

DOI: 10.16516/j.gedi.issn2095-8676.2018.01.005

基于特征聚类分析的大规模发电数据异常辨识方法

李智勇¹, 周剑¹, 俞蕙¹, 郭少青²

(1. 中国南方电网有限责任公司电力调度控制中心, 广州 510623; 2. 北京清大科越股份有限公司, 北京 100084)

摘要: 目前电厂发电量数据集抄已在各电网公司得到了广泛应用。针对电量集抄中首要解决的异常数据辨识问题, 深入分析了不同类型电源的自身特性, 基于特征聚类分析技术, 提出了不同类型电源日发电量异常的辨识条件, 由此构建了基于特征聚类分析的大规模发电数据异常辨识方法, 最后基于某省网的实际数据, 验证了所提出方法的有效性。

关键词: 发电量集抄; 特征聚类分析; 异常数据辨识; 电源特性

中图分类号: TM715

文献标志码: A

文章编号: 2095-8676(2018)01-0035-05

An Abnormal Data Identification Method of Large-scale Generation Data Based on Cluster Analysis

LI Zhiyong¹, ZHOU Jian¹, YU Hui¹, GUO Shaoqing²

(1. CSG Power Dispatching & Control Center, Guangzhou 510623, China;

2. Beijing QU Creative Technology Co., Ltd., Beijing 100084, China)

Abstract: Centralized collection technology of generation data is widely used in many power grid companies at present. To solve the abnormal data identification problem, this paper analyzed the different type of power generation. Then an abnormal data identification method of large-scale generation data based on cluster analysis was proposed. At last, the effectiveness of the proposed method was verified based on actual data from a province grid company.

Key words: centralized collection technology; cluster analysis; abnormal data identification; characteristics of generation

近年来, 电量集抄已成为各大电网企业电量采集的重要方式。然而受通信不稳定等因素影响, 电量集抄数据中可能存在异常数据问题。随着集抄技术的快速发展, 如何自动高效的辨识异常数据, 已成为集抄技术推广应用中面临的重要问题。

异常数据辨识的基本原理在于依据数据自身规律, 定位数据序列中的异常数值。在电力系统生产运行的众多技术领域均需要解决异常数据辨识问题, 然而由于数据自身规律差异, 使得不同领域不同类型的异常数据辨识其方法并不一致^[1-4]。能量管理系统(Energy Management System, 简称 EMS 系统)中主要采用状态估计方法来分析各状态量的

数据异常^[1]。在负荷预测业务中, 主要通过负荷数据与气温、节假日等数据的相关性分析, 来对预测或量测结果校验, 提出其中的不良数据^[2-4]。在设备状态监测领域, 则基于设备运行状态的特征分析, 通过分析同类设备采集量之间的数据关系, 来实现对设备状态监测量的异常辨识^[5]。文献[6]介绍了用电类型数据自动辨识模拟机的原理和实践情况, 用电数据异常辨识一方面需要从时间角度出发分析数据时序关系, 另一方面需要分析负荷、气象等因素的相关性, 通过综合分析得到辨识结果。文献[7-9]则从数据特征分析方法出发, 将基于 B 样条小波分析方法、流式计算方法、数据挖掘方法等新的数据分析方法引入异常辨识, 以提升数据辨识效率。

本文研究的重点是发电企业日发电量数据异常辨识问题。首先基于日发电量数据的自身特征, 运

收稿日期: 2017-05-01 修回日期: 2017-09-05

基金项目: 南方电网技改项目“电力系统驾驶舱-Web 展示服务模块建设”(000000GS61160041)

用特征聚类的分析方法,研究不同类型电源的运行特征,提出不同类型电源不良数据的辨识条件;接着基于该判别条件构建了基于特征聚类分析的大规模发电数据异常辨识方法;最后以南方电网某省区的实际数据,介绍了本文所提出方法的有效性。

1 电力系统发电数据的聚类特性

1.1 特征聚类分析

所谓特征聚类分析,是指在发电数据辨识过程中,从发电数据自身特性出发,根据不同类型电源的运行特征将具有同一类运行规律的电源汇总成一个类型,提出其特征判定条件的分析方法。

不同类型电源运行特征不同,根据其在时间尺度上的差别,发电数据辨识特征聚类分析可以分为纵向聚类和横向聚类两个方面。

纵向聚类分析,是指通过某一发电资源历史数据变化规律得到的聚类特征;横向聚类分析,是指通过同一类型发电资源在同一时间断面上的相互关系特征分析得到的聚类特征。不同类型电源具有不同的聚类特征,可能同时具有横向和纵向聚类特征,也可能仅有横向或纵向聚类特征。

1.2 传统电源的纵向聚类特性

本文所指的传统电源包括:大型火电、水电和核电。尽管上述电源在生产流程等并不一致,然而从厂用电率上上述电源具有纵向一致性,表现在:

- 1) 不同电厂的厂用电率分布有所不同。
- 2) 同一电厂不同开机方式下厂用电率分布不同。
- 3) 同一电厂相同开机方式下厂用电率分布规律较为显著。

如图1所示,为用股价图表示的单机容量600 MW,机组数均为2台的三个火电厂在不同开机情况下的厂用电率分布。A1、A2、B1、B2、C1、C2分别表示火电厂A、B、C开机数量为1台或2台的情况。股价图的最高、最低、开盘、收盘四个指标分别代表该开机方式下电厂全年统计中最高厂用电率、最低厂用电率和以厂用电率全年统计足最大分布概率为中心上下各推40%概率情况的上下限。

观察图1,不难发现由于装机不同、锅炉特性不同,开机方式不同,厂用电率将有所不同,但是在同一电厂相同开机方式下,其厂用电率分布较为集中,据统计其数据分布一般不超过0.2%^[10],则

上述偏差对于一个装机1.2 GW的火电厂,可近似折算电量40 MWh,占其发电量的0.2%。该数据误差在发电数据辨识中完全满足要求。

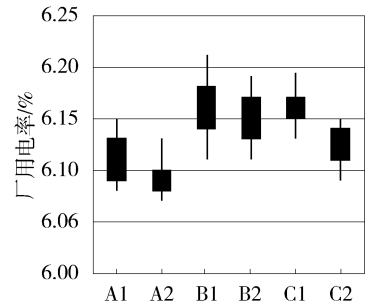


图1 火电厂用电率

Fig. 1 Power utilization rate of thermal power

而大型水电、核电也具有相同的规律特征,这里不再赘述。据此可以统计每一个传统电源不同开机情况下的厂用电率分布,在通过人工分析可以认定数据准确的前提下,假设某电厂*i*日厂用电率最大、最小分别为 $\gamma_{H_i}^{\text{Max}}$ 、 $\gamma_{H_i}^{\text{Min}}$,则对待辨识*j*日该电厂的辨识判据为:

$$\gamma_{H_i}^{\text{Min}} \leq \gamma_{H_i}^j \leq \gamma_{H_i}^{\text{Max}} \quad (1)$$

式中: $\gamma_{H_i}^j$ 为待辨识日该电厂的厂用电率,其计算公式为:

$$\gamma_{H_i}^j = \frac{W_{F,H_i}^j - W_{S,H_i}^j}{W_{F,H_i}^j} \times 100\% \quad (2)$$

式中: W_{F,H_i}^j 为该电厂的发电量; W_{S,H_i}^j 为该电厂的上网电量。

需要特别说明的是,由于大型水电机组在一日内存在多次启停,难以严格给出其开机数量情况,故可直接对其不同发电量下的厂用电率进行统计,其结果是一致的。

1.3 新能源的横向聚类特性

本文所指的新能源包括:风电、光伏、小水电为代表的电源类型。与传统电源相比,其发电量受自然条件影响很大,来风、降水、气温等都会对上述电源的发电量产生较大影响,导致其厂用电率指标也会发生较大波动,难以通过纵向聚类特性进行发电量数据辨识。而从横向来说,由于同一地区气候条件相似,则在同一地区的风电、光伏、小水电其电量往往呈现高度的相似性。

如图2所示,为位于某省的三个风电场某年1月份逐日利用小时数变化趋势图。可以发现虚线表

示的风电场 A 和实线表示的风电场 B 逐日利用小时数非常接近, 而尽管处于同一个省, 点画线表示的风电场 C 逐日利用小时数与风电场 A 和风电场 B 相差较大。其原因在于风电场 A 和风电场 B 空间位置比较接近, 处于同一“风带”, 而风电场 C 则相距较远。

如果将各风电场逐日利用小时数作为数据系列, 则风电场 A 与风电场 B 的相关系数为 0.87, 而风电场 A 与风电场 C 的相关系数仅有 0.35。

同一地区的光伏、同一流域的小水电均满足该特性。据此可通过对历史数据分析, 得到各风电场、光伏电站、小水电站出力的相关系数, 人工给定其相关系数一致性门槛值, 若待辨识日的风电场、光伏电站、小水电不满足该相关系数门槛值要求, 则认为该数据有误, 判别公式可表示为:

$$\gamma(\hat{W}) \geq \gamma_{\text{set}} \quad (3)$$

式中: \hat{W} 为考虑待辨识日新能源发电量数据后的发电量数据系列; $\gamma(\hat{W})$ 为取两个数据系统的相关系数; γ_{set} 为给定相关系数门槛值。

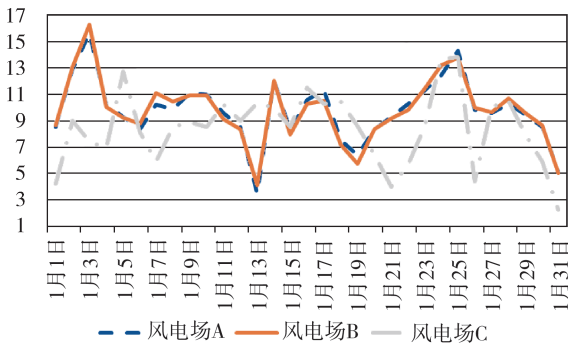


图2 风电场利用小时数

Fig. 2 Wind plant utilization time

2 基于特征聚类分析的大规模发电数据异常辨识方法

2.1 实施流程

基于上述分析, 本文所提出的基于特征聚类分析的大规模发电数据异常辨识方法流程如图 3 所示。其主要流程包括如下五个步骤:

- 1) 统计各电源历史数据, 提取其分布规律。
- 2) 统计待辨识日电源发电量数据。
- 3) 数据合理性初步校验。
- 4) 基于特征聚类分析的数据校验。

5) 给出异常数据辨识结果。

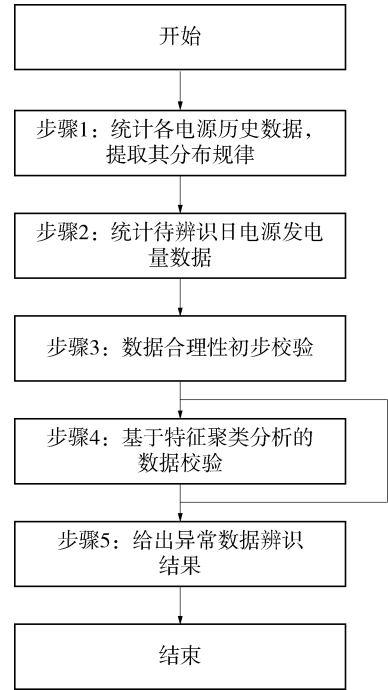


图3 实施流程图

Fig. 3 Implementation flow chart

2.2 历史数据分析

题对大型火电、水电、核电等传统电源, 获取历史逐日发电量、上网电量数据。要求该数据均通过人工检验, 确保数据质量。设某传统电源 i 第 j 日的发电量和上网电量分别为 $W_{C,F}^{i,j}$ 、 $W_{C,S}^{i,j}$, 则上述发电量、上网电量数据系列分别为 $\{W_{C,F}^{i,1}, W_{C,F}^{i,2}, \dots, W_{C,F}^{i,j}, \dots, W_{C,F}^{i,NT}\}$ 、 $\{W_{C,S}^{i,1}, W_{C,S}^{i,2}, \dots, W_{C,S}^{i,j}, \dots, W_{C,S}^{i,NT}\}$ 。据此可以计算得到逐日厂用电率 $\{\gamma_C^{i,1}, \gamma_C^{i,2}, \dots, \gamma_C^{i,j}, \dots\}$ 。其中厂用电率满足式(2)计算要求。对上述不同日的厂用电率分类, 对火电厂、核电厂根据其开机天数分类, 对水电站根据其日发电量分类, 统计每一类的厂用电率分布情况, 人工给定厂用电率变化范围上下限 $\gamma_C^{i,\max}$ 、 $\gamma_C^{i,\min}$ 。

对于风电、光伏、小水电等新能源, 获取其历史逐日发电量。同样对数据质量要求满足人工校验的基本要求。设某新能源 i 第 j 日的发电量为 $W_{X,F}^{i,j}$, 则对任意两个新能源 i 、 $i+1$, 其发电量序列分别为 $\{W_{X,F}^{i,1}, W_{X,F}^{i,2}, \dots, W_{X,F}^{i,j}, \dots, W_{X,F}^{i,NT}\}$ 、 $\{W_{X,F}^{i+1,1}, W_{X,F}^{i+1,2}, \dots, W_{X,F}^{i+1,j}, \dots, W_{X,F}^{i+1,NT}\}$, 可以计算其相关系数 $\gamma_X^{i,i+1}$ 。统计所有新能源电站的相关系数, 人工给定门槛值 $\gamma_{X,\text{set}}$, 规定两两之间相关系数均超过门槛值的新能

源电站为相关电站群。

2.3 获取待辨识日数据

从计量自动化系统中抽取待辨识日每个电站的运行数据,不同类型电站抽取数据不同,如表1所示。

表1 抽取数据内容
Tab. 1 Extracted data

类型	抽取数据
火电	开机台数、单机容量、发电量、上网电量
水电	装机容量、发电量、上网电量
核电	开机台数、单机容量、发电量、上网电量
风电	装机容量、发电量
光伏	装机容量、发电量
小水电	装机容量、发电量

2.4 初步校验

初步校验的目的在于对上述数据的合理性做直观判断,其主要方式是计算各电源的利用小时数,判断明显不合理的异常数据。

各电源的利用小时数可由其发电量和装机容量计算得到。当利用小时数为负值或超过24 h,则上述数据明显异常。

若初步校验即发现发电量数据存在异常,则直接将上述异常结果输出,执行步骤5;否则转入基于特征聚类分析的数据校验,执行步骤4。

2.5 特征聚类分析

对传统电源待预测日的厂用电率可由其发电量和上网电量,参考公式(2)计算得到。对照该开机或发电量下的厂用电率变化范围上下限,若该日厂用电率超过该范围限值,则说明该日数据异常。

对新能源电站,将相关电站群的待辨识日发电量数据汇入其发电量数据序列,计算其扩展待辨识日发电量数据后的新序列间相关系数。若相关电站群中存在某电站待辨识日内发电量扩展后导致扩展序列与其他电站扩展序列相关系数低于给定阈值,则说明该电站发电量数据为异常数据。

3 算例分析

3.1 基础数据

以南方电网某省实际数据为基础构造算例验证本文所提出方法的有效性。这里将选取1个发电厂

和1个包含3个风电场的相关电源群作为对象进行分析。上述电源的基本信息如表2所示。

表2 电厂基础数据
Tab. 2 Generation plants basic data

名称	单机容量/MW	总装机/MW
火电厂 A	600	1 200.0
风电场 A	—	320.0
风电场 B	—	112.5
风电场 C	—	65.5

3.2 历史数据分析

对火电厂 A,选取了经核准的历史30 d发电量、上网电量数据,如图4所示,为其厂用电率和开机数变化趋势图。

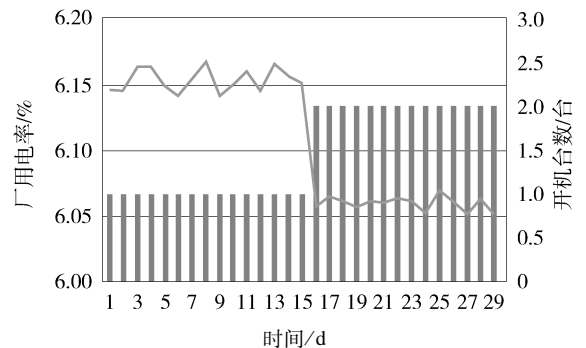


图4 火电厂 A 历史数据

Fig. 4 History data of power plant A

在开机数为1台时,其厂用电率变化范围为 $[6.14\%, 6.17\%]$ 。在开机数为2台时,其厂用电率变化范围为 $[6.05\%, 6.07\%]$ 。据此,规定该电厂厂用电率辨识范围,开机台数为1台时,厂用电率须处于 $[6.13\%, 6.18\%]$,开机台数为2台时,厂用电率须处于 $[6.04\%, 6.08\%]$ 。上述门限值适度放大主要是考虑所选取历史数据量较少。

对三个风电场,其利用小时趋势图如图5所示。三个风电场之间相关系数为: $\gamma(A, B) = 0.95$, $\gamma(A, C) = 0.94$, $\gamma(B, C) = 0.91$ 。设定相关性系数阈值为0.90。

3.3 待预测日辨识

待预测日基础数据如下表3所示。火电厂 A 的厂用电率为0.48%,明显超出厂用电率阈值范围,该数据存在异常。而经过校验,在将待辨识日风电场数据补充进历史数据序列后, $\gamma(\tilde{A}, \tilde{B}) = 0.96$,

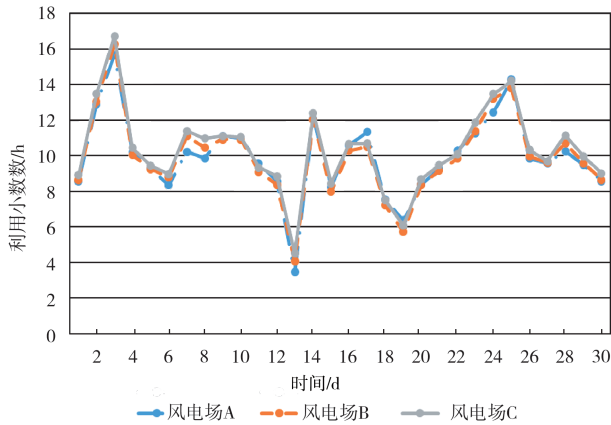


图 5 风电场逐日利用小时数

Fig. 5 Wind plant utilization time

$\gamma(\tilde{A}, \tilde{C}) = 0.94$, $\gamma(\tilde{B}, \tilde{C}) = 0.90$, 说明上述风电场发电量数据合理。

表 3 待辨识日数据

Tab. 3 Data of identification day

名称	开机数/台	发电量/MWh	上网电量/GWh
火电厂 A	2	18 968	18.876
风电场 A	—	120.4	—
风电场 B	—	32.8	—
风电场 C	—	25.7	—

4 结论

针对大规模发电量集抄后的异常数据辨识问题, 本文结合电力系统发电数据异常辨识的客观需要, 分析了传统电源的纵向聚类特征和新能源的横向聚类特征, 提出了适应不同电源的异常判据。基于该特征聚类分析的发电数据异常辨识方法在南方电网得到了广泛应用, 大幅提升数据质量。

参考文献:

[1] 叶锋, 何桦, 顾全等. EMS 中负荷预测不良数据的辨识与修正 [J]. 电力系统自动化, 2006, 30(15): 85-88.
YE F, HE H, GU Q, et al. Bad data identification and correction for load forecasting in energy management system [J]. Automation of Electric Power Systems, 2006, 30(15): 85-88.

[2] 王珂, 王天秀. 基于二次聚类的电力负荷异常数据辨识 [J]. 电气技术, 2014, 6(11): 1-3 + 17.
WANG K, WANG T X. Identification of abnormal power load data based on two times clustering [J]. Electrical Engineering,

2014, 6(11): 1-3 + 17.

[3] 陈亮, 文福拴, 童述林, 等. 基于密度估计的异常电力负荷数据辨识与修正 [J]. 华南理工大学学报(自然科学版), 2012, 40(2): 124-129.
CHEN L, WEN F S, TONG S L, et al. Identification and correction of abnormal electric load data based on density evaluation [J]. Journal of South China University of Technology (Natural Science Edition), 2012, 40(2): 124-129.

[4] 肖白, 徐潇, 宋坤, 等. 空间电力负荷预测中异常数据的辨识与处理 [J]. 东北电力大学学报, 2013, 33(1/2): 45-50.
XIAO B, XU X, SONG K, et al. Abnormal data identification and treatment in spatial electric load forecasting [J]. Journal of Northeast Dianli University, 2013, 33(1/2): 45-50.

[5] 王德文, 杨力平. 智能电网大数据流式处理方法与状态监测异常检测 [J]. 电力系统自动化, 2016, 40(14): 122-128.
WANG D W, YANG L P. Stream processing method and condition monitoring anomaly detection for big data in smart grid [J]. Automation of Electric Power Systems, 2016, 40(14): 122-128.

[6] 王桂兰, 周国亮, 赵洪山, 等. 大规模用电数据流的快速聚类 and 异常检测技术 [J]. 电力系统自动化, 2016, 40(24): 27-33.
WANG G L, ZHOU G L, ZHAO H S, et al. Fast clustering and anomaly detection technique for large-scale power data stream [J]. Automation of Electric Power Systems, 2016, 40(24): 27-33.

[7] 唐艺灵, 杨荣芳, 孙艳英, 等. 基于 B 样条小波的异常数据辨识 [J]. 指挥控制与仿真, 2007, 29(5): 73-77.
TANG Y L, YANG R F, SUN Y Y, et al. Identifying the exceptional data based on B-spline wavelet [J]. Command Control & Simulation, 2007, 29(5): 73-77.

[8] 马亚南. 电力系统异常数据的检测辨识与修正探究 [J]. 中国电业(技术版), 2014, 14(4): 22-25.
MA Y N. Study on detecting and identifying of abnormal data in power system [J]. China Electric Power (Technology Edition), 2014, 14(4): 22-25.

[9] 杨国强. 基于数据挖掘的电力系统异常数据辨识与调整 [J]. 应用能源技术, 2011, 12(1): 25-28.
YANG G Q. Power system detection and adjustment of bad data based on data mining [J]. Applied Energy Technology, 2011, 12(1): 25-28.

[10] 黄湘. 我国火电厂厂用电现状及节电途径 [J]. 中国电力, 2006, 39(9): 7-11.
HUANG X. Present status of service power and electricity saving method of coal-fired power plant in China [J]. Electric Power, 2006, 39(9): 7-11.

- tistics, 2007, 89(1): 75-87.
- [4] SIOSHANSI F P, PFAFFENBERGER W. Electricity market reform: an international perspective [M]. The Netherlands: Elsevier Science, 2006.
- [5] BORENSTEIN S, BUSHNELL J. The US electricity industry after 20 years of restructuring [J]. Annu. Rev. Econ., 2015, 7(1): 437-463.
- [6] GEDDES R R. A historical perspective on electric utility regulation [J]. Regulation, 1992(15): 75.
- [7] GÓMEZ E A, CONEJO A J, CANIZARES C. Electric energy systems: analysis and operation [M]. Newyork: CRC Press, 2009.
- [8] JOSKOW P L. Incentive regulation in theory and practice: electricity distribution and transmission networks [C]//Economic Regulation and Its Reform: What Have We Learned [M]. Chicago: University of Chicago Press, 2014: 291-344.
- [9] BURTRAW D, PALMER K, PAUL A, et al. Reliability in the US electricity industry under new environmental regulations [J]. Energy policy, 2013(62): 1078-1091.
- [10] TER-MARTIROSYAN A, KWOKA J. Incentive regulation, service quality, and standards in US electricity distribution [J]. Journal of Regulatory Economics, 2010, 38(3): 258-273.
- [11] CHAO H P, HUNTINGTON H G. Designing competitive elec-

tricity markets [M]. Germany: Springer Science & Business Media, 2013.

作者简介:



胡际(通信作者)

1987-, 男, 广东广州人, 广州供电局有限公司财务部会计师, 南京大学经济学硕士, 主要从事电价理论与实务、电力系统经济学、工程财务管理研究 (e-mail) leytonhu@sina.com。

HU J



陈雯

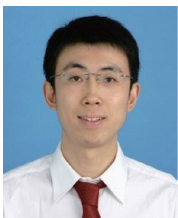
1988-, 女, 江苏兴化人, 中国能源建设集团广东省电力设计研究院有限公司经济师, 南京大学经济学硕士, 主要从事能源经济、电力市场研究, (e-mail) chenwen@gedi.com.cn。

CHEN W

(责任编辑 郑文棠)

(下接第 39 页 Continued from Page 39)

作者简介:



李智勇(通信作者)

1982-, 男, 江苏徐州人, 高级工程师, 博士, 主要从事电力系统运行控制研究 (e-mail) lizhiyong@csg.cn。

LI Z Y



周剑

1982-, 男, 江苏徐州人, 高级工程师, 博士, 主要从事电力系统运行控制研究 (e-mail) lizhiyong@csg.cn。

ZHOU J



愈蕙

1986-, 女, 湖南湘潭人, 工程师, 硕士, 主要从事电力系统控制、电力市场交易管理及研究 (e-mail) yuhui@csg.cn。

YU H



郭少青

1963-, 男, 福建南安人, 北京清大科越股份有限公司总经理, 高级工程师, 长期从事智能电网调度、电力市场交易等应用领域的研究设计与管理工作 (e-mail) guoshaoqing@vip.sina.com。

GUO S Q

(责任编辑 郑文棠)