

DOI: 10.7667/PSPC152053

# 一种基于 Hadoop 的电力大数据属性实体识别算法

齐俊<sup>1</sup>, 曲朝阳<sup>1</sup>, 娄建楼<sup>1</sup>, 王冲<sup>2</sup>

(1. 东北电力大学信息工程学院, 吉林 吉林 132012;

(2. 国网内蒙古东部电力有限公司信息通信分公司, 内蒙古 呼和浩特 010020)

**摘要:** 随着大数据时代的来临, 传统的实体识别技术由于电网数据体积大以及类型复杂等特性已经无法有效地进行数据预处理。近年来兴起的 Hadoop 技术能够对大数据进行较好的处理。因此提出一种基于 Hadoop 的电力大数据属性实体识别算法。该算法利用改进离散化算法选取信息准确率较高的离散点, 并提出了一种离散化评价指标。最后, 在 Hadoop 平台上对某风电机组的监测数据进行了属性实体识别。实验证明, 该算法在实验正确性和断点数目方面表现良好, 并且具有较好的加速比, 适用于电力大数据的属性实体识别处理。

**关键词:** 电力大数据; 实体识别; 离散化算法; 信息准确率

## A kind of attribute entity recognition algorithm based on Hadoop for power big data

QI Jun<sup>1</sup>, QU Zhaoyang<sup>1</sup>, LOU Jianlou<sup>1</sup>, WANG Chong<sup>2</sup>

(1. School of Information Science and Engineering, Northeast Dianli University, Jilin 132012, China; 2. Information & Telecommunication Branch Company, State Grid East Inner Mongolia Electric Power Co., Ltd., Hohhot 010020, China)

**Abstract:** With the coming of the era of big data, traditional entity recognition technologies have been unable to effectively finish data pre-processing because of the large scale of power grid data and volume complex type features. The rising of the Hadoop technologies in these years can deal with the big data processing better. Therefore this paper proposes a power big data entity recognition algorithm based on Hadoop. This algorithm uses the discretization algorithm to select higher information accuracy discrete points and puts forward a discretization evaluation indicator. In the end, the entity recognition of the monitoring data of wind turbines is finished on Hadoop platform. Experimental results show that the proposed algorithm performs well in terms of correctness and breakpoint number experiments and it has a good speed-up ratio. The proposed algorithm can be applied to power large data entity recognition processing.

This work is supported by National Natural Science Foundation of China (No. 51277023).

**Key words:** power big data; entity recognition algorithm; discretization; information accuracy

## 0 引言

随着信息通信技术的不断进步, 数字化、信息化已经深入渗透到我们生活的方方面面, 电力企业的信息化进程也得到了长足的发展。下一代智能电网的全面建设过程中产生的温度、时间等数据的关联分析也使得电力大数据的类型不断增加, 对电力大数据中有效信息的分析处理要求也不断提高。如何从电力大数据中获取电力企业决策时需要的有效信息, 是在大数据时代电网企业进行数据预处理时

的一个重要难题<sup>[1]</sup>。实体识别一直以来都是数据质量管理研究的一项关键技术, 对能否提高数据预处理的质量起着至关重要的作用。在电力大数据中, 数据类型复杂、数据表现不一致的现象更为普遍。因此, 实体识别技术在电力大数据中也具有更为广泛的应用<sup>[2]</sup>。

电力大数据中的属性实体识别是在给定的大数据集中, 精准地识别属于同一种实体的不同实体名和属性并将其聚类, 使得每个实体在电网决策中能够被更为有价值地识别出来的过程。与传统的中文实体识别技术不同, 电力大数据具有更为复杂的结构, 如 XML 数据等。因此, 尽可能地寻找一种有效的电力大数据实体识别方法对降低数据处理规

**基金项目:** 国家自然科学基金资助项目(51277023); 吉林省科技厅社发处重点科技攻关项目(20150204084GX)

模, 提高数据分析效率具有重要作用, 例如文献[3]利用基于并行机群的大数据实体识别算法。此算法借助 n-Gram 算法辅助解决同一物体有不同属性的问题, 在短时间内对大数据高效地进行实体识别并取得了较好的效果。传统的实体识别技术有很多, 但主要集中在文本形式的词组或关系数据上, 针对不同类型数据的实体识别的研究还刚刚兴起。文献[4]提出了一种两阶段的关联实体识别模型, 考虑了实体的模式特征与属性特征, 并提出了一种增量式验证算法。

当前已有的研究方法主要是面向识别的有效性, 即重点在于如何能够准确地识别出描述同一实体的对象, 面向大数据中的实体识别效率的技术仍然较少。而且这些方法大多针对字符串和关系元组, 对 XML 数据、图数据等非结构化数据的判别方法研究仍然较少<sup>[5-6]</sup>。同时, 这些算法缺少有效评估大数据实体识别结果质量的理论及公共测试数据集合。

Hadoop 是一种能够对大数据进行分布式处理的基础架构平台。其架构底层 HDFS 的上层是 Map-Reduce 执行引擎, 该引擎由单独运行在主节点上的 JobTracker 和多个运行在集群节点上的 TaskTracker 组成。MapReduce 是一种编程模型, 用于大规模数据集的并行运算。其主要概念包括“Map(映射)”和“Reduce(归约)”两个过程。当前的算法实现是指定一个 Map(映射)函数, 用来把一组键值对映射成一组新的键值对, 指定并发的 Reduce(归约)函数。

本文深入研究了 Map-Reduce 编程模型, 在电力大数据的背景下, 给出了一种基于信息准确率的电力大数据实体识别算法(Entity Recognition in Big Data Based on Information Accuracy, ERBIA)。该算法首先通过离散化方案计算类属性的分布与属性值的相似程度, 从而通过 ERBIA 算法选出信息准确率较高的离散点, 而后提出一种改进的离散化评价指标得到了最终决策结果。最后, 在 Hadoop 平台上对真实数据集和随机产生的数据进行了多组对比试验, 得到了具有更好的有效性和高效性的电力大数据处理方案。

## 1 电力大数据实体识别离散化方案描述

数据处理的首要问题是知识的表达。为了便于数据的集成处理, 提高数据预处理的效率, 本文采用列联表对电力大数据属性进行形式化定义, 每组数据的属性形式化定义如式(1)。

$$S = (U, V, C, f) \quad (1)$$

在表达式中:  $U = \{a_1, a_2, \dots, a_n\}$  定义为数据的非空有限集合, 称为属性域;  $V = \bigcup_a (\forall a \subset C)$  定义为有效信息函数  $f$  的值域;  $C$  称为属性域, 并且有  $C \neq \emptyset$ ;  $f = \{f_a : V \rightarrow V_a\}$  表示关联列表的信息函数, 其中  $f_a$  为属性  $a$  的信息函数。

根据上述定义, 电力大数据集可以表示为属性域中元素个数为  $N$  的关系列表, 即电力大数据集  $S$  中有  $N$  个属性值, 对于其中的第  $i$  个属性值有  $a_i \subset V$ , 值域为  $C_i$ ,  $S$  中  $a_i$  的值的集合可以表示为  $a^i(u) = \{a_{11}^i, a_{21}^i, \dots, a_{n1}^i\}$ , 其中  $n$  表示  $a_i$  在  $U$  中的数量。假设  $a$  为数据集中的任一连续属性, 且在每一个连续属性上都存在一种离散化方案  $R$ , 将阈值为  $N$  的属性域集合划分为  $n$  个交集为零的区间,  $R: \{(c_0, c_1), (c_1, c_2), \dots, (c_{n-1}, c_n)\}$ , 其中属性  $a$  的值域  $V_a = [c_0, c_n]$ , 方案  $R$  中的值按顺序排列并组成相应的断点集  $\{c_0, c_1, \dots, c_n\}$ 。由于断点集与所提出的离散化方案相对应, 因此可以采用两者中的任意一种对属性离散化进行表述。根据以上定义可以建立某种属性  $a$  对应的离散化方案  $D$  的对应表, 如表 1 所示。

表 1 某种属性  $a$  对应的离散化方案  $D$  的对应表

Table 1 Corresponding table of some attribute corresponding to discretization scheme

属性类别	离散化区间划分					决策属性
	[[ $(c_0, c_1)$ ], [ $(c_{i-1}, c_i)$ ], [ $(c_{n-1}, c_n)$ ]]					
$a_1$	$q_{11}$	...	$q_{1i}$	...	$q_{1n}$	$d_1$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_i$	$q_{i1}$	...	$q_{ii}$	...	$q_{in}$	$d_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_m$	$q_{m1}$	...	$q_{mi}$	...	$q_{mn}$	$d_3$

从上述定义可以看出, 本文提出的离散化算法对电力大数据集进行实体识别<sup>[7-8]</sup>, 实质上是通过选取适当的间断区间对数据的属性集进行划分。这样可以避免传统数据实体识别方法通常采用的针对单一实体模式特征或者基于单一类型实体属性特征的方法对数据间的关联性进行衡量时, 难以对两者进行有效地整合的问题。下面给出一种 Hadoop 平台上一种基于信息准确率的大数据实体识别算法。

## 2 Hadoop 平台上一种基于信息准确率的大数据实体识别算法

传统的属性离散化算法主要用于知识发现以及知识决策等领域, 检验离散化效果优劣的指标主要由信息熵来表现。信息熵作为用于度量信息量的概

念,可以更为细致地对离散化区间进行划分,并使离散化后的区间上的信息更加明确。但是基于信息熵的评价指标的弊端在于,虽然分类的信息量所包含的内容越来越多,但是由于离散区间的划分过于细致,会导致计算过程中产生的内存过大,对算法的效率和硬件的消耗均有影响,并且不利于后续数据的处理<sup>[9]</sup>。因此,本文在信息论的基础上,针对电力大数据属性提出了一种基于信息准确率的大数据实体识别算法 ERBIA,并提出了一种新的离散化评价指标使算法的效率得到提高。

### 2.1 电力大数据的来源及特点

电力系统不断运行的过程中所产生的数据也具有大数据的典型特征。电力系统大数据地理位置分布广、通信调度高度可靠、实时运行从不停止的特点决定了电力系统运行时产生的数据数量庞大、增长快速、类型丰富,完全符合大数据的描述特征,是典型的大数据。因此在电力系统数据高速增长的形势下,传统的数据处理技术不能满足从海量电力大数据中快速获取知识与信息的要求,大数据技术在电力行业中的应用是电力行业信息化、智能化发展的必然要求<sup>[10]</sup>。

由于智能电网的不断深入推进,电力系统的数字化、信息化带来了更多的数据源,例如智能电表收集到的家庭和企业终端用电数据,电力设备状态监测系统从数以万计的变压器、发电机、开关设备、高压电缆等设备中获取的不断累积的监测数据,光伏和风电功率预测过程中所需的大量历史运行数据等。这些运行数据会有若干种属性,然而在监测机器是否正常运行时无关属性过多会导致决策时的效率低下。因此,将若干相似属性进行离散化处理可以简化决策过程,提高决策效率。电力大数据属性离散划分点的选取就成为提高离散化算法处理效率需要处理的首要问题。

### 2.2 信息准确率定义

电力大数据属性离散化的本质是在属性的值域范围内设置离散划分点,将属性的值域划分为区间,最后用整数值点代表每种划分的属性值<sup>[11]</sup>。因此首先要对划分点的选取进行研究。在本文中,将划分点选取的标准定义为信息准确率,假设存在信息表  $S$ ,存在属性  $a_i (i=1, 2, 3, \dots, n)$  的信息准确率  $Q_i$ , 则

$Q_i = \frac{1}{n} \sum_{l=1}^n \frac{1}{l_i}$  表示决策系统中属性  $a_i$  的信息准确率,  $l_i$  表示当属性  $a_i$  取值为  $i$  时的值,即决策属性可能取到的数目值。从定义可以看出,当  $Q_i$  的值越大,说明属性  $a_i$  在离散化点的选取中越准确。在文献<sup>[10]</sup>

中已经得到证明,属性的重要程度之间在概率统计上是相互独立的,则信息表离散点总准确率可以定义为  $Q_0 = \prod_{i=1}^n Q_i$ , 其中当数据量趋于无限大时,可以认为各个  $Q_i$  相等,将其标记为  $Q$ , 则有  $Q_0 = Q^n$ , 可得  $Q = \sqrt[n]{Q_0}$ 。

### 2.3 改进的离散化评价指标

在确定了信息准确率的定义之后,本文提出了一种改进的基于信息熵的离散化评价指标,用来度量电力大数据中某种属性  $a$  所对应的离散化方案  $D$  的离散效果。传统的信息熵定义表示如下。

$$H(X) = -\sum_{i=1}^n p_i \log_2 p_i, p_i = n_i / |X| \quad (2)$$

式中:  $|X|$  表示  $X$  的基数; 属性  $i$  的实例个数为  $n_i$ 。

在本文的离散化方案中,将每一个区间  $d$  上的信息熵表示为  $H_D(X) = -\sum_{i=1}^d p_i \log_2 p_i$ , 如果离散点  $d$  可以将集合  $X$  分成两个子集  $X_{d-x}$  和  $X_{d+x}$ , 则可以将点  $d$  对于集合  $X$  的信息熵定义为

$$H(X) = (|X_{d-x}|(H(X_{d-x})/|X| + |X_{d+x}|(H(X_{d+x})/|X|)/\log_2(n) \quad (3)$$

对所提出的改进的离散化评价指标定义如下:

(1) 当  $H(X)$  越大时,说明电力大数据连续属性离散化后的信息准确率越高,离散化方案的划分质量较高;

(2) 本文使用  $\log_2(n)$  作为调节算子,尽量将离散区间数限定在一个合理的范围内,避免由于区间划分得过大或者过小导致的离散结果不理想;

(3) 当区间  $X$  取零时,表示所有划分区间类分布均匀,  $H(X)$  取最小值。

### 2.4 算法描述

算法:一种基于信息准确率的电力大数据实体识别算法。

输入:训练样本监测数据集  $D$ ;

输出:决策表  $S$ ;

令  $S = (U, V, C, f)$ , 属性数目为  $N$ , 决策属性集为  $d$ 。

Step1: 通过聚类生成需要进行电力大数据预处理的初始离散点集合  $S_D$ ;

Step2: 计算  $S_D$  中各个离散点的信息准确率  $Q_i$ , 并对所得到的结果进行升序排列,组成信息准确率  $Q_i$  的集合  $S_Q$ ;

Step3: 对离散点集合  $S_D$  进行初始化;

Step4: 将  $S_D$  中的值  $Q_i$  进行赋值, 把得到的新集合命名为  $S'$ ;

Step5: 设定划分区间数目初始值为 1;

Step6: 对于下一个  $S_Q$  中的属性值返回执行 Step1, 循环执行直到所有的属性完成;

Step7: 根据评价指标判断当前离散化方案是否满足  $H(X)$  在值域范围内, 如果满足则将决策结果在决策表  $S$  中表示出来, 即将最后的决策结果输出并保存在决策表中, 若不满足则将该离散区间的方案排除执行下一区间, 直至所有区间完成;

Step8: 将评价后的结果用数字进行编码表示, 得到决策结果。

### 3 实验分析

为了验证本文提出的基于信息准确率的大数据实体识别算法的有效性, 本文以辽宁某风电公司并网风电机组在线监测数据为例, 从算法的正确性、断点数目分析以及加速比为例进行分析<sup>[12]</sup>。

#### 3.1 正确性

本文选取 2013 年 12 月辽宁某风电公司风电机组部分运行监测数据, 选取数据中的几种运行参数作为类别属性, 以离散效果作为决策条件, 选取风电机组的 6 种不同温度作为输入数据, 分别为 NCC300 温度  $a_1$ , NCC320 温度  $a_2$ , 机侧半导体温度  $a_3$ , 环境温度  $a_4$ , 网侧半导体温度  $a_5$  和齿轮箱轴承温度  $a_6$ 。为了方便表示, 本文将决策结果用三种编码表示, 分别是用 00 表示正常, 用 10 表示合格, 用 11 表示不合格。表 2 是从监测数据中截取的部分属性的数值(单位为摄氏温度)。

表 2 部分监测数据属性值

$S_D$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$d_1$
$x_1$	25.8	27.8	32.8	21.6	30.8	22.8	00
$x_2$	26.3	29.3	32.9	22.6	31.6	24.1	10
$x_3$	27.9	29.7	33.6	23.6	32.4	24.8	00
$x_4$	27.1	30.6	31.8	23.4	33.1	24.1	00
$x_5$	30.5	33.5	40.6	26.4	34.3	29.1	11
$x_6$	31.2	34.5	46.1	28.6	35.6	28.0	00

在 Eclipse 环境下采用 ERBIA 算法对上述属性进行离散化后的数据如表 3 所示。

可以看出, 本文提出的 ERBIA 算法与常规算法所得出的计算结果具有一致性, 而常规算法采用的整体计算均值的算法会使个别属性的评判具有较大偏差, 使得决策结果与实际运行情况偏差较大。

表 3 经过 ERBIA 算法离散化后部分监测数据属性值

Table 3 Part of the values of monitoring data of attribute

$S_D$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$d_1$
$x_1$	3	3	2	4	3	2	00
$x_2$	2	2	3	1	0	0	10
$x_3$	4	4	3	3	4	2	00
$x_4$	4	5	5	4	3	3	00
$x_5$	5	2	1	2	0	0	11
$x_6$	5	6	5	4	5	3	00

#### 3.2 断点数目分析

对数据属性进行离散化的目标就是在保持原数据信息的前提下简化断点区间数, 断点数目越少, 说明算法对数据类别信息的完整性保持的越好。在电力系统大数据中, 断点数目越少说明需要合并的数据属性越少, 对原数据的破坏性越小。对于上述实验数据使用 CAIM 算法进行离散化处理, CAIM 算法是一种全局静态的、自上而下的有监督离散化算法, 该算法的优点是以实现属性相关度最大化和最少断点数为目标, 因此将本文提出的 ERBIA 算法与 CAIM 算法在断点数目方面进行比较。表 4 是两种离散化算法在断点数目方面的比较。由表 4 可以看出, ERBIA 算法的断点数显著减少。

表 4 两种算法在断点数目方面的比较

Table 4 Two kinds of algorithm comparing in breakpoint number

算法	断点数目					
	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
CAIM	91	105	217	335	372	312
ERBIA	6	4	3	5	6	7

#### 3.3 加速比

本实验采用两台 DELL PowerEdge R720 服务器。在每台服务器上安装 VMware ESXI 5.5 虚拟机软件。此外, 本实验 JDK 采用 JDK1.6 版本, 程序集成开发环境采用 Eclipse-SDK-4.2.2, Hadoop 版本是 2.4.0。

加速比定义为同一个任务在单处理器系统和并行处理器系统中运行时间的比率, 用来衡量程序并行化的性能和效果, 加速比也可以定义为在单机中运行的时间与在集群中运行的时间的比率。本文提供的测试数据集为某风电机组的运行监测数据, 其中包含若干监测属性, 数据大小为 2 G, 分别在节点数为 2, 4, 6, 8 的不同集群上进行加速比试验, 得到的实验数据如表 5 所示。

表 5 不同节点集群的加速比

Table 5 Speedup on different nodes of the cluster

节点数	2	4	6	8
加速比	1.57	3.32	4.17	5.08

本文对数据进行加速比测试的主要目的是证明 ERBIA 算法在大数据并行处理环境下有较好的表现。从实验结果可以看出,随着节点数目的增加,算法运行时间明显减少,算法的运算速度得到提高,本文提出的算法获得了良好的加速比。

#### 4 结语

传统的实体识别算法只能对简单的命名等关系进行识别,而随着电力大数据时代的来临,对电力大数据中关系复杂的数据属性进行实体识别的问题已经迫在眉睫<sup>[13-15]</sup>。本文提出的 ERBIA 算法,针对现有实体识别算法的不足,利用离散化方法的优点,提出了一种基于信息准确率的离散化方案,并提出了一种改进的离散评价指标对算法进行评价,最后在 Hadoop 集群上进行了实验。实验证明了本文算法的正确性以及在离散断点数目和加速比方面的优势。下一步的工作重点是研究对大数据集冗余和相关分析的问题<sup>[16]</sup>,以便更好地对大数据集进行预处理,从而为电网中的最终决策提供支持。

#### 参考文献

- [1] 李明达, 王宏志, 张佳程, 等. PEIF: 基于并行机群的大数据实体识别算法[J]. 计算机研究与发展, 2013, 50(增刊 1): 211-220.  
LI Mingda, WANG Hongzhi, ZHANG Jiacheng, et al. PEIF: parallel entity resolution on big data[J]. Journal of Computer Research and Development, 2013, 50(S1): 211-220.
- [2] 寇月, 申德荣, 刘恒, 等. 异构网络中关联实体识别模型及增量式验证算法研究[J]. 计算机学报, 2013, 36(10): 2096-2108.  
KOU Yue, SHEN Derong, LIU Heng, et al. Research on related entity identification model and incremental verification algorithm for heterogeneous networks[J]. Chinese Journal of Computers, 2013, 36(10): 2096-2108.
- [3] 覃雄派, 王会举, 杜小勇, 等. 大数据分析——RDBMS 与 MapReduce 的竞争与共生[J]. 软件学报, 2012, 23(1): 32-45.  
QIN Xiongpai, WANG Huiju, DU Xiaoyuan, et al. Big data analysis-competition and symbiosis of RDBMs and Mapreduce[J]. Journal of Software, 2012, 23(1): 32-45.
- [4] 李玲娟, 张敏. 云计算环境下关联规则挖掘算法的研究[J]. 计算机技术与发展, 2011, 21(2): 43-46, 50.  
LI Lingjuan, ZHANG Min. Research on algorithm of mining association rule under cloud computing environment[J]. Computer Technology and Development, 2011, 21(2): 43-46, 50.
- [5] LI L, WANG H, GAO H, et al. EIF: a framework of effective entity identification[J]. Lecture Notes in Computer Science, 2010: 717-728.
- [6] 曲朝阳, 朱莉, 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 Zhaoyang, ZHU Li, ZHANG Shilin. Data processing of Hadoop-based wide area measurement system[J]. Automation of Electric Power Systems, 2013, 37(4): 92-97.
- [7] 王宏志, 樊文飞. 复杂数据上的实体识别技术研究[J]. 计算机学报, 2011, 34(10): 1843-1852.  
WANG Hongzhi, FAN Wenfei. Object identification on complex data: a survey[J]. Chinese Journal of Computers, 2011, 34(10): 1843-1852.
- [8] 董雷, 张昭, 蒲天骄, 等. 基于 PMU 量测信息的电压稳定在线评估指标[J]. 电力系统保护与控制, 2015, 43(11): 1-6.  
DONG Lei, ZHANG Zhao, PU Tianjiao, et al. An online voltage stability index based on measuring information of phasor measurement unit[J]. Power System Protection and Control, 2015, 43(11): 1-6.
- [9] 周国亮, 朱永利, 王桂兰, 等. 实时大数据处理技术在状态监测领域中的应用[J]. 电工技术学报, 2014, 29(增刊 1): 432-437.  
ZHOU Guoliang, ZHU Yongli, WANG Guilan, et al. Real-time big data processing technology application in the field of state monitoring[J]. Transactions of China Electrotechnical Society, 2014, 29(S1): 432-437.
- [10] 孙英娟. 基于粗糙集的分类方法研究[D]. 长春: 吉林大学, 2011.  
SUN Yingjuan. Research on classification methods based on rough set[D]. Changchun: Jilin University, 2011.
- [11] 曲朝阳, 陈帅, 杨帆, 等. 基于云计算技术的电力大数据预处理属性约简方法[J]. 电力系统自动化, 2014, 38(8): 67-71.  
QU Zhaoyang, CHEN Shuai, YANG Fan, et al. An attribute reducing method for electric power big data

- processing based on cloud computing technology[J]. Automation of Electric Power Systems, 2014, 38(8): 67-71.
- [12] 李辉, 胡姚刚, 唐显虎, 等. 并网风电机组在线运行状态评估方法[J]. 中国电机工程学报, 2010, 30(33): 103-109.
- LI Hui, HU Yaogang, TANG Xianhu, et al. Method for on-line operating conditions assessment for a grid-connected wind turbine generator system[J]. Proceedings of the CSEE, 2010, 30(33): 103-109.
- [13] 李俊刚, 张爱民, 张杭, 等. 广域保护系统数据网络可靠性评估[J]. 电工技术学报, 2015, 30(12): 344-350.
- LI Jungang, ZHANG Aimin, ZHANG Hang, et al. Reliability evaluation of the wide area protect system[J]. Transactions of China Electrotechnical Society, 2014, 2015, 30(12): 344-350.
- [14] 刘巍, 黄墨, 李鹏, 等. 面向智能配电网的大数据统一支撑平台体系与构架[J]. 电工技术学报, 2014, 29(增刊 1): 486-491.
- LIU Wei, HUANG Zhao, LI Peng, et al. Summary about system and framework of unified supporting platform of big data for smart distribution grid[J]. Transactions of China Electrotechnical Society, 2014, 29(S1): 486-491.
- [15] 王茂海, 赵玉江, 齐霞, 等. 电网实际运行环境中相量测量装置性能在线评价方法[J]. 电力系统保护与控制, 2015, 43(6): 86-92.
- WANG Maohai, ZHAO Yujiang, QI Xia, et al. Proposal of index for PMU field performance evaluation in actual operating environment[J]. Power System Protection and Control, 2015, 43(6): 86-92.
- [16] 金小明, 吴鸿亮, 周保荣, 等. 电网规划运行数据库与集成管理平台的设计与实现[J]. 电力系统保护与控制, 2015, 43(15): 126-131.
- JIN Xiaoming, WU Hongliang, ZHOU Baorong, et al. Design and implementation of integrated database management platform for power grid planning and operation[J]. Power System Protection and Control, 2015, 43(15): 126-131.

---

收稿日期: 2015-11-25; 修回日期: 2016-01-19

作者简介:

齐俊(1989-), 男, 通信作者, 硕士研究生, 研究方向为智能电网信息工程; E-mail: qij0427@163.com

曲朝阳(1964-), 男, 博士, 教授, 主要研究方向为电力信息化、计算机网络技术;

娄建楼(1972-), 男, 副教授, 硕导, 主要研究方向为云计算、物联网应用、大数据处理等。

(编辑 魏小丽)