

基于大数据分析的暂态电能质量综合评估方法

张华赢¹, 朱正国¹, 姚森敬¹, 高田², 曹军威², 韩蓄², 王淼²

(1. 深圳供电局有限公司, 广东 深圳 518020; 2. 清华大学信息技术研究院, 北京 100084)

摘要: 运用基于大数据处理架构的 Naive Bayes 分类方法提出了暂态电能质量评估方法, 将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, 并将评估结果按严重程度分为暂态正常状态、短时电压暂降状态、短时深度电压暂降状态、短时电压失压状态。基于 MapReduce 架构, 设计分布式 Naive Bayes 算法实现状态分类。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。试验结果表明, 所提出的基于大数据分析的暂态电能质量评估方法是可行, 在准确率和处理速度上都取得了较好的效果。

关键词: 大数据; MapReduce; 分布式数据挖掘; 朴素贝叶斯(Naive Bayes)分类

Comprehensive Evaluation Method of Transient Power Quality Based on Big Data Analysis

ZHANG Huaying¹, ZHU Zhengguo¹, YAO Senjing¹, GAO Tian², CAO Junwei², HAN Xu², WANG Miao²

(1. Shenzhen Power Supply Co., Ltd., Shenzhen, Guangdong 518020, China;

2. Research Institute of Information Technology, Tsinghua University, Beijing 100084, China)

Abstract: A transient power quality assessment method is proposed based on Naive Bayes classification in the architecture of big data processing. The data sources are extended to power grid monitoring data, power customer data and public data, and the assessment severities are classified into normal state, abnormal state, critical state, and failed state according to the results of Naive Bayes classification. A Naive Bayes classification method based on MapReduce to realize power quality assessment is designed. In the classifier training phase, massive historical data are used as the distributed learning object, and assessment rules are generated periodically. In the state assessment phase, each assessment node updates the assessment rules generated by the training phase, generates real-time evaluation of the samples from the stream processing framework, and evaluates the power quality state according to the current rule. Experiment results show that the transient power quality evaluation method based on big data analysis presented in this paper is feasible, and achieve good results both in classification accuracy and processing speed.

Key words: big data; MapReduce; distributed data mining; Naive Bayes classification

0 引言

电能质量扰动主要指供用电设备正常运行时频

率、电压、电流等指标偏离额定值的程度。长期以来, 人们对电能质量稳态问题比较关注, 主要集中于对电压偏差、频率偏差和三相不平衡等问题的研究和治理^[1-3]。但随着敏感设备的大量使用, 暂态扰动已经成为用户电能体验问题的主要原因, 尤其是电压暂降和短时间中断严重威胁敏感用户的用电可靠性^[4-6]。因此, 暂态电能质量综合评估的目的在于得出对电能质量状况的一个准确描述, 使供用电双方都能准确地评估系统中暂态电能质量问题的严重程度, 从而为电能质量责任确认和电能质量扰

基金项目: 国家重点基础研究发展计划(973 计划)(2013CB228206); 国家自然科学基金(61233016); 中国南方电网公司 2012 年科技项目(K-SZ2012-026)。

Foundation item: Supported by the National Basic Research Program of China (973 Program) (2013CB228206); National Natural Science Foundation of China (61233016); Science & Technology Projects of CSG (K-SZ2012-026).

动治理提供重要依据。

当前信息通信技术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管理的深度融合,其最终表现形式必将是电力数据的爆发性增长。在电力大数据背景下,贯穿输电、配电、管理等各个环节的数据都可用于暂态电能质量综合评估。这些数据从类型上可以分为海量历史数据和实时流数据,具有多源、海量、异构、时空连续等特点,共同构成了评估所需要的大数据^[7-8]。基于大数据的分布式计算能力和流计算能力,可以提高暂态电能质量综合评估中数据分析的广度和深度,从而得到更加实时和可靠的评估结果。

本文研究了目前电网中暂态电能质量监测和处理的大数据处理架构,并提出了一种基于分布式朴素贝叶斯(Naive Bayes)分类方法的暂态电能质量评估方法,将结果按扰动的严重程度分为暂态正常状态、短时电压暂降状态、短时深度电压暂降状态、短时电压失压状态。基于大数据框架的处理方法,不仅可以克服原始数据分散、模型训练速度慢的缺点,而且可以有效地提高状态评估的效率和准确性。

1 暂态电能质量评估研究现状

目前,暂态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^[9]。国内外对于暂态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^[10]。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。

分类算法的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。本文的暂态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。

本文将直接或间接影响暂态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暂态正常状态、短时电压暂降状态、短时深度电压暂降状态、短时电压失压状态四个等级,并分别定义为:

1)暂态电压正常状态(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)短时电压暂降状态(voltage dip condition):单个或多个电网运行特征项发生较小的暂态变化,暂态(电压暂降)阈值为固定参考电压的85%~90%,可能会对敏感电力用户产生用电风险。

3)短时深度电压暂降状态(deep voltage dip condition):单个或多个电网运行特征项发生幅值较深的暂态变化,暂态(电压暂降)阈值为固定参考电压的70%~85%,会对所有敏感电力用户和部分非敏感电力用户产生用电风险。

4)短时电压失压状态(less voltage condition):单个或多个电网运行特征项发生极大的暂态变化,严重超过了标准限值,暂态(电压暂降)阈值为固定参考电压的35%~70%,会对全部电力用户产生用电风险。

2 基于大数据的暂态电能质量综合评估框架

暂态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,本文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。Hadoop是目前使用最广的大数据处理框架,具有良好的数据存储和数据处理能力,并能应对批处理和实时处理的多种业务需求。MapReduce是Hadoop架构的核心组件之一,它把数据处理过程分为"Map(映射)"和"Reduce(规约)"两个阶段,一个MapReduce作业(job)通常会把输入的数据集切分为若干独立的数据块,由map任务(task)以完全并行的方式处理它们。框架会对map的输出先进行排序,然后把结果输入给reduce任务,reduce任务

接受 map 阶段的输出数据, 将这组数据进行合并产生一组规模更小的值。其中 Job Tracker 负责作业的管理和调度, Task Tracker 负责 map 作业和 reduce 作业的执行^[11-12]。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, 基于朴素贝叶斯(Naive Bayes)分类算法实现暂态电能质量状态的分类(暂态正常状态、短时电压暂降状态、短时深度电压暂降状态、短时电压失压状态等类别), 从而进行状态评估和故障预警。Naive Bayes 分类的数学理论基础是贝叶斯定律, 即根据输入样本的先验概率求样本类别后验概率, 从而获得特征空间和输出值之间的联合概率分布, 从而获取数据样本的分类结果。本文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 1 所示。

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 MapReduce 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

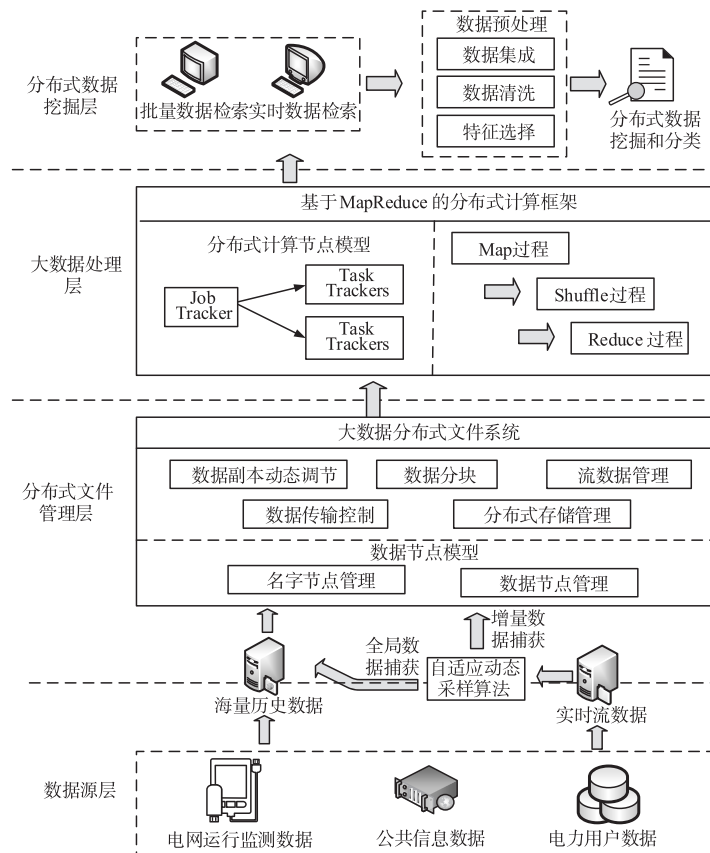


图 1 面向暂态电能质量综合评估的大数据框架

Fig. 1 Big data structure for comprehensive evaluation of transient power quality

数据挖掘算法,并实现暂态电能质量状态分类。

3 基于分布式 Naive Bayes 分类的评估算法

本文基于 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 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。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,传统数据挖掘方法且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 MapReduce 提供的可伸缩性和并行性优势就变得明显。

3.1 分布式 Naive Bayes 分类建模

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。本文在 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 $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$,其中 x_1, \dots, x_n 为数据的特征,如电压暂降的深度、持续时间、用户重要度、敏感度等,样本为监测点采集的在某次电压暂降发生时的数据。同时需要将样本划分为暂态正常状态、短时电压暂降状态、短时深度电压暂降状态、短时电压失压状态四个类别,化为 C_1, C_2, C_3, C_4 ,根据如式(1)所示的贝叶斯定理:

$$P(C_i | \mathbf{X}) = \frac{P(\mathbf{X} | C_i)P(C_i)}{P(\mathbf{X})} \quad (1)$$

贝叶斯分类问题可以转化为求得 $P(C_i | \mathbf{X})$ 最大的类标号 C_i 的过程,如式(2)所示。

$$f(x) = \max P(C_i | \mathbf{X}), i = 1, 2, 3, 4 \quad (2)$$

其中, $P(C_i | \mathbf{X})$ 为 \mathbf{X} 发生时的条件概率。 $P(C_i)$ 为先验概率可由训练样本直接得出,如式(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。

$$P(C_i) = \frac{S_{C_i}}{S} \quad (3)$$

在 Naive Bayes 分类中,条件概率 $P(\mathbf{X} | C_i)$ 可通过公式(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。

$$P(\mathbf{X} | C_i) = \prod_{j=1}^n P(X_j | C_i) \quad (4)$$

式(3)中,如果样本 \mathbf{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(如用户投诉次数、用户负荷容量等),则使用如式(5)所示的归一化方法进行处理,然后将连续型属性转换为离散型属性。

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

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。本文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2所示。

3.2 分布式 Naive Bayes 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备天然的并行性。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 的计算涉及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;

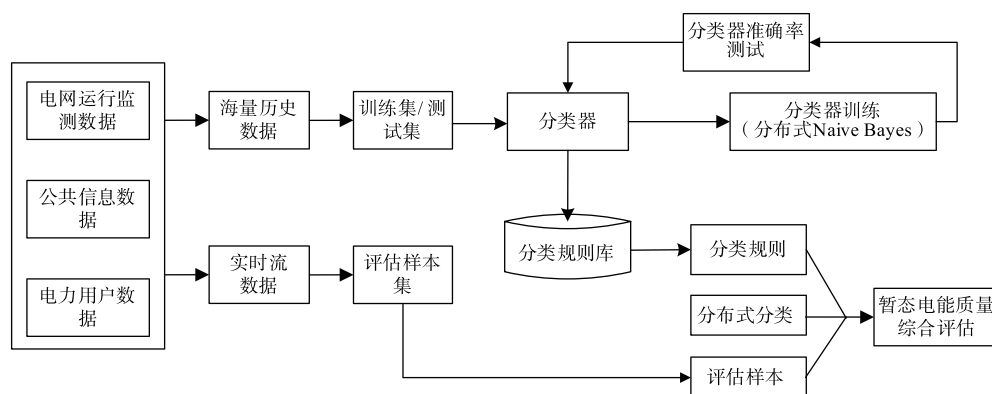


图2 暂态电能质量评估流程图

Fig. 2 Flowchart of transient power quality evaluation

(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) 任务分解阶段。为了分布式地实现贝叶斯分类的计算, 需在各分布式节点上计算 $P(C_i)$ 和 $P(X|C_i)$ 。因此, 在此阶段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 $\langle \text{Key}, \text{Value} \rangle$ 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: 分布式地计算先验概率 $P(C_i)$ 。即基于 MapReduce 方法在分布式节点先计算各自节点的 $P(C_i)$, 然后在集中式节点中计算全部数据集的 $P(C_i)$ 。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 $P(C_i)$ 。

4) Job 2: 计算条件概率 $P(X|C_i)$ 。即基于

MapReduce 方法在分布式节点先计算各自节点的 $P(X|C_i)$, 然后在集中式节点中计算全部数据集的 $P(X|C_i)$ 。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 $(C_i, x \downarrow f)$ 。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 3 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 $\max(P(C_i|X))$, 确定样本属于的类别。

4 仿真试验和结果分析

本文针对中国某电网内部 2010 年、2011 年、2012 年内 3 条 10 kV 电力用户专线进行暂态电能质量评估。实现的过程包括数据采集、数据预处理、特征项选择、算法执行、试验结果评估。

数据采集阶段: 从该电网生产管理系统中采集 2010、2011 和 2012 三年内的系统故障时间、故障原因; 从电能质量监测系统中获取电能质量监测数据, 包括: 电压暂降、电压暂升、短时中断等; 获取此 3 条专线中用户的敏感度、重要度、负荷容量等数据; 最后获取当前的公共数据, 包括温度、湿度、天气等。

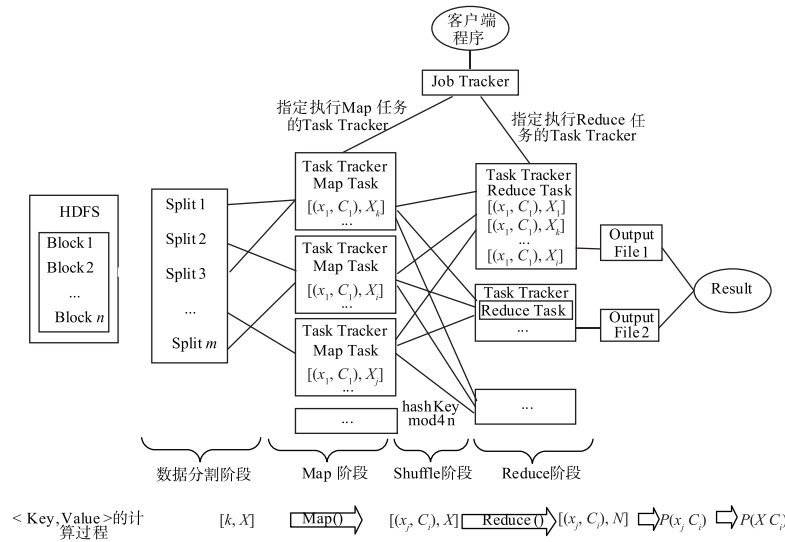


图3 Job2 的 MapReduce 计算过程

Fig. 3 Programming model of MapReduce for Job2

数据预处理阶段, 主要进行数据清理、数据集和数据格式变化。数据清理具体包括去除噪声数据、识别或删除离群点等工作; 数据集成将多个数据源的数据结合起来并统一存储; 数据格式变化主要是指数据的归一化、平滑化, 或根据其概率分布进行转换, 以用于数据挖掘。

特征项选择阶段, 通过多次协方差计算和分类试验, 选取对于评估结果影响较大的数据特征, 形成如表 1 所示的特征项, 并根据这些特征项采集某时刻 t_i 的训练样本和测试样本。具体的特征项包括: 这 3 条 10 kV 线路上电压暂降的深度、持续时间、电流变化范围; 相关用户的敏感度级别、重要度级别、负荷容量级别; 当时的湿度、温度和天气等公共信息等。仿真试验中, 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, 并将采集的结果实时转换到如表 1 所示的特征项空间内, 对结果进行分布式分类处理。

算法执行阶段: 根据算法模型和数据集情况,

本文使用 5 台主机搭建分布式计算环境。集群中包括 5 个节点, 1 个 Master, 4 个 Slave, 节点之间通过局域网连接。5 个节点上分别安装 CentOS6.4 操作系统, 并部署了 Hadoop 分布式计算环境。Master 主机配置 NameNode 和 JobTracker 的角色, 负责总管分布式数据和分解任务的执行, 4 个 Slave 机器配置 DataNode 和 TaskTracker 的角色, 负责分布式数据存储以及算法的执行。试验中, 分别按照流程分布式地运行了 Job1 和 Job2, 并在每个 Slave 机器上计算 $\max(P(C_i | X))$, 以获得状态评估额类别。

算法结果评估阶段: 算法评估的指标主要根据分类准确率, 其定义如下:

分类准确率 =

$$\frac{\text{正确分类的样本数目}}{\text{正确分类的样本数目} + \text{错误分类的样本数目}}$$

评估采用如下方案: 分别将 2010、2011 和 2012 年的数据作为训练集, 将 2013 年的数据作为

表 1 训练/评估数据集结构

Tab. 1 Structure of training/evaluation sets

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

测试集, 计算分类的准确率结果如图4所示。

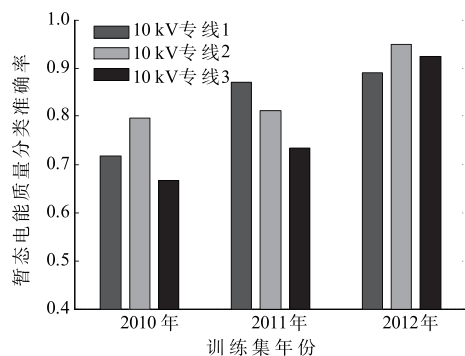


图4 暂态电能质量评估结果准确率对比

Fig. 4 Accuracy comparison of evaluation results of transient power quality

根据如图4所示的试验结果, 在使用2012年的数据作为分类训练集时, 分类算法可以取得较好的准确性, 证明基于分布式朴素贝叶斯算法的方法可以较准确地评估暂态电能质量状态。同时, 训练集的产生时间对于分类准确率的影响较大, 训练集和测试集的产生时间越接近, 其分类准确率越高, 从而表明周期性更新训练集并产生分类器是必要的, 能够提高状态评估的准确率。其原因是因为电网和用户的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。

5 结语

随着电力行业信息化的快速发展, 数据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。本文采用基于大数据处理架构的贝叶斯分类算法, 综合电网运行监测数据、电力用户数据、公共数据等异构的海量数据, 实现了对暂态电能质量的状态评估, 将电能质量分类至暂态正常状态、短时电压暂降状态、短时深度电压暂降状态、短时电压失压状态。试验表明, 基于Hadoop处理架构中的MapReduce进行数据处理, 实现分布式的Naive Bayes分类算法, 能够取得较好的分类准确率, 可以有效地进行状态评估。

参考文献

[1] 林广明, 史训涛, 曹小冬. 逐层序关系组合赋权法在电能质量综合评估的应用[J]. 南方电网技术, 2013, 7(2): 110 - 114.

LIN Guangming, SHI Xuntao, CAO Xiaodong. Application of the stepwise order relation based combination weighing method in comprehensive evaluation of power quality [J]. Southern Power System Technology, 2013, 7(2): 110 - 114.

[2] GNACINSKI P, MINDYKOWSKI J, TARASIUK T. A new concept of the power quality temperature factor and its experimental verification [J]. IEEE Transactions on Instrumentation and Measurement, 2008, 8(57): 1651 - 1660.

[3] DUGAN R, MCGRANNAGHAN M, SANTOSO S, et al. Electrical power systems quality [M]. New York: McGraw-Hill, 2002.

[4] BANERJEE S, AGARWAL N. Analyzing collective behavior from blogs using swarm intelligence [J]. Knowledge and Information Systems, 2012, 12(33): 523 - 547.

[5] AHMED R, KARYPIS G. Algorithms for mining the evolution of conserved relational states in dynamic networks [J]. Knowledge and Information Systems, 2012, 12(33): 603 - 630.

[6] WU Xindong, ZHU Xingquan, WU Gongqing, et al. Data mining with big data [J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 1(26): 97 - 107.

[7] CHANG E Y, BAI H, ZHU K. Parallel algorithms for mining large-scale rich-media data [C]. Proceedings of 17th ACM International Conference on Multimedia, 2009: 917 - 918.

[8] 宫宇, 吕金壮. 大数据挖掘分析在电力设备状态评估中的应用[J]. 南方电网技术, 2014, 8(6): 74 - 77.

GONG Yu, LÜ Jinzhuang. Application of big data mining analysis in power equipment state assessment [J]. Southern Power System Technology, 2014, 8(6): 74 - 77.

[9] BOLLEN M H J. Understanding power quality problems: voltage sags and interruptions [M]. New Jersey: John Wiley & Sons, Inc, 2000: 1 - 34.

[10] 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, 2011, 26(4): 1161 - 1171.

[11] ZHANG S B, HAN J Z, LIU Z Y et al. Spatial queries evaluation with MapReduce [C]. Proceedings of the 2009 Eighth International Conference on Grid and Cooperative Computing, 2009.

[12] EKANAYAKE J, LI H, ZHANG B, et al. A runtime for iterative MapReduce [C]. Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing. 2010.

收稿日期: 2014-11-05

作者简介:

张华赢(1981), 男, 工程师, 硕士, 从事高电压技术、电能质量综合治理技术研究, zhyszpower@163.com;

朱正国(1977), 男, 高级工程师, 从事配网技术监督、电能质量监测与治理技术研究;

姚森敬(1969), 男, 高级工程师(教授级), 从事高电压试验研究、电网运行管理、电能质量综合管理等工作。