

电能质量监测数据清洗技术研究

张华赢¹, 阳子婧², 段绍辉¹, 姚森敬¹, 曹军威², 许延详²

(1 深圳供电局有限公司, 广东 深圳 518000; 2 清华大学信息技术研究院, 清华大学信息科学与技术国家实验室, 北京 清华大学, 100084)

摘要: 随着电能质量问题得到高科技企业和电网公司越来越多的重视, 大量电能质量监测系统得到部署和应用。然而, 基于监测数据的高级分析并未得到广泛应用, 其重要原因在于电能质量监测数据没有得到有效的梳理。本文提出数据清洗 (Data Cleaning) 技术是使电能质量高级分析的关键和前提, 描述了电能质量监测数据清洗的具体步骤和软件实现。本文以深圳电网电能质量监测事件的具体数据为例, 说明了数据清洗技术应用于电能质量数据分析中的效果, 指出清洗后的电能质量监测数据避免了无效和缺失等问题, 在反应实际情况中更具有实用性, 为进一步的电能质量高级分析奠定了基础。

Data Cleaning for Power Quality Monitoring

Huaying Zhang¹, Zijing Yang², Shaohui Duan¹, Senjing Yao¹, Junwei Cao², Yanxiang Xu²

(1 Shenzhen Power Supply CO., LTD, Shenzhen 518000, China, 2 Research Institute of Information Technology, State Key Laboratory of Information and Technology of Tsinghua University, Tsinghua University, Beijing 100084, China)

Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China

Abstract: Power quality issues are becoming more critical for high-tech enterprises and grid companies. Many power quality monitoring systems are deployed in recent years. Advanced analysis of monitoring data is not widely applied due to the lackness of data management. In this work, data cleaning technology is introduced to enable advanced study of power quality data, with detailed procedures and software implementation. With power quality monitoring data from Shenzhen Power Company, the effectiveness of data cleaning technology applied for power quality data analysis is demonstrated. Cleaned data that avoid voidness and lackness is more feasible in actual usage, as a good basis for further advanced analysis of power quality.

一、介绍

随着计算机、信息设备、精密仪器、高端制造业等对电源质量敏感设备的应用, 电能质量问题受到用户的广泛关注。对敏感用户 (如半导体制造企业) 而言, 几十毫秒的电压暂降就可能造成设备损坏、生产线停产, 造成巨大经济损失。例如作为全国第四个负荷过千万的

城市，深圳市的用电量甚至可与一个普通省份用电量相比较，其中大部分为高新技术企业。近年来，各类微电子、半导体、生物医药、精密制造等企业，以及医院、金融业、通信行业、大型数据中心等敏感用户对电网的供电电能质量提出了更高要求。

目前国内有部分省网已陆续开始了电能质量监测系统的建设，有的已初具规模。建设模式主要分为引进成熟系统和自主研发两类。上海、华北、云南等电力公司采用了美国电科院成熟的 PQView3.2 平台并进行二次开发，已相继建立了网络化的电能质量监测平台；广东、浙江、江苏电力公司采用自主研发的方式，开发了区域电能质量监测网。深圳电网现有电能质量监测体系已经初具规模，目前已在主网 146 个变电站安装了电能质量监测终端 651 台，电能质量监测系统具备的主要功能为巡检、远程操作、数据查询分析、报表统计等，缺乏基于数据挖掘功能支撑的高级分析应用。

为更好地支持高级分析应用，需要对电能质量监测数据进行整理和清洗，主要基于以下几方面的原因：电能质量监测系统的开发人员不熟悉电能质量相关标准，将本不属于电能质量问题的事件进行了记录；电能质量监测系统的开发人员对输配电系统的运行原理缺乏了解，对关联事件进行了重复记录；由于通信等方面的系统错误，记录的部分数据可能存在不合理性或非一致性，需要剔除。

本文的主要贡献在于将数据清洗技术应用于电能质量监测数据的梳理。以深圳电网电能质量监测事件的具体数据为例，说明数据清洗技术应用于电能质量数据分析中的效果。

二、数据清洗技术

电能质量监测数据高级分析的前提和基础是数据的正确性、一致性和有效性，这需要通过数据清洗 (Data Cleaning) 技术来实现。数据清洗是指将不一致的或错误的从记录集、表格或数据库中检测出并进行修正的过程。由于用户的误输入、传输或存储过程中的出错等原因，数据集中可能存在非完整的、错误的或不相关的信息。因此，需要通过数据清洗，来有效识别、替换、修正或删除这些“脏数据”，使该数据集与系统中其他的相似的数据集保持一致。

国外对数据清洗的研究最早出现在美国，是从对全美的社会保险号的纠错开始^[1]。随后美国信息业和商业的发展，有力推动了数据清洗技术的研究，并主要集中在检测及消除数据异常、检测及消除近似重复记录、数据集成和特定领域的的数据清洗四个方面。在国内，对数据清洗的研究起步较晚，目前主要应用于数据仓库、决策支持、数据挖掘和全面数据质量管理等领域，但对于商业性的数据清洗工作则主要是针对各自的具体应用，理论性不强^[2-4]。而如何结合实际应用，运用有效的模型和方法对脏数据如不完整数据、错误数据和重复数据进行数据清洗，从而提高数据质量，仍是有待深入研究的难点问题^[5]。

三、电能质量监测数据高级分析

电能质量监测数据为深入研究实际运行电网的电能质量问题提供了基础，当前已经存在

大量面向该类数据的分析和研究，按研究手段和应用目的可将这些研究分为两大类：一类是以数据统计为手段，以电能质量监测优化、评估和运行管理为目的的基础分析^[6-8]；另一类是以数据挖掘为手段，以从大规模、高维的电能质量监测数据中提取出隐藏的模式和规则，为电力系统规划和决策提供支持依据的高级分析^[9-12]。其中，电能质量监测数据高级分析的主要过程如图 1 所示。从图中可发现：数据清洗是进行电能质量数据高级分析的前提和基础。

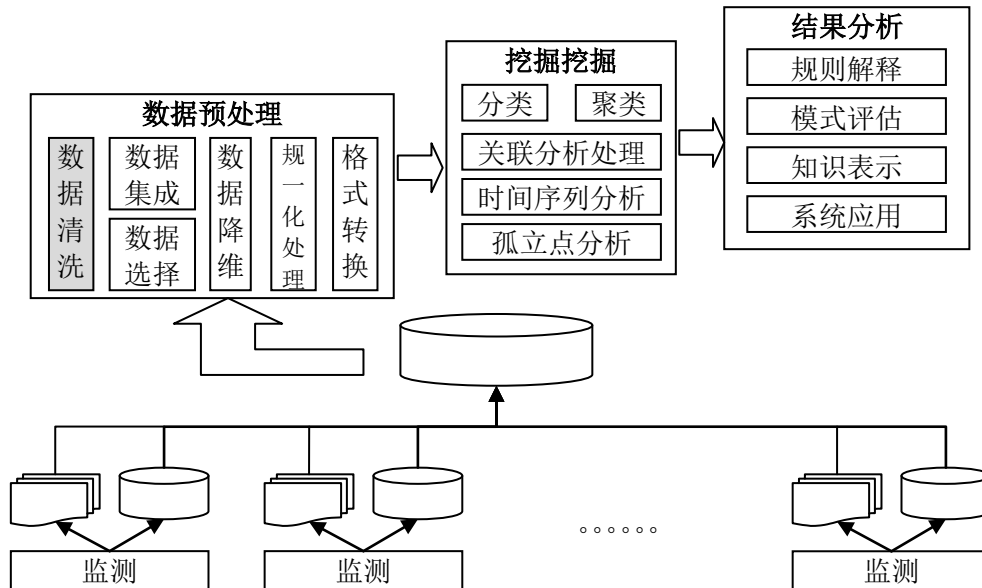


图 1 电能质量数据高级分析流程图

四、电能质量监测数据清洗

为说明数据清洗技术在电能质量数据分析中的重要性，本文重点以电能质量监测系统所记录的暂态事件的对应数据为例，对其进行具体的梳理和清洗。

4.1 具体步骤

针对电能质量问题，对采集到的监测数据进行清洗，可通过以下三个主要步骤来实现：

(1) 数据剔除：即将不符合电能质量暂态事件定义的数据予以剔除。根据由 IEEE SCC 22 提出并已被 IEC 采纳的电能质量的分类标准，将持续时间在半个周波到一分钟以内的电压变化（含电压暂升、电压暂降和短时电压中断）定义为电能质量暂态事件。因此，对持续时间在半个周波到一分钟范围以外的监测数据予以剔除。

(2) 事件统计：即对电能质量暂态事件的发生次数进行准确统计。目前，输、配电是通过三相交流电即由三个频率相同、电势振幅相等、相位差互差 120° 角的交流电路组成的电力系统来实现的。在统计过程中，为保证统计结果的准确性，将同一变电站的同一母线下在同一时刻发生的两相或三相的暂态事件均按照分析目标（即电压暂升、电压暂降或短时电压中断）记为零次或一次。

(3) 数据再处理：即对经过以上数据剔除和事件统计两个步骤的数据作进一步的处理。在预期分析目标的要求下，部分数据可能存在不合理性或非一致性。因此，需根据相关信息如事件的关联分析、用户需求等对监测数据进行再处理。

4.2 软件实现

数据清洗软件是实现行业中数据进行高效预处理的有力工具。近年来，市场上的数据清洗软件，既有商业开发的，也有由大学和科研机构开发的。而针对电能质量监测数据的相应软件的研发，目前还处于初步探索阶段。本研究开发了电能质量暂态事件分析软件（见图 2），数据清洗是其中的重要组成部分。

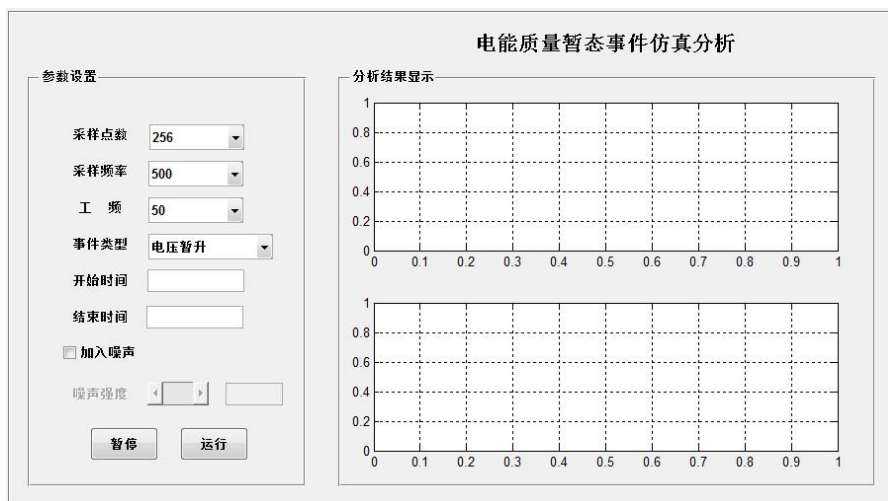


图 2 电能质量暂态事件分析软件实现

软件实现是基于美国 MathWorks 公司出品的 MATLAB 平台而开发。利用 MATLAB 的计算功能和友好的人机交互界面，软件实现包括参数设置、数据清洗、小波分析、噪声分析等多个模块。

五、实验验证

本节采用深圳电网某中心站 2010 年至 2012 年电能质量暂态事件的详细记录，作出其 ITI (CBEMA) 幅值—时间分布图，并对比数据清洗前后的效果，分别如图 3、图 4 所示。其中，CBEMA 曲线是由美国计算机和商用设备制造商协会 (CBEMA) 对电能质量的要求提出的电压容限曲线。ITI 曲线是美国信息技术工业协会 (ITIC) 在 CBEMA 曲线的基础上发展起来，根据计算机等信息工业设备对暂态电能质量的抗扰度水平形成的。该曲线是目前评估暂态电能质量事件影响的一个重要依据，被 IEEE 引用为美国标准^[13]。

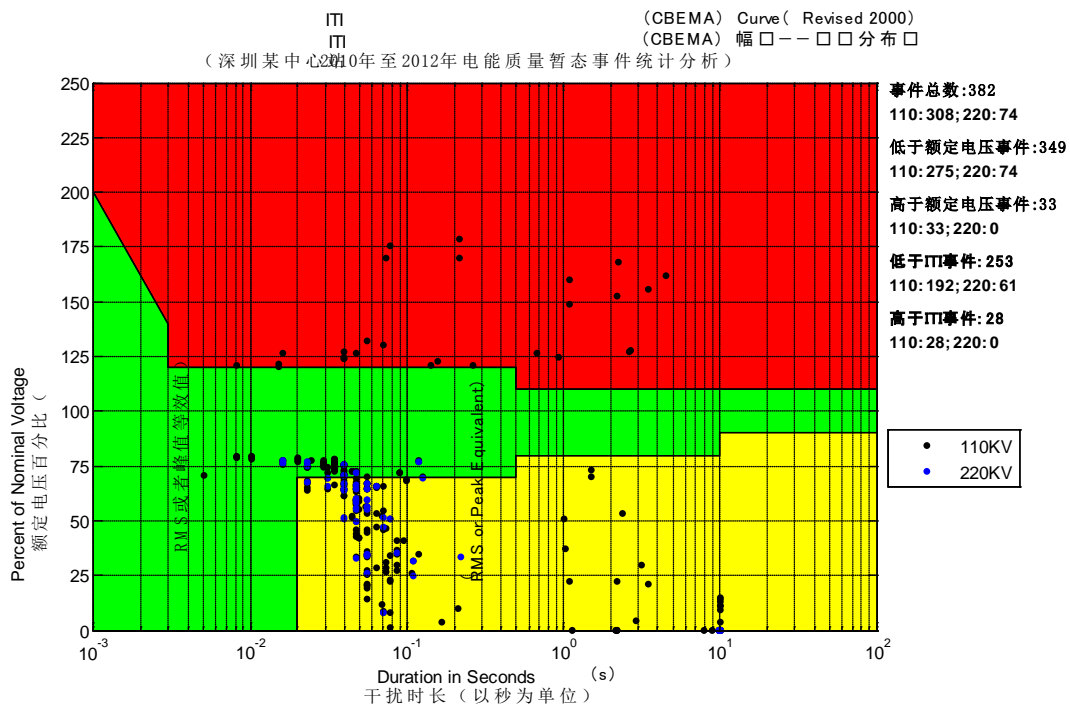


图3 电能质量监测数据清洗之前的统计分析结果

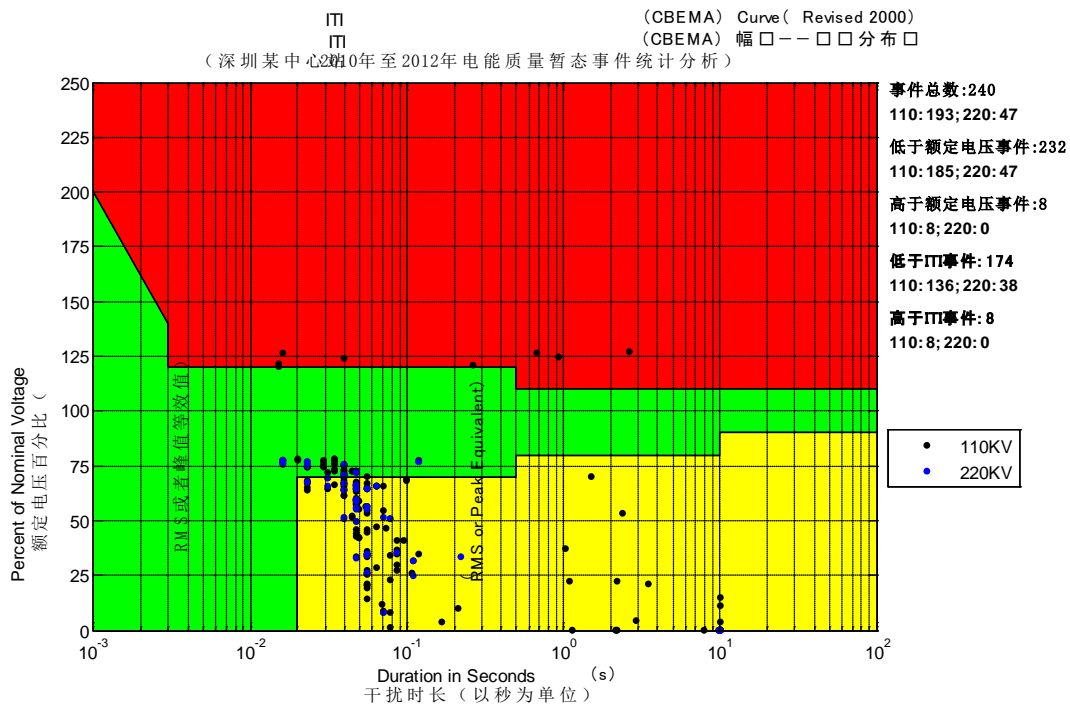


图4 电能质量监测数据清洗之后的统计分析结果

六、结论

本文主要工作在于将数据清洗技术应用于电能质量的监测数据,为未来进一步的高级分析应用奠定了基础。电能质量监测系统产生的海量数据因多种原因在正确性和一致性等方面存在种种问题,甚至可能产生误导。若不能对电能质量监测数据进行有效的整理和剔除,就无法准确反映实际电能质量问题的严重性。

本工作结合深圳电网的实际情况,汇集了电能质量监测系统的相关数据,对基础数据进行了深入研究,提出数据清洗的必要性,并通过实际案例给出了清洗前后的效果对比。深圳作为高科技产业密集的城市,对电能质量的要求非常高,本工作的研究成果对深圳电能质量问题的进一步分析和治理具有重要的指导意义。

致谢

国家 973 基础研究计划(2013CB228206); 国家自然科学基金(61233016)。

参考文献:

- [1] Galhardas H, Florescu D. An Extensible Framework for Data Cleaning [C]. In: Proceedings of the 16th IEEE International Conference on Data Engineering. San Diego, California, 2000: 312 - 312.
- [2] 王曰芬, 章成志, 张蓓蓓, 吴婷婷. 数据清洗研究综述[J]. 情报分析与研究. 2007, 12: 50-56.
- [3] 曹建军, 刁兴春, 汪挺, 王芳潇. 领域无关数据清洗研究综述[J]. 计算机科学. 2010, 37 (5): 26-29.
- [4] 叶鸥, 张璟, 李军怀. 中文数据清洗研究综述[J]. 计算机工程与应用. 2012, 48 (14): 121-129.
- [5] 蒋勇青, 杨奕虹, 杨贺. 论数据清洗对信息检索质量的影响及清洗方法[J]. 中国索引. 2012, 1: 16-20.
- [6] 肖湘宁, 韩民晓, 徐永海等. 电能质量分析与控制[M]. 北京: 中国电力出版社. 2004.
- [7] 唐会智, 彭建春. 基于模糊理论的电能质量综合量化指标研究[J]. 电网技术. 2003, 27(12): 85-88.
- [8] 江辉, 彭建春, 欧亚平等. 基于概率统计和矢量代数的电能质量归一量化与评价[J]. 湖南大学学报(自然科学版). 2003, 30(1): 66-70.
- [9] 欧阳森, 宋政湘, 王建华等. 基于信号相关性和小波方法的电能质量去噪算法[J]. 电工技术学报. 2003, 18(3): 111-116.

- [10] 黄文清, 戴瑜兴。电能质量扰动的 Block-Thresholding 去噪方法[J]。电工技术学报。2007, 22(10): 160-166。
- [11] 刘志刚, 曾怡达, 钱清泉。多小波在电力系统信号消噪中的应用[J]。中国电机工程学报, 2004, 24(1): 30-34。
- [12] 唐良瑞, 杨雪。基于三角模融合算子的电能质量去噪算法[J]。电工技术学报。2007, 22(9): 154-158。
- [13] IEEE Application Guide for IEEE Std 1547™, IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems[S]. 2009.

收稿日期: 2013-06-30

阳子婧 (1984), 女, 湖南湘潭人, 博士后, 从事信号处理、数据挖掘技术及其在电力系统中的应用研究。(Tel): 010-62797752。(e-mail): zijing396@sina.com。

曹军威 (1973), 男, 河北乐亭人, 博士, 研究员, 从事分布式计算技术及其在能源电力行业的应用研究。

许延祥 (1975), 男, 河北抚宁人, 博士后, 从事数据挖掘技术及其在电能质量领域应用研究。