

基于 MapReduce 的电力设备并行故障诊断方法

王德文, 刘晓建

(华北电力大学 控制与计算机工程学院, 河北 保定 071003)

摘要: 以智能电网中电力变压器故障诊断为例, 给出了基于 MapReduce 的电力变压器并行故障诊断过程, 其应用 4 个 MapReduce 过程执行故障诊断算法的训练阶段, 并得出分类模型, 应用 1 个 MapReduce 过程完成对电力设备状态信息数据的故障诊断。建立了电力设备状态信息并行故障诊断实验平台, 基于海量变压器油中溶解气体分析数据进行并行故障诊断实验, 实验结果表明并行故障诊断速度高于传统单机环境下的诊断速度, 满足智能电网环境下对海量电力设备状态信息快速故障诊断的要求。

关键词: 智能电网; 电力设备; 状态信息; 并行处理; 故障诊断; MapReduce; 电力变压器

中图分类号: TM 407; TM 507

文献标识码: A

DOI: 10.3969/j.issn.1006-6047.2014.10.019

0 引言

电力设备是电力系统的基本组成部分, 电力设备故障将直接影响智能电网的安全运行。数据挖掘是实现电力设备故障诊断的关键技术之一, 是分析电力设备运行状态的主要方法。为了保证电力设备的安全运行, 各种故障诊断方法得到应用。文献[1]根据模糊粗糙集算法建立信息决策系统, 解决变压器故障诊断问题。文献[2]根据神经网络算法建立了一个针对断路器的快速智能诊断系统, 利用断路器状态特征量进行故障类型的识别。这些方法都是在单机环境下进行的, 在现有的数据级别范围内能达到较快的故障诊断速度, 但没有考虑海量数据对诊断速度的影响。

智能电网中电力设备状态信息数据具有分布广、种类多、数据量巨大的特征。随着智能电网的发展, 电力设备故障诊断速度的提高将会成为智能电网建设的瓶颈。

a. 数据分布广。我国现有电网分布在不同地域, 电力设备分布点多面广, 造成电力设备状态信息数据分布广泛。

b. 数据种类多。智能电网中电力设备状态信息数据不但包含电力设备在线状态数据, 而且包含电力设备运行情况记录数据、电力设备故障诊断数据、电力设备检修试验数据、电力设备缺陷信息数据等^[3]。

c. 数据量巨大。例如, 输电线路中绝缘子泄漏电流数据采样频率高, 假设 10 ms 采集一次数据, 一个杆塔在一个月内的数据量就达到了 2.5 亿条^[4]; 智能电网中断路器数量较多, 断路器振动信号数据采集

频率高, 产生的数据量巨大^[5]。

针对智能电网中电力设备状态信息数据的特点, 需要引入新的快速计算技术以解决传统故障诊断方法计算能力不足的问题。电力设备并行故障诊断被提出, 且在智能电网中具有重要的研究价值。

MapReduce 是一个针对大数据并行运算的软件框架, 最大的优势在于隐藏繁琐的底层细节, 简化并行程序的设计与开发工作^[6-8]。文献[9]初步设计了电力系统仿真计算云, 采用并行计算技术对电力系统中海量电网数据进行高效仿真计算。文献[10]设计了基于 MapReduce 的 MPApriori 算法, 在 Hadoop 平台上实现了对海量 WAMS 电网数据的并行数据挖掘, 且并行数据挖掘效率高于传统数据挖掘平台。总之, 在智能电网中借助 MapReduce 的高性能的计算能力, 能够使并行故障诊断算法高效地执行。

本文提出采用 MapReduce 技术对智能电网中海量电力设备进行并行故障诊断的方法。以对智能电网中海量变压器油中溶解气体分析(DGA)数据用朴素贝叶斯算法进行故障诊断为例, 给出了基于 MapReduce 的算法并行化过程。最后, 通过 Hadoop 平台对变压器进行并行故障诊断实验, 实现了对智能电网中海量电力设备状态信息数据的快速故障诊断。

1 基于 MapReduce 的电力设备故障诊断并行化方法

1.1 朴素贝叶斯算法在电力设备故障诊断的应用

基于 MapReduce 框架的并行故障诊断算法主要为回归、聚类与分类等算法。目标是对智能电网中变压器、断路器、输电线路等电力设备产生的海量状态信息数据进行并行故障诊断, 解决传统单机环境下对海量状态数据计算能力不足的问题。

本文中智能电网电力设备并行故障诊断以产生的海量变压器 DGA 数据为例, 也可以对智能电网中

收稿日期: 2013-09-13; 修回日期: 2014-07-10

基金项目: 国家自然科学基金资助项目(61074078); 中央高校基本科研业务费专项资金资助项目(12MS113)

Project supported by the National Natural Science Foundation of China(61074078) and the Fundamental Research Funds for the Central Universities(12MS113)

断路器状态数据、输电线路波形数据等海量电力设备状态数据进行故障诊断。变压器中应用的故障诊断算法很多,如贝叶斯算法、粗糙集理论算法、支持向量机算法等,其中贝叶斯算法在变压器故障诊断中应用比较普遍^[11-15]。

对变压器海量 DGA 数据的故障诊断以并行朴素贝叶斯分类算法为例。朴素贝叶斯算法具有方法简单、准确率高与速度快的优势,虽然原理比较简单,但是具有很好的实际应用效果。目前,使用朴素贝叶斯分类算法对变压器进行故障诊断的研究很多是单机情况下进行的,将来智能电网中需要用到并行化的朴素贝叶斯算法对海量状态数据进行快速故障诊断。

1.2 电力设备故障诊断并行化流程

Hadoop 是一个方便对大数据集进行计算的分布式平台,它在集群中通过并行的工作方式加快计算速度。Hadoop 集群由负责资源调度与管理文件目录结构的 NameNode 和若干负责管理数据存储的 DataNode 构成。MapReduce 是 Hadoop 的一个子项目,用于大数据集的并行计算,可以为电力设备的故障诊断领域提供高效的并行计算能力与简单、通用的并行算法设计环境。MapReduce 的执行流程如图 1 所示。

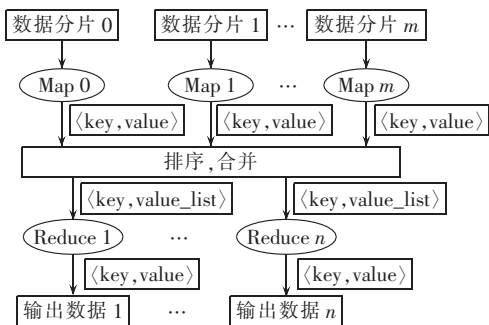


图 1 MapReduce 执行流程
Fig.1 Flowchart of MapReduce execution

MapReduce 运行模式可以自动处理数据的分割、负载均衡、容错处理等问题。MapReduce 程序设计过程中只需将数据并行计算问题分解成许多可并行执行的子问题,并设计对应的 Map 函数与 Reduce 函数,就能将程序运行在分布式系统上。

MapReduce 模型中输入数据文件分割成许多相同大小的数据片段,并保存在 Hadoop 集群中不同的存储节点。数据片段通过执行设计的 Map 函数,得出运算结果<key,value>,通过排序、合并,将 key 值一样的运算结果传送到相同的 Reduce 函数进行数据归并。最后根据 Reduce 函数得到结果,并将最终计算结果数据保存在分布式集群上。MapReduce 模型使许多种计算在此框架下达到相当高的执行效

率,并且还具有可扩展的特点,对大规模数据的计算优势更加明显。

1.3 电力设备故障诊断算法并行化过程

以 MapReduce 为基本的模型,对朴素贝叶斯算法进行并行化操作。以智能电网中海量变压器 DGA 数据为例对变压器进行故障诊断。首先对变压器 DGA 数据进行预处理使其转化成 Bayes M/R job 读入数据要求的格式,即每一行为 1 个变压器 DGA 数据样本,第 1 个字符是变压器故障类别,剩余的是离散化后的 DGA 数据,DGA 属性数据之间用空格分开。

利用 4 个 MapReduce 作业类执行算法的训练过程,目标是得出 4 个值的输出结果:计算每个故障类中 DGA 数据属性特征词出现的频率值,计算变压器故障类中的词频反文档频率(TFIDF)值,对故障类中的 TFIDF 值进行求和,计算变压器各故障类中属性特征词出现的概率。其中前一个作业类的运算结果数据作为后一个作业类的输入数据。训练过程结束后得出贝叶斯分类模型并保存在分布式集群中,并行化的朴素贝叶斯算法处理 DGA 数据训练集的过程如下。

a. 第 1 个 MapReduce 过程。Map 函数接收来自训练文档的 DGA 样本数据块,读取每个 DGA 数据中的属性特征词,计算在每个属性特征词在所在故障类中出现的词频(TF)值。Reduce 函数统计 Map 的数据结果,统计出训练数据集中每个变压器故障类的样本数据总数、TF 值与故障数据中出现属性特征词的样本数据总数。

b. 第 2 个 MapReduce 过程。根据第 1 个 MapReduce 的输出结果计算每个变压器故障类中每个特征属性的 TFIDF 值。Map 过程计算每个特征属性的反文档频率(IDF)值。Reduce 过程根据 TF 值乘以 IDF 值计算每个变压器故障类中特征属性的 TFIDF 结果,同时计算训练文档特征属性总数。

c. 第 3 个 MapReduce 过程。Map 函数调用每个故障类的样本数据总数与训练集 DGA 数据总数计算各故障类的先验概率。Reduce 函数进行汇总,计算每个变压器故障类中所有特征属性的 TFIDF 总和,计算变压器故障类的先验概率。

d. 第 4 个 MapReduce 过程。通过计算得到变压器各故障类中特征属性出现的概率,最终得出贝叶斯分类模型。

并行化的朴素贝叶斯算法对 DGA 数据测试集的故障诊断过程中,先对变压器 DGA 测试集进行预处理,将其组织成 Bayes 模型输入文本格式,然后用 1 个 MapReduce 过程完成对测试集的故障诊断。Map 函数计算待测试的变压器 DGA 数据中属性特征词

出现的次数,并加载训练过程得到的贝叶斯分类模型计算出测试集所属的故障类型。Reduce 函数进行归并,最终获得诊断结果。朴素贝叶斯算法并行化流程图如图 2 所示。

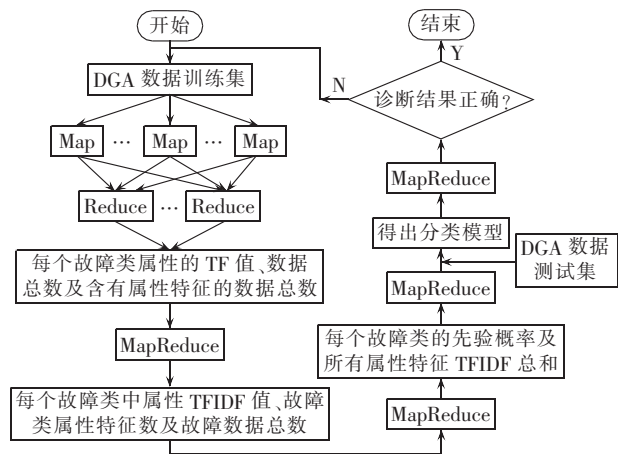


图 2 贝叶斯算法并行化流程图

Fig.2 Flowchart of Bayesian algorithm parallelization

2 实验过程与测试结果分析

2.1 实验环境

本文在实验室构建了一个电力设备状态信息并行故障诊断实验平台。实验环境由 10 台配置相同的普通 PC 机组成,其中一台 PC 机作为 NameNode,其他 PC 机均作为 DataNode。每台 PC 机 CPU 均为双核 Interi5-2400,主频 3.10 GHz,4.00 GB 内存,80 GB 硬盘。Hadoop 所用版本为 0.20.2。

实验中以额定电压在 220 kV 及以下的变压器作为研究对象。根据相关文献和电力公司中有关变压器的历史资料收集到具有明确故障的 237 条变压器故障数据与 369 条健康状态下运行的变压器数据^[16-18]。实验过程经过反复多次的测试,取平均值作为最终实验结果。

2.2 变压器故障类型及离散化标准

根据 IEC 标准和我国电力行业标准 DL/T722—2000^[19],变压器故障类型如表 1 所示,部分变压器故障数据如表 2 所示(φ 表示气体含量)。

朴素贝叶斯分类算法是一种符号化分析算法,算法中的属性值都看作定性数据,因此需要对变压器 DGA 数据属性值进行离散化处理。考虑到在实际变压器故障诊断中,不同气体含量值和气体间比值,其定量标准差异很大。所以,可行的离散化方法是依

表 1 变压器故障种类

Table 1 Categories of transformer fault

编号	故障类型	编号	故障类型
C ₁	低能放电	C ₄	高温过热
C ₂	高能放电	C ₅	局部放电
C ₃	中低温过热	C ₆	无故障

表 2 部分变压器故障样本

Table 2 Partial of transformer fault samples $\mu\text{L/L}$

故障编号	$\varphi(\text{H}_2)$	$\varphi(\text{CH}_4)$	$\varphi(\text{C}_2\text{H}_6)$	$\varphi(\text{C}_2\text{H}_4)$	$\varphi(\text{C}_2\text{H}_2)$
C ₁	170	7	24	17	54
C ₂	109	108.9	11.1	201	183
C ₃	120	120	33	84	0.55
C ₄	766	993	116	665	4
C ₅	650	53	34	20	0
C ₆	5.26	1.62	5.66	0	0

据一定的规则,对不同属性值设定阈值进行分段,每一段作为定性数据,从而保证变压器故障诊断顺利进行。根据 DL/T722—2000^[19]和相关文献^[20-21]以及专家经验,变压器 DGA 数据离散化如表 3 所示。

表 3 DGA 数据离散化

Table 3 Discretization of DGA data $\mu\text{L/L}$

属性变量	离散化标准		
	0	1	2
$\varphi(\text{H}_2)$	<140	≥ 140 且 <700	≥ 700
$\varphi(\text{CH}_4)$	<60	≥ 60 且 <400	≥ 400
$\varphi(\text{C}_2\text{H}_6)$	<100	≥ 100 且 <150	≥ 150
$\varphi(\text{C}_2\text{H}_4)$	<120	≥ 120 且 <200	≥ 200
$\varphi(\text{C}_2\text{H}_2)$	<5	≥ 5 且 <35	≥ 35
$\varphi(\text{总烃})$	<150	≥ 150	—
$\varphi(\text{C}_2\text{H}_2)/\varphi(\text{C}_2\text{H}_4)$	<0.1	≥ 0.1 且 <3	≥ 3
$\varphi(\text{CH}_4)/\varphi(\text{H}_2)$	≥ 0.1 且 <1	<0.1	≥ 1
$\varphi(\text{C}_2\text{H}_4)/\varphi(\text{C}_2\text{H}_6)$	<1	≥ 1 且 <3	≥ 3

2.3 朴素贝叶斯算法并行化的验证

本实验变压器 DGA 样本数据的训练集与测试集按 2:1 随机划分,在小规模 DGA 数据集的情况下,在集群上应用并行化的朴素贝叶斯分类算法对变压器进行故障诊断,实验结果如表 4 所示。

表 4 变压器故障诊断结果

Table 4 Result of transformer fault diagnosis

故障编号	训练样本数	测试样本数	正判数	正确率/%
C ₁	28	14	11	78
C ₂	24	12	9	75
C ₃	24	12	12	100
C ₄	34	17	15	88
C ₅	48	24	21	87
C ₆	236	123	120	97

如表 4 所示,依据文献^[21],本实验对存在故障的变压器的诊断准确率是正常的故障诊断率,在变压器数据样本充足的情况下诊断正确率会进一步提升。并行化的朴素贝叶斯算法对变压器 DGA 数据显示出预期的诊断效果,并对算法的诊断结果取得了初步验证。

2.4 Hadoop 集群与单机运行时间对比实验

本实验要在 Hadoop 实验平台上模拟对智能电网中海量变压器在线监测 DGA 数据集的故障诊断,必须增大测试集的数据规模。由于条件限制无法得到大规模测试数据,实验中采取随机复制已有数据

集的方法,来得到大规模的测试数据。在实验中,测试在相同大小 DGA 数据集下朴素贝叶斯算法在 Hadoop 平台与单机平台上的运行时间,实验结果如表 5 所示。

表 5 测试时间对比
Table 5 Comparison of test time

DGA 数据大小/MB	单机测试时间/s	集群测试时间/s
17	10	29
130	78	43
520	249	61
1040	582	83

由表 5 可见,当测试的变压器 DGA 数据集规模较小时,集群的运行时间大于单机运行时间,这是因为算法的执行过程中集群计算节点之间需要信息交互,消耗了一定的时间,而这部分时间占总运行时间的比例较大,导致集群的运行时间大于单机环境下;当测试数据规模逐渐增加时,集群的计算时间效率越来越高于单机环境下,体现了 Hadoop 集群对大数据集进行运算的优越性。

2.5 集群加速比性能分析

加速比为处理相同数据规模时处理器数量从 p 增加为 q 时集群的加速效果,是评判一个算法并行化效果的重要指标^[22]。本实验主要测试并行化的朴素贝叶斯算法处理不同大小 DGA 数据集时的加速比,结果如图 3 所示。

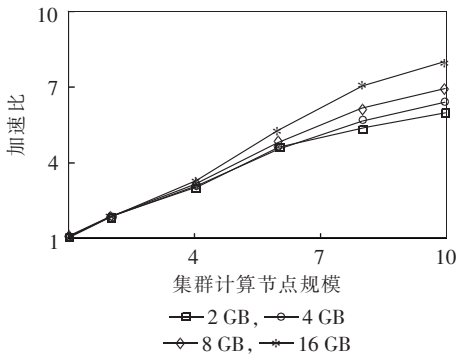


图 3 加速比测试结果
Fig.3 Result of speedup ratio test

从图 3 中可知,并行化的朴素贝叶斯算法的加速比是接近线性的,这是由于相同数据规模下集群规模越大,算法的计算时间越短。随着 DGA 数据集规模的增大,朴素贝叶斯算法加速比性能会越来越好,也说明并行计算方法能够很好地解决大规模数据情况下计算时间过长的问题。在相同的 DGA 数据规模下,随着集群计算节点规模的增加,加速比的增长速率逐渐减缓,这是由于集群规模的扩大导致计算节点之间通信量增加,减缓了加速比的增长速率。考虑到 16 GB 大小的 DGA 数据集相对于集群而言并不是很大,集群性能没有得到全面的体现,图中曲线并没

有靠近对角线,这再次说明集群更适合大规模数据集。

3 结论

本文针对传统电力设备故障诊断算法在对海量状态数据进行计算时出现的计算能力不足的问题,研究了 MapReduce 并行计算框架。利用 Hadoop/MapReduce 平台给出了朴素贝叶斯算法的并行故障诊断过程,实现了以变压器为例的电力设备快速并行故障诊断。实验表明,当变压器 DGA 数据集达到一定规模后,Hadoop 集群可以对变压器进行快速并行故障诊断,并显示出了较好的加速比性能。

目前,本文只是对电力设备的并行故障诊断做了实验,后期工作将逐步应用到电力设备的故障诊断中,并对现有的诊断方法进行改进,进一步提高故障诊断的速度与精度。

参考文献:

[1] 熊浩,李卫国,畅广辉,等. 模糊粗糙集理论在变压器故障诊断中的应用[J]. 中国电机工程学报,2008,28(7):141-147.
XIONG Hao,LI Weiguo,CHANG Guanghui,et al. Application of fuzzy rough set theory to power transformer faults diagnosis[J]. Proceedings of the CSEE,2008,28(7):141-147.

[2] 王小华,荣命哲,吴翊,等. 高压断路器故障诊断专家系统中快速诊断及新知识获取方法[J]. 中国电机工程学报,2007,27(3):95-99.
WANG Xiaohua,RONG Mingzhe,WU Yi,et al. Method of quick fault diagnosis and new knowledge obtainment for high voltage circuit breaker expert system[J]. Proceedings of the CSEE,2007,27(3):95-99.

[3] 王德文,宋亚奇,朱永利. 基于云计算的智能电网信息平台[J]. 电力系统自动化,2010,34(2):7-12.
WANG Dewen,SONG Yaqi,ZHU Yongli. Information platform of smart grid based on cloud computing[J]. Automation of Electric Power Systems,2010,34(2):7-12.

[4] 王灿林. 基于紫外光脉冲检测技术的绝缘子电晕特性研究[J]. 中国电机工程学报,2007,27(36):19-25.
WANG Canlin. Characteristics of insulator corona based on UV pulse detection[J]. Proceedings of the CSEE,2007,27(36):19-25.

[5] 杜凌艳,王振浩,王刚,等. 高压断路器运行状态实时监测系统设 计[J]. 电力自动化设备,2006,26(1):58-61.
DU Lingyan,WANG Zhenhao,WANG Gang,et al. Design of real-time operating state monitoring system for HV breaker[J]. Electric Power Automation Equipment,2006,26(1):58-61.

[6] 宋亚奇,周国亮,朱永利. 智能电网大数据处理技术现状与挑战 [J]. 电网技术,2013,37(4):927-935.
SONG Yaqi,ZHOU Guoliang,ZHU Yongli. Present status and challenges of big data processing in smart grid[J]. Power System Technology,2013,37(4):927-935.

[7] YANG H,DASDAN A,HSIAO R,et al. Map-reduce-merge:simplified relational data processing on large clusters[C]//Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data. Beijing,China:ACM,2007:1029-1040.

[8] GUNARATHNE T,WU T L,QIU J,et al. MapReduce in the

- clouds for science[C]//Cloud Computing Technology and Science, 2010 IEEE Second International Conference on Digital Object Identifier. Indianapolis, USA:IEEE, 2010:565-572.
- [9] 沐连顺,崔立忠,安宁. 电力系统云计算中心的研究与实践[J]. 电网技术, 2011, 35(6):170-175.
- MU Lianshun, CUI Lizhong, AN Ning. Research and practice of cloud computing center for power system[J]. Power System Technology, 2011, 35(6):170-175.
- [10] 曲朝阳,牛莉,张士林. 基于Hadoop的广域测量系统数据处理[J]. 电力系统自动化, 2013, 37(4):92-97.
- QU Chaoyang, NIU Li, ZHANG Shilin. Data processing of Hadoop-based wide area measurement system[J]. Automation of Electric Power Systems, 2013, 37(4):92-97.
- [11] 赵文清. 基于选择性贝叶斯分类器的变压器故障诊断[J]. 电力自动化设备, 2011, 31(2):44-47.
- ZHAO Wenqing. Transformer fault diagnosis based on selective Bayes classifier[J]. Electric Power Automation Equipment, 2011, 31(2):44-47.
- [12] 杨廷方,李景禄,曾祥君,等. 基于多方法组合诊断模型的大型变压器故障诊断[J]. 电力系统自动化, 2009, 33(20):92-108.
- YANG Tingfang, LI Jinglu, ZENG Xiangjun, et al. Large power transformer fault diagnosis based on a diagnosis model combined with several methods[J]. Automation of Electric Power Systems, 2009, 33(20):92-108.
- [13] 尹金良,朱永利,俞国勤. 相关向量机及其在变压器故障诊断中的应用[J]. 电力自动化设备, 2012, 32(8):130-134.
- YIN Jinliang, ZHU Yongli, YU Guoqin. Relevance vector machine and its application in transformer fault diagnosis [J]. Electric Power Automation Equipment, 2012, 32(8):130-134.
- [14] 刘卫华,廖瑞金,杨丽君. 基于点密度加权核模糊聚类的变压器故障诊断方法[J]. 电力自动化设备, 2012, 32(6):66-69.
- LIU Weihua, LIAO Ruijin, YANG Lijun. Power transformer fault diagnosis based on dot density weighted fuzzy kernel clustering[J]. Electric Power Automation Equipment, 2012, 32(6):66-69.
- [15] 司马莉萍,舒乃秋,李自品,等. 基于SVM和D-S证据理论的电力变压器内部故障部位识别[J]. 电力自动化设备, 2012, 32(11):72-76.
- SIMA Liping, SHU Naiqiu, LI Zipin, et al. Identification of interior fault position based on SVM and D-S evidence theory for electric transformer[J]. Electric Power Automation Equipment, 2012, 32(11):72-76.
- [16] 孙才新,陈伟根,李俭,等. 电气设备油中气体在线监测与故障诊断技术[M]. 北京:科学出版社, 2003:31-87.
- [17] 李俭,孙才新,陈伟根,等. 灰色聚类与模糊聚类集成诊断变压器内部故障的方法研究[J]. 中国电机工程学报, 2003, 23(2):112-115.
- LI Jian, SUN Caixin, CHEN Weigen, et al. A method of synthesis based on the grey cluster and fuzzy cluster about internal fault diagnosis of transformer[J]. Proceedings of the CSEE, 2003, 23(2):112-115.
- [18] 田质广,张慧芬. 基于遗传聚类算法的油中溶解气体分析电力变压器故障诊断[J]. 电力自动化设备, 2008, 28(2):15-18.
- TIAN Zhiguang, ZHANG Huifen. Dissolved gas analysis based on GA and FCM for power transformer fault diagnosis [J]. Electric Power Automation Equipment, 2008, 28(2):15-18.
- [19] 贾瑞君,范玉华,薛辰东,等. DL/T722—2000变压器油中溶解气体分析和判断导则[S]. 北京:中国电力出版社, 2001.
- [20] 耿兰芹,王芳,赵文清. SVM回归与朴素贝叶斯分类相结合的变压器故障诊断[J]. 华北电力大学学报, 2006, 33(6):28-32.
- GENG Lanqin, WANG Fang, ZHAO Wenqing. Transformer diagnosis by naive Bayesian classifier combining with SVM regression [J]. Journal of North China Electric Power University, 2006, 33(6):28-32.
- [21] 赵文清. 基于数据挖掘的变压器故障诊断和预测研究[D]. 保定:华北电力大学, 2009.
- ZHAO Wenqing. Study for transformer fault diagnosis and forecast based on data mining[D]. Baoding:North China Electric Power University, 2009.
- [22] LEUNG Y, LI D Y. Maximal consistent block technique for rule acquisition in incomplete information systems [J]. Information Science, 2003, 153:85-106.

作者简介:



王德文

王德文(1973—),男,黑龙江齐齐哈尔人,副教授,博士,主要研究方向为电力系统自动化和智能信息处理;

刘晓建(1987—),男,河北定州人,硕士研究生,主要研究方向为数据挖掘与电力系统故障诊断(E-mail:hblxiaojian@126.com)。

Parallel fault diagnosis based on MapReduce for electric power equipments

WANG Dewen, LIU Xiaojian

(School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China)

Abstract: As an example, a parallel fault diagnosis algorithm based on MapReduce is given for diagnosing the power transformer faults in smart grid, which includes four MapReduce procedures in its training stage to obtain the classification model and one MapReduce procedure to complete the fault diagnosis based on the status data of electric power equipment. An experimental platform is built and a parallel fault diagnosis experiment based on the massive DGA (Dissolved Gas Analysis) data of power transformer is carried out. Results show that, the diagnosis speed of parallel fault diagnosis is faster than that of mono-computer environment, meeting the requirement of rapid fault diagnosis based on the massive status data of electric power equipments in smart grid.

Key words: smart grid; electric equipments; status information; parallel processing; fault diagnosis; MapReduce; power transformers