

DOI:10.3969/j.issn.1674-1951.2019.11.010

“云 + 端”大数据驱动的预见性电厂设备维护

Predictive maintenance of power plant equipment driven
by “cloud + client” big data

王佳梓¹, 王桂玲²
WANG Jiazi¹, WANG Guiling²

(1. 中国华电集团有限公司山东分公司, 济南 250014; 2. 北方工业大学
大规模流数据集成与分析技术北京市重点实验室, 北京 100144)
(1. Shandong Company, China Huadian Corporation Limited, Jinan 250014, China;
2. Beijing Key Laboratory on Integration and Analysis of Large-scale Stream Data,
North China University of Technology, Beijing 100144, China)

摘 要:电厂设备可靠性的维护和保障对发电集团具有重要意义。提出预见性电厂设备维护系统架构,在前端实现设备的数据采集、本地实时处理和计算,在云端实现数据的汇集、存储、批处理及实时处理和分析。分析了该系统的基于 HBase 的云端数据库设计及写入、海量感知数据的并行化处理以及预见性维护等关键技术。该架构既节约了前端资源,又可以发挥电厂海量感知数据的潜在价值,通过基于海量感知数据的故障预测,实现基于运行实际状态的电厂设备的及时维护,便于决策和维护人员及时制定防范措施、防止和控制可能的故障出现,保障电厂运行可靠和稳定。

关键词:预见性维护;大数据;云-端融合;故障诊断;设备维护

中图分类号:TP 311 **文献标志码:**A **文章编号:**1674-1951(2019)11-0040-06

Abstract: Guarantee for the stability of power plant equipment is very important for power generation enterprises. An architecture of a predictive maintenance system for power plant equipment is proposed, which can collect, process and computing data on the client side, and aggregate, store, batch process and real-time process data on the cloud side. The key technologies of writing, parallel processing of massive sensor data, and predictive maintenance for the system based on HBase cloud database were analyzed. The architecture not only saved front-end resources, but also showed the potential value of the massive sensor data. Through predicting fault based on massive sensor data, the maintenance on the equipment could be done timely according to actual situations, which helped personnel in decision-making and making counter-measures in advance. Preventing and controlling the potential fault will ensure the reliability and stability of power plants.

Keywords: predictive maintenance; big data; cloud + client; fault diagnosis; maintenance of equipment

0 引言

电厂机组在运行过程中,设备会发生各种故障。这些故障既有可能是设备本身发生了故障,如齿轮断齿、轴承故障、风机结冰等,也有可能是传感器故障,如传感器量程漂移、元件损坏、取样管堵塞等。当前,电厂设备故障的常用检测方法,通常是对设备进行周期性的检查和维护,主要依靠人工经验和基于历史数据。这种方法存在着一定的弊端:一是维

护效果受限、成本较高。固定周期的巡检若周期过于频繁则不仅产生巨大的人力成本,还可能会产生一些不必要的,甚至有损害性的维护活动,若周期不够频繁,则又会带来较大的设备故障风险,带来更大的损失。二是设备故障预测准确率受限。当前设备故障的预警及判断,以及设备故障的时间预估和原因判断,往往受限于巡检人员的经验。这种人为估测出来的故障时间和原因往往与事实有较大出入。现实中,故障往往在已经发生一段时间后才会引起人们的注意。而且,基于历史经验的故障数据并不能完全反映出设备的全部故障,与一些隐藏的故障相关的数据没有记录。因此,用以进行设备故障预

测的数学分析模型也建立在未必准确及完整的数据基础之上,预测准确率受限于记录数据的准确性及完整性。

电厂设备的运行状态监控产生了大量的感知数据。例如,一个 2.5 MW 的风力涡轮机有超过 120 个传感器,每秒产生上万条数据记录;一个普通燃气轮机每天产生的感知数据数量超过 520 GB^[1]。这些感知数据表示了电厂设备的运行状态,具有极大的潜在分析价值。近年来,随着物联网、大数据分析技术的快速发展,大数据驱动的“预见性维护(predictive maintenance)”(又称预测性维护)成为在工业物联网中逐渐被广泛接受的概念^[2]。设备故障通常伴随着传感信号(传感设备实时传回的数据)的异常变化。与“预防性维护(preventive maintenance)”不同,大数据驱动的预见性维护是通过各种传感器对设备进行监控,基于设备运行过程中的监控数据及与设备维护相关的其他数据以及海量的历史数据,通过预测可能的失效模式来避免设备故障的活动。预见性维护要进行数据和信号的采集、分析和判断设备的劣化趋势、故障部位、原因并预测变化发展、提出防范措施,防止和控制可能的故障出现。与现有的设备维护方法相比,预见性维护的关键是根据机器实际状态而非根据固定的周期进行监控。基于预见性维护生成的预测模型,工厂就可以基于设备的实际运行状态智能地对维护活动进行计划,避免设备故障带来的巨大损失。预见性维护需要大规模的传感器历史数据。过去,由于存储技术的限制,一般工厂存储的传感器历史数据跨越数个星期甚至更少,源源不断产生的传感器数据很多被丢弃,也没有被充分利用。大数据技术则使得存储、管理和分析大规模的数据成为可能,过去无法存储的长期历史数据,如今可以被无限制地存储起来;过去利用普通的单个服务器在短时间内完成计算任务,在大数据技术的帮助下,也可以由普通开发人员完成^[3]。

本文对“云+端”大数据驱动的电厂设备预见性维护系统的架构、功能进行介绍,对其关键技术进行研究,并结合应用案例进行分析和探讨。

1 系统架构

目前,我国各省市已经逐步开展数字电厂的建设,可以预见,电厂数据的来源将越来越复杂,数据量也将日益增大。在大数据环境下,资源应该在云端和终端(或前端)上灵活、合理地分布^[4-9]。对于大规模的历史数据,只在前端处理和存储存在着无法大规模进行扩展,且不利于数据分析和挖掘的缺

陷;而且,多个传感设备的长期历史数据以及其他数据来源的集成,对预测模型准确率的提升具有较大意义,这是将数据在云端进行存储和处理的优势。

鉴于此,设计了如图 1 所示的“云+端”大数据驱动的预见性电厂设备维护系统架构。该系统中,与设备运行状况有关的数据在“云”(数据中心、基于服务器集群,又称为云端)和“端”(位于前端传感器所在物联网,又称为前端)都有相应的处理。

前端主要实现传感器、智能硬件以及设备的数据采集,并对数据进行必要的转换、过滤等预处理,之后实时上传到云端的数据中心。其中,物联网网关要实现传感网络与通信网络以及不同类型传感网络、异构设备之间的协议转换。出于实时性及节约资源的考虑,感知数据的实时处理与计算要求尽量在本地被满足,或者在部分网络节点上进行处理。感知数据库负责采集并存储本地的感知数据,一方面满足本地感知数据的存储、管理及实时计算需求,另一方面为云端大数据处理提供有价值密度的数据资源。

在云端,采用“关注点分离 SOC(separation of concerns)”^[10]的逐层抽象设计原则,每一层功能独立,以提高每层软件模块的可复用性,每层可独立于其他功能进行优化,并避免其他层次的潜在故障。

云基础设施层提供对中央处理器(CPU)、网络和存储等云基础设施资源的管理、部署和维护。云数据管理层提供对感知数据的接入、存储、管理、预处理、查询及计算。数据服务层提供设备的状态查询、预警模型生成、状态实时预警等各种定制化的分析服务。业务应用层面向业务人员提供可视化及用户交互功能,包括设备健康评估操控盘、事件通知、故障预警以及交互式分析等业务功能。其核心是云数据管理层及数据服务层,下面主要介绍这 2 个。

云数据管理层通过基于 Kafka 的分布式消息队列^[11]接收来自前端不同消息系统、数据源的感知数据。Kafka 是 Apache 下的一个子项目,是一个高性能分布式发布/订阅消息队列系统,它具有快速持久化、高吞吐、支持分布式、自动实现负载均衡等特点。云数据管理层的感知数据发布及订阅基于 Kafka 实现,利用 Kafka 来保证所有消息都会按照写入分区的顺序进行处理,支持感知数据持久化,支持消息的恢复、丢弃等管控,以避免数据丢失,并提供可分割、可重部署、高容错的能力。

感知数据从消息队列接收后,分布式并行计算框架及流数据处理引擎可对其进行数据的批处理计算以及实时处理。采用基于 Spark 及 Spark Streaming 的数据处理引擎。

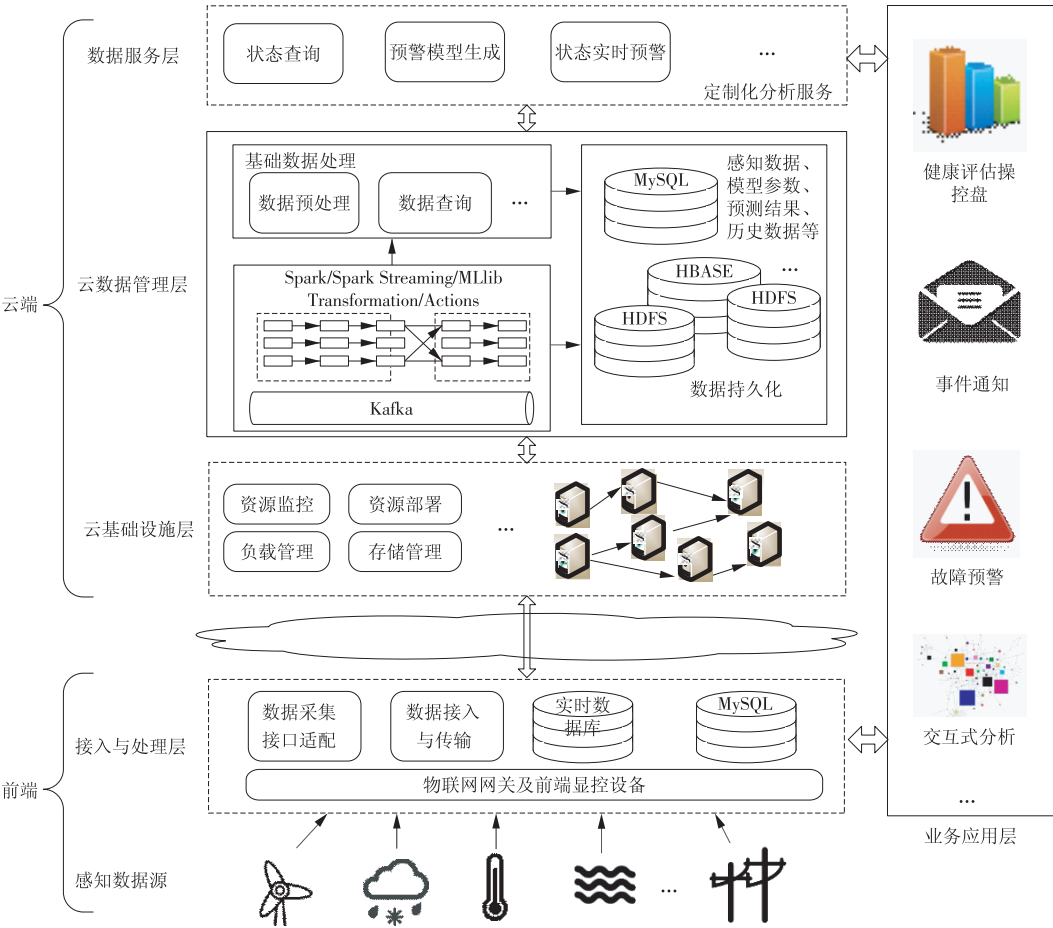


图 1 系统架构
Fig. 1 System architecture

Spark 是 UC Berkeley AMP lab 所开源的类 Hadoop MapReduce 的通用并行框架,拥有 Hadoop MapReduce 所具有的优点,不同于 MapReduce 的是,Job 中间输出结果可以保存在内存中,从而不再需要读写 HDFS,因此 Spark 除了能够提供交互式查询外,能更好地适用于数据挖掘与机器学习等需要迭代的 MapReduce 的算法^[12]。Spark Streaming 是构建在 Spark 上处理流数据的框架,它将流数据分成小的时间片断(几秒),以类似批处理的方式来处理这小部分数据,因此可以同时兼容批量和实时数据处理的逻辑和算法,方便了一些需要历史数据和实时数据联合分析的特定应用场合。它还具有很好的容错性、可扩展性,能运行在 100 + 的结点上,并达到秒级延迟。对于电厂设备维护来说,往往需要将实时感知数据与从历史数据挖掘出的模式进行对比,Spark Streaming 的采用方便了此类应用需求的满足。Spark MLlib 是基于 Spark 的开源机器学习库,提供了包括常见分类算法、回归算法、频繁项集挖掘及关联分析算法、聚类算法、协同过滤算法、降维算法等在内的丰富算法及相应的应用程序编程接口 API(application programming interface),从而方便

数据服务层基于 Spark MLlib 的 API 实现特定的数据分析服务。

海量的实时感知数据、历史数据、模型参数、预测结果以及与故障预测相关的各类业务数据根据其特点分别存储在关系数据库、HBase^[13] 及直接存储在 Hadoop 分布式文件系统 HDFS (Hadoop distributed file system) 之上。其中,原始遗留历史数据直接存储在 HDFS 之上较为方便;实时感知数据由 Kafka 接收后按照与预先定义的数据模型存入 HBase,方便后续进行复杂的查询;模型参数、预测结果以及各类业务数据由于其数据量少、具有较强的结构化特征,且有些本身就来自于关系数据库,存入关系数据库更加方便传统软件系统的访问。

虽然感知数据在前端已经经过部分预处理及价值密度提升,但要作为分析算法的输入,还需要经过一些必要的过滤、转换、关联等预处理。其次,上层的状态监控及实时预警等服务的实现也需要感知数据的查询接口。因此,云数据管理层还基于 Spark/Spark Streaming 以及数据库提供特定的感知数据预处理以及感知数据查询操作接口,例如基本查询操作、聚集查询操作以及连接查询操作等。

数据服务层基于云数据管理层提供的数据查询接口提供了各类电厂设备的单测点、多测点状态查询与分析服务,基于 Spark MLlib 的算法接口及历史数据训练集,可生成预警模型,状态实时预警服务在生成的预警模型之上,对实时到来的感知数据进行判断,可输出相应的预警通知。

2 业务与数据流程

大数据驱动的预见性电厂设备维护系统针对设备故障诊断与健康管理的流程如图 2 所示。在数据采集阶段,可供采集的数据包括设备全生命周期的数据,例如传感器信号、状态监测数据、维护历史记录等。在数据处理阶段,主要是对原始传感器信号进行信号处理,提高数据质量。不同于互联网数据,电厂设备维护相关数据分析对数据质量的要求较高。互联网大数据在进行预测与决策时,考虑的仅仅是两个属性间的关联是否具有统计显著性,其中噪声和个体间的差异在样本量足够大时可以被忽略,虽然预测结果的准确性会因此大打折扣,但不会造成太严重的后果,典型的例子如商品推荐结果准确性对用户的影响。但在电厂这类工业环境中,如果仅仅通过统计的显著特性给出分析结果,哪怕仅仅一次失误都可能造成严重的后果。因此,在这一阶段需要对数据质量进行预判和修复。数据预处理之后就可以用特征提取的方法进行处理来得到衰退性的特征,基于性能特征,就可以对设备的健康状况进行评估和量化。还可以预测特征在某一个未来时间段的值,从而可以预测性能的衰退趋势和设备的有效剩余寿命。最后,这些诊断和预测的结果和部分关键过程以可视化的方式展现出来,供普通业务人员理解、配置和使用。

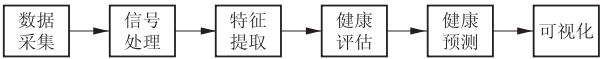


图 2 业务流程
Fig. 2 Business Process

系统的数据流程如图 3 所示。来自电厂设备传

感器的设备感知数据首先上传到物联网网关,进入实时数据库进行处理。前端设备传感器和物联网网关之间的通信协议可以采用 ZigBee 技术,这是一种当前被认为高性价比、低功耗的物联网技术,适合在前端设备应用。其次,在能耗允许的前提下,也可以针对特定的设备采用 Wifi 或有线连接的手段。云数据中心采集的感知数据则可能来自于不止一个(视实际应用需求可进行扩展)物联网网关,物联网网关和数据中心之间通过 Kafka 自身提供的基于传输控制协议 TCP(transmission control protocol)层的协议进行数据传输。感知数据除进入 Spark/Spark Streaming 的批处理及流处理引擎之外,也被持久化到数据库中后续使用。故障预测及状态查询的功能以 RESTful 服务或 Web API 的形式提供接口给业务应用,业务应用中的健康评估模块、事件通知模块、故障预警模块、交互式分析模块调用这些接口实现与业务人员的结果呈现与人机交互。

3 关键技术分析

3.1 基于 HBase 的云端数据库设计及写入技术

利用 HBase 存储和管理海量的电厂设备感知数据,首先考虑的问题是 RowKey(行键)的设计与选择。由于 HBase 的索引建立在行键之上,因此,行键的设计将影响到数据写入和查询的性能。为提高效率,需要尽量将查询的维度或者信息放在行键中。电厂设备“传感器 ID”和“系统时间”2 项组合起来可以唯一标志 1 条数据,即新标志列可以是 [传感器 ID/系统时间] 或 [系统时间/传感器 ID]。HBase 以行键的排序大小来分配集群存储节点,如果“系统时间”在“传感器 ID”的前面,则会导致同一时刻接收的数据被分配到同一节点存储的概率很大,这样容易使得部分节点压力过大,因此,可以选择 [传感器 ID/系统时间] 作为行键。这样设计具有明显的优点:首先时间和 ID 是查询常用的维度,将其加入行键,提高了大部分查询的效率;其次,虽然

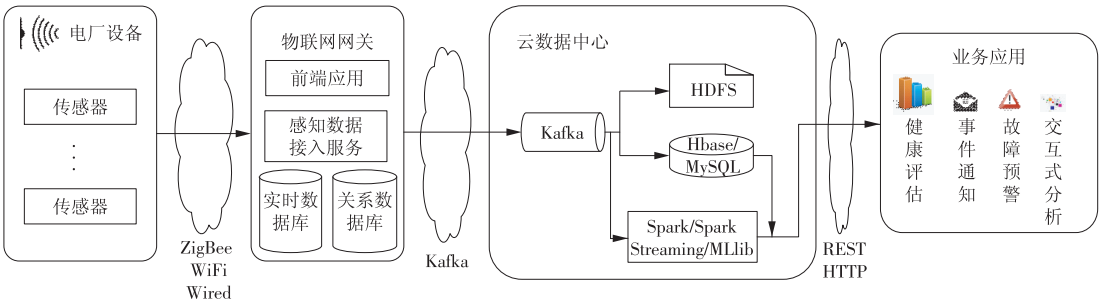


图 3 系统数据流程示意
Fig. 3 Data flow of the system

同一时间会写入大量的传感器数据,但是客户端会根据传感器数据的 ID 前缀,将数据分流到不同的 HBase 服务器上,不会出现写入热点,从而使得集群写入负载较为均衡。

其次,由于感知数据量巨大,给 HBase 造成了很大的写入压力,网络 IO 成为瓶颈。为了进一步提高写入性能,根据传感器类别标志对数据进行分类,然后将分类后的数据存入缓冲区中再批量、多线程并发写入 HBase 可显著提高写入性能及稳定性。

3.2 海量感知数据的并行化处理技术

云端汇集了电厂多个机组、多个传感器的大量数据,当数据量很大时,单台机器的内存便不足以支撑计算,采用 Spark 框架对海量感知数据进行并行计算,需要对传统的算法进行并行化改进。在改进时,为提升并行化效率,需根据感知数据的自身特点合理进行数据划分,尽可能保障数据处理均匀地分布在不同的服务器上;有效利用 Spark 自身的内部排序中间结果进行处理,避免重复的排序操作;通过本地聚合等手段,减少中间数据的产生,尽可能减少大量数据持久化及在不同服务器之间的传输开销。

3.3 预见性维护技术

预见性维护是通过各种传感器对设备进行监控,基于设备运行过程中的监控数据及与设备维护相关的其他数据,通过预测可能的失效模式以避免设备故障的维护模式。和传统维护方法相比,预见性维护的目标是在故障发生之前采取措施,采集、分析和判断设备的劣化趋势、故障部位、原因并预测变化发展,提出防范措施,防止和控制可能的故障出现。

预见性维护方法针对设备的历史故障,提取发生故障前的设备运行状态特征,应用机器学习算法来寻找故障发生的模式,得出故障预测模型。再将模型应用在新产生的数据上,预测设备未来发生故障的概率。不同种类故障的概率往往需要分别进行模型的训练。

对感知数据的交互式分析也是预见性维护的常见手段。业务人员通过对可视化分析结果的交互操作,来一步步对数据的分析方法进行完善,最终得出分析结果。例如,当把数据以合适的方式呈现到坐标轴上时,那些偏离其他大多数数据的离散点就直观地呈现出来了。“健康状态折线图”“健康地图”和“健康雷达图”等数据可视化工具,可以用来表示部件或设备的性能衰退情况、识别健康模式以及设备或系统的健康风险分布。

4 应用探讨

以某燃煤电厂的历史数据为例,系统可监测设

备感知数据中的异常模式(如孤立点),产生预警事件。例如,若某编号密封油系统油泵电流数据突然出现大幅度的下降,表现为显著不同于其他数据对象,即出现了孤立点。历史数据印证,此时段发生了“发电机氢压下降快”的故障,此为单测点异常模式检测。多测点异常模式检测,例如某抗燃油系统中,油箱油温和送风机进口风温历史数据变化趋势基本保持一致,而在故障发生的时候,这种变化趋势发生了显著异常。这些显著异常点可在健康评估操控盘界面中查看,同时,系统在检测到此类异常点时,也将生成预警事件。

在历史数据及历史故障类别已知的条件下,系统可基于历史数据提取合适的特征构造特征向量,支持向量机、神经网络、随机森林、提升树(boosted tree)等监督学习方法进行故障的预测和预警。图 4 给出了一个故障预测的样例,该样例基于模拟数据训练集和测试集进行说明。由于大部分电厂设备的故障具有明显的累积性特征,大部分故障是由于长期累积工作负荷造成的,遵从从运行良好、性能衰退到故障的规律,这类故障的发生可以进行预测和预测性维护。本文利用开源工具模拟生成的设备从运行良好、性能衰退到故障发生的时间序列数据集,包含时间周期(cycle)、故障标志(failure)、最大良好运行时间(max_life)以及 2 种设备性能测量等属性^[14]。图 4 预测值折线是基于 LSTM 神经网络生成的模型在测试集上的预测结果,真实值为测试集的真实故障情况,由图 4 可以看出,模型在故障未发生前预测到可能发生故障的事件以及时间周期,该信息可呈现并通知给决策或维护人员,技术人员就可以对设备进行检测、维修,排除故障因素,从而有效减少故障真正发生的几率。

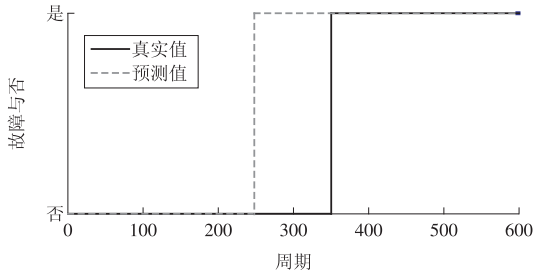


图 4 故障预测
Fig.4 Fault prediction

图 5 是在测试集上预测设备“剩余使用寿命 RUL(remaining use life)”的结果,由图 5 可以看出预测值和真实值比较吻合。业务人员可以通过该曲线的查看进行设备维护时间的合理决策和计划。

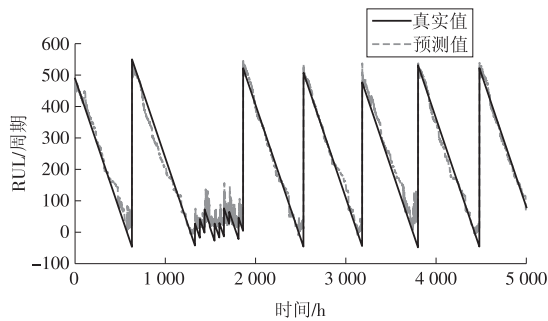


图 5 RUL 预测
Fig.5 RUL prediction

5 结论

本文介绍的“云+端”大数据驱动的预见性电厂设备维护系统架构,能够适应未来数字电厂的特点,该系统综合了电厂设备前端及云端处理的优势,通过分布式并行处理及机器学习方法,本着“治未病”“防患于未然”的原则,通过预测模型的生成及实时判断,在故障发生之前采取措施,防止和控制潜在故障出现。该系统旨在为电厂的安全、可靠运行提供有力的保障。

参考文献:

[1] BESSA R J. Future trends for big data application in power systems[M] // Big data application in power systems. San Diego:Elsevier Incorporated,2018:223 – 242.

[2] 李杰. 工业大数据:工业 4.0 时代的工业转型与价值创造[M]. 邱怡华译. 北京:机械工业出版社,2015.

[3] 王桂玲,王强,赵卓峰,等. 物联网大数据处理技术与实践[M]. 北京:电子工业出版社,2017.

[4] 黄罡,刘譞哲,张颖. 云-端融合应用模型与运行平台[J]. 中国计算机学会通讯,2016,12(11):23 – 27.

[5] 杨亚东. 火电基建质量管理大数据系统应用[J]. 华电技术,2018,40(12):24 – 26.

(上接第 17 页) 系统自动化,2015,39(24):103 – 105.

[12] 严小青. 印尼 BOT 燃煤发电项目商务谈判风险分析与应对[J]. 中国电力企业管理,2019(15):28 – 30.

(本文责编:齐琳)

[6] 朱永利,石鑫,王刘旺. 人工智能在电力系统中应用的近期研究热点介绍[J]. 发电技术,2018,39(3):204 – 212.

[7] 徐建利. 设备精密诊断技术在电厂的实践与应用[J]. 华电技术,2017,39(5):55 – 57.

[8] 张帆. 智慧电厂一体化大数据平台关键技术及应用分析[J]. 华电技术,2017,39(2):1 – 3,7.

[9] 白德龙,倪永中,司兴忠. 新型燃煤电站数字化建设关键技术分析[J]. 华电技术,2017,39(1):10 – 13.

[10] HÜRSCH W L, LOPES C V. Separation of concerns[R]. Boston:Northeastern University,1995.

[11] KREPS J, NARKHEDE N, RAO J. Kafka: A distributed messaging system for log processing[C] // Proceedings of the NetDB,2011:1 – 7.

[12] ZAHARIA M, CHOWDHURY M, FRANKLIN M J, et al. Spark:cluster computing with working sets[C] // Proceedings of the 2Nd USENIX conference on hot topics in cloud computing. HotCloud' 10. Berkeley:USENIX Association, 2010:10.

[13] VORA MN. Hadoop-HBase for large-scale data[C] // Proceedings of 2011 IEEE international conference on computer science and network technology,2011:601 – 605.

[14] MapR for predictive maintenance, data set[Z/OL]. 2019. <https://github.com/mapr-demos/predictive-maintenance/tree/master/notebooks/jupyter/PM3>.

(本文责编:白银雷)

作者简介:

王佳梓(1973—),女,山东济南人,工程师,从事电厂生产、基建技术管理方面的工作(E-mail:w15153577735@163.com)。

王桂玲(1978—),女,北京人,副研究员,博士,从事大数据、流数据处理与分析方面的研究。

作者简介:

郭乐(1974—),男,河北饶阳人,正高级工程师,从事综合能源服务、电力市场研究方面的工作(E-mail:xn.sac@163.com)。

张丹丹(1983—),女,江苏建湖人,工程师,从事综合能源服务方面的工作。