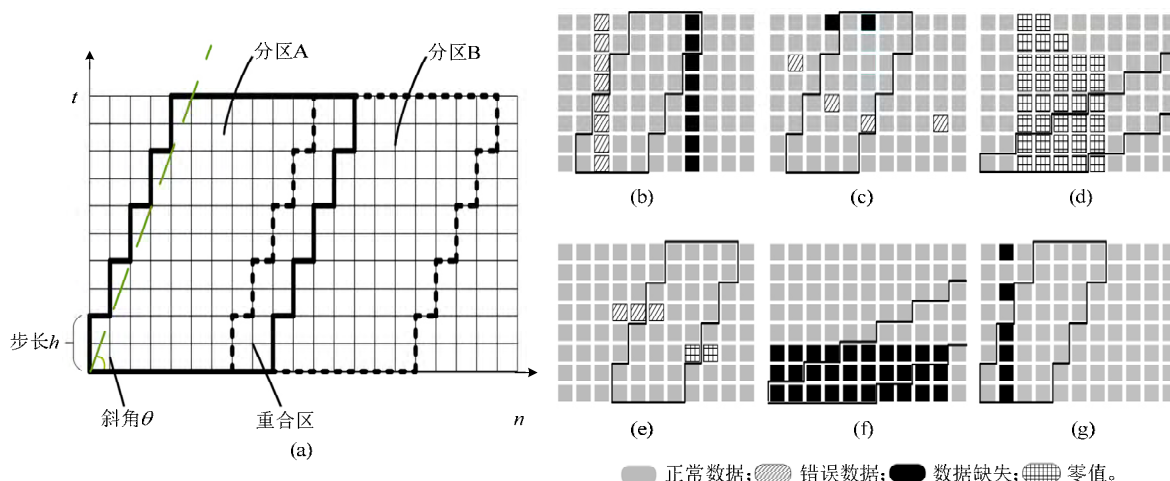


图5 并行数据清洗中“阶梯型”分片策略

Fig.5 Segment strategy 'Stairs Type' in parallel data cleaning

分片间留有重合区的目的是提高处于分片边缘位置数据的辨识正确率,避免因负荷实际位置的不同造成数据差异性误判。参与计算的数据量 E_{SIZE} 与分片斜角、纵向步长间的关系为

$$E_{\text{SIZE}} = \frac{hl}{\tan \theta} \quad (7)$$

式中 l 为参与清洗的周期数。对于图 5(a)(b)所示的常见异常数据分布均能实现辨识,图 5(c)(e)所涉及的异常数据少且分布较为分散,分片斜角与纵向步长较大可减少分区中的异常数据数量,而图 5(b)(g)虽然异常数据连续出现,但数据量相对较小,上述方法同样适用。图 5(d)(f)属于连续性异常,需要降低 θ 、 h 值并提高 l 值,最大限度提高样本中的正常值百分比。这种分区策略可以考虑数据在时序上的一致性,将异常数据隔离在不同的分区中,增加辨识和修复的鲁棒性,实现不同类型异常数据的清洗。

3) 数据清洗,将分片后的负荷数据传输到各分布式节点中并行计算,基于本文所述的异常负荷数据实时辨识算法实现数据辨识,然后在单台计算机上实现数据合并,采用协同过滤推荐算法实现异常负荷数据修正,最后更新数据准备模块清洗窗口中的异常值同时将“干净”数据送至数据库归档。

5 算例分析

为验证本文提出的在线清洗与修复的分布式并行框架,在 5 台高性能服务器上搭建配电网负荷数据清洗 Hadoop 平台(配置为 4*6 核 2.0GHz CPU、16G 内存、1000Mbit/s 网卡、1TB 硬盘),1 台数据库服务器,1 台模拟负荷数据采集计算。

试验样本集为某中型城市用电信息采集系统 1 年的数据,1400 余条馈线约 23000 个负荷监测点,每日数据采集 96 个点,数据维度包括有功功率和视在功率。为测试架构和算法的性能,将原数据集横向拼接为 10 个不同大小的样本集。

图 6 所示异常数据辨识的 5 个步骤平均用时比约为 30:3:70:25:0.1,在预处理过程中时序负荷数据“阶梯”分片和转置时间较长。前 5 个样本中的清洗时间差异性较小,因为 Hadoop 平台处理小文件的效率较低,随着样本集规模的扩大,清洗速度与清洗量近似正相关。图 6 所示实验中最大样本集中有 10 万个监测点共计 200 万条数据,清洗时间约为 11s,其规模可以满足当前以及未来一段时间的配电网系统采集数据量的要求。

图 7 是某一典型日中 14680 个负荷采集点的清洗效率,由于数据传输、分片阶段所消耗时间与异常数据规模无关,同时基于密度的异常数据辨识算法效率本身对数据质量好坏没有依赖性,为验证算法在不同错误率下异常数据的修复时间,在样本的 24 个时间段中随机生成若干异常数据(每个时间段

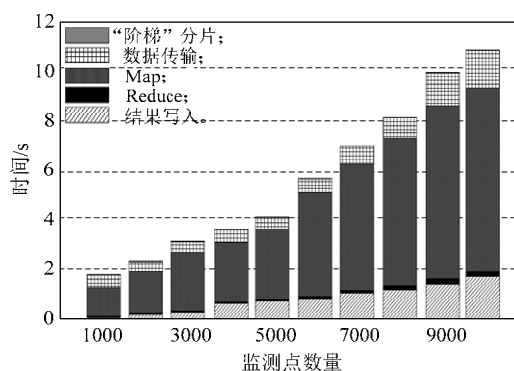


图6 不同规模负荷数据量清洗时间

Fig.6 Cleaning time of load data with different sizes

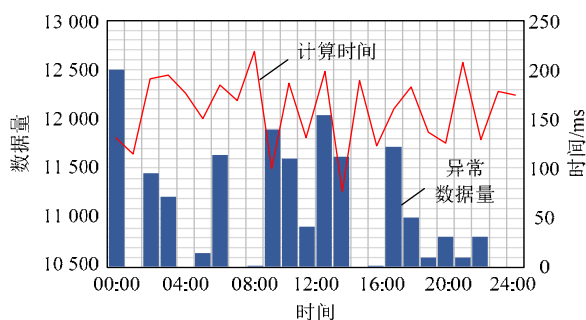


图 7 典型日负荷清洗状况

Fig.7 Typical daily load wash status

小于 200 个),经分别测试表明 20 万条数据的清洗时间在 11272~12679ms 之间,而且不同异常数据量下负荷清洗时间变化较为稳定。

用典型日负荷数据进行数据修复实验,得到协同过滤推荐算法得到的结果范围如图 8 所示,图中虚线为相似负荷时序数据曲线,实线为异常数据曲线。可以看出正确值落在估计范围内。经过对 200 万条负荷数据的测试,数据辨识成功率接近 100%,修复正确率也能达到 90%。

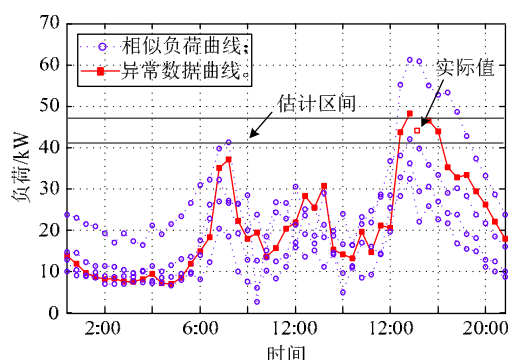


图 8 典型日负荷数据修复

Fig.8 Typical daily load Data Recovery

6 结论

本文提出了大规模配电网负荷数据在线清洗与修复方法,结合当前配电网运行中的实际负荷数据,使用 5 台高性能服务器对文中提出的异常数据在线辨识算法与数据恢复算法进行验证,结果表明:1) 异常负荷数据实时辨识算法能够很好地实现对时序负荷数据的清洗,以无监督的方式实现缺失值、错误值和空值的辨识。2) 基于协同过滤推荐的负荷数据修复算法,根据配电网区域范围内负荷变化的横向关联性计算负荷范围推荐度,实现了异常负荷快速数据修正。

本文提出的在线清洗与修复的分布式并行框架,可有效提高数据清洗修复效率,满足当前和未来一段时间配电网数据预处理的要求,为配电网负荷预测、高级分析与设计规划等提供数据基础,在

地市公司配电网调度、规划和运检等工作中能够发挥实际作用。

参考文献

- [1] 赵腾,张焰,张东霞.智能配电网大数据应用技术与前景分析[J].电网技术,2014,38(12):3305-3312.
Zhao Teng, Zhang Yan, Zhang Dongxia. Application technology of big data in smart distribution grid and its prospect analysis[J]. Power System Technology, 2014, 38(12): 3305-3312(in Chinese).
- [2] 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,2010,35(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, 2010, 35(2): 287-293(in Chinese).
- [3] 张东霞,苗新,刘丽平,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):2-12.
Zhang Dongxia, Miao Xin, Liu Liping, et al. Research on development strategy for smart grid big data[J]. Proceedings of the CSEE, 2015, 35(1): 2-12(in Chinese).
- [4] 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(in Chinese).
- [5] 毛李帆,姚建刚,金永顺.中长期负荷预测的异常数据辨识与缺失数据处理[J].电网技术,2010,34(7):148-153.
Mao Lifan, Yao Jiangang, Jin Yongshun, et al. Abnormal data identification and missing data filling in medium-and long-term load forecasting[J]. Power System Technology, 2010, 34(7): 148-153(in Chinese).
- [6] 张乐.年度负荷预测工作中的数据还原与修正[J].电网技术,2007,31(S1):233-234.
Zhang Le. Data recovery and correction in annual load-forecasting[J]. Power System Technology, 2007, 31(S1): 233-234(in Chinese).
- [7] 王雁平,乐春峡.电力系统负荷建模的数据预处理技术[J].电网技术,2007,31(S2):292-294.
Wang Yanping, Yue Chunxia. A data pretreatment technique about power system load modeling[J]. Power System Technology, 2007, 31(S2): 292-294(in Chinese).
- [8] 毛李帆,江岳春,龙瑞华,等.基于偏最小二乘回归分析的中长期电力负荷预测[J].电网技术,2008,32(19):71-77.
Mao Lifan, Jiang Yuechun, Long Ruihua, et al. Medium and long term load forecasting based on partial least squares regression analysis[J]. Power System Technology, 2008, 32(19): 71-77(in Chinese).
- [9] 张沛,吴潇雨,和敬涵.大数据技术在主动配电网中的应用综述[J].电力建设,2015,36(1):52-59.
Zhang Pei, Wu Xiaoyu, He Jinghan. Review on big data technology applied in active distribution network[J]. Power Construction, 2015, 36(1): 52-59(in Chinese).
- [10] 严英杰,盛戈皞,陈玉峰,等.基于大数据分析的输变电设备状态数据异常检测方法[J].中国电机工程学报,2015,35(1):52-59.
Yan Yingjie, Sheng Gehao, Chen Yufeng, et al. An method for anomaly detection of state information of power equipment based on big data analysis[J]. Proceedings of the CSEE, 2015, 35(1): 52-59(in Chinese).
- [11] 王宁玲,付鹏,陈德刚,等.大数据分析技术在厂级负荷分配中的应用[J].中国电机工程学报,2015,35(1):68-73.

- Wang Ningling ,Fu Peng ,Chen Degang ,et al .Application of big data analytics in plant-level load dispatching of power plant[J] . Proceedings of the CSEE , 2015 , 35(1) : 68-73 (in Chinese) .
- [12] 张智晟,孙雅明,张世英,等.基于数据挖掘多层次细节分解的负荷序列聚类分析[J].电网技术,2006,30(2):51-56.
Zhang Zhisheng, Sun Yaming, Zhang Shiyong, et al . Clustering analysis of electric load series using clustering algorithm of multi-hierarchy and detailed decomposition based on data mining[J] . Power System Technology , 2006 , 30(2) : 51-56(in Chinese) .
- [13] 高静,段会川. JSON 数据传输效率研究[J]. 计算机工程与设计, 2011 , 32(7) : 2267-2270 .
Gao Jing , Duan Huichuan . Research on data transmission efficiency of JSON[J] . Computer Engineering and Design , 2011 , 32(7) : 2267-2270(in Chinese) .
- [14] 严英杰,盛戈峰,陈玉峰,等.基于时间序列分析的输变电设备状态大数据清洗方法[J].电力系统自动化,2015,39(7):138-144.
Yan Yingjie ,Sheng Gehao ,Chen Yufeng ,et al . Cleaning method for big data of power transmission and transformation equipment state based on times sequence analysis[J] . Microcomputer Information , 2015 , 39(7) : 138-144(in Chinese) .
- [15] 顾民,葛良全,熊文贤,等.基于神经网络的电力负荷不良数据的清洗[J].微计算机信息,2005,23(7-3):60-64.
Gu Min , Ge Liangquan , Xiong Wenxian , et al . Cleaning for bad electric load data based on neural network[J] . Microcomputer Information , 2005 , 23(7-3) : 60-64 (in Chinese) .
- [16] 张晓星,程其云,周淦,等.基于数据挖掘的电力负荷脏数据动态智能清洗[J].电力系统自动化,2005,29(8):60-64.
Zhang Xiaoxing ,Cheng Qiyun ,Zhou Quan ,et al .Dynamic intelligent cleaning for dirty electric load data based on data mining[J] . Automation of Electric Power Systems , 2005 , 29(8) : 60-64(in Chinese) .



刁赢龙

收稿日期:2015-09-01。

作者简介:

刁赢龙(1988),男,硕士,工程师,主要研究方向为智能配电网信息化,E-mail: diaoyinglong@126.com;

盛万兴(1965),男,博士,教授级高级工程师,研究方向为电力系统及自动化、可再生能源发电并网技术等;

刘科研(1979),男,博士后,高级工程师,主要研究方向为配电网仿真建模与智能分析;

何开元(1987),男,硕士,工程师,研究方向为人工智能在配电网中的应用;

孟晓丽(1973),女,硕士,高级工程师,研究方向为配电网自动化与信息技术的研究与管理。

(责任编辑李兰欣)

“配用电大数据技术”专题征稿启事

大数据技术是当前国际信息科技前沿热点,已引起世界范围内的广泛关注。目前电力公司在配电系统规划、调度自动化、配电自动化、电能质量监测、电能信息采集、95598 客户服务、电能服务管理等配用电专业信息系统中积累了大量数据资源;营配数据贯通及与电力公司外部的社会公共数据集成,丰富了配用电数据资源;配用电数据呈现数据量大、数据类型多、价值密度低、增长速度快的特点,为开展大数据分析奠定了坚实的数据基础。同时,分布式计算、高性能计算、实时数据库等先进信息技术在电力公司也得到初步应用。为挖掘配用电数据价值,拓展其应用的深度,亟需探索解决大数据技术关键技术。

大数据技术在互联网、科学研究工作中得到了规模应用;国内外学术界和电力工程界对这一问题关注的热度也持续不减;国家 973、863 及电力公司也立项支持大数据基础和应用研究。为促进配用电大数据技术研究,分享国内外该领域最新进展和发展趋势,共享最新学术和技术成果,《电网技术》编辑部特邀国家千人计划专家栾文鹏博士、中国电力科学研究院田世明教授作为特约主编,主持“配用电大数据技术”专题,拟以专辑形式于 2015 年 11 月出版(正刊)。

一、专题征稿范围(包括但不限于)

- 1) 配用电系统大数据技术架构
- 2) 配用电系统营配贯通大数据技术
- 3) 配用电系统大数据集成技术
- 4) 配用电系统大数据存储、查询技术
- 5) 配用电系统大数据数据分析元及可视化分析技术
- 6) 配电系统规划、负荷预测、风险管理、状态估计等方面的大数据应用技术
- 7) 用户用电行为、价值挖掘、客户服务、需求响应等方面的大数据应用技术
- 8) 基于售电量的经济预测技术

二、投稿要求

- 1) 引言中研究目的清晰明确,详细介绍国内外研究背景,对现有其他研究者的工作进行正确的评述;阐述自己的观点,并对自己的研究思路做一总体介绍,引言字数不少于 1000 字。
- 2) 研究设计和方法叙述清楚,数据合理并被正确地分析和解释;比较所提出的方法和现有方法的优缺点。
- 3) 重点突出,论述严谨,文字简练,避免长篇公式推导(必要的推导可列入附录),字数以不超过 6000 字(包括图表)为宜。
- 4) 来稿请用 Word 排版,格式尽量与《电网技术》一致。
- 5) 对英文稿件的要求与对中文稿件的要求相同,请同时提供作者中文信息。

三、投稿截止日期:2015 年 8 月 31 日。

四、投稿方式:请登录 <http://www.dwjs.com.cn> 注册用户名和密码投稿,投稿栏目选择“配用电大数据技术”。

真诚欢迎国内外科研教育、电力企业及产业公司相关领域的专家学者踊跃投稿!

联系人:田世明李兰欣

联系电话:010-82812543

邮箱:laotian@epri.sgcc.com.cn; lilanxin@epri.sgcc.com.cn