



大数据技术在电力行业的应用研究

赵云山¹, 刘焕焕²

(1. 神华国华(北京)电力研究院有限公司 北京 100025;

2. 烟台南山学院计算机与电气自动化学院 烟台 265713)

摘要:随着无处不在的传感器芯片、高速通信网络以及高性能分布式计算系统的广泛应用,随时产生、海量存储、异构形态、即时处理的海量数据正把人类社会带入崭新的大数据时代。电力行业大数据综合了电力企业的产、运、销及运营和管理数据,如何对其采集、传输、存储、分析并最终指导生产、服务终端用户将是电力行业信息化的重要研究课题。从大数据定义、发展趋势、关键技术3方面宏观上概述了大数据时代的特征,在此基础上总结了现有大数据技术在电力行业的研究与典型应用,最后分析了大数据时代电力企业信息化发展所面临的若干挑战。

关键词:大数据;电力信息化;智能电网;数据挖掘

doi: 10.3969/j.issn.1000-0801.2014.01.009

Research on Application of Big Data Technique in Electricity Power Industry

Zhao Yunshan¹, Liu Huanhuan²

(1. Shenhua Guohua (Beijing) Electric Power Research Institute Co., Ltd., Beijing 100025, China;

2. School of Computer and Electric Automation, Yantai Nanshan University, Yantai 265713, China)

Abstract: With the rapid development of information technology, the ubiquitous sensor network, high-speed communication network, and the high performance distributed computing system have been commonly implemented. The massive data, which is generated anytime, with the mass storage, heterogeneous structure, and requiring on-time process, is bringing our society into the age of big data. The electricity big data includes the information of production, transportation, marketing, enterprise operation and management in the electricity power enterprise. Therefore, how to collect, transfer, store, and analyze the big data so as to conduct the production, and serve the electricity end users turns to an important research subject for the power industry. An overview was given about the features of the age of big data from three aspects: the definition, development trend, and the key technology of big data. Then, a summary of the research and typical application of the existing big data technology in the electric power industry was described. Finally, the challenges of big data technology in the future for the electric power information were concluded.

Key words: big data, electric power information, smart grid, data mining

1 引言

从电力行业信息化发展的阶段看,经过近几十年的发

展,其目前已处于“向信息资源应用转变、向资源经营管控转变、向提升经济效益转变”的阶段。与此同时,各种新兴技术已日趋成熟,云计算、物联网、大数据、移动通信、社交



网络等信息技术正支撑着智能电网、绿色电力、电力信息系统集成等的发展。新形势下,无论是国家政策的要求,外部环境的竞争形势,还是企业自身的发展战略规划,都促使电力行业更加重视信息化的合理规划和建设,通过信息化手段提高企业管理和经营水平,如图1所示。

根据麦肯锡(McKinsey)2011年5月发布的《大数据:创新、竞争和生产力的下一个新领域》报告^[1],数据已经渗透到每一个行业和业务职能领域,逐渐成为重要的生产因素。对于电力行业而言,电力生产涉及的运行工况、参数、设备运行状态等实时生产数据,现场总线系统所采集的设备监测数据以及发电量、电压稳定性等方面的数据,电力企业运营和管理数据,如交易电价、售电量、用电客户信息、ERP综合数据等,共同构成了“电力大数据”。由于电力大数据综合了电力企业的产、运、销及运营和管理数据,因此已成为电力企业深化应用、提升应用层次、强化集团企业管控的有力技术手段。

电力企业作为资产密集型和技术密集型行业的典型,信息技术已覆盖基建、生产、经营、办公、管理各个领域,不仅历史数据量巨大,而且实时采集及产生的数据量正迅速增长。目前面临的问题不仅是如何采集、存储和检索数据,关键是如何分析和利用此类数据并提供有价值的信息,进而全面管控企业业务,为企业管理和战略决策提供数据支撑^[2]。因此,对电力大数据的产生、应用方式、关键技术进行系统性分析,为电力企业提供应对大数据分析和处理的解决方案,显得尤为必要。

2 电力大数据概述

一般意义上,大数据是指无法在可容忍的时间内用传统信息技术和软硬件工具对其进行感知、获取、管理、处理和服务的数据集合。从大数据的特征出发,大数据一般具有“5V”特征^[3]。根据电力行业特征,电力大数据主要来源于电力生产、管理运营、智能电网3个领域。

(1) 电力生产大数据

电力生产是大数据产生的主要源头之一,其信息系统

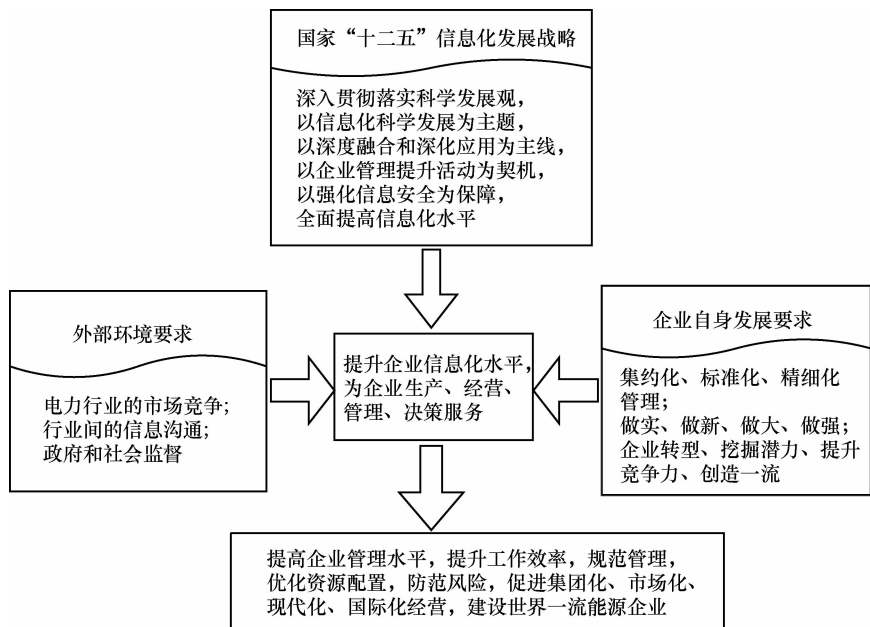


图1 信息化发展的推动因素

覆盖了发电、检修、安全3项主要业务领域。其数据类型主要分为实时生产数据和设备全生命周期数据两大类,主要信息系统包括企业资产管理(EAM)系统、技术监督、性能分析及耗差计算、实时数据库、锅炉寿命管理、巡点检、安全监察、设备可靠性系统等。对此类数据的分析主要侧重于如何利用历史信息指导发电生产及设备检修。

(2) 电力管理运营大数据

电力企业的经营决策需要大量的生产和经营报表支撑,需要跨单位、跨专业、跨业务进行数据分析与挖掘,并以多种方式实现用户友好的数据可视化效果。因此,此类数据的分析利用主要采用商务智能(BI)工具,以多维度、易理解的方式呈现数据视图,为企业的各种经营活动提供决策信息。

(3) 智能电网大数据

随着电力行业工业化与信息化的深度融合,智能电网将承载着电力流、信息流、业务流,电网和电力的信息通信网的用户将发生叠加,电力企业的整体价值将不断跃升,使电力行业具备在大数据时代进行深度数据挖掘和分析的先天优势。智能电网的数据源主要是无处不在的各种传感器网络,通过高速通信网络集中到运营调度中心,实现电能使用的可测可控,使电力系统更加清洁、高效、安全、可靠。

3 大数据的发展趋势

从数据库时代的结构化存储,到大数据时代海量存储和数据分析,在数据规模、数据类型、数据模式、处理对象

和处理工具等方面均存在很大差异^[4]。

3.1 大数据的生成阶段

如图 2 所示,大数据的产生经历了被动、主动、自动生成 3 个阶段:第 1 阶段是运营式系统阶段,此阶段的数据产生方式是被动,数据往往伴随着一定的运营活动而产生并保存在数据库中;第 2 阶段是用户原创内容阶段,随着互联网的诞生尤其是 Web2.0 时代的到来,以博客、微博为代表的新型社交网络出现并快速发展,同时,以智能手机、平板电脑为代表的新型移动设备的出现都使得数据产生速度呈现爆炸性增长,此阶段的数据产生方式是主动的;第 3 阶段是感知式系统阶段,由于感知式系统的广泛使用,此类带有处理功能的传感器广泛布置于社会各个角落,并源源不断地产生新数据,此阶段的数据产生方式是自动的。

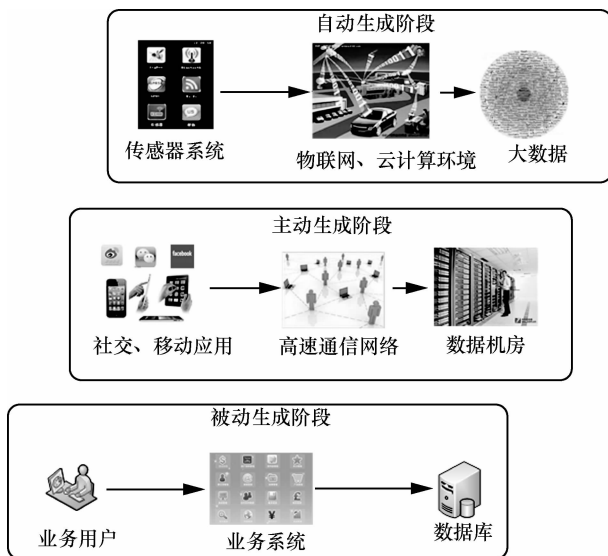


图 2 大数据的产生

3.2 大数据的应用方式

如图 3 所示,对应大数据生成经历各个阶段,其分析处理方式也经历了 4 个阶段的发展:最初是展示阶段,人们并未意识到大数据的产生,仍旧以原有的方式进行存储和展示,此阶段的目的是描述“发生了什么(what happened)”；逐渐地,在数据量进一步增长的情况下,对数据的有效分析显得尤为重要,此阶段更加注重的是“为什么发生(why did it happen)”；第 3 阶段,随着数据挖掘和分析技术的进一步深入,数据处理能力进一步提高,使得海量数据的在线监测成为可能,人们面对的不再是历史数据而是“正在发生、进行中(what's happening)”的

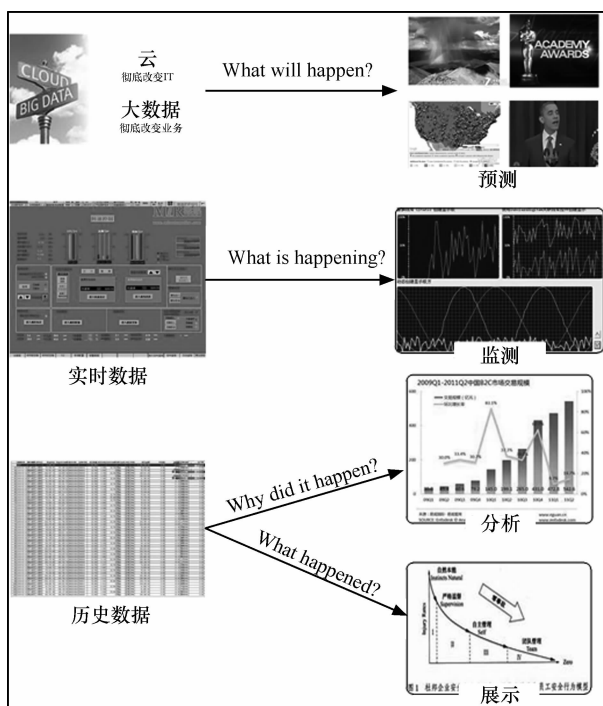


图 3 大数据应用方式的发展

数据生产过程;当前以及未来阶段,大数据分析正朝着有效预测这一目标发展,届时分析“将要发生什么(what will happen)”将成为可能^[5]。

4 大数据关键技术

大数据关键技术包括数据抽取与集成、数据分析、数据解释 3 个领域^[6]。

大数据的一个重要特点是多样性,意味着其数据来源极其广泛,数据类型极为繁杂,这种复杂的数据环境要求必须对数据源中的数据进行抽取和集成,并采用统一定义的结构来存储这些数据。为保证数据质量,需要在数据抽取与集成后进行数据清洗。现有的数据抽取与集成方式主要有基于 ETL(extract-transform-load,提取、转换、加载)引擎的方式和基于搜索引擎的方式等。

数据分析是大数据发挥其核心价值的重要流程,传统的分析技术有数据挖掘、机器学习、统计分析等,分析的结论可用于推荐系统、专家系统、商业智能和决策支持系统等。

在数据分析的基础上,终端用户往往最关心的是数据的展示方式,如果分析结果没有采用适当的解释方法,所得到的分析结论往往让用户难以理解,极端情况下甚至会误导用户。传统的解释方法仅是文本、图表等电脑终端上



的直观显示,未来提升数据解释能力可以引入标签云(tag cloud)等可视化技术解决,甚至可以采用人机交互技术,在交互过程中逐步引导用户进入分析流程,达到最佳的数据解释效果。

大数据价值的完整体现需要多种技术的协同。为了便于数据管理,需要建立云存储系统等存储结构化、非结构化、半结构化类型的数据。为了加速数据处理,需要采用云计算等分布式计算技术。通过索引与查询技术,提供搜索引擎服务,提高用户对大数据的使用效率。

5 大数据技术在电力行业的研究与典型应用

5.1 IBM 智慧电力解决方案

面对环境污染、能源危机、新经济增长方式的种种挑战,IBM 给出了其利用智能(intelligent)和科学(science)的智慧电力解决方案,如管理及优化企业停电计划的“智能停电管理系统”,帮助电网企业优化建设改造投资计划的“智能电网评估与投资优化决策系统”,可智能感知电网实时运行状态并辅助监管人员决策的“电网状态智能感知与报警系统”等。

5.2 大数据支撑智能电网发展

“2012 电力行业信息化年会”的主题为“大数据与宽带中国”。与会专家一致认为:大数据将给电力企业带来新一轮商业模式转变和价值创新,宽带中国战略更为电力信息化发展提速。作为大数据的典型应用,智能电网再次成为讨论的热点。

智能电网的理念,是通过获取更多的如何用电、怎样用电的信息来优化电的生产、分配以及消耗。IBM 早在 2006 年就提出了智能电网的概念,在智能电网中引入了“信息流”的概念,即电网要能够把电能流、信息流结合在一起,实现传输能源的同时实现数据的采集。智能电网还通过优化模型对数据进行深度挖掘和分析,预测电能流的情况,最终实现清洁发电、高效输电、动态配电、合理用电的智慧电力的目标。

5.3 电力企业一体化整合集成平台

随着“十八大”提出了“推动信息化和工业化深度融合”,各大电力企业都在积极地采用信息化手段来改造和提升整个企业的核心业务模式。国家电网公司 2006 年启动实施“SG186 工程”,与智能电网建设相辅相成,电网信息化已经逐渐显示出极强的生命力和发展潜力;华能

集团于 2008 年制定信息化规划,目标是建立集团统一的生产管理平台;华电集团在华电国际信息化基础上,逐步将信息化应用统一到集团层面,建立了统一的生产系统和多业务平台。

各大电力集团和电网公司所筹划的上述系统集成平台或一体化平台,从本质上来说,均属于大数据背景下的数据集成和分析、处理平台。如图 4 所示,从系统架构来看,无论是涉及数据存储的数据架构,还是涉及数据分析处理的技术和应用架构以及上层业务架构,均与大数据的抽取与集成、分析、解释技术密切相关。因此,大数据时代的电力企业信息化建设需要分别从上述角度进行深入考虑,使大数据集成和分析平台高效、易用并且可扩展。

数据管理方面,发电企业在其生产过程中存在大量实时数据的存储、处理和集成问题,仅靠集散控制系统(DCS)和关系数据库技术并不能完全应对。开放结构的实时数据库系统能够提供高速、及时的实时数据服务,有效地集成异构控制系统,提供分布式的数据服务,并在工厂控制层(现场总线、PLC、DCS 等)与管理信息系统(MIS)之间建立实时的数据连接,使企业全生产过程控制与业务管理得以结合。目前国外的主流实时数据库产品如 PI、InfoPlus 等普遍存在价格昂贵、系统维护困难、技术保密、工程实施和应用复杂等问题,而国内实时数据库产品,由于研发时间短、工程应用面不广、性能尚不能完全达到要求、功能不够完善等因素,通常难以满足企业的具体需求。因此,对实时数据库的理论研究需要进一步深入,以支撑对大数据的存储和检索。

5.4 大数据挖掘与分析处理技术在发电企业的应用研究

我国发电企业对于大数据挖掘(data mining,DM)与分析处理技术的应用研究仍处于探索阶段,迫切需要对大数据技术开展全面深入的研究,突破电力大数据处理的核心关键技术,抢占未来数据领域技术制高点,为构建国际先进水平的电力大数据处理平台奠定技术基础。具体研究内容包括:

- 在大数据的预处理方面,提出面向电力生产大数据的数据采集、清洗、转换、集成及数据质量度量方法;
- 在大数据计算方面,研究电力生产大数据的建模方法,设计分布式的大数据计算系统架构,研发大数据计算系统;
- 在大数据存储方面,研究电力大数据的存储模型,

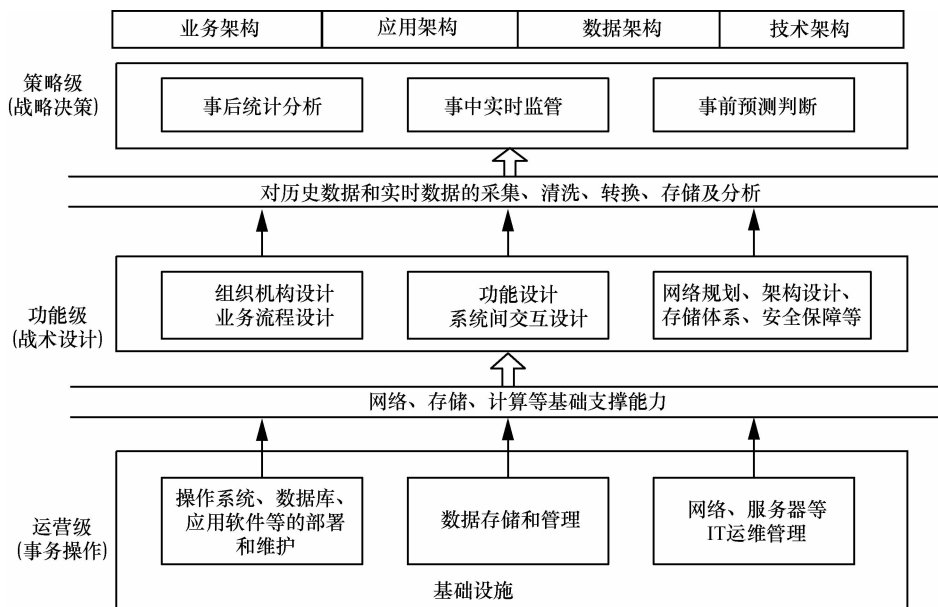


图4 信息系统集成平台架构

设计电力大数据存储系统架构,研发电力大数据存储系统;

- 在大数据传输方面,研究电力大数据的采集和传输方法,设计面向大数据的网络系统架构,研发大数据网络传输系统;
- 在突破电力大数据处理的核心关键技术的基础上,研发针对电力大数据的开发环境、分析挖掘工具和决策支持平台。

基于上述研究,电力生产大数据的挖掘和分析处理技术可以指导发电企业更好地进行设备运行状态评估及故障诊断、发电生产决策与控制等^[7]。

(1)安全评估及故障诊断

基于数据挖掘技术的设备运行状态监视^[8]:基于DCS、PI等系统的实时数据和海量的历史数据,结合其他第三方系统数据(例如天气状况、电网调度历史数据等),建立机组安全运行状态模型,数据挖掘技术可以自动发现某些不正常的分布,从而暴露设备运行中的异常变化,分析潜在的不安全因素,协助运行和检修人员预测机组运行状态,并迅速找出问题发生的范围及时检修和采取对策。

(2)发电生产决策与控制

基于数据挖掘技术的决策支持和控制^[9-11]:决策支持方面,发电企业门户系统的主要功能是统计和展示,并没有提供决策信息(例如某电厂的月度发电量指标、年度经营指标等如何制定);控制方面,当机组出现异常情

况时,目前仍然是基于专家系统的控制方式,即依赖经验丰富的专家(值长),此时呈现在专家面前的数据量从几十条/min瞬时上升为几十条/s,数据量的激增使得专家在应对异常状况时也有较大压力,因此这种控制方式也已无法适应生产要求。数据挖掘技术具有定性分析能力,从大量数据中去除冗余信息,可将每一种状态的故障特征提取出来,成为判断机组状态、如何快速处理故障、准确决策的依据。

(3)设备检修策略改进

基于数据挖掘技术的电力设备状态检修^[12]:首先收集设备的基础信息、历史运行数据、设备缺陷信息等,通过对历史运行数据和缺陷信息进行数据挖掘,得到设备缺陷状态下特征值及关联参数值,将挖掘得到的信息与设备当前运行监测值进行对比分析,即可以判断设备当前运行状态是否正常。例如,通过关联规则分析,往往可以发现A设备振动报警后,B设备也会有较大概率出现振动报警,该关联规则可以提供早期故障预测及原因分析。

6 大数据时代电力企业所面临的挑战

大数据给电力企业带来新发展机遇的同时,也面临着若干挑战:

- 数据中心规模越来越大,拥有的服务器数量越来越多;
- 针对具体应用的需求进行软硬件资源配置而建立



起来的 IT 基础设施和数据中心存在着一系列缺点,其资源利用效率普遍较低;

- 服务器系统需要通过大量人工操作来管理,不仅容易出错,而且不能实现整个数据中心的高可用性和稳定性,企业业务和关键数据安全也得不到有效保障;
- 对现有 IT 资源的使用情况没有有效的监控和管理工具;
- 对 IT 基础架构资源中的中间件、数据库、服务器、存储等的使用没有相应的平台支持和配套的自动化管理手段;
- 对 IT 资源的使用、管理及运维方式,多采用人工方式,费时费力且效率不高;
- 隐私、安全性问题。

针对上述问题,业界已给出了一系列解决方案,例如从标准化、虚拟化、自动化入手,构建企业整体的虚拟化 IT 资源池,并提供统一的运维管理平台,实现 IT 资源的智能使用等。从电力企业管理的角度,改变 IT 系统建设方式、提升运维水平也变得更加紧迫。

7 结束语

大数据时代,数据蕴含着巨大的价值,如何利用好产、运、销等大数据信息对企业的发展具有重要的战略意义。随着物联网、云计算、大数据“三驾马车”的高速驶来,物联网连接了泛在的数据源^[13],云计算提供了基础设施架构和高性能服务计算^[14],大数据则转变为企业的灵魂资产,各种数据分析与处理技术为人们更深入地感知、认识和控制纷繁复杂的物理世界,提供了前所未有的丰富信息。

大数据时代,对数据的分析和挖掘仅仅是技术手段,发现和预测才是最终目标。展望电力行业信息化的未来,“集约化、精细化、信息化”的集团管理将成为进一步提高电力行业运营效率和管理水平的主要方向,而大数据技术的应用势必成为这一进程的焦点。如何利用好大数据,充分挖掘企业数据资产,更好地服务电力行业发展和广大电力用户将成为电力企业持续发展的重要研究课题。

参考文献

- 1 Manyika J, Chui M, Brown B, *et al.* Big Data: the Next Frontier for Innovation, Competition, and Productivity. McKinsey Global

Institute Report, May 2011

- 2 丁杰, 徐建, 王春毅等. 迎战大数据. 电力信息与通信技术, 2013, 11(2):5~10
- 3 IBM. What is big data. <http://www-01.ibm.com/software/data/bigdata>, 2012
- 4 Mayer-Schonberger V, Cukier K. Big Data: A Revolution That Will Transform How We Live, Work, and Think. New York: Eamon Dolan/Houghton Mifflin Harcourt, 2013
- 5 亓开元, 赵卓峰, 房俊等. 针对高速数据流的大规模数据实时处理方法. 计算机学报, 2012, 35(3):477~490
- 6 孟小峰, 慈祥. 大数据管理: 概念, 技术与挑战. 计算机研究与发展, 2013, 50(1):146~169
- 7 陈星莺, 张晓花, 瞿峰等. 数据挖掘在电力系统中的应用综述. 电力科学与技术学报, 2007, 22(3):51~56
- 8 李勋, 龚庆武, 杨群瑛等. 基于数据挖掘技术的保护设备故障信息管理与分析系统. 电力自动化设备, 2011, 31(9):88~91
- 9 李智勇. 电力系统运行信息的数据挖掘研究. 浙江大学博士学位论文, 2009
- 10 王培红, 陈强, 董益华等. 数据挖掘及其在电厂 SIS 中的应用. 电力系统自动化, 2004, 28(8):76~79
- 11 周鲲鹏, 方仍存, 颜炯等. 电网规划智能辅助决策系统的设计与实现. 电力系统自动化, 2013, 37(3):77~82
- 12 杨国庆, 张宇. 数据挖掘技术在电力设备状态检修中的应用. 上海电力学院学报, 2012, 28(2):176~180
- 13 胡雯, 孙云莲, 杨成月等. 基于物联网的智能电网信息化建设研究. 电力信息化, 2013, 11(4):51~54
- 14 王德文. 基于云计算的电力数据中心基础架构及其关键技术. 电力系统自动化, 2012, 36(11):67~71

[作者简介]



赵云山, 男, 博士, 神华国华(北京)电力研究院有限公司工程师, 主要研究方向为电力信息系统架构。



刘焕焕, 女, 烟台南山学院讲师, 主要研究方向为数据挖掘。

(收稿日期: 2013-08-01)