
分类号 _____
UDC _____

密级 _____
编号 _____

清华大学

博士后研究报告

电能质量综合评估技术研究

高田

清华大学 (北京)

2014 年 09 月

电能质量综合评估技术研究

RESEARCH ON ASSESSMENT TECHNOLOGIES OF POWER QUALITY

博 士 后 姓 名: 高田

流动站（一级学科）名称: 控制科学与工程

专业（二级学科）名称: 控制理论与控制工程

合 作 导 师: 曹军威（研究员）

研究工作起止日期 2012 年 11 月 - 2014 年 09 月

摘要

随着电力行业信息化的快速发展,特别是以物联网和云计算为代表的信息通信技术 (ICT) 在智能化电网中的广泛应用, 电力数据的信息量呈爆发式的增长, 形成了贯穿输电、配电、管理等各个环节的电力大数据。同时, 伴随着计算机、电力电子设备等敏感负荷在各个行业广泛应用, 导致电力用户对电压暂降、电压骤升、短时中断、等暂态电能质量扰动非常敏感, 因此电能质量状态评估对于供用电双方非常重要。本报告在电能质量综合评估的主要工作包括:

(1) 干扰源与敏感用户分类供电方案: 主要从电力用户的负荷特性进行分析, 提出了干扰源与敏感用户分类的方案, 包括负荷敏感度、污染度的分析和评估, 并提出了最终在电力企业内部落地实施的具体方案。在此基础上, 提出了干扰源与敏感用户分类供电方案, 减少干扰源用户对电网的污染, 并提高敏感用户的电能质量。

(2) 从电能质量到体验质量。提出了电能体验质量的概念和评估流程, 并对电能体验质量的合理性进行了实例验证。现有电能质量标准体系主要以供电质量为基础, 通过电网电压、电流稳态和暂态等方面的量化指标进行描述, 而实际工作中电能质量的好坏更多取决于用户的敏感程度和实际体验。本报告提出电能体验的概念, 侧重用户敏感度的定量分析, 并以此为基础综合现有电能质量标准和相关用户信息, 提出电能体验质量的指标体系。结合国内某电网的具体案例, 说明电能体验质量在实际情况中更能综合供电质量和用户体验等多方面因素, 对电网实际工作更加具有指导意义。

(3) 基于分类的电能质量综合评估。传统的暂态电能质量评估研究的焦点是科学、客观地将多指标问题归并成单一量化的指标, 使用加权系数反映各指标的相对重要程度, 但加权系数受限于主观赋值和客观赋值的局限性。本课题提出了基于分类的电能质量综合评估方法, 将评估的数据来源扩展至电网运行监测数据、电力用户数据和公

共信息数据等方面，将综合评估的结果按严重程度分为正常状态、异常状态、严重状态、故障状态。本研究设计了用于电能质量综合评估的数据样本模型，用于构建分类器学习的训练集和测试集，并分别使用支撑向量机、决策树、贝叶斯分类方法进行电能质量状态评估。

(4)分布式状态评估方法。在此基础上，将分类器的训练和测试过程推广至MapReduce框架下。本研究采用的大数据处理架构具备分布式处理能力和流数据处理能力，既保证分类器中规则库的定期更新，又能保证状态评估的实时性。在分类器训练阶段，对海量历史数据进行分布式学习，周期性地生成评估规则库并部署到所有评估节点。在状态评估阶段，各评估节点基于流处理框架快速生成实时评估样本，并根据当前规则库实时地得出评估结果。在此基础上，设计了基于MapReduce处理的Naive Bayes分类方法，在分类器训练阶段实现了先验概率 $P(C_i)$ 、条件概率 $P(X|C_i)$ 的分布式计算的分布式计算。

关键词：电能质量；电能体验质量；干扰源与敏感用户分类；电能质量综合评估；分类算法；大数据分析；分布式贝叶斯

ABSTRACT

With the development of information and communication technology (ICT) in power industry, especially the sensor networking and cloud computing which are widely used in smart grid, the amount of power information is experiencing an explosive growth. The concept of big power data appears through all the aspects of the power industry, such as power transmission, power distribution, and power management. Meanwhile, the computer, electronics equipment and other sensitive loads are widely used in various industries, leading to high sensitivity of power customers in the condition of transient power quality disturbances, such as voltage sags, voltage swells, and short-time interruptions. The main research work and innovations are as follows:

(1) Classification of sensitive power user and disturbance power user. This research analyzes the load features of power users, and proposes the classification method to identify the sensitive user and disturbance user, which is based on the sensitivity degree and pollution degree of the load. It also proposes a specific solution for the implementation of Power Company. On this basis, the different scheme of power supply is proposed, which reduces the pollution from disturbance power user, and improves power quality of sensitive power user.

(2) From power quality to experience quality. Existing power quality standards are described by stable or transient states of current and voltage, from a perspective of power supply. In practice, it is up to customers' sensitivity and actual experiences for evaluation of power quality. In this work, power experience is proposed as a composite criterion of both existing power quality standard and related customers'

information. From a perspective of customers' experience, the concept, features, and standards are proposed for power quality of experience. A case study from a certain power company of China is given, demonstrating that power quality of experience is more feasible in actual work for power quality evaluation.

(3) The comprehensive assessment of power quality based on the classification algorithm. The traditional research on transient power quality assessment focus on how summarize the multiparameter indicators into a quantitative indicator, which use weighting coefficients to reflect the relative importance of each indicator. A transient power quality assessment method is proposed in this paper, using Naive Bayes classification method which is based on big data processing architecture. In this research, data sources will be extended to the aspects of power grid monitoring data, the power customer data and the public data, and the assessment severity will be classified into the normal state, the abnormal state, the critical state, and the failed state. In this reseach, we designed a sample model for power quality assessment, and generated the training set and testing set for classifier learning. Then, the SVM, DT and NV classifier are used to assess the power quality.

(4) Distributed assessment of power quality. This research also extends the method of classifier training and testing into MapReduce framework. Based on the data type of power quality assessment, big data processing architecture used in this paper can be able to process distributed data and streaming data, so that it can ensure not only updates classifier rules regularly, but also the real-time condition assessment. In the classifier training phase, we use the massive historical data as the distributed learning object, and generate assessment rules periodically. In the state assessment phase, each assessment node will update the assessment rules generated by training

phase, generate real-time evaluation of samples from stream processing framework, and evaluate the power quality state according to the current rule. On this basis, this paper designs a Naive Bayes classification method based on MapReduce processing, and realizes the *map* and *reduce* process method to compute the priori probability $P(C_i)$ and the conditional probability $P(X|C_i)$ in distributed way. Experiments show that the transient power quality evaluation method based on the big data analysis presented in this paper is feasible, and achieve good results both in classification accuracy and processing speed.

.
KEYWORDS: Power Quality; Power Experience Quality; Data mining; Comprehensive Assessment of Power Quality; Classification algorithm; Big Data; Distributed Bayes Classification.

目录

| | | |
|-------|-------------------------|----|
| 1 | 绪论 | 1 |
| 1.1 | 课题来源 | 1 |
| 1.2 | 研究背景 | 1 |
| 1.3 | 电能质量综合评估意义 | 2 |
| 1.4 | 报告的组织结构 | 3 |
| 2 | 电能质量评估体系 | 5 |
| 2.1 | 电能质量定义 | 5 |
| 2.2 | 电能质量扰动分类 | 7 |
| 2.3 | 电能质量评估研究 | 11 |
| 2.3.1 | 电能质量单项评估 | 11 |
| 2.3.2 | 电能质量综合评估 | 13 |
| 2.4 | 国内某供电公司电能质量现状 | 16 |
| 3 | 电能质量综合评估数据来源 | 19 |
| 3.1 | 电能质量信息化现状 | 19 |
| 3.2 | 电能质量监测装置 | 20 |
| 3.3 | 电能质量监测数据处理 | 23 |
| 3.3.1 | 电能质量在线监测平台 | 23 |
| 3.3.2 | 电能质量监测数据格式 | 26 |
| 3.3.3 | PQDIF 数据格式 | 27 |
| 3.4 | 国内某供电公司电能质量在线监测平台 | 31 |
| 4 | 干扰源与敏感用户分类供电方案 | 38 |
| 4.1 | 术语和定义 | 38 |
| 4.2 | 干扰源与敏感用户分类方案 | 38 |
| 4.2.1 | 用户负荷敏感特性分析 | 38 |
| 4.2.2 | 用户负荷敏感度评估 | 40 |
| 4.2.3 | 用户负荷敏感度实测 | 42 |
| 4.2.4 | 用户负荷污染性评估 | 46 |
| 4.3 | 干扰源与敏感用户分类实施和算例 | 47 |

| | |
|--------------------------------|-----|
| 4.3.1 干扰源与敏感用户分类实施方案 | 48 |
| 4.3.2 干扰源与敏感用户分类算例 | 52 |
| 4.4 干扰源与敏感用户供电方案 | 53 |
| 4.5 从电能质量到电能体验质量 | 57 |
| 4.5.1 电能体验质量的提出 | 57 |
| 4.5.2 电能体验质量评估流程 | 58 |
| 4.5.3 电能体验质量实例验证 | 60 |
| 5 基于分类算法的电能质量状态评估 | 62 |
| 5.1 数据挖掘基本任务 | 62 |
| 5.2 分类算法研究综述 | 65 |
| 5.2.1 SVM 算法综述 | 65 |
| 5.2.2 贝叶斯分类 | 70 |
| 5.2.3 决策树分类 | 71 |
| 5.3 电能质量综合评估建模 | 74 |
| 5.4 基于分类算法的电能质量评估实例 | 76 |
| 6 基于 Hadoop 的分布式电能质量评估 | 80 |
| 6.1 大数据分析 | 80 |
| 6.1.1 Hadoop 分布式计算 | 80 |
| 6.1.2 流数据的处理和框架 | 83 |
| 6.1.3 基于 Mahout 框架的分布式分类 | 89 |
| 6.2 基于分布式分类的电能质量综合评估框架 | 92 |
| 6.3 基于分布式贝叶斯的电能质量综合评估算法 | 94 |
| 6.3.1 分布式评估模型 | 94 |
| 6.3.2 分布式贝叶斯算法模型 | 95 |
| 6.3.3 分布式贝叶斯分类算法 | 97 |
| 参考文献 | 100 |
| 博士后在站期间取得的研究成果 | 109 |
| 致谢 | 110 |

1 绪论

1.1 课题来源

本报告的研究主要源于以下课题：深圳供电公司项目《深圳电网电能质量综合治理研究与工程示范》。江苏省电力公司项目《电能质量高级分析关键技术研究及应用》。澳门能源办项目《澳门电网规划方案论证及配电网稳定性研究》。

1.2 研究背景

近年来，随着现代工业的工艺制造水平发展和设备精度要求、高科技需求的增加，生产力和产品竞争力的提高越来越依赖高质量的电力供应。但是，目前的城市电网中同时存在着非线性、冲击性和不对称性等大型工业负荷，特别是冶金、化工、半导体加工等领域，大容量电力整流设备广泛使用，使电网电能质量问题日益突出，已严重危害系统和其他用户设备安全经济运行。对于计算机、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，电能质量扰动问题特别是谐波和电压暂降的频繁发生越来越严重地影响这些敏感用户的经济利益。谐波和电压暂降扰动的影响主要体现在以下方面：(1)谐波：当衡量谐波电流引起同步电机的附加损耗和发热时，往往折算成等效的基波负序电流来考虑。(2)电压暂降：当电压暂降的电压幅值和持续时间超过设备的敏感曲线时，虽然并未出现电气连接的断开，但可能对用户造成比断电还严重的后果。

随着电力系统信息化的发展，电力数据的来源涉及电力生产和电能使用的发、输、变、配、用及调度各个环节。这些数据涵盖了结构化、非结构化数据，具有多源、海量、异构、时空连续等特点，从而使得基于大数据技术进行电能质量分析评估成为可能。目前，虽然已有一部分大数据分析在电力行业应用的研究与尝试，但在数据的全面性、实时性、数据一致性、隐私性和安全性方面仍有不少的挑战，同时在数据广度和应用深度上还有不足，需要找出相应的解决方法，有待去研究和解决。

目前已有很多国际标准和国家标准规定了电能质量的单项指标及其评估标准。但是从实际物理过程来看,所有电能质量的基本特征量是共同作用在同一电气设备上的,设备的工作状态和性能是由该时刻的这些电能质量多项指标共同决定的。由此可见,虽然电能质量特征是多元化的,但是其对设备性能的影响结果是需要综合多项特征指标来评估的。但是,针对电能质量综合评估的研究难度较大。其原因在于,当电能质量的多个指标共同作用到一个系统时,不同等级的各项评估指标的组合对电力系统的影响,以及对系统中运行设备的影响甚至是损坏差别较大。不同的用电设备对电能质量各评估指标的重视程度也不相同,所以实现对电能质量的综合评估是非常复杂和困难的。因此,对于电能质量的综合评估除了考虑各单项指标外,还需要考虑用户负荷的实际情况。电能质量扰动对于用户设备和负荷的影响决定了综合评估的结果。

电能质量综合评估的研究焦点是如何科学、客观地将一个多指标问题综合成单一量化的指标问题,从而评定电能质量的等级。综合评估建立在单项评估基础上,综合评估方法把部分或全部电能质量扰动或某项电能质量扰动的多个特征指标按属性量化后,采用归一、综合的算法得到一个综合指标或综合等级。这种综合评估方法的主要问题在于各电能质量指标具有明显的不相关性,从而导致它们相应的加权系数难以确定。目前,大量学者分别基于层次分析法(Alytic Hierarchy Process, AHP)、模糊综合评判法(Fuzzy Comprehensive Evaluation Method, FCEM)、机器学习法(Machine Learning Method, MLM)、人工神经网络评价法(Artificial Neural Network Method, ANNM)以及灰色综合评价法(Grey Comprehensive Evaluation Method, GCEM)等方法对电能质量综合评估问题进行了研究,取得了较好的效果。

1.3 电能质量综合评估意义

电能质量扰动对于电力系统、用电企业和普通居民都造成了严重影响:对于电力系统来说,电能质量扰动会降低变压器寿命、使继电保护装置误动作、增加线损,干扰电力系统的正常通信,直接影响智能电网的广域测量、电能计量。对

于用电企业来说，由于大量精密仪器和电力电子设备的使用，用户对于各种电磁干扰都极为敏感，轻微的电压暂降都有可能影响其电子控制系统的正常工作，甚至导致生产线停止、废品率增加等严重后果，直接影响到企业的经济效益；对于居民用户来说，电能质量污染会降低电器的使用寿命、使保护熔丝经常熔断，影响居民的日常用电质量。

因此，进行电能质量治理，提高电能质量已经成为供用电双方的共同愿望，而进行电能质量评估是电能质量治理的前提条件。电能质量综合评估结果可以作为供用电双方制定供电合同的依据，有助于供用电双方明确电能质量责任，并可作为电能商品分质计价的重要参考。同时，电能质量评估是进行电能质量治理的前提，是对电网供电点电能质量的量化指标，对于干扰源与敏感用户的供电方案制定、电源点的选择有重要意义。

1.4 报告的组织结构

本报告的电能质量综合评估主要是利用数据挖掘和机器学习技术，对电能质量问题进行高级分析，综合用户特性、电力运行情况、公共信息等方面的数据，实现对电能质量的综合评估。从机器学习的角度，分类算法是一种有监督的学习方法，其目标是将未知样本归类到不同的分组中，通过学习各种统计数据中类别和样本之间的关系创建模型（通常称为分类器），用于归类未标识的样本。因此，通过学习历史电能质量数据训练分类器，能够根据电能质量监测系统的数据较好地预测电能质量状态，从而成为有效的电能质量评估方法。报告的组织结构如下：

第 1 章是绪论，包括课题来源、研究背景、研究意义，并介绍了整个报告的组织机构。

第 2 章是电能质量评估体系，介绍电能质量的定义，对电能质量扰动的类型进行了分类，并介绍现有的电能质量评估的单项指标，综述了现有电能质量综合评估技术。最后，结合项目情况，介绍了国内供电公司的电能质量现状。

第 3 章是电能质量综合评估数据来源, 分别从电能质量的信息化现状、在线监测平台、监测装置、数据格式等方面进行阐述, 并介绍了国内某供电公司的电能质量在线监测平台的功能和实际使用情况。

第 4 章是干扰源与敏感用户分类供电方案。本章主要从电力用户的角度分析, 提出了干扰源与敏感用户分类的方案, 包括负荷敏感度、污染度的分析和评估, 并提出了最终在电力企业内部落地实施的具体方案。在此基础上, 提出了干扰源与敏感用户分类供电方案, 减少干扰源用户对电网的污染, 并提高敏感用户的电能质量。本章还提出了电能体验质量的概念和评估流程, 并对电能体验质量的合理性进行了实例验证。

第 5 章是基于分类算法的电能质量状态评估, 将评估的数据来源扩展至电网运行监测数据、电力用户数据和公共信息数据等方面, 将综合评估的结果按严重程度分为正常状态、异常状态、严重状态、故障状态。提出了用于电能质量综合评估的数据样本模型, 用于构建分类器学习的训练集和测试集, 并分别使用支撑向量机、决策树、贝叶斯分类方法进行电能质量状态评估。

第 6 章将为了更好地适应电能质量评估的实际情况, 设计了基于 MapReduce 的分布式贝叶斯分类, 可以分布式地训练、测试分类器, 并近乎实时地在分布式评估节点上得到评估结果, 实现了计算资源和存储资源的统一。

2 电能质量评估体系

随着现代工业的工艺制造水平发展和设备精度要求、高科技需求的增加，生产力和产品竞争力的提高越来越依赖高质量的电力供应。近来，大型城市电网中具有非线性、冲击性和不对称性的大型工业负荷不断增加，特别是冶金、化工、半导体加工等领域，大容量电力整流设备广泛使用，使电网电能质量问题日益突出，已严重危害系统和其他用户设备安全经济运行^[1-3]。同时，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。

2.1 电能质量定义

电能质量是指导致用电设备故障或不能正常工作（或运行）的电压、电流或频率的偏差，通常包括电压偏差、电压波动与闪变、电压暂降与短时中断、三相不平衡、暂时或瞬时过电压、波形畸变、频率偏差、长时间电压中断等。目前，国际电工委员会（International Electrotechnical Commission， IEC）IEC61000 系列电磁兼容标准和国际电气与电子工程师协会（Institute of Electrical and Electronics Engineers， IEEE）IEEE1159 就现代电能质量的概念定义、指标推荐和测试方法等方面逐渐趋于一致，IEEE 已正式采用“Power Quality”作为电能质量的术语，并采用电压幅值、电压波形、频率偏差多种指标来衡量电能质量。

自 20 世纪 80 年代以来，电力用户对电能质量的要求越来越高，事实上已经要求供电方能够为敏感设备提供合适的电力设置和接地系统，因此，提高电能质量已经成为供用电双方的共同愿望。对于电力系统来说，电能质量扰动会降低变压器寿命、使继电保护装置误动作、增加线损，干扰电力系统的正常通信，直接影响智能电网的广域测量、电能计量。对于用电企业来说，由于大量精密仪器和电力电子设备的使用，用户对于各种电磁干扰都极为敏感，轻微的电压暂降都有可能影响其电子控制系统的正常工作，甚至导致生产线停止、废品率增加等严重后果，直接影响到企业的经济效益；对于居民用户来说，电能质量污染会降低电器的使用寿命、使保护熔丝经常熔断，影响居民的日常生活质量^[4]。

电力系统应向用户供应质量（电压和频率）合乎规定的电能。电能质量问题在发达国家早已引起高度重视，在美国、日本、加拿大、以色列、欧洲等国家，

各类控制器，包括 LC 滤波器、APF、DVR、DSTATCOM 和 UPQC 均有大量应用，应用行业扩展到了食品加工、造纸、精密加工、半导体生产、医疗和公共配电网中等。针对电能质量问题，国际上都制定了相应的电能质量标准以保证电网安全经济运行、保护电气环境、保障电力用户正常使用^{[5][6]}。目前，国际上关于电能质量的标准主要由 IEC 标准，欧洲电工技术标准制定机构 CENELEC 标准，以及美国 IEEE 标准，分别根据各国电力系统实际情况，规定了电能质量的含义、分类、限值、指标的测量和计算方法等技术性规范。其中，IEEE519 规定了电能系统谐波控制规范和要求，IEEE1159 给出了电能质量监测推荐规范；IEC 把电能质量归为电磁兼容范畴，制订了 IEC61000 系列标准，在电能质量测量方法、谐波与间谐波测量、闪变仪功能与设计等方面进行了规定^[7]。CENELEC 发布了《公用配电系统供电电压特性》标准，作为欧洲关于配电网电能质量的统一标准。以电压偏差为例，表 2-1 列举了一些国家和地区对供电电压允许偏差的规定。

表 2-1 一些国家和地区对供电电压允许偏差的规定

| 国家 | 供电电压允许偏差范围 (%) | | | |
|-----|----------------|----------|----------|--------|
| | 照明用电 | 动力用电 | 农村用电 | 事故后运行 |
| 美国 | ±5 | ±10 | 无规定 | 各类增加 5 |
| 前苏联 | +5; - 2.5 | +10; - 5 | ±5 | |
| 德国 | ±3 | ±5 | ±10 | |
| 日本 | ±6 | ±10 | +9; - 7 | |
| 英国 | ±6 | ±12 | 无规定 | |
| 加拿大 | +10; - 3 | +10; - 3 | +10; - 3 | +10 |

随着电能质量问题的日渐严重，我国制定并颁布了一系列标准。其中，《电能质量电压暂降与短时中断》、《电能质量供电电压偏差》、《电能质量电力系统频率偏差》、《电能质量三相电压不平衡度》、《电能质量电压波动和闪变》、《电能质量公用电网谐波》、《电能质量公用电网间谐波》、《暂时过电压和瞬态过电压》等规定了电能质量相关的术语和定义，并规定了公用电网在电压、频率、电流、谐波等方面应达到的指标。此外，一些国家标准还规定了用户注入公网的各种干扰

限值：《电能质量公用电网谐波》中规定了用户向公用电网注入谐波的限值；《三相电压允许不平衡度》中规定了用户引起不平衡度的限值；《电压波动和闪变标准》中规定了由用户干扰性负荷引起的电压波动和闪变的限值；《电能质量电力系统频率偏差》中规定了用户冲击负荷引起的电力系统频率变动限值。《电能质量监测设备通用要求》中规定了电能质量各项指标的监测方法和精度。根据以上国家标准的规定，我国对用电单位的供电额定电压及容许偏差规定为：①低电压 220V/380V，用于照明用户时允许偏差+5%~-10%；用于其他为±7%。②高电压 10kV 及以下，允许偏差为±7%；对特殊用户有 35kV、110kV 供电的，允许偏差为±5%。此外，我国还规定了 300 万 kW 以上的电力系统的频率偏差不得超过±0.2Hz；而 300 万 kW 以下的小容量电力系统的频率偏差规定不得超过±0.5Hz。我国还规定了用户向电网注入各种污染的限值，如表 2-2 所示。

表 2-2 我国运行用户向电网注入污染的限值

| 扰动类型 | 相应标准 | 规定限值 |
|----------|------------------------------|--|
| 电压波动和闪变 | GB/T 12326-2008《电能质量电压波动和闪变》 | 低压 PCC 点电压波动限值为 4%。 10min 内闪变值不超过 1.0，2h 内闪变值不超过 0.8。 |
| 三相电压不平衡度 | GB/T 15543-1995《三相电压允许不平衡度》 | 每个用户在 PCC 点引起的电压不平衡度允许值一般为 1.3%。 |
| 频率偏差 | GB/T 15945-1995《电力系统频率偏差》 | 冲击负荷引起的系统频率变动一般不得超过±0.2Hz。 |
| 谐波 | GB/T 14549-1993《电能质量公用电网谐波》 | 10kV 公共电网谐波电压总畸变率不高于 4%，奇次谐波电压含有率不高于 3.2%，偶次谐波电压含有率不高于 1.6%。 |

2.2 电能质量扰动分类

为了能系统地分析和研究电能质量现象，按照电能质量扰动现象的两个重要表现特征（变化幅值和持续时间），IEEE 将电能质量问题分为稳态问题和暂态问题两大类，如图 2-1 所示。稳态电能质量扰动是指电压或电流的波形(此指曲线的形状)、幅值、频率，主要以波形畸变为典型代表，主要包括电网谐波、电压

波动和闪变、电压偏差、频率偏差、三相不平衡等^{[8][9]}。稳态电能质量指标的物理含义是在一段较长的时间内电压、电流的实际波形(包含曲线的形状、幅值、频率、相位等)与期望波形的差异程度,因此又被称为连续型电能质量问题。暂态电能质量扰动是指电压、电流突然发生短暂的严重偏离期望波形的情况,其主要包括电压暂降、电压暂升、短时电压中断、脉冲及振荡等,因此又被称为事件型电能质量问题。暂态电能质量指标的物理含义是电压、电流的波形突然发生短暂剧烈变化的严重程度。此外,从一段时间监测统计上看,暂态电能质量扰动的严重性还表现在事件的发生频次上^[10]。

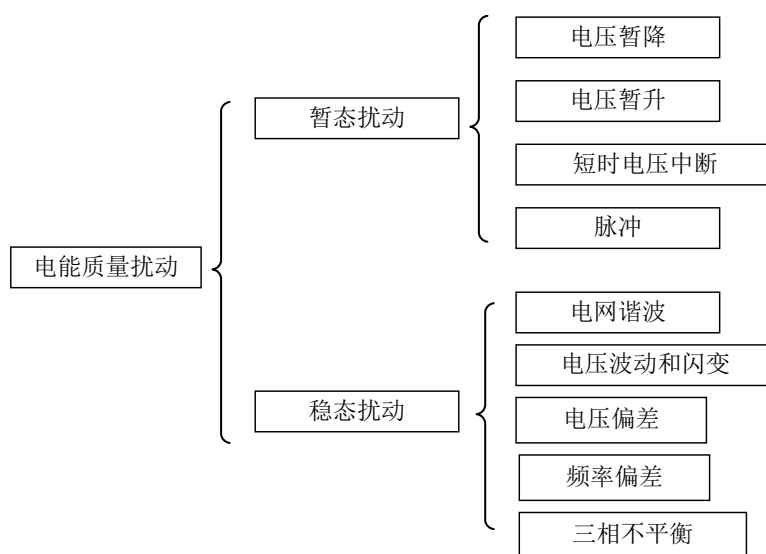


图 2-1 暂态和稳态电能质量扰动归类

在上述电能质量扰动划分的基础上,国际电工委员会(IEC)和美国电气电子工程师协会(IEEE)都对典型的电能质量扰动进行了定义和分类。本报告主要根据 IEEE 的电能质量标准,将典型的电能质量单一扰动描述如下:

表 2-3 IEEE 规定的电磁现象特性及分类

| 类别 | | 典型频谱成分 | 典型持续时间 | 典型电压幅值 |
|----|------|--------|-----------|----------|
| 暂态 | 脉冲暂态 | 纳秒级 | 5ns 上升沿 | <50ns |
| | | 微秒级 | 1μs 上升沿 | 50μs~1ms |
| | | 毫秒级 | 0.1ms 上升沿 | >1ms |
| | 振荡暂态 | 低频 | <5KHz | 0.3~50ms |

| | | | | | |
|--------|------|----|---------------------|----------------|------------|
| | | 中频 | 5~500KHz | 20us | 0~8pu |
| | | 高频 | 0.5~5MHz | 5us | 0~4pu |
| 短期变化 | 断电 | | | 0.5cycles~1min | <0.1pu |
| | 电压骤降 | | | 0.5cycles~1min | 0.1~0.9pu |
| | 电压骤升 | | | 0.5cycles~1min | 1.1~1.8pu |
| 长期变化 | 持续停电 | | | >1min | 0.0pu |
| | 欠电压 | | | >1min | 0.8~0.9 pu |
| | 过电压 | | | >1min | 1.1~1.2 pu |
| 电压不平衡 | | | | 稳态 | 0.5~2% |
| 波形畸变 | 直流偏移 | | | 稳态 | 0~0.1% |
| | 谐波 | | 0~100 th | 稳态 | 0~20% |
| | 间谐波 | | 0~6KHz | 稳态 | 0~2% |
| | 陷波 | | | 稳态 | |
| | 噪声 | | 宽带 | 稳态 | 0~1% |
| 电压波动 | | | <25Hz | 断续 | 0.1~7% |
| 电源频率变化 | | | | <10s | |

(1) 电压暂降(Voltage sag)。电压暂降是指电力系统运行中，电压信号的工频电压有效值降至 0.1~0.9 p.u.，持续时间 0.5T~1min 之内的一种电磁扰动现象，也称作电压凹陷、电压跌落或者电压骤降。引起电压骤降的主要原因包括有电力系统的短路故障、变压器的投运、感应电动机的启动、启动电流较大的大型负荷的接入等方面^{[11][12]}。

(2) 电压暂升(Voltage swell)。电压暂升是指电力系统运行中，电压信号的工频电压有效值增至 1.1~1.8pu，持续时间 0.5T~1min 之内的一种电磁扰动现象，这种电磁扰动现象也可以被称为电压凸起或者电压骤升。一般来说，引起电压骤降的主要原因包括有中性点不接地系统发生一次单相接地短路故障以及大型负荷的切除或大容量电容器组的充电等方面。

(3) 电压中断(voltage interruption)。电压中断是指电力系统运行中，电压信号

的工频电压有效值减至小于 0.1 p.u.，其持续时间在 0.5T~1min 之内之内的一种电磁扰动现象。电压中断主要原因是由雷击、异物等导致的电力系统瞬时故障，电力系统控制装置误动作和设备失效也可能导致电压中单。

(4) 脉冲暂态和振荡暂态。脉冲暂态是指电力系统运行中，电压信号或电流信号在稳定条件下出现了突发的、非电源频率的单极性变化的一种电磁扰动现象。脉冲暂态的原因主要为雷击产生。振荡暂态指指电力系统运行中，电压信号或电流信号在稳定条件下出现突发的、双极性变化的一种电磁扰动现象。与脉冲暂态的区别在于扰动波形的极性，脉冲扰动为单极性者，振荡扰动为双极性，根据其频谱成分，振荡暂态又分为高、中、低频三种类型。

(5) 谐波(Harmonie)。谐波是指电力系统运行中，电压或电流信号含有基波的整数倍频率的正弦波成分的一种电磁扰动现象。谐波通常是由整流器等电力电子设备以及具有非线性特性的用电负荷引起的。通常使用谐波电压总畸变率 THD_u 作为谐波等级划分的指标，谐波电压总畸变率的定义如公式 2-1 所示。

$$THD_u = \frac{\sqrt{\sum_{n=2}^N U_n^2}}{U_1} \quad \text{式 2-1}$$

(6) 电压波动与闪变(Voltage fluctuation and flicker)。电压波动是指电能质量电压信号在其包络线内发生有规则变动，或发生幅值不超出 0.9~1.1p.u.范围之间的一系列随机变化的一种电磁扰动现象。电压闪变通常是指由于电压波动所引起的电光源灯光照度不稳定的人眼反应现象。

(7) 三相不平衡(Unbalance)。三相不平衡是指三相电压信号在幅值和相位角方面发生的偏移。三相不平衡的原因主要由事故型和正常型两类，如一相或两相短线或接地故障都会引起三相不平衡。三相不平衡度是指负序分量比正序分量的百分数，通常用于作为衡量指标，如公式 2-2 所示。

$$\varepsilon_U = \frac{A_2}{A_1} \times 100\% \quad \text{式 2-2}$$

其中， A_1 为三相电压正序分量的均方根值； A_2 为负序分量的均方根值。

2.3 电能质量评估研究

在电能质量评估领域，目前的国际和国家标准在电能质量评估领域并没有给出相应的检测和计算方法^{[13][14]}。国际上一般常采用用户供电可靠性指标来评估电压长时间中断，而对其它暂态电能质量扰动大多只是描述现象。电能质量评估是以检测的基本参量为基础，进行统计计算的过程。评估电能质量需要规范整个评估流程和指标计算方法，这样才能在相同的平台上得到一致的可对比的评估结果。现有的电能质量国际、国内标准往往只是针对单个或部分电能质量现象提出指标或限值，导致对电能质量的评估多是采用单项评估指标^[15-17]。其中，仅有IEC61000系列标准对于稳态电能质量扰动给出了评估统计方法的推荐。根据电能质量评估方式的不同，可划分为如图2-2所示的评估体系。

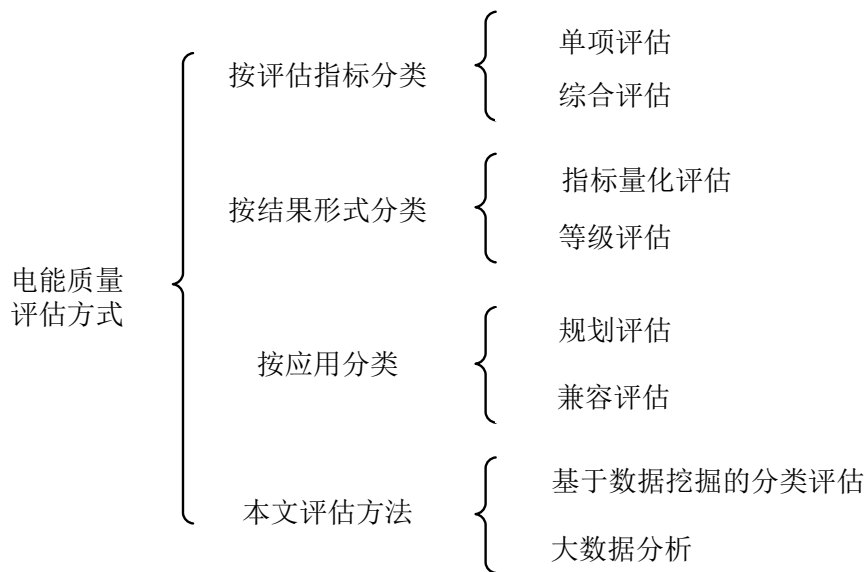


图 2-2 电能质量评估方法体系

2.3.1 电能质量单项评估

根据现有的国内外标准，按照电能质量所包含的各种不同指标进行单项评估是普遍采用的方法。电能质量的单项评估是指，针对某一个电能质量问题或对其某个特征量进行量化从而得到衡量值的过程，评估结果的表现形式就是该项电能

质量的单项指标，评估结果就是指标值。各类电能质量扰动的单项评估指标如表 2-4 所示。

表 2-4 电能质量单项评估指标

| 电能质量扰动类型 | 评估指标 |
|----------|-------------------------------------|
| 电压暂降 | 能量指标, $SARFI_x$ 指标, $SARFI_{curve}$ |
| 电压三相不平衡 | 电压不平衡度 (%) |
| 电压偏差 | 电压合格率 (%), 电压偏差率 (%) |
| 谐波 | THD |
| 电压波动和闪变 | d , P_{st} , P_{lt} , 闪变合格率 |
| 供电中断 | 供电可靠性 (%), 用户平均停电次数和时间 |
| 频率偏差 | 频率合格率 (%), 偏差值 (Hz) |

电力系统中，导致电能质量发生的原因是随机出现的，多种原因同时发生的概率很小。因此，在某一时刻或某段时间内，电气设备（元件）承受的电能质量问题往往是某个方面特别突出。由电力系统设备故障引起电压暂降的同时，往往还伴随着高频振荡和电压不对称，但在此过程中的电气设备的运行状态主要由电压暂降的程度决定，而暂时的波形畸变和电压不平衡等产生的热效应对设备的影响很小。此外，随着高新技术产业的发展以及计算机、电力电子和微电子技术的广泛应用，电压暂降和短时中断的影响较为严重。据统计，在欧美发达工业国，因电压暂降或短时中断引起电力用户对供电的投诉量占到整个电能质量问题投诉量的 80% 以上^{[18][19]}。因此，电压暂降成为电能质量扰动中最受关注的问题。

对电能质量按现象和特征进行单项评估时，监测仪器分别检测和分析单项电能质量的参数，将多元化的电能质量问题分解，能够迅速发现电力系统中的主要电能质量问题，有助于迅速找到引起扰动发生的原因，以解决实际问题。但是，实际应用中，各种电能质量扰动通常是同时发生且互相影响的，单项评估无法从

整体上衡量电能质量的综合指标。

表 2-5 国家标准中规定的电能质量单项评估指标

| 标准编号 | 标准名称 | 电能质量指标允许限值 | | | | | |
|------------------|------------|---|----------|-----------|--------------------|--------|-----|
| GB12325-2003 | 供电电压允许偏差 | 10kV 及以下三相供电电压允许偏差为额定电压的±10%； 220V 单相供电电压允许偏差为额定电压的+7%、-10%； 35kV 及以上供电电压正、负偏差的绝对值之和不超过额定电压的 10%。 | | | | | |
| GB/T14549-1993 | 公用电网谐波 | 各级电网谐波电压限值(%) | | | | | |
| | | 电压(kV) | 电压总谐波畸变率 | 各次谐波电压含有率 | | | |
| | | | | 奇次 | 偶次 | | |
| | | 0.38 | 5.0 | 4.0 | 2.0 | | |
| | | 6、10 | 4.0 | 3.2 | 1.6 | | |
| | | 35、66 | 3.0 | 2.4 | 1.2 | | |
| | | 110 | 2.0 | 1.6 | 0.8 | | |
| 注入公共连接点的谐波电流允许值 | | | | | | | |
| GB/T15543-1995 | 三相电压允许不平衡度 | 电力系统公共连接点正常电压不平衡允许值为 2%， 短时不得超过 4% | | | | | |
| GB/T15459-1995 | 电力系统频率允许偏差 | 正常：0.2 Hz 系统容量较小时：0.5 Hz | | | | | |
| GB12326-2008 | 电压波动与闪变 | 电压变动限值 | | | | | |
| | | r, h ⁻¹ | d, % | | r, h ⁻¹ | d, % | |
| | | | LV, MV | HV | | LV, MV | HV |
| | | r≤1 | 4 | 3 | 10<r≤100 | 2 | 1.5 |
| | | 1<r≤10 | 3 | 2.5 | 100<r≤1000 | 1.25 | 1 |
| | | 电压闪变限值 | | | | | |
| | | 注：本标准中 P _{lit} 每次测量周期取为 2h | | | | | |
| 电压等级 | | ≤110kV | | >110kV | | | |
| P _{lit} | | 1 | | 0.8 | | | |

2.3.2 电能质量综合评估

长期以来，实现定量的、全面的电能质量综合评估，一直是电能质量工作者共同追求的目标。目前虽然有许多国家标准和文献都对电能质量各单项指标的评估方法做了研究，但是针对电能质量综合评估的研究较少。主要是由于当电能质量的多个指标共同作用到一个系统时，不同等级的各项评估指标的组合对电力系统的影响，以及对系统中运行设备的影响甚至是损坏差别较大^{[20][21]}。同时不同的用电设备对电能质量各评估指标的重视程度也不相同，所以实现对电能质量的综合评估是非常复杂和困难的。综合评估结果可以作为供用电双方制定供电合同

的依据,有助于供用电双方明确电能质量责任,并可作为电能商品分质计价的重要参考。同时,电能质量评估是进行电能质量治理的前提,是对电网供电点电能质量的量化指标,对于干扰源与敏感用户的供电方案制定、电源点的选择有重要意义^[22]。

电能质量综合评估的研究焦点是如何科学、客观地将一个多指标问题综合成单一量化的指标问题,从而评定电能质量的等级。综合评估建立在单项评估基础上,综合评估方法把部分或全部电能质量扰动或某项电能质量扰动的多个特征指标按属性量化后,采用归一、综合的算法得到一个综合指标或综合等级^[23]。目前,大量学者分别基于层次分析法(Antalytic Hierarchy Process, AHP)、模糊综合评判法(Fuzzy Comprehensive Evaluation Method, FCEM)、人工神经网络评价法(Artificial Neural Network Method, ANNM)以及灰色综合评价法(Grey Comprehensive Evaluation Method, GCEM)等方法对电能质量综合评估问题进行了研究,取得了较好的效果^{[24][25]}。但是,由于各单一综合评估方法的机理不同,属性层次相异,导致其评价结论存在差异。

虽然电能质量现象及其指标种类繁多,但是从实际物理过程来看,所有电能质量的基本特征量是共同作用在同一电气设备上的,设备的工作状态和性能是由该时刻的这些电能质量多项指标共同决定的。由此可见,虽然电能质量特征是多元化的,但是其对设备性能的影响结果是需要综合多项特征指标来评估的。例如,连续型的谐波和负序对同步电机造成的影响很类似,当衡量谐波电流引起同步电机的附加损耗和发热时,往往折算成等效的基波负序电流来考虑。这就相当于把实时的谐波和负序电流进行综合评估,然后与限值比较以判断对出力的影响。此外,当电压暂降的电压幅值和持续时间超过设备的敏感曲线时,虽然并未出现电气连接的断开,但可能对用户造成比断电还严重的后果。因此,对于电能质量的综合评估除了考虑各单项指标外,还需要考虑用户负荷的实际情况。电能质量扰

动对于用户设备和负荷的影响决定了综合评估的结果。

综合评估方法可以简化评估结果，增强了执行的可操作性和大众的可接受性。其主要的评估方法包括量化评估和等级评估，分别属于定量数值计算和定性等级计算的范畴。

（1）电能质量量化评估

量化评估是将各电能质量指标数值化的计算过程，得到的质量水平是个数值，其优点在于可以直接与标准规定的兼容水平、规划水平或合同规定的限值水平相比较。这是目前普遍采用的评估结果形式，能够直接地反映问题严重性，并且量化值可以作为电能质量运行控制函数和治理设施设计参数的输入自变量。其方法主要包括基于短板理论的数值计算和权重理论的数值计算。

使用各单项电能质量指标相对其限值水平的偏离程度，作为电能质量扰动对设备可能造成影响的程度（故障率、经济损失等）。基于各指标的限值水平，将各评估指标值归一化处理，得到该指标相对于其限值水平的标么值，以表征单项（或同类质量的多指标综合）指标可能对设备造成的影响程度。

①权重理论的数值计算：使用权重表示各单项指标对于用户设备的重要程度，将各指标的归一化后的值基于其权重相加得到单一的量化指标，表示整体电能质量相对限值水平的偏离度，如公式 2-3 所示。

$$I_G = \lambda_H \times I_H + \lambda_U \times I_U + \lambda_P \times I_P + \dots + \lambda_{\delta f} \times I_{\delta f} \quad \text{式 2-3}$$

②木桶理论的数值计算：使用单项指标偏离程度最高的项作为综合评估结果，突出反映最可能使电力用户受到影响的电能质量问题，如公式 2-4 所示。

$$I_G = \max(I_H, I_U, I_P, I_R, \dots, I_{\delta f}) \quad \text{式 2-4}$$

木桶理论是指，在同一时刻决定设备运行状态的往往是干扰最大的电能质量问题。基于木桶理论的综合评价方法可以突出决定综合评估结果的主要问题。

（2）电能质量等级评估

定性等级评估是先将电能质量单项指标划分为若干等级区间，然后在量化评估基础上，根据实际评估水平所在的区间确定评估等级。为了确保设备在实际随机运行环境下的这种高概率正常工作，IEC61000 系列标准针对部分电能质量扰动，分别为电力系统和电力设备制造商规定了限值水平，前者是作为系统内部控制目标的规划水平，后者是设备制造时要求达到的设备免疫力（测试）水平。此外，英国 G5/4 标准也规定了电能质量等级的一些推荐限值。以电压暂降为例，依据电压暂降的严重程度将其干扰域划分为 I ~ V 级，如表 2-6 所示。

表 2-6 电压暂降等级划分

| 残压 U(%) | 持续时间 t (s) | | | | | | |
|-------------|------------|-----------------|----------------|---------------|---------------|-------------|-------|
| | t ≤ 0.02 | 0.02 < t ≤ 0.05 | 0.05 < t ≤ 0.1 | 0.1 < t ≤ 0.2 | 0.2 < t ≤ 0.5 | 0.5 < t ≤ 3 | t > 3 |
| 80 ≤ U < 90 | I | I | I | I | I | I | I |
| 70 ≤ U < 80 | I | I | I | I | I | III | III |
| 60 ≤ U < 70 | I | II | II | II | III | III | III |
| 50 ≤ U < 60 | I | IV | IV | IV | V | V | V |
| 40 ≤ U < 50 | I | IV | V | V | V | V | V |
| U < 40 | I | IV | V | V | V | V | V |

对于监测点的电能质量等级评估，通常基于监测点的实际监测数据，计算出该点的电能质量各项指标，在此基础上对监测点的综合等级进行评估。通常也基于权重理论和木桶理论进行评估，分别由加权求和结果或者等级最差的电能质量指标决定电能质量的综合等级。

2.4 国内某供电公司电能质量现状

从国内某供电公司 2009 年至 2011 年统计的电能质量数据来看，在反映稳态

电能质量的五大类指标中，系统频率和三相电压不平衡度的合格率较好，基本合格率都为 100%。但是，暂态电能质量扰动问题严重威胁着该供电公司的供电质量，特别是电压暂降和短时中断问题引起了大量敏感用户的投诉。投诉用户主要集中于半导体生产和精密仪器生产行业，轻微的电压暂降扰动就会严重影响其生产线，导致产能和产品合格率的严重下降。

据统计，该城市暂态电能质量事件发生较为频繁，严重时单月可达近 900 次，特别是电压暂降、短时中断发生的频率比较高，仅以 2011 年 7 月统计表明，在 135 处变电站共发生 819 次暂态事件。如图 2-3 所示，图(a)是整个供电公司电压暂降和短时中断的情况（残压值和持续时间），图(b)是某中心站发生的暂态电能质量扰动情况（包含电压暂升）。此外，由于该城市有大量汽车、建材、电线电缆等制造行业的电力用户，这些工业负荷均是巨大的电能质量扰动源，从而导致谐波、闪变等问题严重。同时由于城市电网规划与建设滞后于城市经济发展，电能质量的问题在目前没有得到妥善的解决且日益严重。此供电公司选取 48 座变电站总计 75 条重点专线用户分别进行了 24 小时的电能质量测试，发现 75 条专线用户中有 17.8% 的测试用户存在功率因数偏低的问题，13.33% 的用户存在电压闪变严重超标的问题，4.11% 的用户存在电压总谐波畸变率超标的问题。

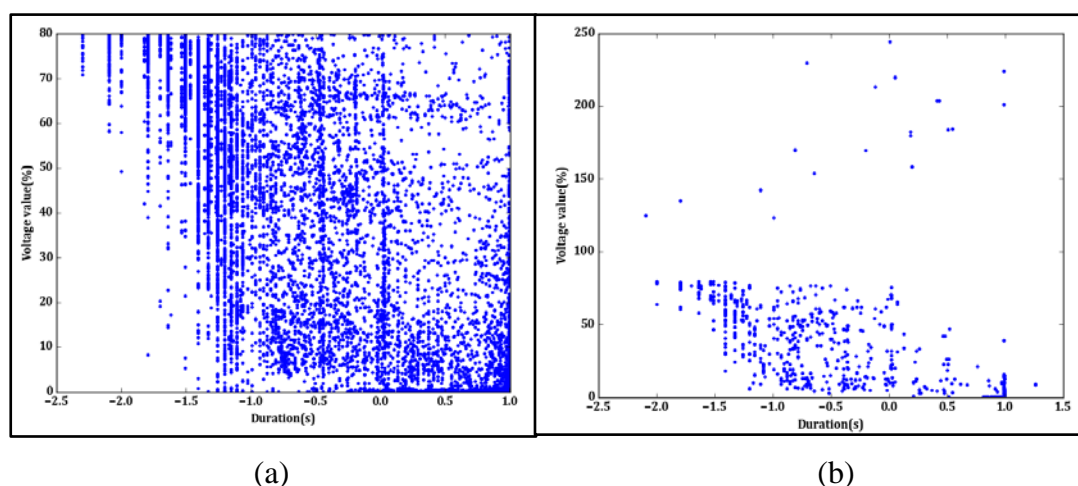


图 2-3 某供电公司 2010-2012 年发生暂态电能质量扰动情况

在实际运行的电力网中，由于各种原因，会导致用户得到的电力供应参数偏

离了其额定数值，甚至导致质量不能满足用户要求，影响用电设备的正常运行。

总结起来，影响电能质量的主要因素在于以下方面：(1)电力系统元件存在的非线性问题。电力系统元件的非线性问题主要包括：发电机产生的谐波；变压器产生的谐波；直流输电产生的谐波等。(2)非线性负荷。工业生产和生活用电中非线性负荷是造成电力系统谐波问题的首要原因。而电弧炉就是其中非线性负载的典型。(3)电力系统故障。随着系统中非线性负荷的不断增长，谐波对电力系统的影响程度逐渐加大，与此同时，系统本身也存在出现故障的风险，这些都是导致电能质量受影响的原因。(4)冲击负荷。这些冲击性的负荷包括电弧炉、弧焊机和轧钢机等。根据对于该供电公司的调研发现，其电能质量现象起因比较复杂，但在总体上分析主要由于雷电、外力、冲击性负荷影响电力系统设备正常运行所导致。本报告对该供电公司 2010 年上半年发生的电压暂降时间进行了原因统计，其结果如图 2-4 所示。

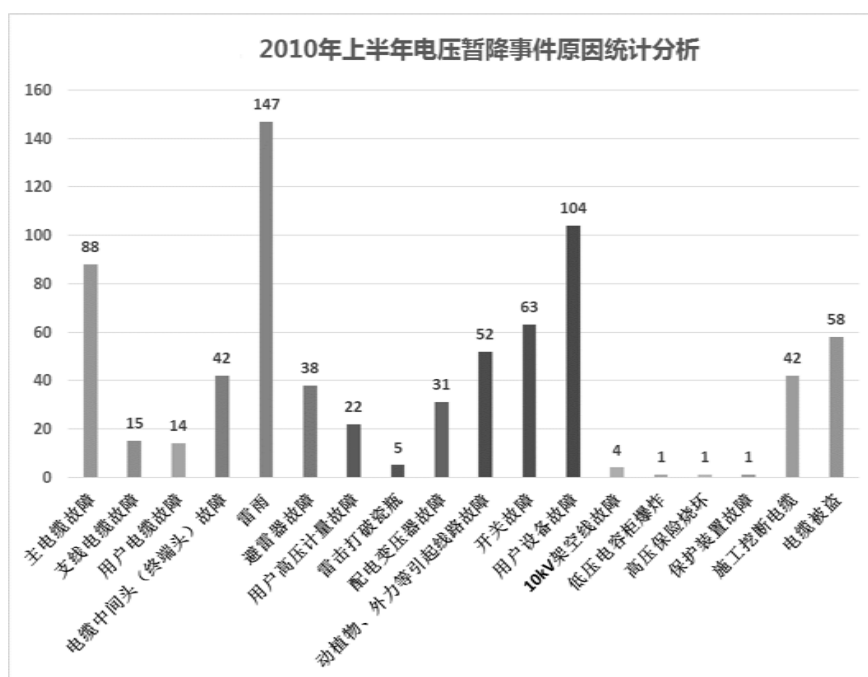


图 2-4 某供电公司电压暂降原因分析

3 电能质量综合评估数据来源

(电能质量监控现状、电网信息化趋势、信息采集、信息集成、信息预处理等)

3.1 电能质量信息化现状

随着信息化技术的飞速发展与应用深化，大数据应用已在公共管理、零售、互联网、电信、金融等领域得到了广泛应用，美国、欧盟、日本等国家在大数据提取、管理、分析、可视化等方面取得了一定成效。我国在大数据研究和应用方面还处于初级阶段，尤其是大数据在智能电网建设的应用还有待研究。目前国内电力企业的信息化水平正日趋成熟，电力企业在生产、经营和管理各个层面都进行了相当规模的信息化建设，取得了一定的成果，主要体现在硬件水平的不断完善，软件水平的不断提高以及应用水平的不断提升。

随着智能电网的全面建设，电力数据在电力生产和使用过程中产生，数据来源涉及电力生产和电能使用的发、输、变、配、用及调度各个环节，这些数据涵盖了结构化、非结构化数据，具有多源、海量、异构、时空连续等特点；特别是以物联网和云计算为代表的新技术在电力行业中的广泛应用，电力数据资源开始急剧增长，电力行业数据已满足体量大、类型多、价值高、速度快的特征。如何从这些海量数据中提取有用信息，是目前电力设备状态评价的一个发展瓶颈。目前，虽然已有一部分大数据应用在电力行业应用的研究与尝试，但在数据的全面性、实时性、数据一致性、隐私性和安全性方面仍有不少挑战，同时在数据广度和应用深度上还有不足，需要找出相应的解决方法，有待去研究和解决。

整体而言，当前信息通信技术（ICT）对中国电力工业的价值贡献正处于量变到质变的关键节点，而变化的本质就是电力信息通信与电力生产以及企业经营的深度融合，其最终表现形式必将是电力数据的爆发性增长。为了解决以上问题，将信息化进一步推进，真正转变为企业发展的助推剂、企业的竞争力和生产力，引入大数据技术建立电力大数据应用平台应用已是大势所趋。电力行业的发展，需贯穿电力生产、经营、管理等各环节的电力大数据应用业务的支撑，同

时进行多类型数据的关联分析，完成数据驱动决策的业务应用，并通过大数据应用驱动电力信息化与电力生产以及企业管理的深度融合。电力大数据将贯穿未来电力工业生产及管理各个环节，起到独特而巨大的作用，是中国电力工业在打造下一代电力工业系统过程中有效应对资源有限、环境压力等问题，实现厚积厚发、绿色可持续性发展的关键。

大数据已在公共管理、零售、互联网、电信、金融等领域得到了广泛应用，且美国、欧盟、日本等经济发达国家和地区在大数据管理、分析、可视化等方面已取得了诸多成效。电力领域，根据美国电力研究协会（EPRI）最近对其国内电力公司大数据应用情况的调查显示，目前大多数电力公司都已着手进行大数据收集和存储，但如何有效筛选和应用这些数据还未取得重大进展。在国内，大数据已被电力企业列入企业战略发展重要指导方向。但是，目前电力企业的信息化也暴露出了诸多问题：现有信息化建设缺乏统一的标准体系，存在大量信息系统重复建设；信息孤岛多，信息集成密度低、无法相互协作发挥整合效益；企业各信息系统相互分离；信息化建设过程中，大量的建设和投入被用于设备的自动控制，数据信息的收集、存储、共享、却忽视了对信息的整理和挖掘。

3.2 电能质量监测装置

电能质量扰动的产生来源于供用电双方，并受外界环境的影响。电能质量监测的目的在于清楚地分析扰动的产生原因和影响范围，从而改善电网内的电能质量，并为供用电双方的污染和治理责任划分提供数字依据。因此，要求电能质量监测装置能够真实有效地实时测量电能质量的各项指标，并及时发现指标偏离正常值的时间、位置、偏离量。电能质量监测装置的功能应包括：数据采集、数据集成、数据转换、数据存储、扰动监测、故障录波，并实现网络化、智能化管理，后台管理分析软件具有强大的分析功能^{[26][27]}。

由于电能质量需要监测的量很多而且大多是高度畸变的，传统的方法是对模

拟信号的分析，监测不同的电能质量指标使用不同的仪表，如传统的测量功率损耗的有功表、无功表，测量电压和电流有效值的电压表、电流表，测量频率的频率表，还有谐波测量仪、三相不平衡度计、电压波动和闪变仪。此类仪器的缺点是可监测的指标少、通用性较差、精度较低、自动化程度较低^[28]。

随着电能质量暂态扰动的影响不断严重，对于电能质量的关注不再仅集中于谐波扰动、电压偏差、频率偏差等稳态指标，还包括电压暂降、暂态谐波等电能质量的暂态指标。这就要求电能质量监测装置能够提高处理的实时性，并提供更为直观的分析结果，以利于对电能质量问题做出决策，如要求系统能够进行故障识别、干扰源识别、故障预测和信息共享等。目前，主流的电能质量监测装置主要采用基于 DSP 技术(Digital Signal Proceeding)、ARM 技术 (Advanced RISC Machines) 的微处理器为核心，通过对电信号进行分析和处理来实现，使监测装置在测量准确性、可靠度和实时性等各方面都有了明显的进步^{[29][30]}。其中，典型的基于 ARM+DSP 数据采集系统架构如图 3-1 所示。

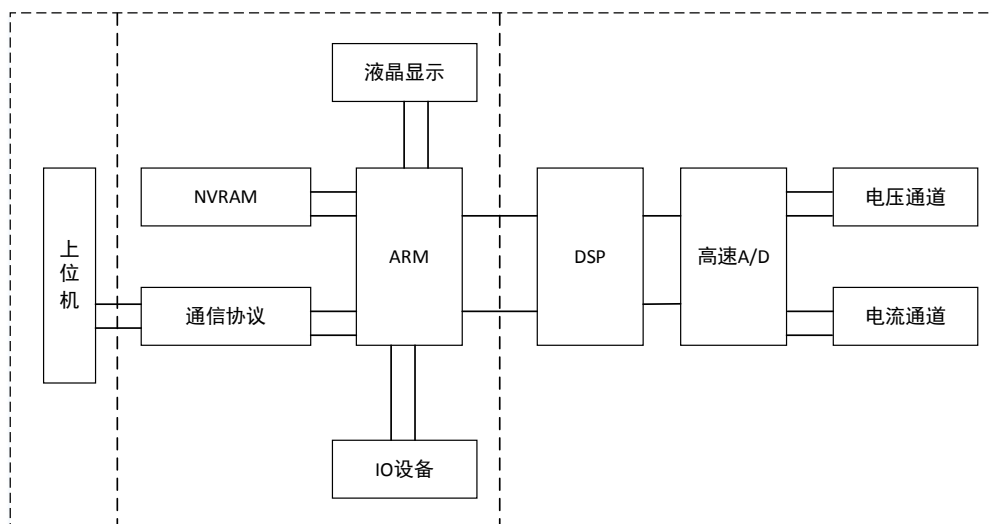


图 3-1 典型的 ARM+DSP 系统架构

此类监测装置通常采用单片机、数字信号处理器作为处理核心，并和计算机相连，构成数据处理能力较强的“PC+微处理器”结构，用数值计算的方法对信号进行采集、解析与识别等加工处理。扰动检测的算法主要包括傅立叶变换 (FT)、短时傅立叶变换 (STFT)、小波变换 (WT)、人工神经网络 (ANN) 等。其主要

优点在于提高电能质量监测的实时性和准确性,并具备多种电能质量扰动的监测功能。“PC+微处理器”类型的监测装置对单个站点的测量有比较好的效果,并积累了大量的历史数据,可作为电能质量分析的数据基础^{[31][32]}。例如福禄克(FLUKE)的345系列电能质量分析仪,可以测量目前几乎所有的电力和电能质量参数,精度可达0.1%。同时,由于电能质量在线监测系统的需求,电能治理的测控技术正在向网络化发展,利用网络协议对监测数据进行通信传输,实现对电能质量的远程监测和网络化管理。

目前电能质量监测除了具备传统的监测功能外,在功能和技术上更加强调各种数据的采集、转换、集成、分类和管理,并使用高效的扰动检测算法,基于监测数据实现对电网运行性能和条件的评估^[33]。通常的电能质量检测方式分为连续监测、定期或不定期监测和专门测量。1)连续监测:一般用于重要变电站的公共供电点的监测。监测指标包括:供电频率、电压偏差、三相电压不平衡度、谐波等。2)定期或不定期监测:适用于需要掌握供电质量而不需要连续监测或不具备连续监测条件的监测方式。一般的公共供电点的供电质量通常只需定时监测。对于无冲击性负荷的电网,一般不存在明显的电压波动和闪变,这两个指标只需一二年测量一次即可。而对于有冲击性负荷的电网,往往一个月或一个季度测量一次,监测周期和每次的测量时间视具体情况而定。对于负荷稳定、用电量不大的电力用户,电能质量指标比较稳定,则可一周或一个月定时监测一次。3)专项测量:主要适用于干扰源设备接入电网(或容量变化)前后的电能质量监测,用以确定电网电能质量指标的背景状况和干扰发生的实际量或验证技术措施效果。

以国内某供电公司为例,该公司正扩展电能质量监测装置的布置范围,并升级在线监测系统。该供电公司已在131个变电站安装了电能质量监测终端578台,安装于变电站110kV与10kV母线(220kV及以上电压等级由省级公司监测系统监测)。其布置情况如表3-1所示。这些电能质量监测装置的功能是对

电能质量数据进行采集、处理和分析，并根据业务系统需求上传至上位机。

表 3-1 某供电公司电能质量监测点布置情况

| 片区 | | 变电一部 | | | | | | 变电二部 | | | | | 总计 |
|----------|-------|------|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|
| 中心站站名 | | 站 1 | 站 2 | 站 3 | 站 4 | 站 5 | 站 6 | 站 1 | 站 2 | 站 3 | 站 4 | 站 5 | |
| 变电站个数 | | 1 | 9 | 22 | 20 | 8 | 12 | 11 | 15 | 12 | 12 | 9 | 131 |
| 监测 点数 | 110kV | 1 | 9 | 22 | 21 | 8 | 12 | 11 | 15 | 12 | 12 | 10 | 133 |
| | 10kV | 4 | 37 | 77 | 74 | 30 | 42 | 38 | 47 | 36 | 31 | 29 | 445 |
| | 总计 | 5 | 46 | 99 | 95 | 38 | 54 | 49 | 62 | 48 | 43 | 39 | 578 |

3.3 电能质量监测数据处理

3.3.1 电能质量在线监测平台

电能质量监测装置利用各种传感元件获取被测对象各种类型的数据和参量，通过信号处理电路对转换后的信号进行采样、调理、编码、量化，最后将这些处理好的信号传送到控制器中实现信号的存储记录与数据处理，实现了对数据的采集和集成。电能质量数据采集是实现电能质量监测的基础，其安装于变电站现场，可实现对各电压等级母线的数据采集。其工作方式通过 PT、CT 完成信号接入，具备数据采集、AD 转换、现场显示、通讯控制、数据处理等功能。这些电能质量监测装置的数据采集通常只能在同一个地点进行现场测量，难以进行对同一供电系统不同地点相关的监测点进行综合在线评估^[34]。因此，在电能质量监测数据采集的基础上，如果需要实电能质量实时监测、远程操作、实时数据查询、监测可视化等高级应用，需要基于 Modem 或以太网等通信方式建立电能质量在线监测平台，其功能描述如图 3-2 所示。

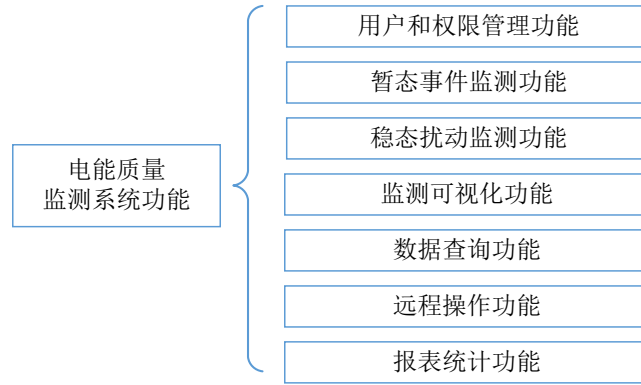


图 3-2 电能质量在线监测平台功能

电能质量监测终端的任务是将原始电压、电流信号采集并转变成数字信号，完成电能质量的数据采集和本地分析，并通过以太网传输将数据传输到监测中心，同时数据浏览和数据存储。由于电能质量的监测、评估、改进需要省级电力单位和各地市级电力单位的配合工作，电能质量在线监测平台的体系架构需要支持多级单位使用，并实现实现数据的分层管理，提供灵活的权限控制功能。电能质量在线监测系统需要对整个电网不同电压等级、不同区域的电能质量进行分析和评估，因此需要实现不同监测装置和服务器的数据融合，并不同级别电力公司之间的数据共享^[35]。出于经济型的考虑，电能质量在线监测平台应能够兼容已有已有的业务系统(包括各地市电力单位已经建立信息化系统和电能质量监测系统)，并满足未来电网发展的需要。

电能质量监测系统的初级形式是分布式监测装置，可以实时监测、分析一个变电站多条出线的状态，进而利用网络技术，逐步发展为分布式监测系统。Internet 和通信技术的发展，为信息共享和数据交换提供了便利，也为电能质量监测的网络化创造了条件。网络化的电能质量监测系统也在不断发展完善。GPS 技术被引入到电能质量监测系统中，用以保证采样数据的同步性和准确性。设计了专门的电能质量网络监测仪的上网接口，改善了联网传输性能。有人提出一种基于互联网地理信息系统(WebGIS)的电能质量监测系统,将电能质量信息与地理分布信息结合起来提供给电网管理人员，且用户终端只需一个支持 Java 编程语

言的 Web 浏览器就可以完成监测的基本操作，降低了系统维护和管理费用。虚拟仪器技术也应用到电能质量监测系统中 [36][37]。也有人以虚拟仪器技术为平台构建了电能质量网络化监测系统，通过 DataSocket 技术实现了对监测数据的动态传输。

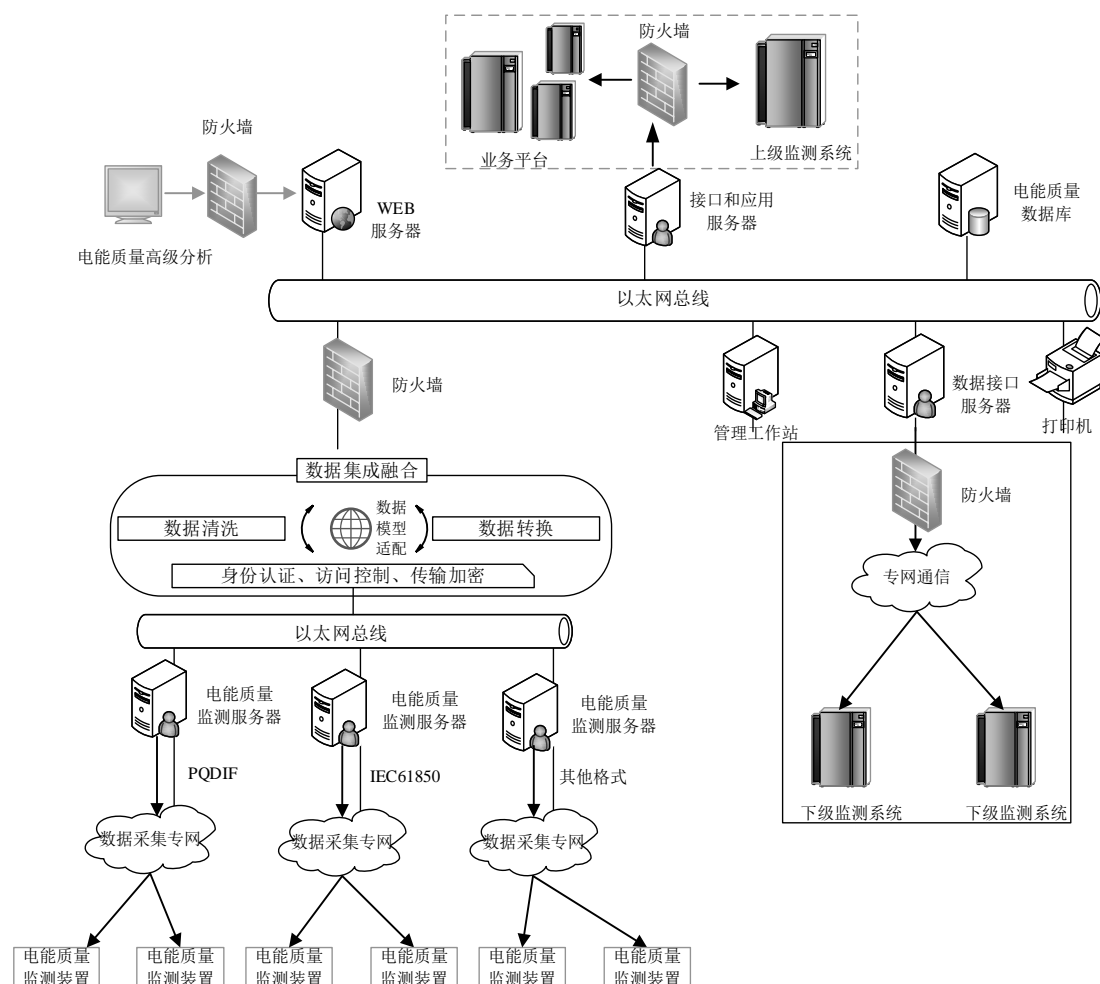


图 3-3 电能质量在线监测平台架构

在线监测平台的架构如图 3-3 所示，架构通常包括数据采集层、数据传输层、数据分析层、应用层组成，通常由监测装置、数据库服务器、以太网总线、数据分析工作站、应用服务器和客户端组成，共同实现对电能质量的远程监控和分析。其中，电能质量监测设备用于收集电能参数并进行简单的分析，形成 PDIIF 格式、IEC61850 格式或者其他格式的数据，并负责上传至数据采集服务器。然后，通过身份认证、数据转换、数据清洗等实现数据的集成和融合，并将处理后的数

据存入电能质量数据库。监控主机通过局域网与下层的监测设备进行通信，监控主机与监测设备的通信规约采用统一的控制格式，各监测设备可以主动上传数据给监控主机，监控主机也可以通过召唤传输任一监测设备的数据。

现代电网规模越来越大，监测点越来越多，未来电能质量的监测要实现不同供电点甚至多个供电系统的集中监测。在功能上，更强调智能化，除具有计算、显示功能外，还要有一定的判断、决策功能，例如能进行事件预测、故障辨识、干扰源识别和实时控制等，初步具有自动的、实用先进的智能评估功能。电能质量在实现了在线监测、实时分析的基础上，正向着网络化、信息化、标准化和智能化的方向发展完善^[38]。网络化、信息化、标准化和智能化已成为电能质量监测系统的必然发展趋势，它为电网的优化和事故分析提供实时可靠的数据，为电能质量综合评估提供切实依据，也是电力企业面向市场，适应竞争的强有力手段，可以进一步保障各级用户的正常用电秩序，为其提供优质的电能。

因此，电能质量在线监测平台采用分级的设计，数据的采集是分布式地获取电网的电能质量报数据，并向上级系统汇总。该设计既可以反映本地区电能质量的长期运行趋势走向，也可以反映整个区域在某一段时间内的电网运行状况^[39]。这对于供电企业了解和掌握电网运行参数、建立切实可行的管理计划提供了有力参考。在后台数据库的基础上可以建立综合信息查询系统，建立能对电能质量数据进行统计分析处理、评估和预测的专家系统以及基于国标的规则库，能分区域、分时段、分电压等级、分性质对电能质量指标进行分析比较、趋势比较等，强大的数据处理功能是该类系统的一个特点。

3.3.2 电能质量监测数据格式

目前国内大部分厂家使用的数据格式各异，指标含义不同意，从而导致了电能质量监测设备采集的数据不能相互兼容。这种情况存在着明显不足：第一，各监测仪器之间的数据不能相互交换，无法实现数据的共享；第二，随着监测指标

不断增加，特别是对于各种暂态过程，都需要采用高速数据采集系统，以反映信号的完整性，采样频率的提高对监测装置的处理速度和平台的存储量提出了一定的要求。电能质量监测和分析不仅需要海量数据，而且涉及广泛的数据来源、多种类型的数据内容和描述。因此，为了适应电能质量有关测量数据和计算数据的存储，必须制定统一的数据指标体系，作为数据采集、交换和分析标准。

为了适应电能质量有关测量数据和计算数据的存储，必须制定统一的数据存储体系，作为数据采集、交换和分析的标准。因此，IEEE 标准委员会提出了一种电能质量数据的交换格式 PQDIF(Power Quality Data Interchange Format)。它完全独立于监测设备的软、硬件，不仅可以较好地解决多数据源数据的兼容问题，还可以实现电能质量物理属性的多角度观察功能，满足了电能质量监测技术的发展需要。PQDIF 是一种平面文件结构，其中的各记录链接而成，通过标记元素结构将数据的物理属性和逻辑属性分离，使多数据源具有良好的兼容性，便于实现数据共享，为分析电网的电能质量问题以及实现电能质量管理和分析系统提供一种很好的数据结构^{[40][41]}。PQDIF 主要优势在于具有良好的通用性、压缩性和数据转换处理能力。通过它可以将原先多对多的数据关系减少到一对多，将外部各种分散孤立的数据源整合起来，然后通过一种简单的方式对整合后的数据进行处理。PQDIF 的设计格式是完全公开的，用户可免费下载到它的应用程序并安装在自己的操作平台下，无需更多繁琐程序就可实现各种功能并对保存在其中的文件进行处理。并且逻辑层和物理层分离，完全可以通过定义新的标识来扩展记录，具有可扩展性，利用标记元素还可实现部分静态信息的转换和共享。

3.3.3 PQDIF 数据格式

PQDIF 由一系列逻辑相关的一记录链接而成，可以分为物理层和逻辑层。其中物理层描述文件的物理结构，通过唯一性标识符来区分文件的内部元素类型，而不管文件实际存储的内容；逻辑层则利用物理层已定义好的结构，指定元素的

标识符，描述文件的内容。

(1) PQDIF 的物理层结构

物理层结构就是指 PQDIF 的物理组成方式，它不涉及具体内容，由一系列相互关联的记录组成。这些关联由记录在文件中的绝对地址或者偏移地址组成，并全部记录在记录头中。这种结构相当于一个链表结构，每个记录头包括本身及其下一个记录在文件中的位置，相当于链表中的指针，可以很方便的在文件中的记录进行添加、插入和删除操作。

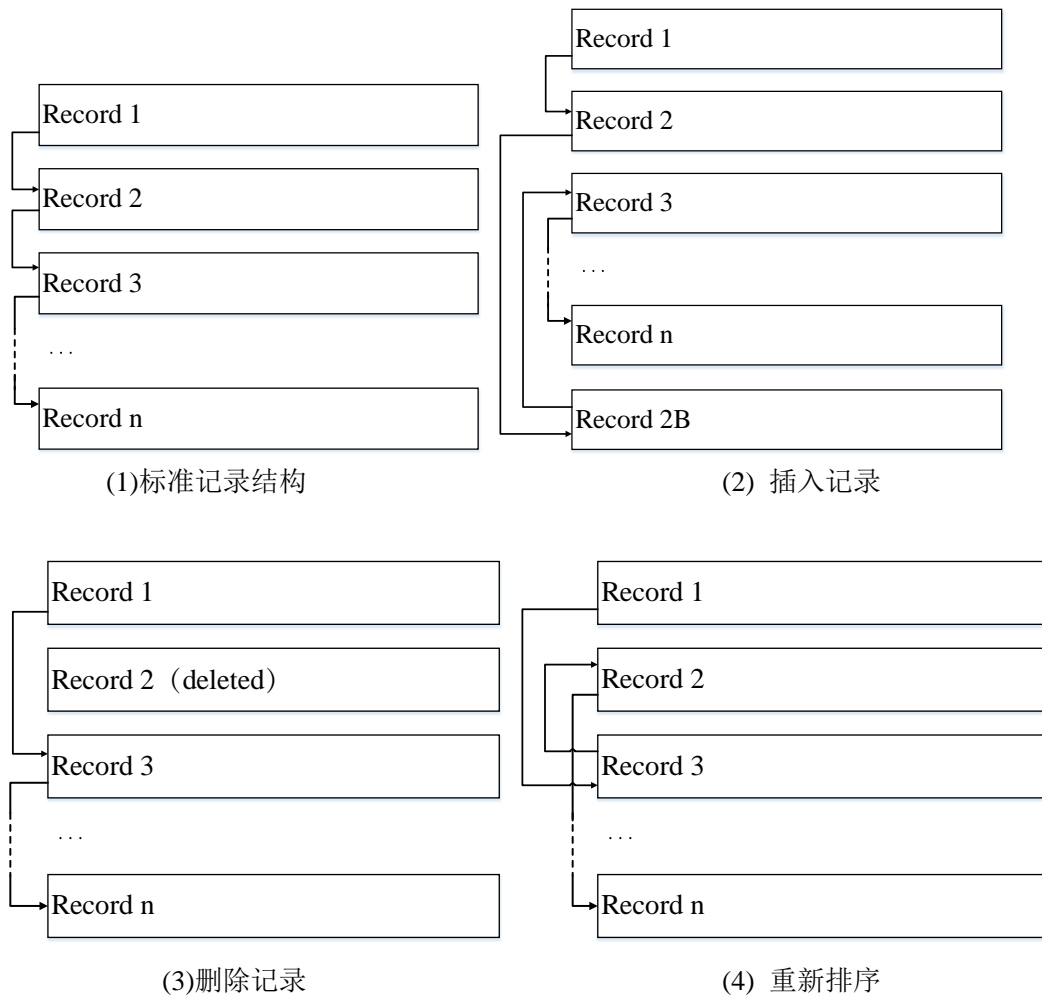


图 3-4 PQDIF 记录的添加、插入和删除操作

在 PQDIF 文件中，每个记录类型都有相同的基础结构，它包括记录头和记录体两部分。如表 3-2 所示，记录头是每个记录的唯一标识符 GUID (Globally Unique Identifier)，指定了记录的标签、大小、状态信息，以及指向下一个记录

的绝对链接。记录体有一系列的元素组成，包括集合、标量、矢量等，记录题以一个集合开始，集合中每一个元素对应着一个标签，该标签由元素在记录中的位置所产生，并由记录头中的标签所标识。

表 3-2 记录结构说明表

| 包含项目 | 举例 |
|--|---|
| 记录头 <ul style="list-style-type: none"> ● 标识符：PQDIF 信息 ● 标识：记录类型 ● 记录头的大小 ● 记录体的大小 | { 4a111440-e49f-11cf-9900-505144494600 } 标识信息包 (tagContainer) 64 byte 512 byte |
| 记录体 <ul style="list-style-type: none"> ● 以一个集合开始 ● 集合元素通过它在记录中的位置来标识，它与记录头中的标识是相对应的 | 集合 <i>Collection</i> <ul style="list-style-type: none"> ● 集合元素数据 12 元素 0 <ul style="list-style-type: none"> ● 标识符：文件名 ● 类型：向量数据或等级值 ● 物理类型：CHAR1 ● 链接(Link) ● 大小 (16 byte-padded from 13) |
| | 向量 <i>Vector</i> <ul style="list-style-type: none"> ● 向量元素 13(包括 NULL 终止符) ● 数据：“FILENAME.PQD” |

(2) PQDIF 的逻辑结构

逻辑结构由一系列记录组成，基本的逻辑结构包含 4 种类型的记录：容器记录 (Container record)、一个或多个数据源记录 (Data source record)、监测设置记录 (Monitor setting record) 和观测记录 (Observation record)。通过定义新的标识可以扩展记录，并且记录的顺序是由每个记录头文件中的绝对链接定义好的，其逻辑结构层次图如图 3-5 所示。容器记录比较特殊，每个 PQDIF 文件都是从一个

容器记录开始，作为查询目录记录了文件的信息和摘要。

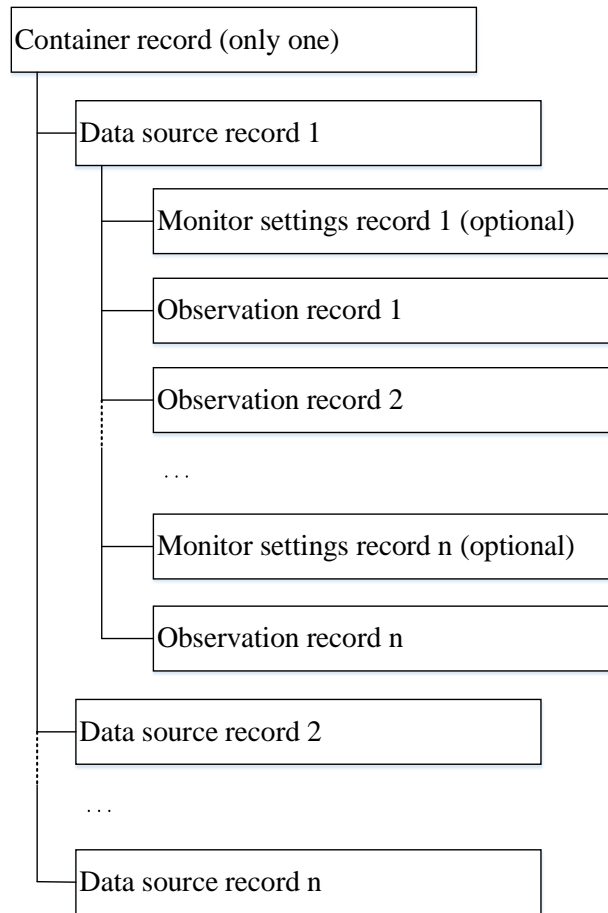


图 3-5 PQDIF 物理结构示意图

容器记录包含一个或多个监测设置记录和观测记录；数据源记录和观测记录的内部结构是并行的，并且相互链接为隐性链接，如图 3-6 所示。数据源记录可以包含许多通道定义，每个通道定义又可以包含许多数列定义。同时，观测记录可以包含许多通道实例，每个通道实例又有许多数列实例。这些定义和实例的结构类似于面向对象语言中的类和对象，并且设计的目的也相同。(1)通道定义：包括通道名称、相别、各种质量类型(波形、趋势等)及其它的名称信息，并在每个观测记录中无需重复说明。(2)数列定义：表达了测量量(电流、电压、功率等)、单位量(伏、安培)、特征量(瞬时、尖峰、有效值等)、数列值类型(最小、最大、平均等)、显示数据方式等数列信息，并在所有观测记录中都有效。(3)通道实例：通道实例主要提供与其它数列实例的链接。由于一个观测记录不是所有的通道定

义都能用到，通常只用到一部分，所以通道实例数要少于通道定义数。(4)数列实例：数列实例与通道实例有不同之处，因为每个数列实例必须与每个数列定义直接匹配，如果数列的拓扑结构有差异，则要为每个数列建立一个新的数列定义来与之相匹配。

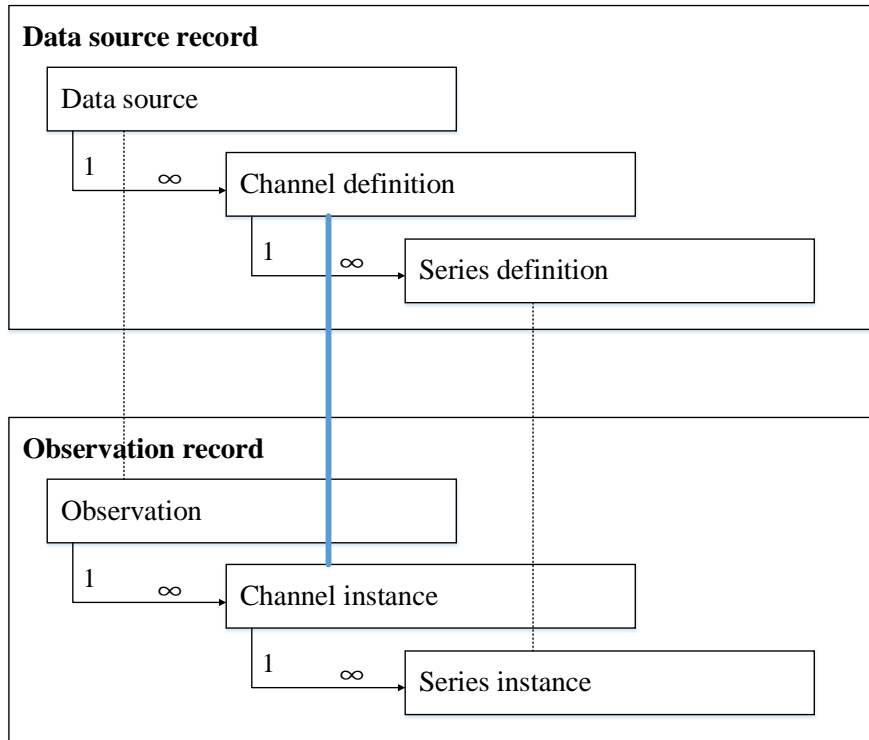


图 3-6 数据源记录和观测记录的并行结构

3.4 国内某供电公司电能质量在线监测平台

国内供电公司使用的软件监测的电能质量指标包括以下内容：稳态指标，包括频率、电压偏差、谐波、三相不平衡、闪变等；暂态指标，包括电压暂降、电压暂升、电压中断等。

具体地，该供电公司现有的电能质量监测系统可以对以下稳态指标进行实时监测、记录查询及图形分析。

(1) 实时监测：将当前监控设备采集的实时数据上传到数据库中并在程序中实时显示出来，默认每 3 秒钟刷新一次数据。

(2) 记录查询：对一定时间范围内的某一数据类型（综合数据、正常数据、

超限数据、实时数据) 的稳态数据进行查询。

(3) 图形分析：可显示为曲线、点状、柱状三种格式，同时提供保存、打印、三相叠加等功能。

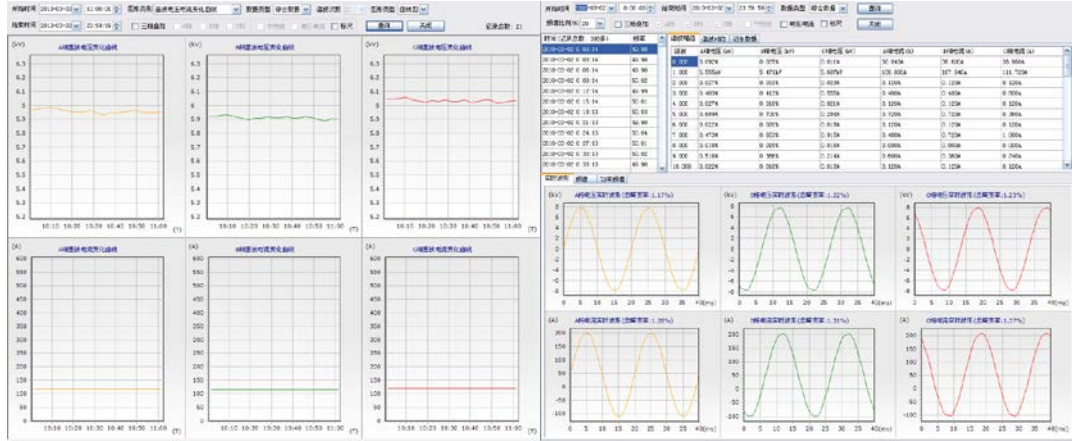
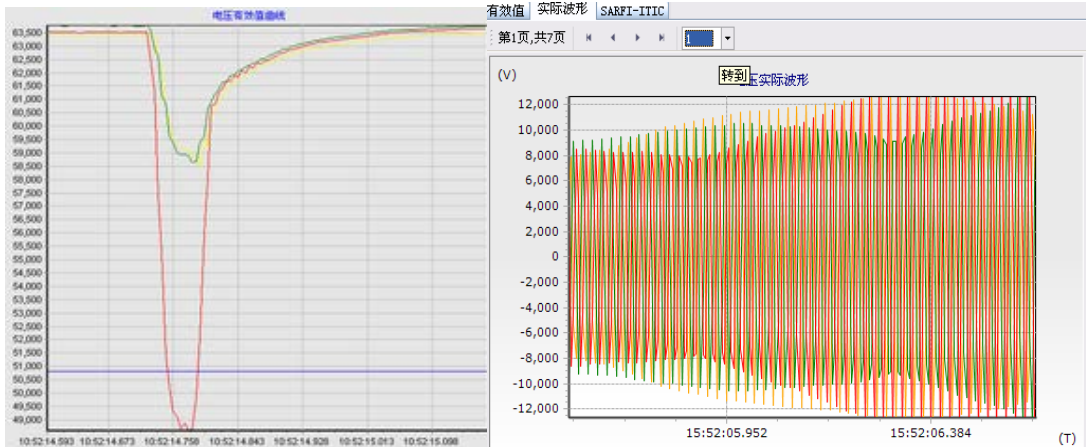


图 3-7 某供电公司电能质量监测稳态分析功能

该供电公司的电能质量监测系统还可以对暂态指标进行单次事件的电压、电流有效值、实际值、SARFI-ITIC 图形进行查看分析。目前还有站点的 10kV 侧没有安装具有暂态监测功能的电能质量在线式监测仪。



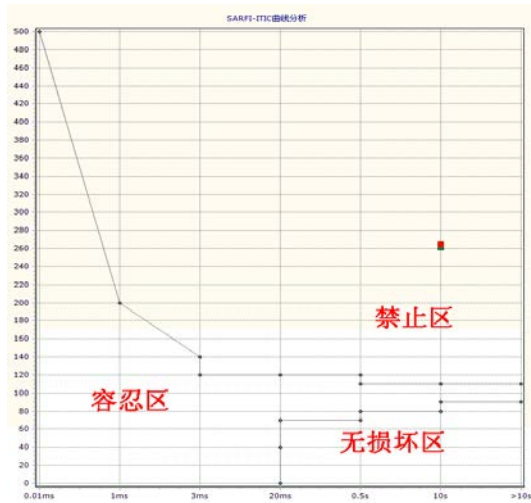


图 3-8 某供电公司电能质量监测暂态分析功能

该在线监测系统还可以按照监测点进行故障录波，记录电能质量扰动发生时的电压有效值的变化情况，如图 3-9 所示。

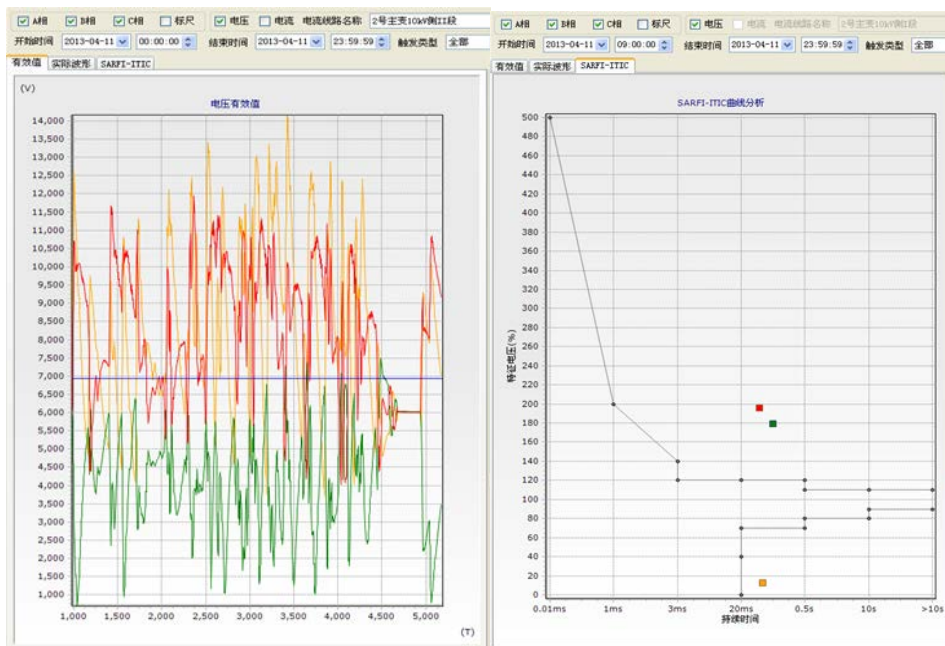


图 3-9 某供电公司电能质量监测点故障录波

该供电公司包含多个大型负荷中心，不仅有大型工业负荷，同时各种现代化的高科技企业密集分布，这些高端工业企业越来越依赖高质量的电力供应。并且该城市目前正面临产业调整和转型，高新产业比例日益扩大，对电能质量要求逐渐提高。近两年来，该供电公司已经接到数十家大型用户针对供电电能质量的应用需求和相关投诉，不完全统计如表 3-3。

表 3-3 某供电公司用户投诉情况

| 投诉时间 | 用户名称 | 投诉内容 | 负荷类型 | 影响情况 | 问题分类 |
|------------|------|---|-------------|--|-----------|
| 2010.05.17 | 公司 1 | 2010 年 5 月 7 日、5 月 16 日两次遭受外部原因造成严重停电事故，恢复时间分别超过 3 小时和 12 小时。 | 半导体制造企业 | 分别造成 50% 和 100% 设备停运。 | 雷击、电压跌落 |
| 2010.06 | 公司 2 | 跳闸 | 商业楼宇 | —— | 电压波动、电压暂降 |
| 2010.11.08 | 公司 3 | 2010 年成立以來，多次（3 月 3 日、5 月 7 日、11 月 6 日）出现电压异常波动造成工厂停电 | 电子产品生产线 | 每次停电造成 10 万片硬盘损失。 | 电压波动、电压暂降 |
| 2011.05.09 | 公司 4 | 2011 年已发生四次电压暂降，5/7、5/9 连续发生两次，造成设备停机。 | 薄膜太阳能电池板生产线 | 5/7 压降 2.87kV/260ms，5/9 压降 3.23kV/130ms；设备停机，影响生产。 | 电压暂降 |
| 2011.05.19 | 公司 5 | 2011 年 5 月 10 日专线跳闸停电 9 小时，5 月 13 日供电瞬时中断，5 月 16 日供电异常。 | 液晶面板生产线 | 供电中断损失 2500 片面板，供电异常造成设备停机。 | 供电中断、电压暂升 |
| 2011.07.25 | 公司 6 | 2011 年 7 月 21 日，测试发现供电三相不平衡 | —— | —— | 三相不平衡 |
| 2011.06.29 | 公司 7 | 2011 年 5 月 30 日至 6 月 25 日，机场 10kV 进线出现 8 次闪变（实为电压跌落） | 机场末端负荷 | 机场末端负荷瞬时失压跳闸，生产设备中断运转 | 电压跌落 |

从投诉情况来看，暂态电压问题（包括电压暂降、暂升和短时中断）是目前该城市电力用户反映最强烈的电能质量问题，暂态电压跌落造成敏感用户的生产线停运，给企业造成大量的经济损失。根据该供电公司对事故原因的分析结果，造成暂态电压跌落的原因包括雷击、系统短路故障、开关操作、大型负荷投退以及冲击性负荷的影响等等。该电网公司 2009 年 8 月到 10 月的电能质量合格率统计结果如表 3-4 和图 3-5 所示。

表 3-4 某供电公司电能质量合格率统计数据

| 月份 | 电压偏差 | 系统频率 | 电压 THD | 三相电压不平衡度 | 电压闪变值 | 监测点数量 | 变电站总数 | 10kV 监测点 | 110kV 监测点 |
|------|--------|---------|--------|----------|--------|-------|-------|----------|-----------|
| 08 月 | 98% | 100.00% | 98% | 98% | 99% | 270 | 67 | 201 | 69 |
| 09 月 | 98% | 100.00% | 98% | 98% | 99% | 488 | 113 | 373 | 115 |
| 10 月 | 98.16% | 100.00% | 96.93% | 100.00% | 98.20% | 488 | 113 | 373 | 115 |
| 11 月 | 97.54% | 100.00% | 96.72% | 100.00% | 96.10% | 488 | 113 | 373 | 115 |
| 12 月 | 97.91% | 100.00% | 97.39% | 100.00% | 97.13% | 575 | 131 | 442 | 163 |

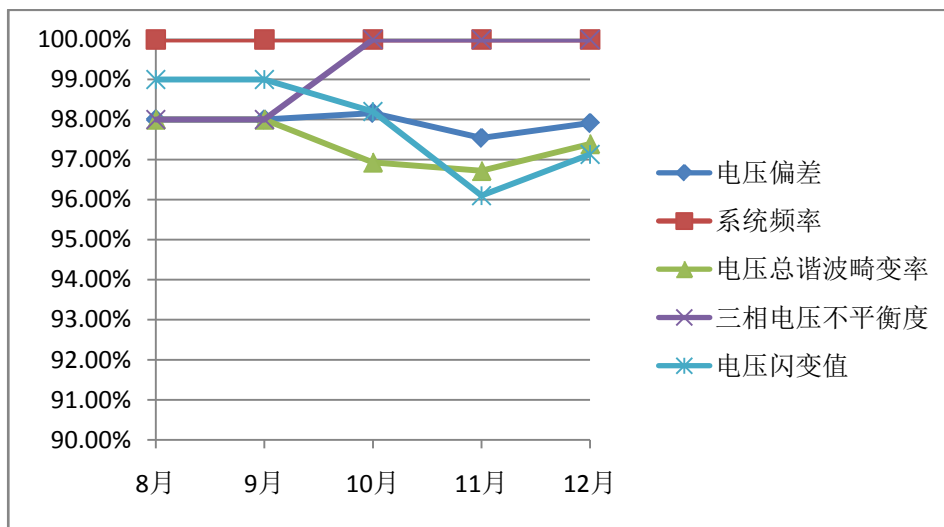


图 3-10 某供电公司稳态电能质量合格率变化趋势图

该电网公司 2009 年 8 月到 10 月的暂态电能质量监测数据统计结果如表 3-5 和图 3-6 所示。

表 3-5 某供电公司暂态电能质量事件统计数据

| 月份 | 变电站数量 | 暂态事件次数 | 暂升次数 | 暂降次数 | 短时中断次数 |
|------|-------|--------|------|------|--------|
| 08 月 | 21 | 121 | 2 | 95 | 24 |
| 09 月 | 22 | 32 | 4 | 22 | 6 |
| 10 月 | 12 | 33 | 3 | 21 | 9 |
| 11 月 | 13 | 26 | 2 | 19 | 5 |
| 12 月 | 18 | 54 | 3 | 35 | 16 |
| 合计 | 86 | 266 | 14 | 192 | 60 |

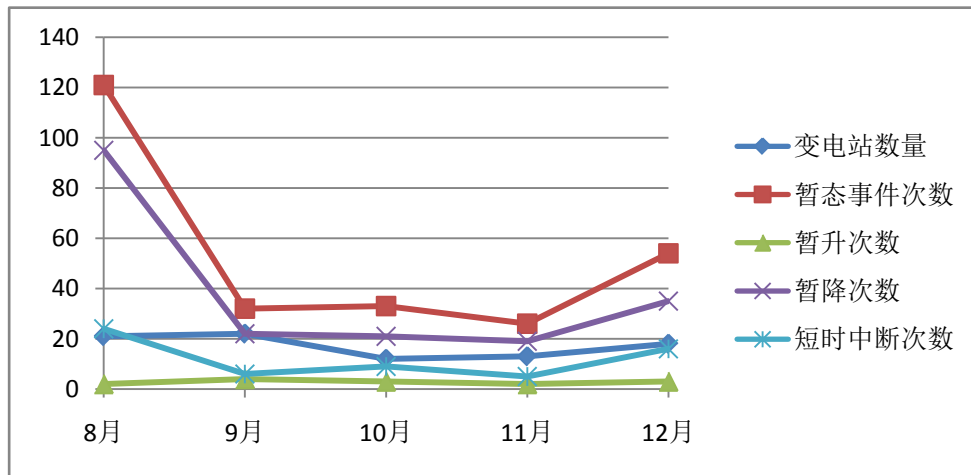


图 3-11 某供电公司暂态事件变化趋势图

引起该供电公司暂态电能质量扰动的原因非常复杂，按照其原因类型可以分为故障类和非故障类，故障类扰动原因主要为各种短路故障；非故障类扰动原因包括变压器激磁涌流、感应电机的启动、电弧炉、轧钢机等冲击性负荷的投运等，与前者相比并不太严重。实际系统中，危害性较大的主要是由系统短路故障引起的，它的传播距离远、跌落幅值大。因此，当主网发生短路故障时，电压暂降会传播渗透到同级或下级电网中。除了短路故障的原因，大容量感应电机启动和雷击也是引起电压暂降的最主要原因。

1) 短路故障引起的电压暂降。当系统发生短路故障时，根据与故障点间的距离，各母线会出现不同程度的电压暂降。由于这类故障发生概率比较大，所以也是敏感设备跳机或误动作的主要原因，二相短路故障引起的电压暂降最为严重。

2) 大容量感应电机启动引起的电压暂降。在电网总负荷中，感应电动机用电量约占 60% 以上。大容量感应电动机启动时，定子电流明显增大，可高达额定电流的 5 至 6 倍，从而引起所接母线发生电压暂降。电压暂降严重程度与感应电动机的容量、启动方式及负荷等因素有关。这种暂降的持续时间较长，暂降程度一般较小，不会对用户造成严重的影响。

3) 雷击时造成的绝缘子闪络或线路对地放电会是保护装置动作，从而导致供电电压暂降，这在多雷区尤为明显。这种暂降的影响范围大，持续时间一般超

过 100ms。

此外，变压器投切会引起电压暂降。由于铁心饱和效应，空载变压器投切时可能产生很大的激磁涌流，从而引起电压暂降。该过程中常伴有二次和四次谐波为主的谐波。由于三相铁心饱和程度的不同，所以三相电压暂降程度也不同。电压暂降程度与开关合闸时刻、电源强度、铁心中的剩磁和网络阻尼有关。通常，由变压器投切引起电压暂降不会低于 85%。

4 干扰源与敏感用户分类供电方案

(分类方法、供电方案、算例、从电能质量到体验质量论文)

4.1 术语和定义

干扰源和敏感用户分类是指以客户的各类用电负荷对于电能质量问题的敏感程度、以及用电负荷对于电网的污染程度作为依据，将用电客户划分为不同档次的干扰用户或敏感用户。分类原则是以客户的各类用电负荷对于电能质量问题的敏感程度、以及用电负荷对于电网的污染程度作为依据的电能质量的干扰源与敏感用户分类供电中使用的术语和定义如下：

- (1) 用户：指可能或已经与供电方建立供用电关系的组织或个人。
- (2) 供电服务：供电方遵循供电质量方面的国家标准和规范，提供合格的电能产品和满意的服务，来满足用户的现实和潜在用电需求。
- (3) 电能质量：供电方提供供电服务的过程中，导致用电设备故障或不能正常工作的电压、电流或频率的偏差。
- (4) 涉及的电能质量扰动包括：电压暂降、谐波、三相不平衡、电压暂升、短时中断、频率偏差、电压偏差、电压闪变等。
- (5) 负荷敏感度：客户的敏感负荷/敏感负荷对于电能质量扰动的耐受程度。
- (6) 负荷污染度：客户的非线性/冲击性负荷对电力系统造成的污染程度。
- (7) 分类供电技术方案：指根据用户的污染度和敏感度制定相应的供电方案，保证敏感用户的供电质量，控制干扰源用户对电网的污染。

4.2 干扰源与敏感用户分类方案

4.2.1 用户负荷敏感特性分析

对敏感用户进行敏感度评估和分类，有利于电力企业和敏感用户降低供用电风险、技术改造、差异化定制电力等有重要意义，是现阶段全面提高城市电能质

量水平的前提条件。在对用户进行负荷敏感特性分析的基础上，结合电能质量单项健康指数和用户电能体验指数，对敏感用户进行敏感度评估和分类，所得分析结果为有针对性地提高敏感用户电能质量提供可靠依据。

负荷敏感度是指客户的敏感负荷/敏感元件对于电能质量问题的耐受程度。负荷敏感度评估即定量地评价各类电能质量问题对客户负荷的影响程度，通常以电能质量问题引起负荷故障的概率来量化。各类电能质量扰动中，对于用电客户敏感负荷影响最大的是电压暂降和谐波问题。因此，本报告主要将负荷敏感度分析和评估集中于这两种电能质量扰动。在实际应用中，部分用电客户对电能质量有较高的要求，满足电能质量国家标准的电能，仍有可能不能满足这些客户的需求。并且对于这类客户，即使出现轻微的电能质量问题，也会导致严重的用电事故和经济损失，此类用户称为敏感客户^[42]。

本报告中用电客户的分类建立在客户负荷敏感特性分析的基础上。通常在供电方与用电客户之间签署的供用电合同中需要详细规定用电客户的设备和负荷信息，如主要设备类型、报装容量、设备工艺等^{[43][44]}。对电能质量要求较高的用电客户还需要详细说明其敏感设备的相关信息，以便供电方评估和分析客户负荷的敏感特性。敏感设备和敏感负荷是指当出现的电能质量扰动时，有很大概率出现故障或停止工作的设备，常见的敏感设备包括计算机、交流接触器、可编程逻辑控制器、可调速电机、高压气体放电灯泡等。

对于已经签署了明确、详尽的供用电合同的客户，直接根据合同中规定的客户敏感设备类型、容量和工艺对客户进行敏感度评估和分类。如果没有签署供用电合同或者供用电合同不够全面详细，并且用电客户无法明确获得其敏感设备的类型、容量和工艺等信息，则需要由供电方或第三方机构进行客户负荷敏感度实测，根据实测结果进行客户负荷敏感度的试验评估和分类。

客户敏感负荷调研的范围主要包括客户的所有敏感负荷，即对客户所有的用

电设备进行全面分析，获取客户各种用电设备的负载、电气参数及使用情况，分析这些设备和元件的负荷特性，确定敏感负荷的敏感元件、敏感类型、敏感程度、负荷容量等，在此基础上进行客户负荷敏感度评估。

4.2.2 用户负荷敏感度评估

当用电客户已经和供电方签署了明确、详尽的供用电合同，并且合同中规定了客户设备类型、容量、工艺以及敏感设备所占的比例等信息，可直接根据供用电合同中相关的信息对客户进行敏感度评估和分类。通过对供电方的调研发现，实际使用中对用电客户影响较大的电能质量问题主要包括电压暂降和谐波两类，且敏感设备往往同时受到多种电能质量的共同影响。因此，本报告主要根据供用电合同进行电压暂降敏感度评估和谐波敏感度评估，并在此基础上进行了综合敏感度评估。**(1) 电压暂降敏感度评估**

电压暂降通常由于电力系统受到干扰所引起。由于目前电力系统中对电压暂降敏感的负荷越来越多,电压暂降已经成为造成电压敏感设备不能正常工作的主要原因,调研表明在各类电能质量问题中由电压暂降造成的用户损失占电能质量损失的 80% 以上^[45]。例如：计算机、可调速电机、交流接触器、可编程逻辑、半导体工厂和轧钢厂等都是电压敏感负荷。

对于常见的敏感设备，综合了 ITIC 曲线和 SEMI 曲线等电压耐受力评估曲线，得到各类敏感设备对应的敏感区域，如表 4-1 所示。电压暂降事件典型的持续时间取 20ms，50ms，100ms，200ms，500ms，3s 及以上等级别。电压暂降事件典型的残压值取 90%，80%，70%，60%，50%，40% 及以下等级别。其中，A 区域代表精密机械设备和控制器会受影响的区域，B 区域代表半导体生产受影响的区域，C 区域代表计算机类设备、PLC、交流继电器以产企业受影响的区域，D 区域代表直流电机、可变速电机以及金属钠灯受影响的区域，E 区域为所有敏感设备的敏感区域。

表 4-1 常见敏感设备敏感区域

| 残压 U (%) | 持续时间 t (s) | | | | | | |
|-------------|------------|-----------------|----------------|---------------|---------------|-------------|-------|
| | t ≤ 0.02 | 0.02 < t ≤ 0.05 | 0.05 < t ≤ 0.1 | 0.1 < t ≤ 0.2 | 0.2 < t ≤ 0.5 | 0.5 < t ≤ 3 | t > 3 |
| 80 ≤ U < 90 | A | A | A | A | A | A | A |
| 70 ≤ U < 80 | A | A | A | A | A | C | C |
| 60 ≤ U < 70 | A | B | B | B | C | C | C |
| 50 ≤ U < 60 | A | D | D | D | E | E | E |
| 40 ≤ U < 50 | A | D | E | E | E | E | E |
| U < 40 | A | D | E | E | E | E | E |

对应于表 4-1 中 A、B、C、D、E 五个区域的设备，其对应的负荷敏感度级别分别为 I、II、III、IV、V。例如，在 A 区域容易出现故障的负荷，其负荷敏感度就是 A。为了对已知敏感设备进行敏感度评估，在表 4-1 的基础上得到如表 4-2 所示的电压暂降负荷敏感度划分。

表 4-2 电压暂降负荷敏感度划分

| 残压 U (%) | 持续时间 t (s) | | | | | | |
|-------------|------------|-----------------|----------------|---------------|---------------|-------------|-------|
| | t ≤ 0.02 | 0.02 < t ≤ 0.05 | 0.05 < t ≤ 0.1 | 0.1 < t ≤ 0.2 | 0.2 < t ≤ 0.5 | 0.5 < t ≤ 3 | t > 3 |
| 80 ≤ U < 90 | I | I | I | I | I | I | I |
| 70 ≤ U < 80 | I | I | I | I | I | III | III |
| 60 ≤ U < 70 | I | II | II | II | III | III | III |
| 50 ≤ U < 60 | I | IV | IV | IV | V | V | V |
| 40 ≤ U < 50 | I | IV | V | V | V | V | V |
| U < 40 | I | IV | V | V | V | V | V |

客户的电压暂降敏感度评估可通过以下步骤完成：

- 通过供用电合同确定客户所使用的电压暂降敏感负荷类型、工艺和容量。
- 根据表 4-2 确定客户各类敏感负荷的敏感度级别。如果设备的敏感度曲线仅穿过表 4-2 的一个区域，则其敏感度则匹配为该敏感区域相同；如果设备的敏感度曲线穿过了表 4-2 的多个区域，则匹配为其中敏感级别最高的区域。

(2) 谐波敏感度评估

谐波是指工频整数倍频率的正弦电压和电流。电力系统的谐波是由供电网络的非线性设备或由供电网络供电的用户设备发射的谐波电流和由其引起的系统电压畸变。谐波对电气设备的影响主要表现在产生附加的功率损耗即热效应、减少绝缘寿命和影响设备安全正常运行上。

设备谐波敏感度的等级划分主要针对 10kV 电力系统而言，使用谐波电压总畸变率 THD_u 作为衡量谐波耐受度指标。谐波敏感负荷对应的敏感度级别分别为 I 至 V 级，其能够耐受的谐波电压总畸变率 THD_u 如表 4-3 所示。

表 4-3 谐波设备敏感度综合等级划分

| | | 可耐受的谐波电压总畸变率 THD_u |
|-----------|-----|-------------------------|
| 谐波设备敏感度等级 | I | $THD_u \leq 3\%$ |
| | II | $3\% < THD_u \leq 4\%$ |
| | III | $4\% < THD_u \leq 8\%$ |
| | IV | $8\% < THD_u \leq 10\%$ |
| | V | $THD_u > 10\%$ |

客户的电压暂降敏感度评估可通过以下步骤完成：

- 通过供用电合同确定客户所使用的谐波敏感负荷类型、工艺和容量。
- 根据表 4-3 的指标确定客户各类敏感负荷的敏感度级别。

4.2.3 用户负荷敏感度实测

在对客户敏感负荷进行调研的基础上，对于供用电合同中没有明确规定敏感设备类型、容量和工艺的客户，需要进行负荷敏感度实测。客户负荷敏感度实测评估是指在模拟电能质量问题下，通过多次试验得出负荷的故障概率，并根据故障概率评估出负荷敏感度。

为了使负荷敏感度标准评估和负荷敏感度实测评估有一个统一的度量标准，首先将电能质量问题按严重程度划分为 5 个级别，然后在等级划分的基础上进行

敏感度评估，将负荷的敏感度也相应地划分为 5 个级别（I~V 级）。对于一个具体的用电客户，其单项电能质量问题的敏感度级别由敏感度最高的设备所决定。

1) 电能质量扰动等级划分

各类电能质量扰动问题中，对于用电客户敏感负荷影响最大的是电压暂降和谐波问题。报告中主要规定了电压暂降和谐波问题的等级划分，并在不同等级上进行负荷的敏感度实测评估。

①电压暂降划分

虽然电压暂降有明确的定义，但是由于电压暂降由电压暂降持续时间、电压暂降幅值多方面的特征所决定，目前没有一个可以得到国际上普遍认可的电压暂降指标。报告中采用统计表格的形式来界定电压暂降的级别划分，表格的纵列代表电压暂降持续时间的范围，横列表示电压暂降残压幅值范围，表内每个元素给出了当前幅值对应的电压暂降事件等级。电压暂降事件等级规定为 A~E 级别区域，反映了电压暂降的严重程度。

该表格的制定主要参考了 UNIPEDA DISDIP 组织推荐的负荷敏感特性，并综合了 ITIC 曲线和 SEMI 曲线等电压耐受力评估曲线，对电压暂降表格设计如表 4-4 所示^[46]。电压暂降事件典型的持续时间取 20ms, 50ms, 100ms, 200ms, 500ms, 3s 及以上等级别。电压暂降事件典型的残压值取 90%, 80%, 70%, 60%, 50%, 40% 及以下等级别。

表 4-4 电压暂降等级划分

| 残压 U(%) | 持续时间 t (s) | | | | | | |
|------------------|---------------|----------------------|---------------------|--------------------|--------------------|------------------|---------|
| | $t \leq 0.02$ | $0.02 < t \leq 0.05$ | $0.05 < t \leq 0.1$ | $0.1 < t \leq 0.2$ | $0.2 < t \leq 0.5$ | $0.5 < t \leq 3$ | $t > 3$ |
| $80 \leq U < 90$ | A | A | A | A | A | A | A |
| $70 \leq U < 80$ | A | A | A | A | A | C | C |
| $60 \leq U < 70$ | A | B | B | B | C | C | C |
| $50 \leq U < 60$ | A | D | D | D | E | E | E |

| | | | | | | | |
|---------|---|---|---|---|---|---|---|
| 40≤U<50 | A | D | E | E | E | E | E |
| U<40 | A | D | E | E | E | E | E |

②谐波等级划分

国家标准 GB/T 14549-1993《电能质量公用电网谐波》中不仅给出了谐波用户谐波电流发射限值，同时也给出了电网谐波电压畸变限值，这是供电公司需要给予保证的谐波电压水平。谐波敏感用户对谐波畸变程度有特殊的要求。

使用谐波电压总畸变率 THD_u 作为谐波等级划分的指标。谐波电压总畸变率的定义如公式 4-1 所示。

$$THD_u = \frac{\sqrt{\sum_{n=2}^N U_n^2}}{U_1} \quad \text{式 4-1}$$

谐波等级划分的依据如下：

A 级： $THD_u \leq 3\%$

B 级： $3\% < THD_u \leq 4\%$

C 级： $4\% < THD_u \leq 8\%$

D 级： $8\% < THD_u \leq 10\%$

E 级： $THD_u > 10\%$

由于国家标准 GB/T 14549-1993《电能质量公用电网谐波》中规定 10kV 公共电网谐波电压总畸变率不高于 4%，即谐波等级为 A 或 B 符合规定的供电标准。此情况下仍可能出现故障或误动的设备定义为谐波敏感设备。

2) 负荷敏感度实测评估方案

对于客户使用的敏感特性未知的负荷，由于敏感负荷设备在不同运行环境、工况和不同电能质量扰动下的敏感特征不同，受供电系统和负荷本身众多不确定因素的影响，不能直接根据负荷特性进行敏感度评估。因此，需要使用实测法评估其敏感度，即通过大量试验样本的监测电能质量问题发生时设备故障和误动的

概率，根据概率高低确定设备敏感度。

①电压暂降敏感度评估

对于客户敏感负荷，其负荷的电压暂降敏感特性无法获知，需通过多次试验的方法进行评估。通过监测负荷供电点的电压暂降特征并与设备电压耐受水平进行比较来确定负荷敏感度。其具体步骤如下：

- 准备可以进行试验测试的敏感设备，模拟负荷运行场景。
- 将供电点和电能计量点设置为供电部门与用户的产权分界处或由供用电协议所规定分界点。
- 试验方案：根据表 4-4 定义的电压暂降等级划分，在每个等级上随机产生大量的试验样本，在供电点模拟这些试验样本对应的电压暂降场景。通过实测获取在当前敏感设备的故障和误动的概率。不同电压暂降等级上试验样本的数目和故障概率计量表如表 4-5 所示。

表 4-5 电压暂降设备敏感度测试方案

| | 电压暂降级别 | | | | |
|----------|--------|-------|-------|-------|-------|
| | A | B | C | D | E |
| 试验次数 | 10 | 20 | 20 | 20 | 10 |
| 故障概率 (%) | 0~100 | 0~100 | 0~100 | 0~100 | 0~100 |

- 根据故障概率评估敏感级别。假设实验中获取的故障概率分别为 P_A 、 P_B 、 P_C 、 P_D 、 P_E ，设备敏感级别与故障率的对应关系如下：

I 级：若 $P_A \geq 10\%$

II 级：若 $P_B - P_A \geq 10\%$ 或者 $P_B \geq 10\%$

III 级：若 $P_C - P_B \geq 10\%$ 或者 $P_C \geq 10\%$

IV 级：若 $P_D - P_C \geq 10\%$ 或者 $P_D \geq 10\%$

V 级：若 $P_E - P_D \geq 10\%$ 或者 $P_E \geq 10\%$

②谐波敏感度评估

对于未知敏感负荷，其负荷的谐波敏感特性无法获知，需通过多次试验的方法进行评估。通过模拟各个电网和用户在公共连接点(PCC)产生的五个等级的谐波水平，测试设备的谐波耐受水平来确定负荷敏感度。

试验方案：将供电点和电能计量点设置为各个电网和用户在公共连接点(PCC点)，根据谐波水平等级划分，在每个等级上随机产生大量的试验样本，在供电点模拟这些试验样本对应的谐波水平场景。通过实测获取在当前敏感设备的故障和误动的概率。不同谐波水平等级上试验样本的数目和故障概率计量表如表 4-6 所示。

表 4-6 谐波水平备敏感度测试方案

| | 谐波水平级别 | | | | |
|---------|--------|-------|-------|-------|-------|
| | A | B | C | D | E |
| 试验次数 | 20 | 20 | 20 | 20 | 10 |
| 故障概率(%) | 0~100 | 0~100 | 0~100 | 0~100 | 0~100 |

I 级：若 $P_A \geq 10\%$

II 级：若 $P_B - P_A \geq 10\%$ 或者 $P_B \geq 10\%$

III 级：若 $P_C - P_B \geq 10\%$ 或者 $P_C \geq 10\%$

IV 级：若 $P_D - P_C \geq 10\%$ 或者 $P_D \geq 10\%$

V 级：若 $P_E - P_D \geq 10\%$ 或者 $P_E \geq 10\%$

4.2.4 用户负荷污染度评估

负荷污染度评估是定量地评价用电客户的电气设备/元件运行中对电网电能质量造成的污染程度。产生电能质量污染的电气设备主要是用户的用电设备，通常称为非线性负荷，即指在正弦对称电压作用下，负荷电流为非正弦或不对称的负荷，以及生产（运行）过程中从电力系统取用快速变动功率的冲击性负荷。非线性负荷一般包括：冶金行业的交流电弧炉、直流电弧炉、钢包精炼炉、中频炉、

轧钢设备、电解设备；交通运输行业的电气化铁路、地铁、轻轨、有轨和无轨电车；其它行业的化工整流负荷、变频调速设备、电焊设备、低压用电设备（开关电源、节能灯、UPS）等。

实际电网中的用户不一定是单一电能质量污染源，有可能存在多种电能质量问题；也许各自电能质量指标都合格，但入网后相互叠加影响，可能引起电网某节点电能质量超标。通常将干扰用户划分为 5 个级别，依据相关国家标准中的规定，通过对用户电气设备在入网 PCC 点处产生的谐波、电压波动和闪变、三相不平衡度等电能质量指标进行实际测试，根据测试结果综合评估得出干扰用户的污染级别程度。

GB/T 14549-1993《电能质量公用电网谐波》中规定 10kV 公共电网谐波电压总畸变率不高于 4%，奇次谐波电压含有率不高于 3.2%，偶次谐波电压含有率不高于 1.6%；而且规定了用户向 PCC 点注入各次谐波电流的限值，可折算各用户注入的谐波电流允许值。

GB/T 12326-2008《电能质量电压波动和闪变》规定中低压 PCC 点电压波动限值为 4%，10min 内闪变值不超过 1.0，2h 内闪变值不超过 0.8，电压波动不考虑冲击性负荷的叠加效应，闪变限值要根据每个用户的协议用电容量占供电容量的比例分配。

GB/T 15543-1995《三相电压允许不平衡度》规定正常电压不平衡度允许值为 2%，短时不得超过 4%；每个用户在 PCC 点引起的电压不平衡度允许值一般为 1.3%，根据实际情况可适当变动。

GB/T 15945-1995《电力系统频率偏差》规定冲击负荷引起的系统频率变动一般不得超过 $\pm 0.2\text{Hz}$ ，也可适当变动。

4.3 干扰源与敏感用户分类实施和算例

在实际操作中，敏感设备和干扰设备同时与多种电能质量扰动相互影响，用

户分类难以直接计算。本节的干扰源与敏感用户分类方法主要根据实际情况制定，充分考虑了分类的可实施性，其分类的主要依据是用户负荷敏感度、用户敏感负荷重要度、用户敏感负荷容量等^{[47][48]}。用户负荷敏感度由供电合同评估或实测评估得出。用户负荷重要度由供电方根据用户负荷所属行业和重要程度决定。用户敏感负荷容量也是用户分类的重要指标之一。

4.3.1 干扰源与敏感用户分类实施方案

干扰源与敏感用户分类的实施方案如下：

- (1) 用户向供电方申报其用电相关的基本信息，包括报装容量、所属行业、注册资金等信息。
- (2) 在供电方的指导下，用户提供其主要用电设备的负荷特性，包括主要敏感负荷、干扰性负荷、正常负荷信息，并确定各类负荷的容量、负荷敏感类型、负荷干扰特性等。
- (3) 现场审核用户提供的基本用电信息，确认用户提供的基本用电信息属实。
- (4) 根据用户基本用电信息分析其使用的干扰性负荷、敏感负荷，初步确定为潜在干扰源用户和敏感用户。
- (5) 对于潜在敏感用户，首先根据表 4-7 确定用户各种敏感负荷的敏感度级别和评分，按照表 4-8 确定用户的重要度级别和评分。

表 4-7 用户负荷敏感度评定标准

| 负荷类型 | 敏感度级别 | 敏感度评分 | 敏感类别 |
|-------------|-------|-------|---------|
| 精密设备控制器 | I | 100 | 电压暂降 |
| 可编程控制器(PLC) | III | 60 | 电压暂降 |
| 交流继电器 | III | 60 | 电压暂降 |
| 计算机类设备 | III | 60 | 电压暂降、谐波 |
| 数控机床 | I | 100 | 电压暂降 |

| | | | |
|----------|-----|-----|-------|
| 光刻机 | II | 80 | 电压暂降 |
| 中束流离子源 | IV | 40 | 电压暂降 |
| 可变速电机 | IV | 40 | 电压暂降 |
| 气体放电类电光源 | IV | 40 | 电压暂降 |
| 超声波铝丝焊接机 | I | 100 | 电压暂降 |
| 核磁共振成像 | III | 60 | 三相不平衡 |
| 高频电刀 | III | 60 | 电压暂降 |

负荷敏感度级别与敏感度评分之间的关系定义为：I级=100分；II级=80分；III级=60分；IV级=40分；V级=20分。

表 4-8 用户行业重要度对照

| 序号 | 行业类型 | 重要度级别 | 重要度评分 |
|----|-------|-------|-------|
| 1 | 半导体制造 | I | 100 |
| 2 | 精密制造 | I | 100 |
| 3 | 金融 | I | 100 |
| 4 | 商业中心 | III | 50 |
| 5 | 通信 | II | 75 |
| 6 | 公共服务 | IV | 25 |
| 7 | 休闲服务 | IV | 25 |
| 8 | 医疗 | I | 100 |

用户重要度级别与用户重要度评分之间的关系定义为：特级为 100 分；一级为 75 分；二级为 50 分；三级为 25 分。

在得到用户负荷敏感度、用户重要度后，按照如式 4-2 所示的公式得到用户的综合敏感度。

$$S = \lambda_1 \times \sum_{i=1}^n \frac{C_i \times \alpha_i \times \gamma_i}{C} + \lambda_2 \times \beta \quad \text{式 4-2}$$

其中，S为客户的综合敏感度评分；n为客户所包含的敏感负荷的种类数；i为

敏感负荷类型标识符； C_i 为敏感负荷类型*i*的容量； C 为客户敏感负荷总容量； λ_1 ， λ_2 为用户负荷敏感度、用户重要度两个参数的权重值，且 $\lambda_1+\lambda_2=1$ ，通常取 $\lambda_1=0.6$ ， $\lambda_2=0.4$ ； γ_i 表示负荷敏感类型的关注度，如电压暂降的关注度为1，谐波的关注度为0.5，三相不平衡的关注度为0.5； α_i 为敏感负荷类型*i*的敏感度评分； β 为用户的重要度评分。

最终，得到的综合敏感度评分与客户综合敏感度分类的对应关系如下：

- I类敏感用户： $S \in [80,100]$
- II类敏感用户： $S \in [60,80)$
- III类敏感用户： $S \in [40,60)$
- IV类敏感用户： $S \in [20,40)$
- V类敏感用户： $S \in [0,20)$

用户负荷敏感度评估的实施流程如图4-1所示。

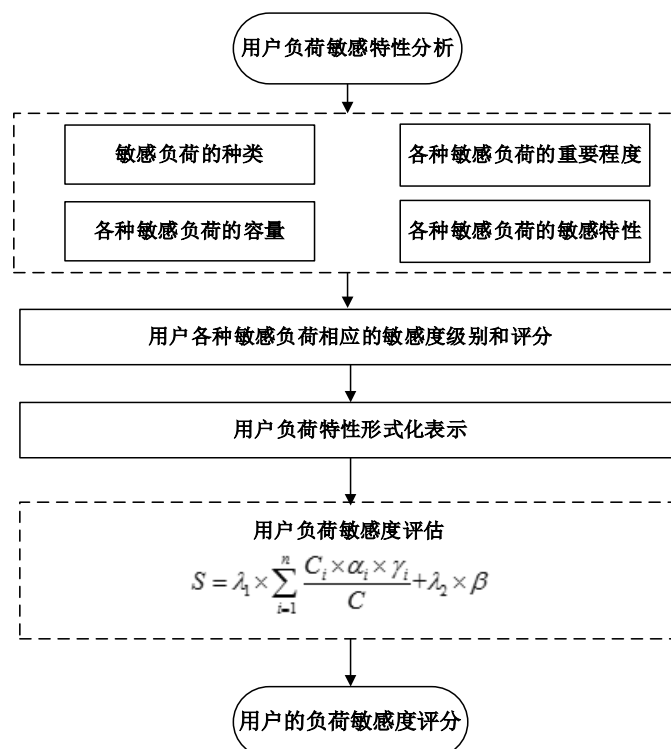


图4-1 用户负荷敏感度评估

(7) 对于潜在干扰源用户，按照以下方法和算例评估其污染度等级：按照表

4-9 确定用户干扰负荷的污染度级别和评分，根据表 4-10 确定用户的干扰负荷容量评分。

表 4-9 用户负荷污染度评定标准

| 负荷类型 | 污染度级别 | 污染度评分 | 干扰类别 |
|----------|-------|-------|-------|
| 电力机车 | I | 100 | 谐波 |
| 铝电解槽 | II | 80 | 谐波 |
| 充电装置 | IV | 40 | 谐波 |
| 逆变电源 | IV | 40 | 谐波 |
| 变频器 | I | 60 | 谐波 |
| 电弧炉、电石炉 | I | 100 | 三相不平衡 |
| 气体放电类电光源 | III | 60 | 谐波 |
| 变压器 | IV | 40 | 谐波 |
| 电抗器 | II | 80 | 谐波 |
| 电容器 | II | 80 | 谐波 |
| 交流弧焊机 | I | 100 | 谐波 |
| 晶闸管整流器 | I | 100 | 谐波 |

负荷污染度级别与污染度评分之间的关系定义为：I 级=100 分；II 级=80 分；III 级=60 分；IV 级=40 分；V 级=20 分。

表 4-10 用户干扰负荷容量级别

| 干扰负荷容量(KVA) | 容量级别 | 容量评分(0-100 分) |
|--------------|------|---------------|
| 5000 以上 | I | 100 |
| (2000, 5000] | II | 75 |
| (500, 2000] | III | 50 |
| (0, 500] | IV | 25 |

在得到用户负荷污染度、用户污染负荷容量后，按照如式 4-3 所示的公式得

到用户的综合污染度。

$$S = \lambda_1 \times \sum_{i=1}^n \frac{C_i \times \alpha_i \times \gamma_i}{C} + \lambda_2 \times C' \quad \text{式 4-3}$$

其中， S 为客户的综合污染度评分； n 为客户所包含的干扰负荷的种类数； i 为干扰负荷类型标识符； C_i 为干扰负荷类型 i 的容量； C 为客户干扰负荷总容量； λ_1 ， λ_2 为用户负荷污染度、用户干扰负荷容量两个参数的权重值，且 $\lambda_1 + \lambda_2 = 1$ ，通常取 $\lambda_1 = 0.6$ ， $\lambda_2 = 0.4$ ； γ_i 表示负荷污染类型的关注度，如谐波的关注度为1，三相不平衡的关注度为0.5； α_i 为干扰负荷类型 i 的污染度评分（0-100分）； C' 为用户干扰负荷的容量评分。

最终，得到的综合污染度评分与用户综合污染度分类的对应关系如下：

- I类干扰源用户： $S \in [80, 100]$
- II类干扰源用户： $S \in [60, 80)$
- III类干扰源用户： $S \in [40, 60)$
- IV类干扰源用户： $S \in [20, 40)$
- V类干扰源用户： $S \in [0, 20)$

4.3.2 干扰源与敏感用户分类算例

(1) 敏感用户算例

例如用户A包含两种敏感负荷，负荷敏感度分别为I级和II级，敏感类型分别为电压暂降和谐波，容量分别为1000KVA和2000KVA，并且其行业为半导体制造行业。根据以上情况，对应的参数值为：

$$C_1 = 1000, C_2 = 2000, C = 3000, \gamma_1 = 1, \gamma_2 = 0.5$$

$$\alpha_1 = 100, \alpha_2 = 80, \beta = 100$$

客户A的综合敏感度评分为：

$$S = 0.6 \times \left(\frac{1000 \times 100 \times 1 + 2000 \times 80 \times 0.5}{3000} \right) + 0.4 \times 100$$

$$= 76$$

根据综合敏感度评分，客户 A 可以评定为 II 类敏感客户。

(2) 干扰源用户算例

例如用户 A 包含两种敏感负荷，负荷敏感度分别为 I 级和 II 级，敏感类型分别为电压暂降和谐波，容量分别为 1000KVA 和 2000KVA，并且其行业为半导体制造行业。根据以上情况，对应的参数值为：

$$C_1 = 1000, C_2 = 2000, C = 3000, \gamma_1 = 1, \gamma_2 = 0.5 \\ \alpha_1 = 100, \alpha_2 = 80, \beta = 100$$

客户 A 的综合敏感度评分为：

$$S = 0.6 \times \left(\frac{1000 \times 100 \times 1 + 2000 \times 80 \times 0.5}{3000} \right) + 0.4 \times 100 \\ = 76$$

根据综合敏感度评分，客户 A 可以评定为 II 类敏感客户。

4.4 干扰源与敏感用户供电方案

干扰与敏感用户分类方案主要包括供电电源配置、用户接入方式选择和电能质量扰动治理等方面。供电电源配置依据用户的负荷重要度、敏感度等级、用电性质、用电容量、当地供电条件等因素进行技术经济比较，与客户协商确定；用户接入方式选择根据用户的负荷等级、敏感度等级、用户需求等方面，以保证用户的供电可靠性，并满足节约投资和便于扩建等要求；电能质量扰动治理则是根据用电客户负荷的污染特性和负荷的敏感特性，按需求为客户定制治理装置，提高敏感用户抵抗电能质量扰动的能力，并减轻干扰用户对电网的污染。电力用户供电方案应遵循以下原则：①安全性，应满足电网运行和用户用电安全的要求，确保对电网电能质量的影响满足国家标准的规定；②可靠性，供电电源、应急电源的供电线路、接线方式、运行方式等选择应合理可靠，满足对用户供电可靠性的要求；③经济性，满足用户近期和远期对电力的需求，变压器容量、台数选用应适当，无功补偿装置配置符合国家和电力行业标准规定，计量方式、计量点设置、计量装置选型配置正确，电力设施维护管理责任划分明确；④合理性，用户

用电性质应分类正确，供电电压选择合理，并根据地形、地貌和道路规划要求就近选择接入电源点，确保用户受电端有合格的电能质量。

干扰源与敏感用户分类供电方案建立在根据客户分类的基础上，根据客户的污染程度、敏感程度制定相应的供电方案，保证敏感客户的电能质量，并控制干扰源用户对电网的污染。其具体实施方案如下：

(1) 选择用户的电压等级。供电方规定用户的电压等级，依据包括用电容量、用电设备特性、供电距离、供电线路回路，并综合用户的敏感度和污染度。

(2) 确定用户的供电电源配置。按照表 4-11 对敏感用户进行电源配置，并按照表 4-12 选择接入点的电能质量等级。干扰源用户对于供电可靠性方面没有特殊需求。

表 4-11 敏感用户的供电电源配置建议

| 供电模式 | | 电源 | 电源点 | 用户类别 |
|------|------------|------|------------|----------------------|
| 三电源 | | 电源 1 | 变电站 1 | 敏感用户 I 级 |
| | | 电源 2 | 变电站 2 | |
| | | 电源 3 | 变电站 3 | |
| 双电源 | 不同变电站 | 电源 1 | 变电站 1 | 敏感用户 I 级、II 级 |
| | | 电源 2 | 变电站 2 | |
| | 同一变电站，不同母线 | 电源 1 | 变电站 1，母线 1 | 敏感用户 II 级、III 级、IV 级 |
| | | 电源 2 | 变电站 1，母线 2 | |
| 单电源 | | 电源 1 | 变电站 1 | 敏感用户 V 级 |

表 4-12 敏感用户接入点的电能质量等级选择

| 用户类别 | 接入点电能质量等级 | | | | | | |
|-----------|-----------|------|------------|-------------|----|----|------|
| | 电压暂降 | 电压偏差 | 长时电压 闪变 | 电压三相不 平衡 | 谐波 | 频率 | 综合指标 |
| I 级敏感用户 | 优 | 优 | 优 | 优 | 优 | 优 | 优 |
| II 级敏感用户 | 优 | ≥良 | ≥良 | ≥良 | ≥良 | ≥良 | 优 |
| III 级敏感用户 | ≥良 | ≥良 | ≥良 | ≥良 | ≥良 | ≥良 | ≥良 |
| IV 级敏感用户 | ≥良 | ≥中 | ≥中 | ≥中 | ≥中 | ≥中 | ≥良 |

| | | | | | | | |
|---------|----|----|----|----|----|----|----|
| V 级敏感用户 | ≥中 | ≥差 | ≥差 | ≥差 | ≥差 | ≥差 | ≥中 |
|---------|----|----|----|----|----|----|----|

(3) 选择用户接入点。基于用户负荷特性、变电站间隔、电网条件限制、施工难度等因素选择用户接入点。敏感用户不应与干扰源用户接入同一接入点，干扰源用户避免接入输电干线上。

(4) 干扰源和敏感用户电能质量治理。根据用户负荷特性，按需求为客户定制治理装置，提高敏感用户抵抗电能质量扰动的能力，并减轻干扰源用户对电网的污染。

(5) 认定电能质量治理责任。将电力系统公共连接点电能质量监控结果作为治理责任认定的标准，并采用分离电能质量责任的方法处理电能质量问题，即供电方主要对电压质量负责，电力用户对电流质量负责的方法进行责任分离。

(6) 用户治理方案选择。敏感用户根据经济损失指标、自身治理意愿等自主选择治理方式。污染性高的干扰源用户需强制其治理注入电网的污染问题。

(7) 选择电能质量治理装置。按照表 4-13 治理装置功能，根据图 4-1 为用户选择合理的治理装置，解决干扰源和敏感用户面临的电能质量问题。

表 4-13 各种电能质量治理装置主要功能

| 电能质量治理装置 | 短时中断 | 三相不平衡 | 电压暂降 | 闪变 | 谐波 | 浪涌 |
|---|------|-------|------|----|----|----|
| 不间断电源 (Uninterruptible Power System, UPS) | 有 | 有 | 有 | 无 | 无 | 无 |
| 动态电压补偿器 (Dynamic Voltage Restorer, DVR) | 有 | 有 | 有 | 有 | 无 | 无 |
| 动态无功补偿装置 (Static Var Generator, SVG) | 无 | 有 | 有 | 有 | 有 | 无 |
| 统一电能质量调节器 (Unified Power Quality Conditioner, UPQC) | 有 | 有 | 有 | 有 | 有 | 有 |
| 有源电力滤波器 (Active Power Filter, APF) | 无 | 有 | 无 | 无 | 有 | 有 |
| 固态切换开关 (Solid State) | 有 | 无 | 无 | 无 | 无 | 无 |

| | | | | | | |
|--|---|---|---|---|---|---|
| Transfer Switch, SSTS) | | | | | | |
| 无源滤波 (Passive Power Filter, PPF) | 无 | 无 | 无 | 无 | 有 | 有 |
| 瞬变电压脉冲抑制器 (Transient Voltage Surge Suppressor, TVSS) | 无 | 无 | 无 | 无 | 无 | 有 |
| 静止无功补偿器 (Static VAR compensator, SVC) | 无 | 有 | 无 | 有 | 有 | 有 |

敏感用户、干扰源用户和需综合治理的用户，应结合其负荷特性，根据图 4-2 选择合适的电能质量治理装置，提高敏感用户抵抗电能质量扰动的能力，并减轻干扰源用户对电网的污染。

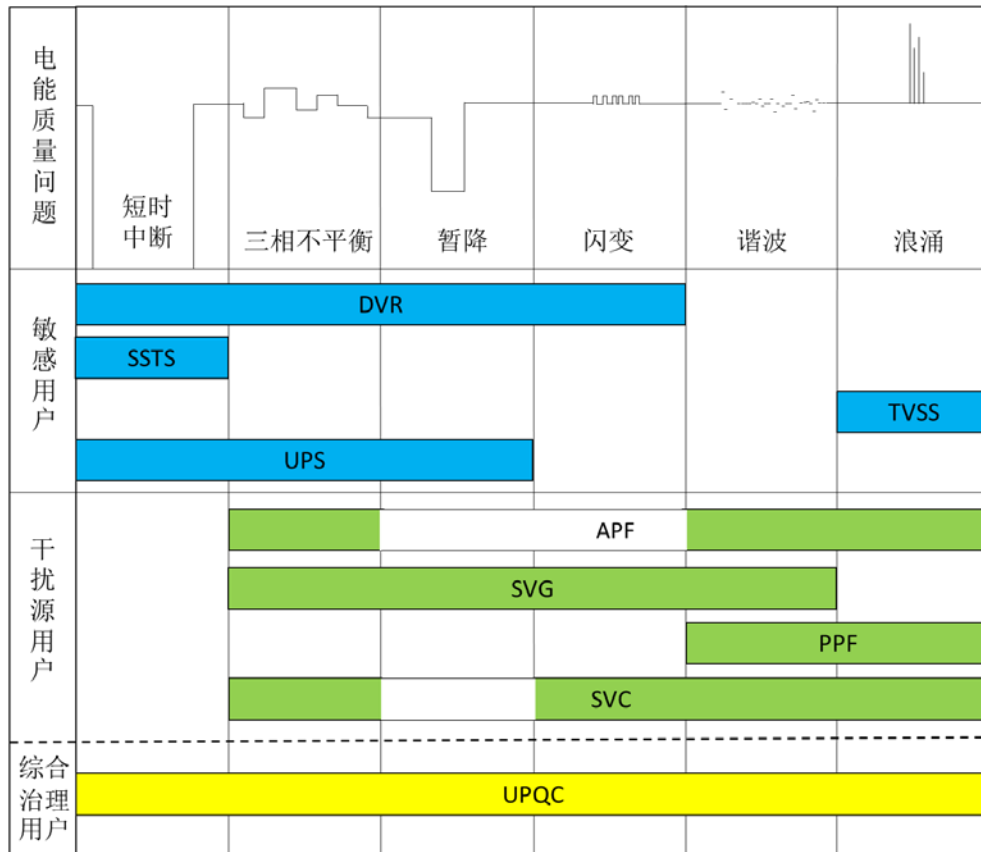


图 4-2 敏感用户、干扰源用户和综合治理用户的治理装置选择

I 级、II 级用户（敏感用户和干扰源用户）要求必须安装电能质量治理装置，III 级、IV 级用户（敏感用户和干扰源用户）建议安装电能质量治理装置。

(8) 定制电能质量治理装置。根据用户负荷的敏感度/污染度以及负荷容量，

定制电能质量治理装置，包括装置容量、补偿方式（串联/并联/综合）、尺寸、安装位置等。

(9) 评估敏感用户和干扰源用户治理效果。

4.5 从电能质量到电能体验质量

4.5.1 电能体验质量的提出

根据对国内某供电公司的统计，电能质量的主要问题有电压偏差、电压波动和闪变，其中暂态电能质量（暂升、暂降、中断）问题发生最为频繁。2010 至 2012 年因电能质量问题引起的大客户投诉及供电公司的反馈，事故原因为电压暂降的事件数量占 15/21，负荷类型大多为半导体、精密模具等敏感用户。电压暂降对于不同用户的影响差异巨大。在实际应用中，根据用电负荷不同的特性以及对电能质量的要求与敏感度，一般将用电负荷分为普通负荷、敏感负荷。一些电力系统用户使用了大量的敏感负荷，对于这类用户即使出现轻微的电能质量问题，也会导致严重的经济损失。因此，单纯根据电力参数评估电能质量并不能很好地刻画电能质量扰动对用户的影响。

作为基础设施来提供服务，服务质量（Quality of Services, QoS）一直是被关注的问题。服务质量是产品生产的服务或服务行业满足规定或潜在要求（或需要）的特征和特性的总和。比如在信息通信技术领域，与服务质量相关的特征量包括通信延迟、丢包率等等。近年来，研究人员发现服务质量并不能在全面刻画服务的特征，尤其是从用户的角度，好的服务质量并不一定导致好的用户体验。于是提出了体验质量（Quality of Experience, QoE, QoX）的概念，是客户对服务提供商整体满意度的衡量。体验质量与服务质量密切相关，但又不尽相同。比如在信息技术领域服务质量体现可测量、改进和有保障的硬件及软件特征。相比而言，体验质量代表用户客观及主观满意度。

一般地说，针对某一具体评估内容，“指标体系+评估方法=评估体系”。电

能质量评估内容多，而每项的评估指标、方法又各不相同。因此应该先对各项电能质量问题分别建立评估子体系，各子体系有自己的指标和评估方法且彼此之间相对独立。同时还应该建立综合评估体系，结合各个单项指标，运用恰当的综合评估方法，作综合评估。最后根据各单项指标的评估子体系的评估结果以及综合评估的结果对电能商品进行经济管理，最终实现电能的按质定价。

报告借鉴信息技术领域的发展经验，将体验质量的理念引入电能质量研究领域并提出电能体验的概念，总结了其特征，并综合现有电能质量评价标准和相关用户敏感度分析，建立电能体验质量的指标体系。电能体验质量不单单取决于供电质量，还和用户敏感程度紧密相关，因此在实际工作中更加具有指导意义。电能体验质量（Power Experience）是电力用户对供电方电能服务整体满意度的衡量。相比而言，电能体验质量代表用户客观及主观满意度。

电能体验是依赖用户的，因为一些客户比另一些更敏感，所以相同的电能质量会导致不同程度的电能体验。体验质量并非总能通过量化进行界定，但任何企业都希望将自身的用户体验质量最大化。电能体验质量评价可以通过对大量用户进行民意测验或抽样调查得出，研究中需要从中抽取可量化的特征。对于电力企业，最有意义的结果还应当包括是用户的投诉次数以及用户主观的电能体验。

4.5.2 电能体验质量评估流程

现有电能质量标准体系主要以供电质量为基础，对于电能质量好坏的判断通过电网电压、电流稳态和暂态等方面的量化指标进行描述。例如，对于电压暂降问题，目前供电方通常是以发生电压暂降事件的次数作为供电质量的评价标准。但是，在实际应用中电能质量的好坏更多取决于用户的负荷敏感程度和实际使用体验。因此，本报告对用户敏感度的定量分析为基础，综合现有电能质量标准和相关用户敏感度信息，建立了电能体验质量的指标体系。

用户电能体验质量的指标体系综合了供电质量指标和用户指标,其中用户指标主要包含用户投诉、用户负荷敏感度和用户电能体验调研等方面。

①用户投诉次数 (M): 是指供电方的客户服务部门收到的客户投诉信息,分析引起客户投诉的供电质量问题。

②用户负荷敏感度 (S): 是指用户对于某种特定电能质量问题的敏感程度。当发生该电能质量问题时,用户负荷敏感度越高则说明它越容易出现故障。用户负荷敏感度代表用户负荷的特性,该特性反映了电能质量出现问题对于客户的影响程度,如表 4-14 所示。

③用户电能体验调研得分 (P): 通过调查问卷的形式对于敏感客户进行电能体验调研,获取某段时间内的电能质量各单项问题的客户主观评分(1%至100%)。调查问卷主要指标包括:用户电能不满意度指标、用户所处行业敏感度指标、用户敏感负荷故障的经济损失指标等。用户电能体验调研中,得分越高说明该用户的电能体验越差。用户电能体验调研得分根据调查问卷结果进行主观评定。

以电压暂降的电能质量问题为例,表 4-14 给出了单独使用供电质量指标体系和使用电能体验质量指标体系来衡量电压暂降严重度的对比。使用电能体验质量指标体系需要对上述电压暂降事件次数和用户投诉次数进行归一化处理,如式 4-4 和式 4-5 所示。

$$N' = \frac{N - N_{\min}}{N_{\max} - N_{\min}} \quad \text{式 4-4}$$

$$M' = \frac{M - M_{\min}}{M_{\max} - M_{\min}} \quad \text{式 4-5}$$

其中, N_{\max} 和 N_{\min} 分别表示一段时间内所有用户供电中发生电压暂降事件次数的最大值和最小值, M_{\max} 和 M_{\min} 分别表示一段时间内所有用户投诉的最大值和最小值。

表 4-14 供电质量评估体系和用户电能体验质量指标体系对比

| 指标体系 | 各项指标 | 指标权重 | 电能质量问题严重程度评分 |
|------------|-----------------|-------------|--|
| 供电质量评估体系 | 电压暂降事件次数 (N) | 100% | N' |
| 电能体验质量指标体系 | 电压暂降事件次数 (N) | λ_N | $\lambda_N \times N' \times S + \lambda_M \times M' + \lambda_P \times P$ 其中, $\lambda_N + \lambda_M + \lambda_P = 100\%$ |
| | 用户投诉次数 (M) | λ_M | |
| | 用户负荷敏感度 (S) | | |
| | 用户电能体验调研得分 (P): | λ_P | |

在用户电能体验质量指标体系中, λ_N , λ_M , λ_P 分别表示供电质量指标、用户投诉指标、以及用户电能体验调研指标在体系中的重要程度。需要说明的, 用户负荷敏感度反映了电能质量问题对于客户的影响程度, 使用 $N \times S$ 表示电压暂降次数对于供电质量影响, 不仅反映了电压暂降次数的多少, 还考虑了用户敏感程度的因素对于供电质量影响。例如, 同样发生 100 次电压暂降问题, 对于负荷敏感度高的用户和负荷敏感度低的用户来说, 其所面临的电能质量问题的严重程度是不同的。

4.5.3 电能体验质量实例验证

本报告采用某供电公司 2011 年至 2012 年两个企业用户 (分别为企业 1、企业 2) 的供电质量中的电压暂降事件为例, 企业 1 和企业 2 相关的包括电压暂降事件次数、用户投诉次数、用户负荷敏感度、电能体验调研等各项评估指标如表 4-15 所示。

表 4-15 企业 1 和企业 2 各项评估指标对比

| 企业 | 电压暂降事件次数 (N) | 用户投诉次数 (M) | 用户负荷电压暂降敏感度 (S) | 用户电能体验调研得分 (P): | 供电质量评估体系评分 | 电能体验质量指标体系评分 |
|----|--------------|------------|-----------------|-----------------|------------|--------------|
| | | | | | | |

| | | | | | | |
|------|----|---|------|------|-------|-------|
| 企业 1 | 5 | 7 | 100% | 0.70 | 0.263 | 0.488 |
| 企业 2 | 12 | 4 | 72% | 0.80 | 0.632 | 0.450 |

根据表 4-15 中企业 1 和企业 2 的指标,分别使用供电质量评估体系和电能体验质量指标体系来计算供电质量,比较两者在衡量供电质量高低方面的区别,两种指标的对比结果如图 4-3 所示。其中, λ_N 、 λ_M 、 λ_P 分别取值为 30%、50%、20%。

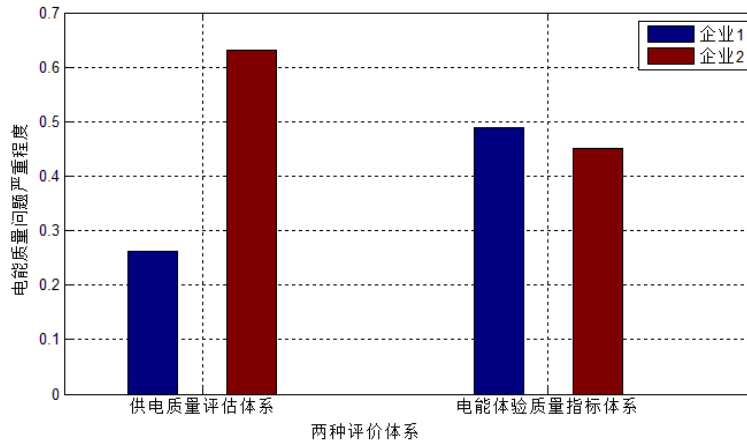


图 4-3 两种评价指标体系下电能质量严重程度对比

根据对图 4-3 的分析,如果只考虑供电质量的标准体系,企业 1 比企业 2 的供电质量要好。然而,使用电能体验质量指标体系,引入用户相关的各项指标信息后,企业 1 比企业 2 的电能质量评估结果反而要差。通过现场用户调研和分析结果证明,使用用户电能体验质量指标体系来衡量电能质量的高低,能够更准确地反映供电对于用户实际工作的影响。因此,用户电能体验质量指标体系能够更全面的反映供用电双方的问题和情况,在实际应用中具有较高地指导意义。

5 基于分类算法的电能质量状态评估

5.1 数据挖掘基本任务

数据挖掘(Data Mining, DM)又称数据库中的知识发现(Knowledge Discover in Database, KDD),是目前人工智能和数据库领域研究的热点问题,所谓数据挖掘是指从数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的非平凡过程^[49]。数据挖掘是一种决策支持过程,它主要基于人工智能、机器学习、模式识别、统计学、数据库、可视化技术等,高度自动化地分析企业的数
据,做出归纳性的推理,从中挖掘出潜在的模式,帮助决策者调整市场策略,减少风险,做出正确的决策。

知识发现过程由以下三个阶段组成:(1)数据准备,(2)数据挖掘,(3)结果表达和解释。数据准备是从相关的数据源中选取所需的数据并整合成用于数据挖掘的数据集;规律寻找是用某种方法将数据集所含的规律找出来;规律表示是尽可能以用户可理解的方式(如可视化)将找出的规律表示出来。

数据挖掘可以与用户或知识库交互。并非所有的信息发现任务都被视为数据挖掘。例如,使用数据库管理系统查找个别的记录,或通过因特网的搜索引擎查找特定的Web页面,则是信息检索(information retrieval)领域的任务。虽然这些任务是重要的,可能涉及使用复杂的算法和数据结构,但是它们主要依赖传统的计算机科学技术和数据的明显特征来创建索引结构,从而有效地组织和检索信息。尽管如此,数据挖掘技术也已用来增强信息检索系统的能力^[50]。数据挖掘面向广泛的应用领域,不同的挖掘目的和数据形式决定了数据挖掘任务的多样性。数据挖掘任务用于指定数据挖掘要找的模式类型。一般而言,数据挖掘任务可以分为两类:描述和预测。描述性挖掘任务描述数据集合中数据的一般性质;预测性挖掘任务对当前数据进行推断,以做出预测^{[51][52]}。

根据目标模式的不同，数据挖掘任务主要可以分为：概念/类描述、频繁模式挖掘、分类与预测、聚类分析、离群点分析和演变分析等几类：

(1) 概念/类描述

数据可以与类或概念相关联。用汇总的、简洁的和精确的方式描述、各个类和概念可能是有用的。这种类或概念的描述称为概念/类描述（Class/concept description）。这种描述可以通过下述方式得到：1) 数据特征化，一般地汇总所研究类的数据；2) 数据区分，将目标类与一个或多个可比较类进行比较；3) 数据特征化的比较。

(2) 频繁模式挖掘

频繁模式（frequent pattern）是在数据中频繁出现的模式。存在多种类型的频繁模式，包括项集、子序列和子结构。通常，频繁项集是指频繁地事务数据集中一起出现的项的集合；频繁子序列是针对同一识别标识（如：用户 ID）下所产生事务按时序上的先后关系经常出现的序列模式；子结构可能涉及不同的结构形式，如图、树或格，如果一个子结构频繁出现，则称它为（频繁）结构模式。

对频繁模式的挖掘称为关联分析，分析结果是一系列的关联规则。通常关联规则是那些同时满足最小支持度阈值和最小置信度阈值的，而不能同时满足则被认为是不令人感兴趣的而被丢弃。根据不同要求，还可以做相关联的属性-值对之间的其它相关分析（如：提升度、全置信度、 χ^2 度量、余弦度量）。

(3) 分类与预测

分类（Classification）是找出能够描述和区分数据类或概念的模型（或函数），用于预测类标号未知的对象类。导出模型是基于对训练数据集（即类标号已知的数据对象）的分析^[53-55]。

预测 (Predication) 是建立连续值函数模型, 用于预测空缺的或不知道的数值数据, 而不是类标号。驾照分析 (regression analysis) 是一种最常用的数值预测统计学方法。预测也包含基于可用数据的分布趋势识别。

(4) 关联分析

关联分析 (relevance analysis) 可能需要在分类和预测之前进行, 它试图识别对于分类或预测过程无用的属性, 这些属性应当排除。

(5) 聚类分析

聚类 (Clustering) 分析数据对象不考虑已知的类标号。一般情况下, 训练数据中不提供类标号, 因为开始并不知道类标号。可以使用聚类产生这种标号^[56]。对象根据最大化类内部的相似性、最小化类之间的相似性的原则进行聚类或分组。也就是说, 对象的簇 (Cluster) 形成簇内对象具有很高的相似性, 面与其它簇中对象很不相似。所形成的簇可看作一个对象类, 由它可导出规则。

(6) 离群点分析

数据集中可能包含一些数据对象, 与数据集一般行为或模型不一致。这些数据对象被称为离群点 (Outlier)。大部分数据挖掘方法将离群视为噪声或异常而丢弃, 而有些应用 (如欺骗检测) 中, 罕见的事件可以比正常出现的事件更令人感兴趣, 因而展开离群点挖掘 (outlier mining)。

可以假定一个数据分布或概念模型, 使用统计检验检测离群点; 或者使用距离度量, 将远离任何簇的对象视为离群点。基于偏差的方法通过考察一群对象主要特征上的差别来识别离群点, 而不是使用统计或距离度量。

(7) 演变分析

演变分析 (Evolution analysis) 描述行为随时间变化的对象的规律或趋势, 并对其建模。尽管这可能包括时间相关数据的特征化、区分、关联和相关分析、

分类、预测或聚类。这类分析的不同特点包括时间序列数据分析、序列或周期模式匹配和基于相似性的数据分析。

5.2 分类算法研究综述

5.2.1 SVM 算法综述

Vapnik^[59]等人提出了一种全新的机器学习方法——支持向量机(Support Vector Machine, SVM), 其核心内容在 1992 年至 1995 年间提出。它建立在结构风险最小化原则(Structural Risk Minimization Inductive Principle)基础之上, 可以实现实际问题中统计学习理论的算法, 在理论基础和应用前景等很多方面都优于传统的学习技术: 其泛化性能是通过训练得到的支持向量集, 而不依赖于全部训练数据, 这引起国内外机器学习领域的广泛重视, 且仍在不断发展中。SVM 分类算法中涉及的相关概念定义如下:

(1) 最优分类线: 将两类样本无错误地分开, 保证经验风险最小(为 0); 且使两类样本之间的分类间隔最大, 置信范围最小, 进而使实际风险最小。

(2) 最优超平面 (Optimal Hyperplane): 是将最优分类线推广到高维空间的情形。如果 P 是 R^n 中非零向量, α 是一个实数, 则集合 $H = \{x | P^T x = \alpha\}$ 为 R^n 中的一个超平面, 向量 P 为该超平面的法向量。

(3) 支持向量 (Support Vectors, SV): 同时满足在两类样本中离分类面最近的样本点, 且平行于最优超平面的训练样本。SV 在概率意义上可以代表整个数据集, 虽然仅占训练集的一小部分即 $N_{sv} \ll N_{train}$, 却具有数据集分类的重要边界信息。

(4) 最大间隔 (Maximal Margin) 原则: 在训练样本集 $T = \{(x_1, y_1), \dots, (x_m, y_m)\}$ 线性可分的分类问题中, 存在超平面 $(\omega \cdot x) + b = 0$, 选择使训练集对线性函数 $(\omega \cdot x) + b$ 的几何间隔取最大值的参数对 (ω, b) 。

(5) 决策函数: $f(x) = \text{sgn}((\omega \cdot x) + b)$ 。其中, $\text{sgn}()$ 是符号函数, $x_i \in R^n$ 和

$y_i \in \{1, -1\}$ 为类别标号, ω 是超平面的法向量, b 是偏置, $(x \cdot y)$ 表示 x 和 y 的内积。

SVM 是用于数据分类问题的一种技术, 可以实现对线性可分数据的最优分类, 一般有线性可分、线性不可分和非线性三种情况。分类问题通常包含一些数据实例的训练数据和测试数据, 训练集中的每个实例包含一个“目标值”(类标签)和一些“属性”(特征)。**SVM** 的目标是产生一个模型, 该模型可以预测在测试集中只给出属性的数据实例的目标值。

给定一个带标签的实例对 $(x_i, y_i), i = 1, \dots, l$ 的训练集, 其中 $x_i \in R^n, y_i \in \{1, -1\}^l$, n 为输入维数。**SVM** 需要求解最优化问题: $\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i$, 使得 $y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0$ 。其中, $C > 0$ 是误差项的惩罚参数 (Penalty Parameter)。函数 ϕ 将训练矢量 x_i 映射到更高维度空间, **SVM** 根据最大间隔原则, 在该空间中找到一个线性的分割超平面 (Linear Separating Hyperplane)。

SVM 的机理可以简单描述为: 寻找一个满足分类要求的最优分类超平面, 满足训练集中属于不同类别的样本点正好位于该超平面的不同侧面, 保证了分类精度; 同时使超平面两侧的空白区域最大化, 因而保证了最终所获得的平面位于两个类别样本的中心, 提高了分类问题的容错性。

对于非线性问题, Vapnik 等人引入核映射方法, 将低维空间中的非线性问题转化为高维空间中的线性可分问题进行解决。

核函数 (Kernel Function) 是用来计算两个数据点在映射下的像的内积, 是影响 **SVM** 性能的关键因素, 它隐式地定义了高维特征空间的结构, 将线性不可分问题转化为线性可分问题, 而且不增加计算量。如果存在某个内积空间(或 Hilbert 空间) H 以及映射 $\phi: X \rightarrow H$, 使得 $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$, 称二元函数 $K: X \times X \rightarrow R$ 是核函数。

不同核函数的使用产生了不同的 **SVM** 算法, 常用的基本核函数有:

(1) 线性核函数: $K(x_i, x_j) = x_i^T x_j$ 。

(2) 多项式核函数: $K(x_i, x_j) = (\alpha x_i^T x_j + \beta)^\gamma, \alpha > 0$ 。

(3) 径向基函数 (Radial Basis Function, RBF):
 $K(x_i, x_j) = \exp(-\alpha \|x_i - x_j\|^2), \alpha > 0$, 与传统的 RBF 方法的基本区别: 每一个基函数的中心对应一个 SV, 由算法自动确定它们以及输出权值。

(4) 多层感知器核 (Sigmoid 核) 函数: $K(x_i, x_j) = \tanh(\alpha x_i^T x_j + \beta)$, 利用该核函数实现了一个两层神经网络, 算法自动确定网络的权值、网络的隐含层节点数目, 而且算法不存在神经网络方法经常出现的局部极小点问题。

其中, α, β 和 γ 是内核参数。

SVM 的核心思想之一是控制推广能力, 即归一化决策函数, 使两类训练样本都满足 $|(\omega \cdot x) + b| \geq 1$, 离分类面最近的样本满足 $|(\omega \cdot x) + b| = 1$, 则分类间隔为 $2/\|\omega\|$, 因此分类间隔最大, 等价于使 $\|\omega\|$ 最小。利用 SVM 计算决策函数, 其形式类似于一个神经网络, 因此被称为支持向量网络, 其中中间层节点是输入样本与支持向量的内积, 输出是中间层节点的线性组合。

训练 SVM 可理解为是一个求解 QP 的问题, 该问题所求解未知数的个数等于训练样本的个数, 经典的算法有积极方集法、对偶方法和内点算法等。但随着训练样本的增多会出现维数灾难, 或者因为内存有限而无法训练样本, 这成为 SVM 实际应用的瓶颈问题。目前, SVM 算法的研究主要集中在以下几个方面:

(1) 理论基础的研究

Vapnik 领导的 AT&T 实验室提出了 SVM, 并成功推广到求解回归问题。SVM^[60] 的理论基础是统计学习理论和 VC 维理论, 通过结构风险最小化 (Structural Risk Minimization, SRM) 原则控制目标函数, 使有限的样本信息在模型的复杂性和学习能力之间寻求折衷, 得到较好的泛化能力, 目前对学习机泛化能力的研究尚未完善。VC 维 (Vapnik-Chervonekis dimension) 可以描述函数集或学

习机的复杂性，是学习能力的重要指标之一，但至今没有一个统一的定义和计算方法。除了以上统计学习理论，正则化理论、贝叶斯理论也被加入 SVM 的研究中。正则化理论强调逼近未知函数时，越“平滑”越好，结合 SVM 出现了正则化的核方法^[61-64]。贝叶斯理论强调利用先验分布和样本的观测值推断出后验分布，是基于概率理论的建模工具，结合 SVM 出现了^[65]，以及相关向量机(Relevant Vector Machine)、高斯过程(Gaussian Process)和拉普拉斯过程(Laplacian Process)。

(2) 模型选择的研究

将实际问题转化为能用 SVM 求解的数学模型的过程称为模型选择。其中包括：数据集的获取、预处理，SVM 类型的选择，SVM 核函数和二次规划参数选择（如：惩罚系数）。核函数影响数据在特征空间中的分布，是研究的重点；惩罚系数影响特征空间中的经验风险水平^[66-68]。核方法用核映射将原空间的线性不可分问题转化成高维空间的线性可分问题，而不增加计算复杂度，如核 Fisher 判别分析(Kernel FDA)、核主成分分析(Kernel PCA)、核独立成分分析 (Kernel ICA)、核聚类分析、核等距映射(Kernel Isomap)等，它们与 SVM 一起构成核方法。它可以解决非线性问题，并推广到任何包含内积运算的算法中。对核函数的研究主要集中在三个方向：理论研究、构造方法和参数选择^{[69][71]}。

(3) 训练算法的研究

许多 SVM 算法的共同思想是“分而治之”：将原问题分解成为若干子问题，按照某种迭代策略求解子问题，最终使结果收敛到最优解。根据子问题的划分和迭代策略的不同，又可分为两类：

① 块选算法(Chunking)：1995 年由 Cortes 和 Vapnik^[72]提出，目标是通过迭代方式逐步排除非支持向量，根据删除 Lagrange 乘子为零的训练样本不影响最终结果的事实，把二次规划 (Quadratic Programming, QP) 问题分解为多个小规模 QP 子问题，删除值为零的 Lagrange 乘子，保留非零值的 Lagrange 乘子。

其优点是当 SV 数目远小于训练样本的数目时，能有效提高训练速度；但当 SV 数目较多时，特别是随着迭代次数的不断增加，训练样本规模的增大会不断降低训练速度。

② 分解算法(Decomposition): 1997 年由 Osuna 等^{[73][74]}解决 SVM 训练速度慢且复杂度高的问题，并应用于人脸检测中，给出了简单的证明。虽然后来有人发现其证明并不严密，但其分解算法却是开创性的，为后来的研究奠定了基础。Joachims 于 1998 年对 Osuna 的分解算法提出了具体的实现方案，并用软件进行了实现。

Platt 于 1998 年提出了序列最小优化(Sequential Minimal Optimization, SMO)的分解算法^[75]，将问题分解成可以解析求解的最小 QP 问题，且可以用核优化的方法提高 SMO 性能。Keerthi 等极大改进了 SMO 算法，证明了 SMO 算法的收敛性。Ronan 将其应用在分类和回归问题中；2000 年又提出只适用于分类问题的最近点算法。O. L. Mangasarian 等在 1999 年提出了逐次超松弛(Successive Over Relaxation, SOR)算法，2000 年提出了基于有效集(Active Set)的 ASVM 算法^[106]，并加入光滑(Smooth)函数提出了 SSVM 算法，2001 年又提出快速且简单的 Lagrangian SVM 算法(LSVM)。不同于非线性 LSVM，线性 LSVM 可以处理上百万样本的大规模数据集。但这些算法是内存驻留算法，即需要将所有的样本驻留内存进行处理，但随着训练样本数量的增多，所占内存的开销会显著增加。

SVM 作为一种新兴的技术存在一些局限性：随着训练样本数目的增多，SVM 的训练速度会明显变慢，容噪性降低，不能支持增量学习并整合先验知识等。

(4) 应用方面的研究

SVM 理论上的突出优势，促使各国学者关注其应用方面的研究。在模式识别方面，最突出的应用是手写阿拉伯数字识别^[76]和麻省理工(MIT)的人脸检测^[77]，其它还有语音识别、文本字体识别、人脸识别、三维物体识别、旅游突发事件的

SVM 等；在图像处理方面有图像的分割、图像的检测、图像的检索、遥感图像的分析等。

5.2.2 贝叶斯分类

贝叶斯定理(Bayes theorem)，它是一种把类的先验知识和从数据中收集的新证据相结合的统计原理，朴素贝叶斯在贝叶斯分类器中应用最为广泛。贝叶斯分类是统计学分类方法，它是一类利用概率统计知识进行分类的算法。它在先验概率与条件概率已知的情况下，预测类成员关系可能性的模式分类算法。如计算一个给定样本属于一个特性类的概率，并选定其中概率最大的一个类别作为该样本的最终判别。假设每个训练样本用一个 n 维特征向量 $X=\{x_1,x_2,\dots,x_n\}$ 表示，分别描述 n 个属性 A_1, A_2,\dots,A_n 对样本的测量。将训练样本集分为 m 类，记为 C_1, C_2,\dots,C_m 。贝叶斯原理通常用下面的公式来表示：

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

其中， X 表示观测数据样本， $P(C_i)$ 是 C_i 的先验概率， $(i,j=1, 2, \dots, m)$ 。 $P(X|C_i)$ 是条件概率，先验概率对条件概率加权平均后，得到条件 X 下， C_i 的后验概率 $P(C_i|X)$ 。上述是朴素贝叶斯的工作过程，也是贝叶斯分类算法的判别准则。贝叶斯分类问题可以转化为求得 $P(C_i|X)$ 最大的类标号 C_i 的过程。

$$f(X) = \max P(C_i|X)$$

其中， $P(C_i|X)$ 为 X 发生时类 C_i 的条件概率。 $P(C_i)$ 为先验概率可由训练样本直接得出，如下式所示，其中 S_{C_i} 表示训练样本中属于类 C_i 的数量， S 为训练样本的总数。

$$P(C_i) = \frac{S_{C_i}}{S}$$

由于贝叶斯定理假设一个属性值对给定类的影响独立于其它的属性值，而此

假设在实际情况中经常是不成立的，因此其分类准确率可能会下降。为此，就出现了许多降低独立性假设的贝叶斯分类算法，如 TAN(tree augmented Bayes network)算法、贝叶斯网络分类器 (Bayesian network classifier, BNC)。

Naive Bayes 分类器以简单的结构和良好的性能受到人们的关注，它是最优秀的分类器之一。在许多场合，朴素贝叶斯 (Naïve Bayes, NB) 分类可以与决策树和神经网络分类算法相媲美，该算法能运用到大型数据库中，且方法简单、分类准确率高、速度快。朴素贝叶斯分类器建立在一个类条件独立性假设(朴素假设)基础之上：给定类节点(变量)后，各属性节点(变量)之间相互独立。朴素贝叶斯分类器可以看作是贝叶斯网络的一种最简化的模型。在 Naive Bayes 分类中，条件概率 $P(X|C_i)$ 可通过下式计算，其中 $P(x_j|C_i)$ 表示属于类 C_i 中样本的第 j 个属性为 x_j 的概率。

$$P(X | C_i) = \prod_{j=1}^n P(x_j | C_i)$$

条件概率 $P(X_1/C_i), P(X_2/C_i), \dots, P(X_n/C_i)$ 可以从训练数据集求得。根据此方法，对一个未知类别的样本 X ，可以先分别计算出 X 属于每一个类别 C_i 的概率 $P(X/C_i)P(C_i)$ ，然后选择其中概率最大的类别作为其类别。朴素贝叶斯算法成立的前提是各属性之间相互独立。当数据集满足这种独立性假设时，分类的准确度高，否则可能较低。

5.2.3 决策树分类

决策树和决策规则是解决实际应用中分类问题的数据挖掘方法。决策树是一个分类预测模型，它代表的是对象属性与对象值之间的一种映射关系。决策树中每个节点表示某个对象，而每个分叉路径则代表的某个可能的属性值，而每个叶结点则对应从根节点到该叶节点所经历的路径所表示的对象的值。决策树仅有单一输出，若需要多个输出，可以建立独立的决策树以处理不同输出。数据挖掘中决策树是一种经常要用到的技术，可以用于数据分析，同样也可以用来作预测。

决策树模型是一颗有向无环树。最为典型的决策树即二叉树，树中根节点没有父节点，所有其他节点都有且仅有 1 个父节点。1 个节点可以包含 0-2 个子节点。若节点没有子节点，称其为叶节点(leaf node)，否则称为内部节点(internal node)，每个叶节点都对应一个类别标识 C 值，每个内部节点都对应一个用于分割数据集的属性 X_i ，称为分割属性(splitting attribute)。每个内部节点都有一个分割判断规则 q_j (splitting predicate)。如果 X_i 是连续属性，那么 q_j 的形式为 $X_i \leq x_i$ ，其中 $x_i \in dom(X_i)$ ， x_i 是节点 n 的分割点(split point)。若 X_i 是离散属性，那么 q_j 的形式 $X_i \in Y_i$ ，其中 $Y_i \in dom(X_i)$ ， Y_i 称为节点 n 的分割子集(split subset)。节点 n 的分割属性和分割判定规则组成了节点 n 的分割标准(split criterion)。

从数据产生决策树的机器学习技术叫做决策树学习。决策树学习也是资料探勘中一个普通的方法。在这里，每个决策树都表述了一种树型结构，它由它的分支来对该类型的对象依靠属性进行分类。每个决策树可以依靠对源数据库的分割进行数据测试。决策树学习一般可分为两个阶段：决策树构建(building)和决策树修剪(pruning)。决策树学习是一个递归式的迭代过程，通过对树进行递归式的修剪，当不能再进行分割或一个单独的类可以被应用于某一分支时，递归过程就完成了。

主要的决策树分类算法有 ID3、C4.5(C5.0)、CART、PUBLIC、SLIQ 和 SPRINT 算法等。它们在选择测试属性采用的技术、生成的决策树的结构、剪枝的方法以及时刻，能否处理大数据集等方面都有各自的不同之处。在当前决策树学习的各种算法中，影响最大的是 J.R.Quinlan 于 1986 年提出的 ID3 算法，他提出用信息增益作为属性的选择标准，以使得在对每一个非叶子节点进行测试时，能获得关于被测试记录最大的类别信息。本报告主要以 ID3 为例介绍决策树分类算法。

ID3 总是选择具有最高信息增益的属性作为当前节点的测试属性。具体方法是：检测所有的属性，选择信息增益最大的属性产生决策树节点，由该属性的不同取值建立分支，再对各分支的子集递归调用该方法建立决策树节点的分支，直到所有子集仅包含同一类别的数据为止，最后得到一棵决策树，它可以用来对新

的样本进行分类。ID3 算法通过不断的循环处理，初步求精决策树，直到找到一个完全正确的决策树。在选择重要特征时利用了信息增益的概念。

设 S 是 s 个数据样本的集合。假定类标号属性具有 m 个不同值，定义 m 个不同类 $C_i (i=1, \dots, m)$ 。设 s_i 是类 C_i 中的样本数。对一个给定的样本分类所需的期望信息由下式给出：

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i)$$

其中 $p_i = s_i/s$ 是任意样本属于 C_i 的概率。注意，对数函数以 2 为底，其原因是信息用二进制编码。

设属性 A 具有 v 个不同值 $\{a_1, a_2, \dots, a_v\}$ 。可以用属性 A 将 S 划分为 v 个子集 $\{S_1, S_2, \dots, S_v\}$ ，其中 S_j 中的样本在属性 A 上具有相同的值 $a_j (j=1, 2, \dots, v)$ 。设 s_{ij} 是子集 S_j 中类 C_i 的样本数。由 A 划分成子集的熵或信息期望由下式给出：

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j} + \dots + s_{mj})$$

熵值越小，子集划分的纯度越高。对于给定的子集 S_j ，其信息期望为

$$I(s_{1j} + \dots + s_{mj}) = - \sum_{i=1}^m p_{ij} \log_2(p_{ij})$$

其中 $p_{ij} = s_{ij} / |s_j|$ 是 S_j 中样本属于 C_i 的概率。在属性 A 上分枝将获得的信息增益是：

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A)$$

ID3 决策树分类算法优点在于，算法的信息熵数学理论基础清晰，计算速度快。算法的搜索空间是完全的假设空间，目标函数就在搜索空间中，不存在无解的危险。全盘使用训练数据，可得到一棵较为优化的决策树。

在实际应用中，对于非增量式的学习任务，ID3 算法通常是建立决策树的很好选择，但该算法不足之处在于：（1）不能增量地接受训练例，这就使得每增加一次实例都必须废除原有的决策树，重新计算信息增益并构造新的决策树，这造成极大的开销；（2）只能处理离散属性，在分类前需要对其进行离散化的处理；（3）在建树时，每个节点仅含一个特征，这是一种变元的算法，特征间的相关性强调不够；（4）对噪声较为敏感，数据质量差将直接导致生成的决策树

过于庞大或决策树中很多分支的信息量很少。(5) 在建树的过程中每当选择一个新属性时, 算法只考虑了该属性带来的信息增益, 未考虑到选择该属性后为后续属性带来的信息增益, 即未考虑树的两层节点; (6) 其信息增益存在一个内在偏置, 它偏袒属性值数目较多的属性。

5.3 电能质量综合评估建模

目前, 电能质量扰动已经成为影响电力系统安全、稳定、可靠运行的重要因素, 暂态电能质量评估的目的在于近乎实时地评估出暂态电能质量扰动的严重程度, 实现对敏感用户的潜在故障预警, 其研究的焦点是科学、客观地将多指标问题归并成单一量化的指标问题。国内外对于电能质量综合评估的研究主要分为定性分析和定量分析两类, 其方法都是通过使用加权系数反映各指标的相对重要程度, 并将多维电能质量扰动向一维归并。这种综合评估方法的主要问题在于各电能质量指标具有明显的不相关性, 从而导致它们相应的加权系数难以确定。

从机器学习的角度, 分类算法是一种有监督的学习方法, 其目标是将未知样本归类到不同的分组中, 通过学习各种统计数据中类别和样本之间的关系创建模型(通常称为分类器), 用于归类未标识的样本。报告中电能质量综合评估采用分布式分类算法, 通过模仿和学习相关大数据中的海量样本来训练分类模型, 然后根据分类模型快速评估和预测电能质量。将直接或间接影响电能质量综合评估的各类数据抽象为以下三类数据: (1) 电网运行监测数据, 如电流、电压、频率等; (2) 公共信息数据, 指来自于电力系统外部但与电能质量评估密切相关的数据, 如气象、经纬度、地质灾害等; (3) 电力用户数据, 如电力用户敏感度、重要度、负荷容量、投诉等方面的信息。将综合评估的结果按严重程度分为正常状态、异常状态、严重状态、故障状态四个等级, 并分别定义为:

(1) 正常状态 **normal condition**: 电网稳定运行, 基本不会对电力用户产生用电风险的状态。

(2) 异常状态 **abnormal condition**: 单个或多个电网运行特征项发生较大变化,

可能会对敏感电力用户产生用电风险。

(3) 严重状态 **critical condition**: 单个或多个电网运行特征项发生较大变化, 可能会对大多数电力用户产生用电风险。

(4)故障状态 **failed condition**: 单个或多个电网运行特征项严重超过了标准限值, 大多数甚至全部电力用户会发生用电设备故障。

对于样本数据中的离散型属性(如用户敏感度等级、天气情况等), 可直接作为样本的特征项进行处理。对于属性为数值型的情况(如电压暂降深度、三项不平衡度), 则将其转化为离散型属性, 通常使用值域平均划分的方法进行离散化处理。对于难以转化的数值型属性(如用户投诉次数、用户负荷容量等), 则使用所示的归一化方法进行处理。

根据公式(3)中, 如果样本 X 的该属性为离散型(如用户敏感度等级), 通过计算 C_i 中第 j 个属性为 x_j 的样本数量和 C_i 中样本总数的比值来计算 $P(x_j|C_i)$ 。当该属性为连续型时(如电压暂降深度), 则通常将其转化为离散型属性后计算 $P(x_j|C_i)$ 。对于难以转化的连续型属性, 则使用如公式 5-1 所示的归一化方法将连续型属性转换为离散型属性。

$$f(x_k) = \frac{x_k - \min(x_i)}{\max(x_i) - \min(x_i)}, \quad i = 1, \dots, n \quad \text{式 5-1}$$

将电能质量监测数据、电力用户数据和公共信息数据三类数据形式化表示为一个 n 维向量的形式。训练集和测试集的形式为 $X = \{x_1, x_2, \dots, x_n, C_i\}$ 。其中, x_1, x_2, \dots, x_n 等特征项为用于电能质量评估的各个属性, C_i 为电能质量的评估结果。分布式贝叶斯分类的任务就是将评估样本 $\{x_1, x_2, \dots, x_n\}$ 划分至正常状态、异常状态、严重状态、故障状态四个类别, 分别记为 C_1, C_2, C_3, C_4 。

5.4 基于分类算法的电能质量评估实例

针对中国某电网内部 2010 年、2011 年、2012 年内 3 条 10kV 电力用户专线进行暂态电能质量评估。电能质量综合评估采用分布式分类算法,通过模仿和学习相关大数据中的海量样本来训练分类模型,然后根据分类模型快速评估和预测暂态电能质量。收集的数据包括:①电能质量监测数据,主要指这 3 条 10kV 线路上电压暂降、暂升、短时中断等的监测数据;②公共信息数据,指来自于电力系统外部但与电能质量评估密切相关的数据,如气象、经纬度、地质灾害等;③电力用户数据,如电力用户敏感度、重要度、负荷容量、投诉等方面的信息。数据的采集和集成过程如下:

①电能质量监测数据集成。电能质量监测数据包括监测数据包括电压、电流和频率的有效值,以及电压合格率(电压偏差)、三相不平衡度、电压波形畸变率及闪变等数据,将其分为历史数据和在线监测数据两类进行处理。对于海量历史数据,按照 IEEE 的电能质量数据交换格式 PQDIF 进行数据统一转换,生成训练集和测试集;对于在线监测数据,数据通过增量数据采样获取,同时将采样结果实时转换为 PQDIF 格式进行存储,结合评估节点的其他数据共同生成实时评估样本集。

②公共信息数据集成和电力用户数据集成。对于离散型属性(如用户敏感度等级、天气情况等),可直接作为样本的特征项进行处理,当该属性为数值型时(如电压暂降深度),则将其转化为离散型属性,通常使用值域平均划分的方法进行离散化处理。对于难以转化的数值型属性(如用户投诉次数、用户负荷容量等),则使用归一化方法进行处理。

在对电能质量监测数据、公共信息数据、电网用户数据三类数据采集的基础上,使用统一的分类模型进行数据转换和集成,生成用于分类的训练集、测试集和评估集。其中,训练集和评估集通过海量历史数据生成,包含了公共信息数据和电力用户数据,以及持久化的电能质量监测数据。分类算法中,使用训练集进行分布式学习并生成电能质量状态分类器,并使用测试集评估分类器分类的准确

率。暂态电能质量评估阶段，对电能质量监测数据进行增量数据采样，结合评估节点的公共信息数据和电力用户数据共同生成实时评估样本集，并根据当前分类规则库在各评估节点分布式地评估暂态电能质量。

仿真实验中，将以上 3 个专线模拟为暂态电能质量评估的分布式训练节点和评估节点，例如某一时刻 t_i 的训练集和测试集的部分结构如表 5-1 所示。

表 5-1 训练/评估数据集结构

| 训练/评估节点 | 暂态事件监测数据 | | | 用户数据 | | | 公共数据 | | | ... | 暂态电能质量评估级别 |
|---------|-----------------|-------------------|------------------|-----------------|-----------------|--------------|------------|------------|-----------|-----|------------|
| | 电压暂降深度 (1~10 级) | 电压暂降持续时间 (1~10 级) | 电流值变化范围 (1~10 级) | 用户敏感度等级 (I~V 级) | 用户重要度等级 (I~V 级) | 负荷容量 (I~V 级) | 温度 (1~5 级) | 湿度 (1~5 级) | 天气 (根据标准) | | |
| 专线 1 | 2 级 | 5 级 | 3 级 | V 级 | II 级 | IV 级 | 适宜 | 适宜 | 雷雨 | ... | 异常 |
| 专线 2 | 3 级 | 9 级 | 7 级 | II 级 | IV 级 | I 级 | 低温 | 湿度大 | 大雪 | ... | 故障 |
| 专线 3 | 8 级 | 2 级 | 7 级 | I 级 | IV 级 | III 级 | 高温 | 湿度极大 | 晴 | ... | 正常 |

为了公正的反应分类结果和好坏，主要使用以下两个评价指标：分类总体准确率和类别 F1-measure 值，其定义如公式 5-2 和 5-3 所示。其中准确度使用的是整个训练集的分类准确率，而不是单个类别的分类准确率。对于单个类 $class_i$ 来说，则使用 F1-measure 作为衡量指标，它综合考虑了当前类 $class_i$ 的准确率和召回率，因而更为准确、合理地反映了当前类别的分类效果。

$$Precision(all) = \frac{\text{分类正确的样本数}}{\text{测试集的样本总数}} \quad \text{式 5-2}$$

$$F1 - measure(class_i) = \frac{2 \times Precision(class_i) \times recall(class_i)}{(Precision(class_i) + recall(class_i))} \quad \text{式 5-3}$$

实验采用如下方案：分别将 2010、2011 和 2012 年的数据作为训练集，将 2013 年的数据作为测试集，分别调用 Weka 工具集中支持向量机(SMO)、决策树(DT)、贝叶斯(NB)分类算法，分类的准确率对比如图 4 所示。

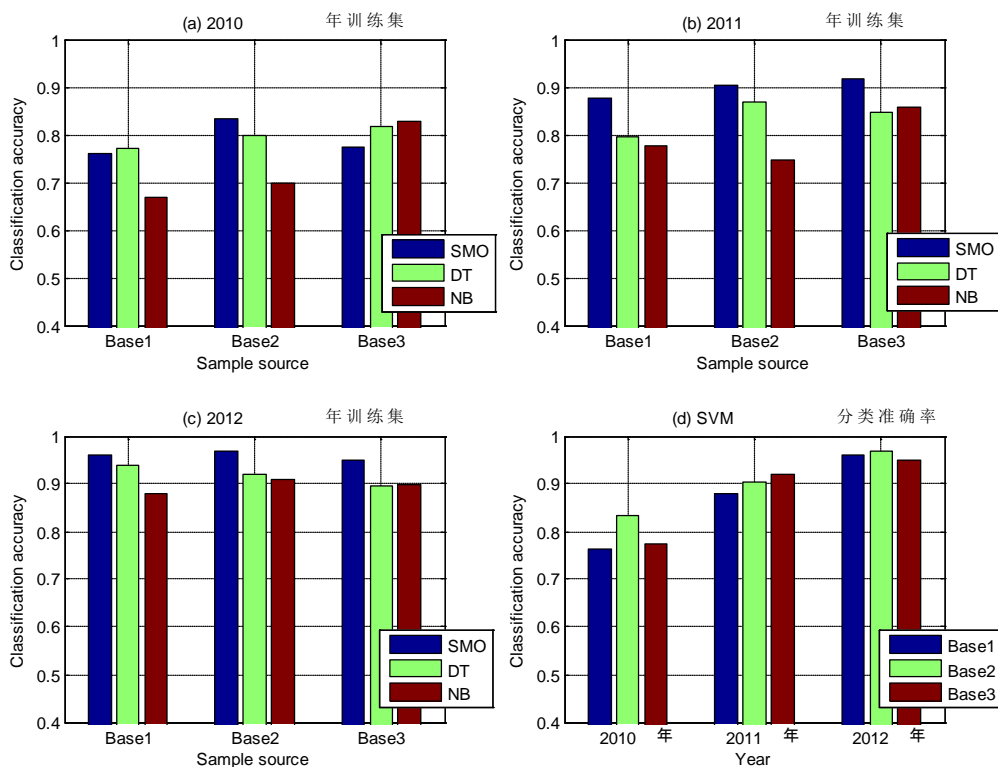


图 5-1 不同训练集下的分类准确率对比

根据如图 5-1 中的图(c)所示的实验结果,在使用 2012 年样本作为训练集时,使用各种分类算法进行电能质量综合评估能取得较好的准确性。由图(a)、(b)和(c)的比较可以得出,支持向量机算法的效果整体优于决策树分类和贝叶斯分类算法。根据图(d)的结果,在使用 2012 年样本作为训练集时,支持向量机分类的准确率均达到了 90% 以上。同时,结果表明训练集的产生时间对于分类准确率的影响较大,使用 2012 年训练集时,得到的分类准确率明显优于 2010 年训练集。这表明训练集和测试集产生时间越接近,其分类准确率越高。其原因在于电网和用户情况都是不断变化的,特别是各类电力电子装置(不管是电网公司安装的还是用户安装的)正显著改变着电能质量的影响。

由于电能质量综合评估与监测装置和区域用户密切相关,电能质量综合评估的难点在于如何融合分布式存储资源和计算资源,既保证分类规则库的定期更新,又能保证状态评估的实时性。本报告采用分布式机器学习和流数据实时处理结合

的数据挖掘架构，使用分布式机器学习算法，将分布式的海量历史数据作为机器学习训练集，周期性的运行分布式学习算法更新分类器，并自动部署到所有电能质量的评估节点。

6 基于 Hadoop 的分布式电能质量评估

6.1 大数据分析

6.1.1 Hadoop 分布式计算

随着信息技术融入人类社会，数据和信息的复杂性和数量都在不断增长，而且增长速度本身也在加快。互联网、物联网、GPS、金融、电信、安全监控等都在不停地产生大量的新数据。人类社会中越来越多的应用涉及到大数据技术和大数据分析方法。大数据分析的内涵主要包括两方面：一是希望利用到全部数据进行数据挖掘和分析，而不是部分抽样数据；二是研究和开发大数据分析工具，在合理时间内达到数据采集、管理、处理、并整理成为帮助企业经营决策的信息。因此，大数据分析应具备主要五个方面的基本能力：预测性分析能力(Predictive Analytic Capabilities)、数据挖掘算法(Data Mining Algorithms)、可视化分析(Analytic Visualization)、语义引擎(Semantic Engines)、数据质量和数据管理(Data Quality and Master Data Management)。大数据具有以下公认的 4V 特征^[80]：

第一，体量巨大 (Volume)。由于大量传感器的应用，并且非结构化数据的快速增长，数据量级正从 TB 级别快速跃升到 PB 级别。

第二，数据多样性 (Variety)。相对于以往便于存储的以文本为主的结构化数据，非结构化数据越来越多，包括网络日志、音频、视频、图片、地理位置信息等，这些多类型的数据对数据的处理能力提出了更高要求。

第三，价值密度高 (Value)。由于大数据中存在大量的不相关信息，价值密度的高低与数据总量的大小成反比。在大数据背景下，利用数据挖掘技术更迅速地获取知识和有价值的信息，提高数据的价值密度。

第四，处理速度快 (Velocity)。大数据时代很多应用要求近乎实时地进行分析而非批量式的分析，要求有新的大数据处理工具或架构，这也是大数据区别于传统的数据仓库技术和 BI 技术的特征之一。

Hadoop 是一个用于分布式计算的基础架构，能够帮助用户在不了解底层分布式细节的前提下，充分利用集群的威力高速运算和存储，并开发分布式程序处理海量异构数据。Hadoop 可以在大量廉价的硬件设备组成的集群上运行分布式应用程序，为应用程序提供了易用的分布式编程接口，从而构建一个可扩展的分布式系统。Hadoop 由许多组件和项目组成，其核心组件包括分布式文件系统 HDFS(Hadoop Distributed Filesystem)和 MapReduce(Google MapReduce 的开源实现)，并包含了列式数据库 HBase、分布式锁服务 Zookeeper、Pig、Hive 等组件^{[81][82]}。

Hadoop 生态圈涉及的项目和组件关系架构图如图 6-1 所示。

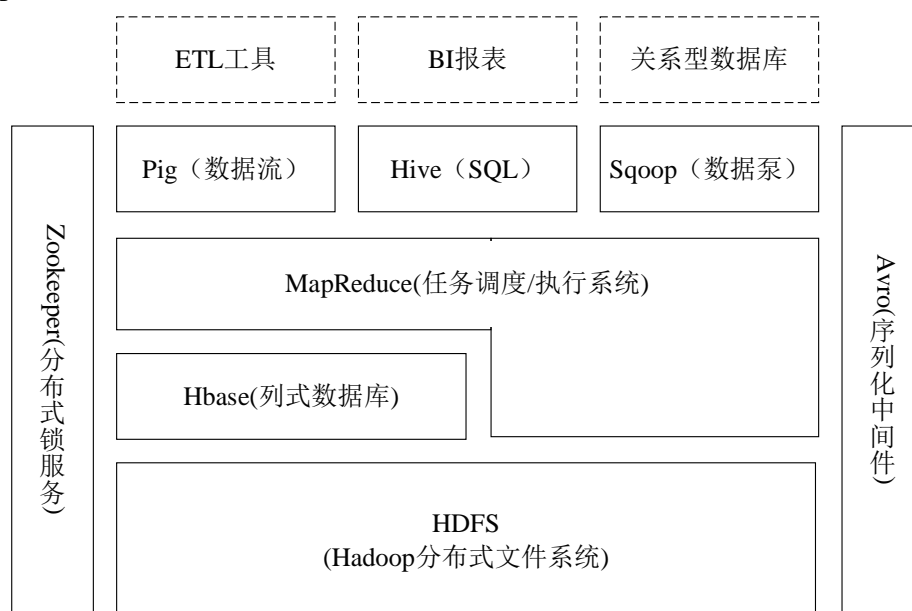


图 6-1 Hadoop 生态圈架构图

HDFS 是 Hadoop 的底层组件，负责存储 Hadoop 集群中所有存储节点的文件。HDFS 架构由一组节点构成，包括 NameNode（仅一个）和多个 DataNode。如图 6-2 所示，HDFS 内部的所有通信都基于标准的 TCP/IP 协议，NameNode 负责管理文件系统名称空间和控制外部客户机的访问，DataNode 响应来自 HDFS 客户机的读写请求，而且响应创建、删除和复制来自 NameNode 关于数据块的命令。存储在 HDFS 中的文件会被划分为块（默认大小为 64MB）存储于 DataNode 中，并将其复制为 3 份以保证文件的可靠性。文件索引时从 NameNode 中查找块的位置。

置，然后到相关的 DataNode 数据节点中访问数据目录。在 HDFS 运行的整个过程中，DataNode 要利用心跳检测机制频繁的向 NameNode 发送心跳，如果 NameNode 未能接收到来自 DataNode 的任何心跳，就将其检测为故障。

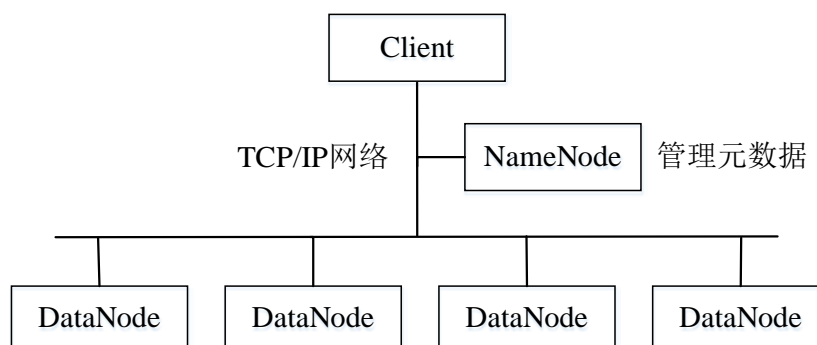


图 6-2 HDFS 节点结构

MapReduce 是用于并行处理大数据集的并行化编程框架，其并行处理模式对任务的处理分为两个阶段：Map（映射）和 Reduce（规约），其基本功能分别是把任务分解为多个子任务和分解后多任务处理结果汇总。在部署上，MapReduce 框架和 HDFS 通常运行在一组相同节点上，计算节点和存储节点通常部署在同一个分布式节点，从而提高集群任务调度的效率。MapReduce 框架由 TaskTracker 和 JobTracker 两个服务组成，其中 TaskTracker 是具体负责任务执行的服务，而 JobTracker 为唯一一个主控服务，负责调度和管理程序运行情况，若 TaskTracker 出现故障，JobTracker 会把该 TaskTracker 负责的任务分配给其他空闲的 TaskTracker。

如图 6-3 所示，在 HDFS 中以 key-value 对形式存储的数据将作为 Map 阶段的输入。在 Map 阶段，根据指定的 InputFormat 类从文件块中读取数据并产生 key-value 对，然后调用自定义的 Map 函数进行处理并产生中间结果存放在本地。在 Reduce 阶段，远程读取 Map 阶段产生的中间结果，调用自定义的 Reduce 函数进行处理，并将最终结果存储到 HDFS。用户只需根据具体情况实现 Map 和 Reduce 函数，这样极大的降低了用户并行化编程的难度。

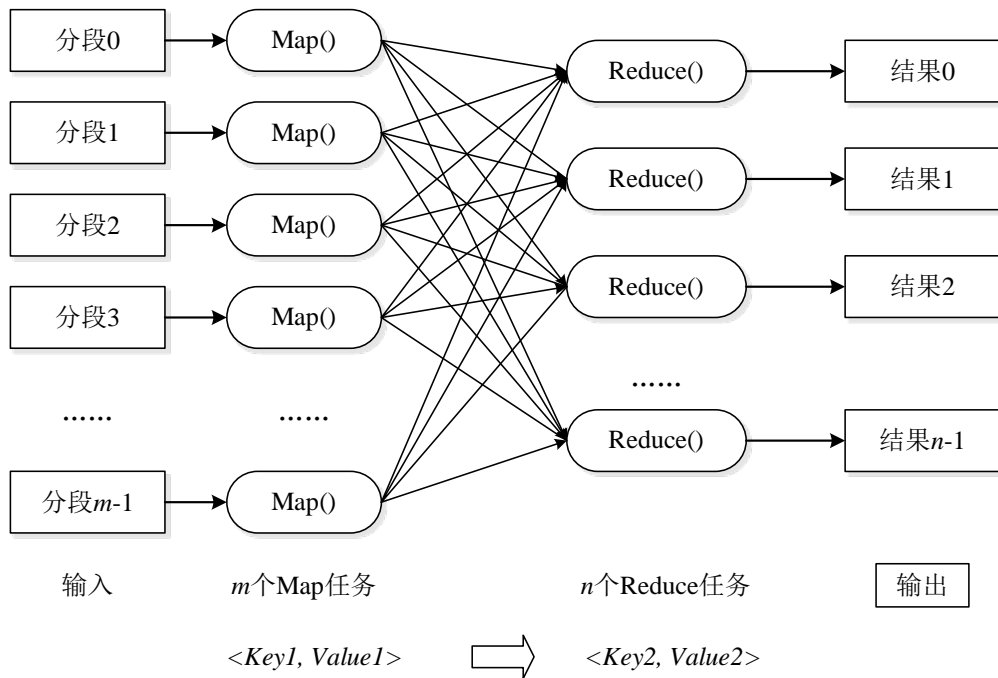


图 6-3 MapReduce 处理流程图

6.1.2 流数据的处理和框架

大数据时代，海量数据实时地、互不兼容地产生于社交网络、网络视频、传感器网络、电子商务等领域，很多新型应用需要长期地、持续地对数据进行流式处理、连续计算和分布式流数据处理等。“流数据”通常是指连续的、不间断的、非结构化的数据队列。因此，流数据可被视为一个随时间延续而无限增长的动态数据集合，通常具有以下特点：数据实时到达，输入顺序无法保证，数据的处理是一次性的，不是静态的存储后处理，而是动态的随到随时处理。数据经过处理后，如果不特意进行持久化操作，一般直接丢弃。

流计算是指一种高效地利用并行和定位，使用流计算处理器，流计算编程语言等多种技术手段处理流数据的新型计算模式^{[79][83]}。不同于大数据中的面向非实时数据的批处理计算框架（例如 Hadoop），流计算面向的数据规模庞大且实时持续不断地到达，数据次序独立且时效性强，同时流数据的价值会随着时间的流

逝而降低，要求数据在产生后必须立即对其进行处理。面对这种“大数据流”，传统的分布式计算模型不再能满足需求，而批处理计算框架在实时性、容错性等方面都有所欠缺。能够实时处理流动数据并作出合适决策的流计算技术应具备以下特点：

(1) 为了使得对数据流的处理延迟尽可能的低，系统必须摒弃将数据存储下来处理的传统做法，实时对数据进行处理然后丢弃。

(2) 流计算系统无法对数据流的产生和到达做任何假设，而且需要兼容静态数据和流数据，应用需要将当前数据与过去数据对比。

(3) 流计算需要具备在分布式节点上拓展的能力，各处理单元对实时数据进行提取、过滤和分析等操作和处理，单元只负责完成自身处理功能，每个处理单元地位平等，不能相互干涉。

(4) 流计算模型多核处理器和多线程的应用，充分发掘系统的计算资源，当负载增大时应当采用均衡技术使得负载转移到相对空闲的节点上去。

(5) 流计算能够实时整合来自多种异构数据源的数据，并对这些数据进行连续实时的处理。

目前，流计算的模型和框架成为业界研究的焦点，并已经形成了一系列分布式流计算框架，如 Yahoo! S4^[84]，Facebook Data Free-way and Puma^[85]，Twitter Storm^[86]等等。

(1) S4 (Simple Scalable Streaming System)

S4 是一个通用的、分布式、可伸缩的、部分容错的、可插拔的平台，该平台上开发人员可以很容易地开发应用程序来处理持续到达的流数据。S4 提供了简单的编程接口来处理流式数据，可以在普通硬件之上部署可扩展的高可用处理集群。S4 处理集群用于处理客户端发送的海量数据流量处理请求，集群包含很多分布式流计算处理节点，各节点之间完全对等，均具有相同的处理能力，同时每

个处理节点内均含用户自定义的处理模块。，这种集群架构简化了部署和维护，提高了系统的可靠性和健壮性。同时，在 S4 框架中使用了可插拔的架构，各个模块之间松耦合，各模块即通用又可定制化。

S4 的分布式流计算处理集群采取流水线处理方式，各节点中不同的处理模块分别承担不同环节的处理工作，每个处理节点均各自独立运行，没有中心节点。各处理节点处理请求完成后，会将产生的中间结果发往其它的分布式节点继续进行处理，直到产生最终的处理结果，并将结构发送至客户端。

在 S4 中，所有发往分布式流计算处理集群的数据流会被封装成不同类型和属性值的事件，供分布式流处理集群进行处理。处理集群中，最基本的逻辑计算单元被称为一个 PEs (Processing Elements, 处理单元)，各处理单元之间通过事件 (events) 进行交互。在一个 PE 收到或消费一个事件后，执行以下操作中的一项或两项：①发出一个或多个 event 给其它 PE。②公布结果。每个 PE 只消费事件类型、属性 key、属性 value 都匹配的事件(events)，并可能会产生输出事件。分布式流计算框架会为每个属性值初始化一个 PE PE 负责监听某些特定类型的事件，可以通过指定更详细的“键-值”对来注册监听某类更细粒度的符合某种“键-值”关系的事件的特定子集。

处理节点 PN(ProcessingNode)是处理单元 PE 的逻辑主机，一个物理节点上可以有多个处理节点。处理节点构建在通讯层的基础上，通信层提供了集群管理和容错功能，实现物理节点和逻辑节点的映射，通信层能够自动检测到硬件故障和相应更新映射。在此基础上，处理节点负责监听事件，在到达事件上执行操作，通过通讯层的协助分发事件，并发送输出事件。S4 通过一个哈希函数将每个事件路由到处理节点，这个哈希函数作用于事件的所有已知属性值上。单个事件可能被路遇到多个 PN 上。所有可能的属性 Key 的集合通过 S4 集群的配置文件获知。PN 中的事件监听器将到来的事件传递给 PE 容器，PE 容器以适当的顺序调

用适当的 PE。

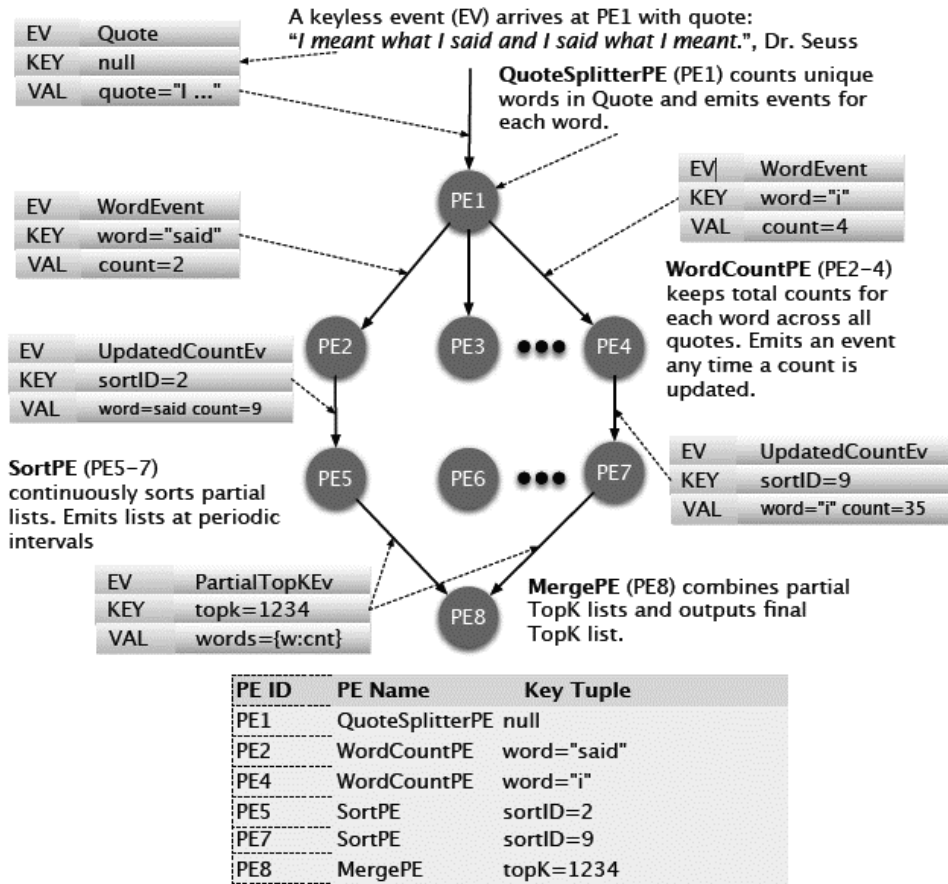


图 6-4 S4 单词计数流计算示例^[84]

在图 6-4 描述的示例中^[84]，输入事件是一系列英文文档，任务是以最小的延迟持续的计算得出所有文档中出现最频繁的 K 个词。Quote 事件发送到 s4，该事件是没有 key 值的；然后，QuoteSplitterPE 对象(如图 2 中的 PE1)侦听系统中的 Quote 事件，QuoteSplitterPE 是一个无 key 值的 PE 对象，处理所有 Quote 事件。对于文档中每个单独的单词，QuoteSplitterPE 对象分配一个数，并发出新类型的事件 WordEvent，以 word 为 key 值。WordCountPE(如图 2 中的 PE2-4)以 word 为 key 值监听 WordEvent 事件。每次发现 key 值和 word 相匹配的单词，如果 WordCountPE 对象存在，PE 对象和计数器将递增，否则新 WordCountPE 对象实例化。当 WordCountPE 对象的计数器发生变化，它会将更新计数 SortPE 对象(如图 2 中的 PE5-7)。SortPE 对象的 key 值是一个[1, n]的随机整数，其中 n 是 SortPE

对象所需的数量。一旦 *WordCountPE* 对象选择一个 *sortID*，那么在它的整个生命周期都会使用该 *sortID* 为发出事件的 Key 值。使用多个 *SortPE* 对象的目的是为了更好地将负载分布在多个处理节点或处理器。当 *UpdatedCountEvent* 事件到达时，每个 *SortPE* 对象更新 topK 列表，并定期地将其所属的 topK 列表发送 *PartialTopKEv* 消息到 *MergePE* 对象(如图 2 中的 PE 8)。*MergePE* 对象监听所有合并所有 *SortPE* 发出的 key 值为 topK 列表的事件，并合并各 topK 列表输出最终的全局 topK 列表。

(2) Twitter Storm

Storm 与其他大数据解决方案的不同之处在于它的处理方式。*Hadoop* 在本质上是一个批处理系统。数据被引入 *Hadoop* 文件系统 (HDFS) 并分发到各个节点进行处理。当处理完成时，结果数据返回到 HDFS 供始发者使用。*Storm* 支持创建拓扑结构来转换没有终点的数据流。不同于 *Hadoop* 作业，这些转换从不停止，它们会持续处理到达的数据。

Storm 是一个分布式的、容错的实时计算系统，遵循 Eclipse Public License，可以方便地在一个计算机集群中编写与扩展复杂的实时计算。*Storm* 为分布式实时计算提供了一组通用原语，可被用于流计算之中，实时处理消息并更新数据库。*Storm* 也可被用于“连续计算” (continuous computation)，对数据流做连续查询，在计算时就将结果以流的形式输出给用户。*Storm* 还可被用于“分布式 RPC”，以并行的方式运行昂贵的运算。在流计算处理过程中，*Storm* 的主要特点如下：

- 1) *Storm* 使用 ZeroMQ 作为其底层消息队列，消除了中间的排队过程，使得消息能够直接在任务自身之间流动。

- 2) 在 *Storm* 框架之上可以使用多种常见编程语言，如 Clojure、Java、Ruby 和 Python 等，并且可以增加对其他语言的支持。

- 3) *Storm* 可以自动管理工作进程和节点的故障，具有很强的容错性。*Storm*

实现了有保障的消息处理，保证每个消息至少能得到一次完整处理，如果发现一个消息还未处理，会实现消息源重试消息。Storm 还实现了任务级的故障检测，在一个任务发生故障时，消息会自动重新分配以快速重新开始处理。

4) Storm 框架下，流计算可以在多个线程、进程和服务器之间并行进行的。

5) Storm 有一个“本地模式”，可以在处理过程中完全模拟 Storm 集群。这让你可以快速进行开发和单元测试。

6) Storm 集群由一个主节点和多个工作节点组成。主节点运行守护进程 Nimbus，工作节点运行守护进程 Supervisor。其中，Nimbus 的作用包括布置任务、分配代码及故障检测，Supervisor 的作用是监听和控制工作进程的启动与销毁。Nimbus 和 Supervisor 都具有很好的健壮性，它们都能够快速的从失败中恢复，而且是无状态的，它们的协调工作是由 Zookeeper 来完成的。

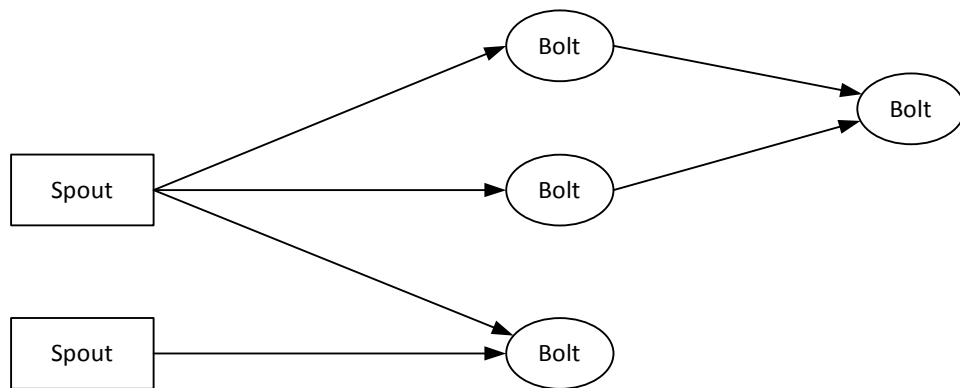


图 6-5 Storm 拓扑结构的概念架构

流计算中，Storm 框架中的角色通常包括 Stream、Tuple、Spout、Bolt、Topology、Stream Grouping 等，如图 6-5 所示。其中，Stream 是指被处理的流数据，这是一个无限的元组 Tuple 组成的序列。Storm 认为每个 stream 都有一个数据源 Spout，数据从外部来源 Spout 流入 Storm 拓扑结构中。Bolt 可以理解为流数据的中间状态转换，Stream Grouping 规定了 Bolt 可接收输入数据的类型，Topology 是由 Stream Grouping 连接起来的 Spout 和 Bolt 节点网络。

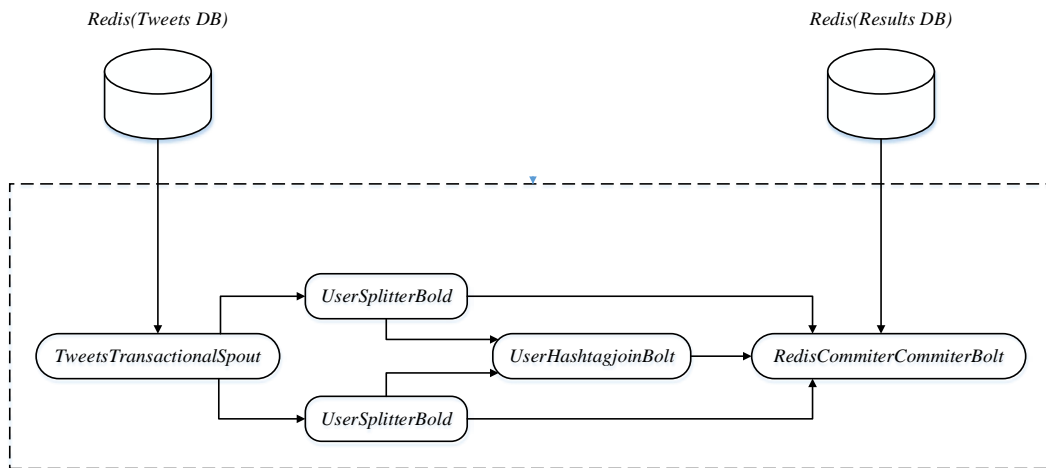


图 6-6 Topology 节点网络^[87]

在图 6-6 描述的示例中^[87]，*TweetsTransactionalSpout* 会连接到你的 tweets 数据库并且在 Topology 节点网络中作为数据源发射批量的元组。Topology 节点网络中包括两个不同的 Bolts，*UserSplitterBolt* 和 *HashtagSplitterBolt*，它们从 *TweetsTransactionalSpout* 接收元组。*UserSplitterBolt* 会分析 tweet 数据元组，查找用户名(以@为开头的词)，并发送这些词到一个叫做 *users* 的自定义流。*HashtagSplitterBolt* 同样分析 tweet 数据元组，查找标签词(以#为开头的词)，并且发送这些词到一个叫做 *hashtags* 的自定义流。Topology 节点网络中还包括第三个 Bolt：*UserHashtagJoinBolt*，它接收 *users* 和 *hashtags* 两个流并且计算在一个命名用户的 tweet 中一个 hashtag 出现了多少次。最后，Topology 中还包括一个 Bolt 叫做 *RedisCommitterBolt*，接收由 *UserSplitterBolt*，*HashtagSplitterBolt* 和 *UserHashtagJoinBolt* 产生的流。在该示例中，它会完成所有计数，并且一旦完成了元组批次的处理就会发送到 Redis 数据库中。

6.1.3 基于 Mahout 框架的分布式分类

Apache Mahout 是 Apache Software Foundation (ASF) 旗下的一个开源项目，它基于 Hadoop 的分布式计算框架，提供了一系列可扩展的机器学习领域经典算法的实现，旨在帮助开发人员更加方便快捷地创建智能应用程序，可以辅助开发

人员更好更快的开发数据分析程序。Mahout 基于 MapReduce 实现了一些数据挖掘算法，解决了部分并行挖掘的问题。Mahout 提供了一套具有可扩充能力的类库，它提供分布式计算框架的同时，还实现了一些可扩展的数据挖掘和机器学习领域经典算法，可以帮助开发人员更加方便快捷地创建智能应用程序 [88][89]。Mahout 中的算法能够高效地运行在分布式计算环境中，通过和 Apache Hadoop 分布式框架相结合，Mahout 可以有效地运行分布式数据挖掘算法。

Mahout 中的分类算法可以广泛地应用于海量数据分类工程中。一般来说，Mahout 分类算法对资源的要求不会快于训练数据和测试数据的增长速度，而且能够转换为分布式处理，从而可以通过扩展分布式处理节点的数量解决大规模数据集分类问题。如图 6-7 所示，当训练例子的数量相对较小时，与 mahout 分类算法相比，传统数据挖掘方法能有相同或更好的性能。但当样本数量增加后，传统不可扩展的分类算法架构所需的处理时间快速增加，此时 Mahout 的可伸缩和并行算法的优势就变得明显。

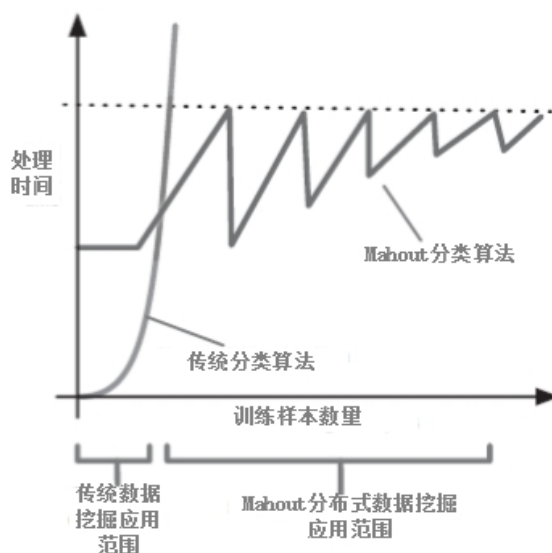


图 6-7 使用 Mahout 分类和传统分类算法的处理时间

Mahout 的主要优势是其强劲的分布式处理能力，可应对不断增长的数据集。Mahout 拥有多种分类算法，其中大多数算法都是为了在 Hadoop 上运行而编写的（除了随机梯度下降法），并具有可伸缩性。表 6-1 列举了 Mahout 中的各种分

类算法和其所适用的数据规模，并对于特定分类问题推荐了最合适的算法。以朴素贝叶斯算法为例说明 Mahout 中的分类算法处理流程。朴素贝叶斯分类算法以速度快和准确性高而著称，但其关于数据的简单（通常也是不正确的）假设是完全独立的，该算法包括两个流程：第一个步骤称作训练（training），它将通过特定的分类算法学习已标记类别的样本来创建一个分类器，并使用部分已标记类别的样本测试该分类器的准确性。第二个步骤称作分类，它将使用在训练阶段中创建的分类器对未标记类别的样本进行归类，实现对于样本类别的预测。因此，要运行 Mahout 的分类器，首先需要训练得到分类器，然后再使用该分类器对新内容进行分类。

表 6-1 Mahout 分类学习算法特点

| 数据集类型 | Mahout 算法 | 处理模式 | 特点 |
|----------------------------|--|----------|--|
| 小型到中型数据集(千万等级以下的训练样本数目) | 随机梯度下降算法(Stochastic gradient descent, SGD) | 序列化, 增量式 | 适用于所有数据类型的分类和预测, 在百万训练样本等级上运行效率高。 |
| 中型到大型数据集(千万等级到数亿等级的训练样本数目) | 支持向量机算法(Support Vector Machine, SVM) | 序列化 | 仍在实验阶段, 可在大型数据集中使用 |
| | 朴素贝叶斯分类算法(Naive Bayes) | 分布式 | 非常适合对文本数据进行分类, 中到高等级的训练开销, 可有效处理对于 SGD 算法来说太大的数据集 |
| | 互补朴素贝叶斯分类算法(Complementary Naive Bayes) | 分布式 | 有效处理对于 SGD 算法来说太大的数据集, 并解决了朴素贝叶斯算法的使用限值(如数据的独立性不符合要求)。 |
| 小型到中型数据集(千万等级以下的训练样本数目) | 随机森林算法(Random forests) | 分布式 | 适用于所有数据类型的分类和预测, 训练开销高, 适用于复杂的分类场景, 能处理数据中的非线性和条件关系 |

Mahout 中朴素贝叶斯和互补朴素贝叶斯算法均可并行运行，在多个有效机器中工作，因此他们可应用于比 SGD 算法更大的数据集中。如果要解决千万级别的训练样本，并且预测变量是单一的、文字性的值，朴素贝叶斯或互补朴素贝叶斯算法是最佳选择。Apache Mahout 官方网站给出了两个非常适合朴素贝叶斯算法解决的实例：Wikipedia Example 和 Twenty Newsgroups Classification Example，

可以通过运行这两个案例来了解 Mahout 分类算法的运行过程。以 Wikipedia Example 为例，在运行训练程序和分类器之前，首先需要下载用于训练和测试的文档样本。在获取 Wikipedia 副本的基础上有 HDFS 将数据划分为存储块，然后对每个存储块内的数据按照国家运行分类算法，训练一个分类器来预测一篇文章所属的国家类别，代表此国家对于该文章没有阅读权限。

6.2 基于分布式分类的电能质量综合评估框架

电能质量综合评估面对的数据对象包括海量历史数据和实时数据，因此本报告中的大数据处理平台需要同时具备分布式数据和流数据处理能力，并基于 Hadoop 框架实现数据转换、数据融合、数据集成。在获取数据的基础上，采用分布式分类学习和流数据实时处理结合的数据挖掘架构，基于 Naive Bayes 分类算法实现暂态电能质量状态的分类（正常、异常、严重、故障等类别），从而进行状态评估和故障预警。本课题采用分层的大数据体系结构，将面向暂态电能质量综合评估的大数据框架分为由数据源层、分布式文件管理层、大数据处理层、和分布式挖掘层构成，如图 6-8 所示。

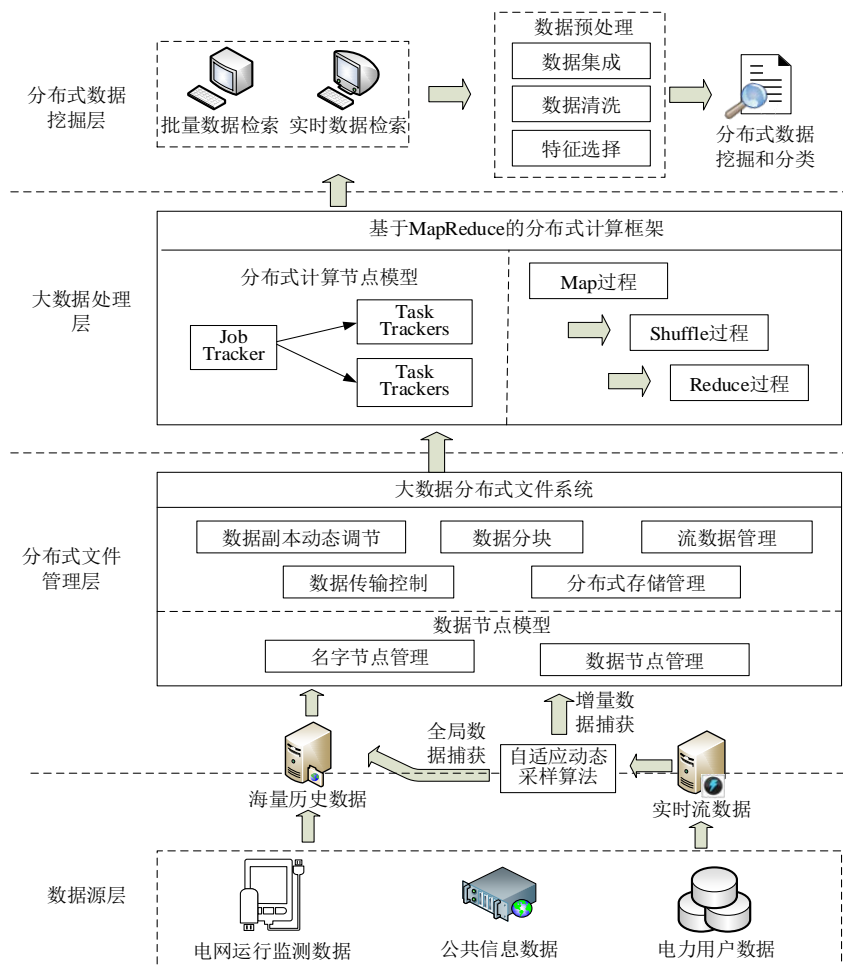


图 6-8 面向暂态电能质量综合评估的大数据框架

(1) 数据源层：通过多源数据融合和数据集成，实现跨业务系统、跨安全区域、多传感器之间的数据获取，并分为海量历史数据和实时流数据进行处理，分别用于分类器训练和实时分类过程。

(2) 分布式文件管理层：使用名字节点和数据节点的主从结构，能提供高吞吐量的数据访问，可实现大规模数据集和流数据的存储管理、副本动态调节、传输控制和数据分块，具有高容错性和可移植性。

(3) 大数据分布式处理层：分布式计算框架拟采用业界通用的 MapRduce 分布式计算框架，Job Tracker 节点通过调度 Task Trackers 节点上的任务协调所有运行在系统上的作业；并针对分布式任务构造“Map（映射）”函数和“Reduce（规约）”函数实现大数据的并行处理，设计适用于分布式处理的 key/value 对，并

在 shuffle 过程中对中间的 key/value 对集快速排序，并使用 HashMap 将具有相同 key 值分发到同个 Reducer 中，Reduce 函数合并所有的具有相同中间 key 的中间 value，并产生输出数据。

(4) 数据挖掘层：在 MapReduce 分布式计算框架的基础上，使用批量数据检索和实时数据检索获取数据，并对数据进行预处理，然后分布式地运行数据挖掘算法，并实现暂态电能质量状态分类。

6.3 基于分布式贝叶斯的电能质量综合评估算法

6.3.1 分布式评估模型

由于电能质量综合评估与监测装置和区域用户密切相关，其综合评估的需求和数据基础都具备天然的并行性。因此，本报告基于 Hadoop 分布式计算框架，把运行于单机上的机器学习算法转化为了 MapReduce 处理模式，解决部分并行数据分析和数据挖掘的问题，提升算法可处理的数据量和处理性能。而且，当训练集和测试集数据量不断扩展时，分布式分类算法可以通过扩展分布式处理节点的数量解决大规模数据集分类问题。Hadoop 架构主要实现的任务包括：

(1) 基于 Hadoop 架构的电能质量分布式文件管理。主要由 Hadoop 中的分布式文件系统(HDFS)实现电能质量数据的分布式管理，HDFS 采用主/从架构，由一定数量的数据节点(Datanode)和一个命名节点(Namenode)组成。将分布式的电能质量监测点作为 Datanode，负责管理该监测点的电能质量监测数据，并在内部将数据划分成一个或多个数据块(Block)，并在 Namenode 的统一调度下进行数据块的创建、删除和复制。通常由电能质量数据中心作为 Namenode，负责执行文件系统的命名空间操作，如打开、关闭、重命名文件等，同时负责数据块到具体 DataNode 节点的映射。使用名字节点和数据节点的主从结构，能提供高吞吐量的数据访问，可实现大规模数据集和流数据的存储管理、副本动态调节、传输控制和数据分块，具有高容错性和可移植性。

(2) 训练集和测试集的生成。将电能质量监测数据、电力用户数据和公共信息数据三类数据形式化表示为一个 n 维向量的形式。训练集和测试集的形式为 $X=\{x_1, x_2, \dots, x_n, C_i\}$ 。其中, x_1, x_2, \dots, x_n 等特征项为用于电能质量评估的各个属性, C_i 为电能质量的评估结果。分布式贝叶斯分类的任务就是将评估样本 $\{x_1, x_2, \dots, x_n\}$ 划分至正常状态、异常状态、严重状态、故障状态四个类别, 分别记为 C_1, C_2, C_3, C_4 。

(3) 分布式分类算法。运行于 Hadoop 上的分类算法能够高效地应用于海量数据分类工程中。一般来说, 随着训练数据和测试数据的增长, 分布式分类算法可以通过扩展分布式处理节点的数量解决大规模数据集分类问题。当训练例子的数量相对较小时, 与分布式分类算法相比, 传统数据挖掘方法能有相同或更好的性能。但当样本数量增加后, 评估样本集的生成和分类需要近乎实时的处理, 传统不可扩展的分类算法架构所需的处理时间快速增加, 此时 MapReduce 提供的可伸缩性和并行性优势就变得明显。

6.3.2 分布式贝叶斯算法模型

Naive Bayes 分类有着坚实的数学基础, 模型所需估计的参数很少, 对缺失数据不太敏感, 在不完备样本的情况下仍能获得相对稳定的分类效率。本课题在 Hadoop 平台上使用 Naive Bayes 分类器实现暂态电能质量综合评估, 使用的数据主要包括电网运行监测数据、电力用户数据和公共信息数据。把上述三类数据形式化表示为一个 n 维向量的形式, 即 $X=\{x_1, x_2, \dots, x_n\}$, 同时需要将样本划分至正常状态、异常状态、严重状态、故障状态四个类别形式化为 C_1, C_2, C_3, C_4 , 根据如公式 6-1 所示的贝叶斯定理:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad \text{式 6-1}$$

贝叶斯分类问题可以转化为求得 $P(C_i|X)$ 最大的类标号 C_i 的过程, 如公式 6-2 所示。

$$f(X) = \max_{i=1,2,3,4} P(C_i | X) \quad \text{式 6-2}$$

其中， $P(C_i | X)$ 为 X 发生时类 C_i 的条件概率。 $P(C_i)$ 为先验概率可由训练样本直接得出，如公式 6-3 所示，其中 S_{C_i} 表示训练样本中属于类 C_i 的数量， S 为训练样本的总数。

$$P(C_i) = \frac{S_{C_i}}{S} \quad \text{式 6-3}$$

在 Naive Bayes 分类中，条件概率 $P(X | C_i)$ 可通过公式 6-4 计算，其中 $P(x_j | C_i)$ 表示属于类 C_i 中样本的第 j 个属性为 x_j 的概率。

$$P(X | C_i) = \prod_{j=1}^n P(x_j | C_i) \quad \text{式 6-4}$$

在分类模型所使用三类数据中，电网运行监测数据是一种实时流数据的形式，公共信息数据和电力用户数据更新频率相对较低。本课题将暂态电能质量评估的数据划分为海量历史数据和实时流数据，分别用于分类器训练和状态分类中。海量历史数据包含了公共信息数据和电力用户数据的，以及持久化的电网运行监测数据。算法中将海量历史数据作为分类训练集和测试集，使用训练集进行分布式学习并生成暂态电能质量状态分类器，并使用测试集评估分类器分类的准确率，然后将分类器自动部署到所有评估节点（通常是变电站或母线）。暂态电能质量评估阶段，对电网运行监测数据进行增量数据采样，结合评估节点的其他数据共同生成实时评估样本集，并根据当前分类规则库在各评估节点分布式地评估暂态电能质量，其处理流程如图 6-9 所示。

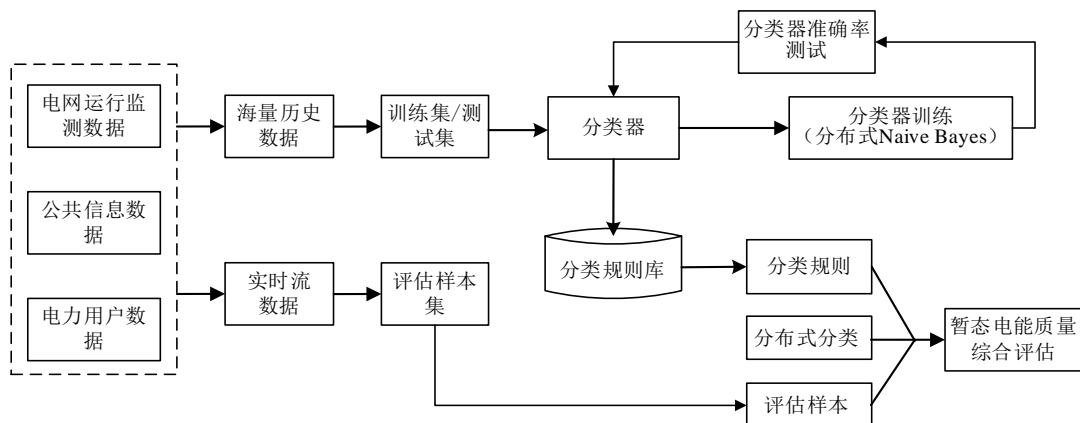


图 6-9 暂态电能质量评估流程图

6.3.3 分布式贝叶斯分类算法

由于暂态电能质量综合评估与监测装置和区域用户密切相关,其综合评估的需求和数据基础都具有天然的并行性。基于 Hadoop 进行大数据分析的理论核心就是分布式数据挖掘方法,通过采用 MapReduce 并行计算模型实现数据挖掘算法的并行化。Naive Bayes 分类算法的数据清洗、特征抽取、分类器训练和分类整个过程都是基于 MapReduce 计算模型分布式进行的。基于 MapReduce 进行涉及 4 个的实体:(1) Client,负责编写 mapreduce 程序,配置任务,提交作业;(2) JobTracker,负责初始化任务,将任务分配给 TaskTracker 并协调整个作业的执行,同个任务只允许有一个 JobTracker;(3)TaskTracker,负责在分配的数据片段上执行 Map 或 Reduce 任务;(4)Hdfs:负责保存任务的数据、配置信息、结果数据等。

在 MapReduce 计算框架下,Naive Bayes 分类器训练算法具体描述如下:

(1) 将分类器中参数的训练和分类任务划分为两个 MapReduce Job: Job1、Job2,分别对应于计算先验概率 $P(C_i)$ 、条件概率 $P(X|C_i)$ 两个任务,由两个 JobTracker 分别来控制它们的执行。输入数据为<Key, Value>对的形式,其中 Key 为样本编号 k , Job1 和 Job2 的 Value 为包括特征项和类别的样本,记作 $[k, X]$ 。

(2) 在 Input Split 阶段,MapReduce 会根据输入文件计算 Input Split, Input Split 中记录了分片长度和数据位置,真正的数据存放于 HDFS 系统中的数据块 (block)。

(3) Job1: 计算先验概率 $P(C_i)$ 。Map 阶段从 Input Split 中获取数据,每个 Mapper 分别计算类别 C_1, C_2, C_3, C_4 中的样本数目和样本总数目,Map 阶段输出数据的 Key 为样本所属的类别 C_i 。Shuffle 阶段将 Mapper 输出的数据进行排序,并使用 $\text{hash}(\text{key}) \bmod 4$ 函数保证相同 key 的 Map 输出被分配至同一个 Reducer 中。Reduce 阶段,将获得所有 Mapper 输出中与之相关的数据,可以容易地计算

C_1, C_2, C_3, C_4 各类别的样本总数及 $P(C_i)$ 。

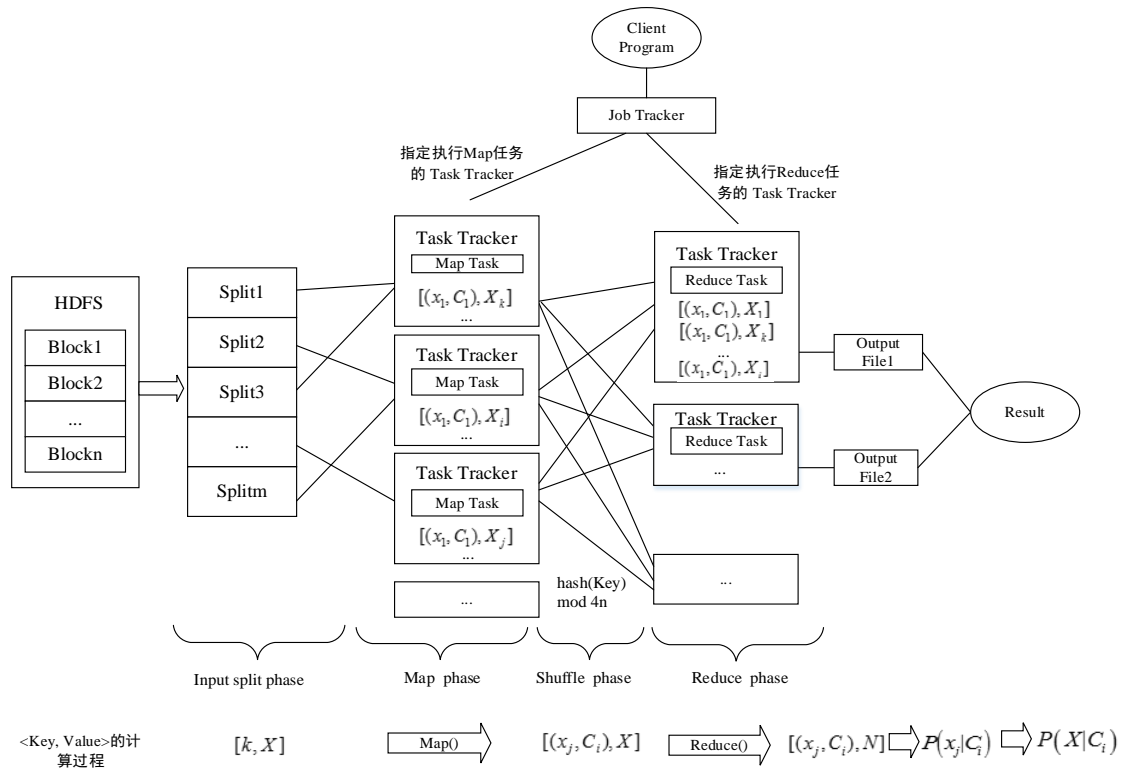


图 6-10 Job2 的 MapReduce 计算过程

(4) Job2: 计算条件概率 $P(X|C_i)$ 。如公式 (4) 所示, 计算 $P(X|C_i)$ 需要计算每个 $P(x_j|C_i)$, 即属于类 C_i 的样本中第 j 个属性为 x_j 的概率。Map 阶段从 Input Split 中获取数据, 每个 Mapper 分别计算属于类 C_i 的样本中第 j 个属性为 x_j 的数量, Map 阶段输出数据的 Key 为 (C_i, x_j) , 并计算 C_i 的样本数量。Shuffle 阶段使用 $\text{hash}(\text{Key}) \bmod 4n$ 函数保证相同 Key 的 Map 输出被分配至同一个 Reducer 中。Reduce 阶段, 首先计算出属于类 C_i 的所有样本中第 j 个属性为 x_j 的总数, 然后根据 Job1 中计算 C_i 样本总数计算 $P(x_j|C_i)$, 最后根据公式 (4) 计算出条件概率 $P(X|C_i)$ 。Job2 的处理过程如图 6-10 所示。

(5) 电能质量状态分类。在每个评估节点, 对没有进行类别标识的样本, 按照公式 (2) 计算样本的 $\max_i P(C_i|X)$, 确定样本属于的类别。

总之, 基于 MapReduce 的分布式贝叶斯分类, 可以用于电能质量的分布式评估, 实现了计算资源和存储资源的统一, 并使业务执行地可以近乎实时地评估

出当前电能质量的综合状态。

参考文献

- [1] Gaouda A M, Salaam M M A, Sultan M R, et al. Power quality detection and classification using wavelet-multiresolution signal decomposition[J]. IEEE Trans on Power Delivery, 1999, 14(4): 1469-1476.
- [2] Wang Min, Gabriel I Rowe, Alexander V Mamishev. Classification of power quality events using optimal time-frequency representations- part 2 : application[J]. IEEE Trans on PowerDelivery, 2004, 19(3): 1496-1503.
- [3] 刘颖英, 徐永海, 肖湘宁. 地区电网电能质量综合评估新方法[J]. 中国电机工程学报, 2008, 28(22): 130-136.
- [4] 徐方维, 杨洪耕, 叶茂清, 等. 基于改进 S 变换的电能质量扰动分类[J]. 中国电机工程学报, 2012, 32(4): 77-84.
- [5] Shen G Q, Xu D H, Cao L P, et al. An improved control strategy for grid-connected voltage source inverters with an LCL filter[J]. IEEE Transactions on Power Electronics, 2008, 23(4): 1899-1906.
- [6] 赵东旭. 电能质量数据中心建设关键问题研究[D]. 硕士学位论文, 北京, 华北电力大学, 2012.
- [7] 余贻鑫, 栾文鹏. 智能电网述评[J]. 中国电机工程学报, 2009, 29(34): 1-8.
- [8] 赵英男. 电能质量的电压闪变检测算法研究[D]. 硕士学位论文, 浙江, 浙江大学, 2013.
- [9] 张全明, 刘会金. 最小二乘支持向量机在电能质量扰动分类中的应用[J]. 中国电机工程学报, 2008, 28(1): 106-110.
- [10] Wang Min, Gabriel I Rowe, Alexander V Mamishev. Classification of power quality events using optimal time-frequency representations- part 2 : application[J]. IEEE Trans on PowerDelivery, 2004, 19(3): 1496-1503.

- [11] Dash P K , Panigrahi B K , Panda G . Power quality analysis using S-transform[J] . IEEE Transactions on Power Delivery, 2003, 18(2): 406-411.
- [12] 刘旭娜, 肖先勇, 汪颖. 电压暂降严重程度及其测度、不确定性评估方法[J]. 中国电机工程学报, 2014, 34(4): 644-658.
- [13] 何吉彪, 程浩忠. 含微网配电网规划中的电能质量综合评估[J]. 电网技术, 2012, 36(8): 209-214.
- [14] 林德清, 顾伟, 王元凯, 等. 基于动态时间弯曲空间距离度量的电能质量综合评估[J]. 电网技术, 2013, 37(2): 562-567.
- [15] 李泓泽, 郭森, 唐辉, 等. 基于改进变权物元可拓模型的电能质量综合评价[J]. 电网技术, 2013, 37(3): 653-659.
- [16] 江辉, 张清联, 彭建春. 基于改进云物元模型的风电场电能质量评价[J]. 电网技术, 2014, 38(1): 206-210.
- [17] 刘俊华, 罗隆福, 张志文, 等. 基于模糊集对分析法的电能质量综合评估[J]. 电网技术, 2012, 36(7): 81-85.
- [18] Lee G J, Albu M, Heydt G T. A power quality index based on equipment sensitivity, cost and network vulnerability [J]. IEEE Transactions on Power Delivery, 2004, 19(3): 1504-1510.
- [19] Guasch L, Córcoles F, Pedra J. Effects of symmetrical and unsymmetrical voltage sags on induction machines [J]. IEEE Transactions on Power Delivery, 2004, 19(2): 774-782.
- [20] 毛丽林, 黎灿兵, 何禹清. 考虑用户需求和专家知识的电能质量综合评估[J]. 电网技术, 2011, 35(10): 135-139.

- [21] Mendes V F, De Sousa C V, Silva S R, et al. Modeling and ride-through control of doubly fed induction generators during symmetrical voltage sags[J]. IEEE Transactions on energy conversion, DEC 2011, 26(4): 1161-1171.
- [22] 康世崑, 彭建春, 何禹清. 模糊层次分析与多目标决策相结合的电能质量综合评估[J]. 电网技术, 2009, 33(3): 113-118.
- [23] 丁泽俊, 朱永强, 陶顺. 事件型电能质量的综合评估[J]. 电网技术, 2011, 35(2): 84-87.
- [24] 张蔓, 林涛, 曹健, 等. 理想区间法在电能质量综合评估中的应用[J]. 电网技术, 2009, 33(3): 33-38.
- [25] Bollen M H J. Understanding power quality problems: voltage sags and interruptions[M]. New Jersey: John Wiley & Sons, Inc, 2000: 1-34.
- [26] 徐文远, 雍静. 电力扰动数据分析学—电能质量监测数据的新应用[J]. 中国电机工程学报, 2013, 33(19): 93-101.
- [27] 曹健, 林涛, 徐遐龄, 等. 一种电能质量扰动监测与识别新方法[J]. 中国电机工程学报, 2011, 31(31): 125-133.
- [28] 焦东升, 陆冬良, 应俊豪, 等. 动态电能质量实时监测系统的设计与实现[J]. 电网技术, 2011, 35(5): 110-114.
- [29] Bowers J S, Sundaram A, Benner C L, et al. Outage avoidance through intelligent detection of incipient equipment failures on distribution feeders[C]//IEEE PES General Meeting, Pittsburgh, USA: IEEE, 2008.
- [30] 王丽颖. 基于 DSP 和 ARM 的电能质量监测系统[D]. 硕士学位论文, 上海, 上海交通大学, 2009.

- [31]梁启权.基于 TMS320F28335 电能质量监测装置的设计与实现[D].硕士学位论文,北京 ,北京交通大学,2011.
- [32]董浩.电能质量监测装置上位机管理系统的研究[D].硕士学位论文,北京 ,北京交通大学,2011.
- [33]卫志农, 吴霜, 孙国强, 等. 多目标电能质量监测器的优化配置[J]. 电网技术, 2012, 36(1): 176-181.
- [34]Kezunovic M. Data analytics: Creating Information and Knowledge [J]. IEEE Power & Energy Magazine, 2012, 10(5): 14-23.
- [35]乔凯庆.新型电能质量装置的数据处理模块和算法的研究与实现[D].硕士学位论文,北京 ,北京交通大学,2009.
- [36]王增平.电能质量在线监测系统的研究[D].硕士学位论文,北京 ,华北电力大学,2008.
- [37]高民.用户电能质量监测技术研究[D].硕士学位论文,北京 ,华北电力大学,2008.
- [38]Lotfifard S, Kezunovic M, Mousavi M J. Voltage sag data utilization for distribution fault location[J]. IEEE Trans. on Power Delivery, 2011, 26(2): 1239-1246.
- [39]Kim J, Grady W M, Arapostathis A, et al. A time-domain procedure for locating switched capacitors in power distribution systems[J]. IEEE Trans. on Power Delivery, 2002, 17(4): 1044-1049.
- [40]张欣欣.电能质量在线监测系统与 PQDIF 的实现[D].硕士学位论文,北京 ,华北电力大学,2008.
- [41]熊志敏.电能质量在线综合监测中电压暂降的识别与评估[D].硕士学位论文,北京 ,华北电力大学,2009.

- [42]朱振伟,方鸽飞,王国福,等. 电网夏季气象敏感负荷特性分析[J]. 华东电力, 2008, 36(9): 53-58.
- [43]康重庆,周安石,王鹏,等. 短期负荷预测中实时气象因素的影响分析及其处理策略[J]. 电网技术, 2006, 30(7): 94-100.
- [44]刘旭, 罗滇生, 姚建刚, 等. 基于负荷分解和实时气象因素的短期负荷预测[J]. 电网技术, 2009, 33(12): 5-10.
- [45]肖先勇,王希宝,薛丽丽,等. 敏感负荷电压凹陷敏感度的随机估计方法[J]. 电网技术, 2007, 33(8): 30-33.
- [46]陈卫东, 肖先勇, 李皖, 等. 敏感负荷电压凹陷敏感度的最佳平方逼近评估法[J]. 电网技术, 2009, 31(22): 55-59.
- [47]Taylor J, Buizza R. Neural network load forecasting with weather ensemble predictions[J]. IEEE Trans on Power Systems, 2002, 17(3): 626-632.
- [48][12] Hor Ching-Lai, Watson S J, Majithia S. Analyzing the impact of weather variables on monthly electricity demand[J]. IEEE Trans on Power Systems, 2005, 20(4): 2078-2085.
- [49]杨育彬,李宁,张瑶. 基于社会网络可视化分析的数据挖掘 [J]. 软件学报, 2008, 19(8): 1980-1994.
- [50]刘大有, 陈慧灵, 齐红, 等. 时空数据挖掘研究进展[J]. 计算机研究与发展, 2013, 50(2): 225-239.
- [51]李锋.面向数据挖掘的隐私保护方法研究[D].硕士学位论文,上海,上海交通大学,2008.
- [52]刘浩熙.数据挖掘在客户关系管理综合决策中的应用[D].硕士学位论文,北京,北京邮电大学,2008.

- [53] Tsoumakas G. Multi-label classification. *International Journal of Data Warehousing & Mining*, 2007, 3(3):1-13
- [54] Hu Yichung. Sugeno fuzzy integral for finding fuzzy if-then classification rules[J]. *Applied Mathematics and Computation*, 2007, 185(1):72-83
- [55] Hu Yichung. Determining membership functions and minimum fuzzy support in finding fuzzy association rules for classification problems[J]. *Knowledge-Based Systems*, 2006, 19(1):57-66
- [56] Bouguessa M, Wang S. Mining projected clusters in high-dimensional spaces. *IEEE Trans. on Knowledge and Data Engineering*, 2009, 21(4):507-522.
- [57] 李雄飞, 李军, 屈成伟, 等. 数据挖掘中平衡偏斜训练集的方法研究[J]. *计算机研究与发展*, 2012, 49(2): 346-353.
- [58] 胡文瑜, 孙志挥, 吴英杰. 数据挖掘取样方法研究[J]. *计算机研究与发展*, 2011, 48(1): 45-54.
- [59] Boser B., Guyon I. and Vapnik V. A training algorithm for optimal margin classifiers[A]. In: *Proceedings of the 5th Annual ACM Conference on Computational Learning Theory*[C], Pittsburgh, PA, USA. ACM Press, 1992, 144-152.
- [60] Wilson M. D. Support Vector Machines[M]. *Encyclopedia of Ecology*. 2008, 3431-3437.
- [61] Andras P.. The equivalence of support vector machine and regularization neural networks[J]. *Neural Processing Letters*. 2002, 15(2): 97-104.
- [62] Vito E. D., Rosasco L., Caponnetto A., et al. Some properties of regularized kernel methods[J]. *Journal of Machine Learning Research*. 2004, 5: 1363-1390.
- [63] Suykens J. Kernel Methods [M]. *Comprehensive Chemometrics*. 2009, Chapter

3.11, 437-451.

- [64]Cristina Giannone, Roberto Basili, Chiara Del Vescovo, et al. Kernel-based relation extraction from investigative data [A]. In Proceedings of the Third Workshop on Analytics for Noisy Unstructured Text Data [C]. New York, USA. 2009.
- [65]Kwok J. T. Y.. The evidence framework applied to support vector machines[J]. IEEE Transactions on Neural Networks. 2000, 11(5): 1162-1173.
- [66]Mika S., Ratsch G., Weston J., et al. Miiller. Fisher discriminant analysis with kernels[A]. In: Proceedings of the 1999 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing IX[C]. 1999, 41-48.
- [67] Baudat G., Anouar F.. Generalized discriminant analysis using a kernel approach[J]. Neural Computation. 2000, 12(10): 2385-2404.
- [68] Mika S., Scholkopf B., Smola A. J., et al. Kernel PCA and de-noising in feature spaces[J]. In: Advances in Neural Information Processing Systems(NIPS). 1998, 536-542.
- [69]张莉, 周伟达, 焦李成. 核聚类算法[J]. 计算机学报. 2002, 25(6): 587-590.
- [70][92] Choi H., Choi S. Kernel Isomap[J]. Electronics Letters. 2004, 40(25): 1612-1613.
- [71][93] Choi H. and Choi S. Robust kernel Isomap[J]. Pattern Recognition. 2007, 40(3): 853-862.
- [72]Rumelhart D. E.and McClelland J. L.. Learning internal representations by error propagation[M]. Parallel distributed processing: explorations in the macrostructure of cognition. Cambridge, MA: MIT Press. 1986, 1-2: 318-362.
- [73]边肇祺, 张学工. 模式识别(第二版)[M]. 北京: 清华大学出版社. 2000.

- [74]Poggio T., Riflcin R., Mukherjee S., et al. General Conditions for predictivity in learning theory[J]. Nature. 2004, 428: 419-422.
- [75]Vito E. D., Rosasco L., Caponnetto A., et al. Some properties of regularized kernel methods[J]. Journal of Machine Learning Research. 2004, 5: 1363-1390.
- [76]Vapnik V. N. Statistical learning theory[M]. New York: John Wiley and Sons. 1998.
- [77]Oren M., Papageorgiou C., Sinha P., et al. Pedestrian detection using wavelet templates[A]. In: Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. 1997, 193-199.
- [78]Osuna E., Freund R., Girosit F. Training support vector machines: an application to face detection[C]. Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1997, 130-136.
- [79]KUO K, RABBAH,R. A productive programmng environment for stream computing[J]. Computer Science and Aritificial intelligence Laboratory,vol 6, no10,2005:198-201..
- [80](美)怀特(White, T.)著; 周敏奇, 王晓玲, 金澈清, 钱卫宁译. Hadoop 权威指南(第 2 版)[M]. 北京: 清华大学出版社, 2011.7。
- [81]HDFS Architecture Guide[DB/OL].
http://hadoop.apache.org/docs/stable1/hdfs_design.html.
- [82]Jeffrey Dean, Sanjay Ghemawat. MapReduce: Simplified Data Processing on Large Clusters [J]. Communications of the ACM, Volume 51 Issue 1, January 2008: 107-113.
- [83]杨学军, 曾丽芳, 邓宇. Imagine 流处理器上流的优化组织方法[J]. 计算机学报, 7(31), 2008: 1092-1100.

- [84] S4: Distributed Stream Computing Platform [DB/OL].(2012-12-28).
- [85] Puma [DB/OL].(2012-12-28). <http://www.facebook.com/Puma>.
- [86] Twitter Storm [DB/OL].(2012-12-28). <https://github.com/nathanmarz/storm>.
- [87] Jonathan Leibiusky, Gabriel Eisbruch, Dario Simonassi. Getting Started with Storm[M]. O'Reilly Media, 2012: 72-75.
- [88] S Owen, R Anil, T Dunning, E Friedman. Mahout in action [M]. Manning Publications Co. , 2011
- [89] k-Means clustering basics. [DB/OL].
<http://mahout.apache.org/users/clustering/k-means-clustering.html>

博士后在站期间取得的研究成果

论文

1. Tian Gao, Junwei Cao, Yanxiang Xu, Huaying Zhang, Peng Yu, and Senjing Yao. From Power Quality to Power Experience. Proc. 4th Int. Conf. on Networking and Distributed Computing, Hongkong, China, 2013. (EI 检索)
2. 张华赢, 高田, 姚森敬, 史帅彬, 许延祥, 黄志伟, 余鹏, 曹军威. 电能体验质量研究. 南方电网技术, 2014.
3. Huang Zhiwei, Gao Tian, Zhang Huaying, Han Xu, Cao Junwei, Hu Ziheng, Yao Senjing, Zhu Zhengguo. TRANSIENT POWER QUALITY ASSESSMENT BASED ON BIG DATA ANALYSIS. The 6th China International Conference on Electricity Distribution. 2014.(已录用)

标准和专利

1. 干扰源与敏感用户分类供电指引, 中国南方电网行业标准, 2014.
2. 一种电压暂降扰动下敏感客户电能体验质量的预警方法, 发明专利, 专利号: 201310419436.X, 2013.
3. 基于分布式贝叶斯分类的电能质量综合评估, 发明专利, 受理中, 2014.

致谢

时间荏苒，珍贵而短暂的博士后工作马上就要结束。在站期间，各位老师和同学给予了我很多无私的帮助和鼓励，清华园里那种务实、严谨、创新的氛围让我受益匪浅。如今，带着沉甸甸的收获即将离开，有一种发自内心的感激，同时也非常自豪曾在这里生活学习过。

首先，衷心感谢我的导师曹军威教授。曹老师睿智幽默，勤勉谦和，极富人格魅力，在科研、工作方面的言传身教使我成长很多、收获很多。

感谢各位同事和好友，是你们陪我一起度过了博士后工作的美好时光，很幸运曾经和你们共同奋斗、分享欢乐，祝福大家。

深深地感谢我的父母和妻子，你们始终坚定地站在了我的背后，你们的理解和支持是我前进的最大动力。

高田

2014年9月