

能源互联网多源数据融合技术研究

黎琦¹, 张涛允²

(1. 上海电力大学 电气工程学院, 上海 200090;

2. 国网甘肃省电力公司 电力科学研究院, 甘肃 兰州 730070)

摘要: 随着我国能源互联网战略的推进, 边缘计算成为解决能源互联网数据处理的有效模式, 针对能源互联网的电力异构配用电数据的存储混乱与融合性差的问题, 提出能源互联网多源数据处理与融合技术。首先, 在边缘计算的基础上设计能源互联网多源数据处理与融合架构; 其次, 提出了智能配用电数据存储资源层负载均衡优化调度策略, 利用 Hadoop 平台实现了多源数据存储和处理; 然后, 通过主成分分析方法得到能源互联网异构数据的主要特征, 构建基于矛盾优化 DS 证据推理的多源数据融合模型, 在边缘计算框架的基础上对多源异构数据进行分组聚合; 最后, 在某配电网模拟系统进行实验分析, 分析结果验证了所提方法的有效性。

关键词: 能源互联网; 边缘计算; 多源数据处理; 云计算

中图分类号: TM933

文献标识码: A

DOI: 10.19457/j.1001-2095.dqed23053

Research on Multi-source Data Fusion Technology of Energy Internet

LI Qi¹, ZHANG Taoyun²

(1. College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China;

2. Institute of Electric Power Science, State Grid Gansu Electric Power Company, Lanzhou 730000, Gansu, China)

Abstract: With the advancement of China's energy internet strategy, edge calculation has become an effective mode to solve the data processing of energy internet. Aiming at the problems of chaotic storage and poor fusion of heterogeneous power distribution and power data in energy internet, the multi-source data processing and fusion technology of energy internet were proposed. Firstly, a multi-source data processing and fusion architecture for energy internet was designed based on edge calculation. Secondly, the optimal scheduling strategy of load balancing for data storage resource layer of intelligent power distribution was proposed, and multi-source data storage and processing were realized by using Hadoop platform. Then, through the principal component analysis method, the main characteristics of the energy internet heterogeneous data were obtained, and the multi-source data fusion model based on the contradiction optimization DS evidence reasoning was constructed, and the multi-source heterogeneous data was grouped and aggregated on the basis of the edge calculation framework. Finally, an experimental analysis was conducted in a distribution network simulation system, the analysis results prove the effectiveness of the proposed method.

Key words: energy internet; edge calculation; multi-source data processing; cloud computing

能源互联网是电力系统智能化的高级表现, 是电力系统各节点紧密联系的新型电力网络。能源互联网的建设可以加强供需联系、整合电力系统交易、提高电力供应的稳定性和实效性。

能源互联网的建设也为国家实现碳中和的目标提供助力^[1-2]。能源互联网属于“新基建”融合基础设施范畴, 智能化、网络化、信息化是其主要特点。能源互联网中应用的智能感知技术, 针

基金项目: 国家自然科学基金资助项目(51707112)

作者简介: 黎琦(1994—), 男, 硕士研究生, Email: 670092925@qq.com

对性地部署于电力系统的各个环节,更有利于相关参数的监视及环境信息的实时收集,可以有效地提高系统抗风险能力。能源互联网范畴如图1所示。

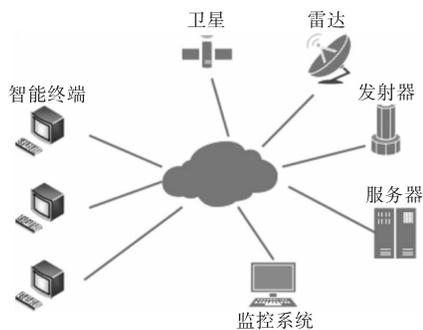


图1 能源互联网范畴图

Fig.1 Energy internet category

伴随着能源互联网战略的不断推进,智能化的感知终端设备产生的数据大量增加,且呈现种类繁多、多源、不确定等特性^[3]。能源互联网实现了云-边缘的计算模式,由于能源互联网建设的不断推进导致数据量、数据种类和数据复杂度急剧增加。能源互联网正常运行时产生的智能终端监测数据和电力系统运行数据等参数信息的传输产生的数据流为通信传输层带来了巨大挑战,也是导致能源互联网智能发展的挑战。如何利用边缘计算提高能源互联网智能化成为研究的重点。利用边缘计算去提高云计算的处理效果成为一种重要方式^[4-5]。但是复杂的多源异构数据的边缘化集成为能源互联网的高效边缘计算带来了新的挑战,因此,如何在边缘计算的基础上运用好云计算技术是提升和保证能源互联网智能化发展的研究重点^[6-7]。

近年来,随着人工智能、数据挖掘、云计算等技术的不断发展,也为能源互联网的发展提供了强大的推力和宝贵机遇。电力系统是一个高度耦合的复杂的非线性动力系统,因此对电力系统运行中产生的海量异构数据的高效处理标志着能源互联网的智能化水平。国内外专家学者关于能源互联网海量数据分析挖掘方法与多源数据处理与融合技术开展了众多有意义的研究,周杨珺等^[8]提出了四层架构的电力系统数据分析系统对电力系统数据进行分析;王伏亮等^[9]提出了一种基于边缘计算的中低压配电网多类型传感器接入技术,将边缘计算引入中低压配电网的数据融合中;Okay F Y等^[10]将云计算应用到了智能

电网中,并针对安全问题进行了阐述;马洲俊等^[11]提出了多源异构数据分析结果辅助配电网故障信息识别定位的策略,为有效挖掘配电网智能化数据应用化价值提供了研究借鉴;Barik Rabindra K等^[12]在云计算的基础上利用雾计算提高智能电网的信息处理能力;常利伟等^[13]在BP神经网络的基础上结合DS(Dempster-Shafter)证据推理理论,实现了多源数据融合,并有效评估了网络威胁状况。

因此,文中提出一种基于边缘计算的能源互联网多源数据处理与融合技术,设计充分考虑边缘计算的数据处理与融合架构。首先,提出基于云计算的能源互联网数据处理方法,提高数据的整合和处理能力;然后,通过构建基于冲突优化DS推理的多源数据融合模型,实现多源异构数据的融合处理。能源互联网多源数据融合技术可有效地实现电力系统运行设备参数、智能终端数据、外部环境参数数据等多源数据的储存和融合,为提高能源互联网智能化水平奠定基础。

1 能源互联网多源数据处理与融合架构

电力系统是由发电节点、输电系统、用户组成的。发电节点通过能源转换将化石能、风能、水能等一次能源转换成电能。然后通过输电系统输电线路,先升压后降压,将电能输送到用户处。输电系统还包括信息监测系统、通信系统、保护系统和调度系统等。而能源互联网是在此基础上将人工智能技术、通信技术、云计算等前沿技术的智能化电力系统,实现电力流全程的智能化控制^[14]。由于能源互联网内融合了大量的智能化设备,导致数据存在体量大、数据结构复杂、数据价值低、数据利用率低的特点。但是这些数据是能源互联网运行的忠实体现,通过深入挖掘可为能源互联网电力调度和系统调控提供数据支撑。

文中在边缘计算的基础上,结合云计算可有效提高能源互联网数据的处理能力,实现数据的储存、处理、分析和计算的全过程。充分发挥云计算的规模大、可靠性高、扩展性强的特点,以及边缘计算速度快、成本低的优势,从而实现能源互联网业务功能实时控制调整、终端扩展灵活的目标^[15-16]。通过对能源互联网产生的数据进行处

理变化,实现异构数据的融合,从而服务于能源互联网。

能源互联网多源数据处理与融合架构如图2所示。

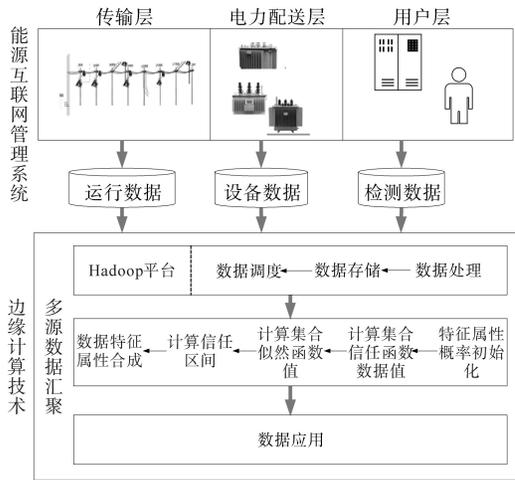


图2 多源数据处理与融合技术架构
Fig.2 The architecture of multi-source data processing and fusion technology

2 基于云计算的能源互联网多源数据存储和处理

能源互联网的海量数据根据数据的来源不同分为:电力系统数据、用户数据和其他数据。这些数据集成收集在平台数据库中,如图3所示。

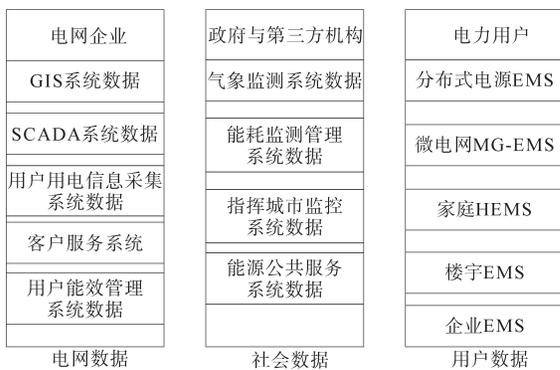


图3 能源互联网用电数据组成
Fig.3 Data components of smart power distribution and utilization

图3中,电力系统数据包括:数据采集与监视控制系统数据(supervisory control and data acquisition, SCADA)、用电信息数据等电力系统数据。用户数据包括:分布式电源管理系统(energy management system, EMS)、微电网电源管理系统(micro-power grid energy management system, MG-EMS)、家庭电源管理系统(home energy management

system, HEMS)、楼宇电源管理系统、企业电源管理系统(enterprise energy management system, EEMS)等用户数据;其他数据包括:天气监测数据、能耗数据、城市服务数据等相关信息。

2.1 能源互联网多源数据调度策略

利用电力企业构建的专用通信网络和5G通信网络作为连接纽带,将变电站自动化信息站连接。

以现有的配电自动化主站平台构建智能配用电的大数据资源层的主节点,由于企业内外网的关系,将主节点分为电力系统内部主节点(负责电力系统的资源管理和调用,简称主内节点)和外网主节点(负责电力系统的资源管理和调用,简称主外节点),并建立备用节点,提高系统的稳定性。主节点均配置有互为备用的冗余热备用节点,其他应用平台和资源作为从节点服从主节点的调度和分配,从而组成一个多源异构的能源互联网多源数据中心集群,其MapReduce的实现流程如图4所示。

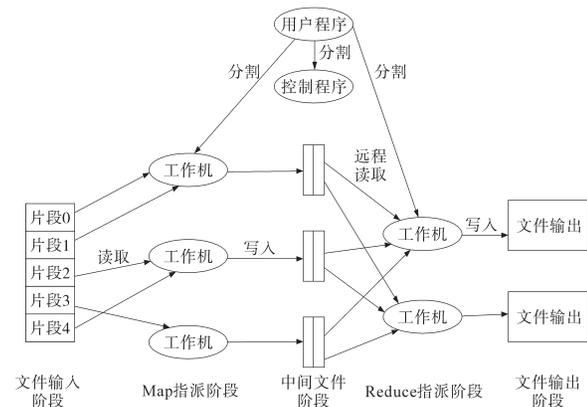


图4 MapReduce的实现流程
Fig.4 Implementation flow of MapReduce

为提高能源互联网多源数据中心集群的处理性能,提出了负载均衡优化调整策略进行动态、静态相结合的负荷调整。电力系统内部由主内节点负责进行内部的调度和管理,使用权重轮循-静态负载调节法,进行负载均衡优化,如图4所示。从节点节点代表的是权重系数,用圆圈表示,并进行电力系统内部的资源调度和使用;电力系统外部由主外节点进行外部的调度和管理,使用最快响应-动态负载调节法,进行复杂均衡优化。

图5为智能配用电数据存储资源层负载均衡优化调度策略示意图。图5中,从节点代表了该点的响应时间,并根据响应时间的长短进行外

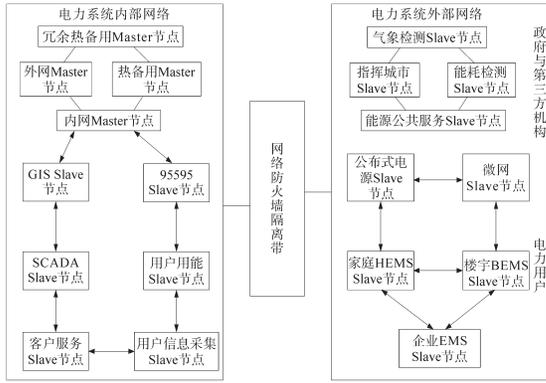


图5 智能配用电数据存储资源层负载均衡优化调度策略
Fig.5 The strategy of load balancing optimization scheduling for data storage resource layer in the intelligent power distribution

部系统资源的调用。

2.2 能源互联网多源数据预处理

由于能源互联网多源数据的特点,将数据分为两类:

1)结构化数据。是组织结构整齐的数据,可以直接进行数据加工的数据类型。

2)半结构化数据和非结构化数据。半结构化数据和非结构化数据需要对数据进行相应的预处理。最终将结构化和非结构化数据以可扩展标记语言(extensible markup language, XML)格式存入云计算节点中。XML非常适合网络快速传输,适用于超文本传输协议(hypertext transfer protocol, HTTP)网络传输协议,并支持应用程序接口(application programming interface, API)扩展。然后基于Hadoop平台通过数据库技术对能源互联网多源数据进行高效地分布式存储。

3 能源互联网多源数据融合模型

信息融合是通过对能源互联网多源数据处理与融合来实现的,只需要涉及极少简略的数据计算,关键数据计算处于信息融合后的阶段之中。文中在云计算的多源数据标准化处理的基础上,构建基于冲突优化DS推理的能源互联网多源数据融合模型,从海量多源异构的电力信息系统日志数据中通过剔除无效、过滤冗余、数据插补、整合数据等方式重新对各原始数据进行处理,从而进行异构数据集间的特征提取、特征匹配等操作,以便下一步进行大数据融合和决策分析。

D-S证据推理法(Dempster-Shafter, DS)是一种处理不确定性问题的人工智能推理方法,实现

了贝叶斯概率论的进一步改进,避免了先验概率的计算,能够很好地表示“不知道”,被广泛应用于各种领域的数据融合。但由于DS推理法处理冲突子集时,因组合规则中的归一化过程会出现违背不同数据源融合常理的结果,因此,应用主成分分析算法对数据进行降维提高多源数据融合的效果。

1)主成分分析算法可以建立 $m(m < n)$ 个新成分的目标,保留原有信息特征,提高主要特征的影响程度,成分可信度函数:

$$\tilde{s} = \sum_{k=1}^n \varepsilon_i AV_{i,t} \quad (1)$$

式中: \tilde{s} 反映了数据之间的矛盾程度; n 为最大维度; ε_i 为矛盾信息中的不同成分; i 为成分的分类; AV 为对应维度; t 为维度个数。

由于能源互联网用用电数据具有数据种类多、数据来源广、不确定等特点,因此将各数据源或监测终端数据信息抽象为特征属性子集,实现能源互联网多源异构数据的特征级融合。

2)能源互联网数据初始化, P 为能源互联网数据融合模型框架,则函数 $n: 2^P \rightarrow [0, 1]$ 如下要求:

$$\begin{cases} n(\varphi) = 0 \\ \sum_{A \subset U} n(A) = 1 \end{cases} \quad (2)$$

式中: $n(A)$ 为信任度取值; $n(\varphi) = 0$ 为数据集 A 的初始值, φ 为初始状态。

3)信任函数定义,公式如下:

$$Bel(A): 2^u \rightarrow [0, 1] \quad (3)$$

式中: $Bel(A)$ 为对数据集 A 的信任可信度; u 为所有可能假设的集合, 2^u 为其幂集合。

4)似然函数定义,公式如下:

$$\begin{cases} pl(A) = 1 - Bel(\bar{A}) = \sum_{B \subset U} n(B) - \sum_{B \subset \bar{A}} n(B) \\ \tilde{t}_{A \cap B} = \sum_{i=1}^n e_i BV_{i,t} \end{cases} \quad (4)$$

式中: $pl(A)$ 为似然函数的下限函数; $Bel(\bar{A})$ 为似然函数的上限函数; $n(B)$ 为融合数据集集中相交数据子集的概率; $\tilde{t}_{A \cap B}$ 为冲突成分可信度值。

5)根据信息函数对数据进行信任度计算。

$$\begin{cases} Bel(A): 2^u \rightarrow [0, 1] \\ Bel(A) = \sum_{B \subset A} n(B) \quad \forall A \subset U \end{cases} \quad (5)$$

6)计算数据融合的信任空间。根据能源互联网多源数据的信任函数和似然函数,计算实际

应用数据的信任值,可以表示为

$$\eta(A) = pl(A) - Bel(A) \quad (6)$$

式中: $\eta(A)$ 为集合 A 中不确定性特征属性度量。

7)多源异构数据特征属性合成。利用能源互联网框架中可分离 mass 函数进行不同类型的数据特征 $n_i(i=1,2,\dots,n)$ 融合。合成规则为

$$(n_1 \oplus n_2 \oplus \dots \oplus n_n)(A) = \frac{1}{s} \sum_{A_1 \cap A_2 \cap \dots \cap A_n = A} n_1(A_1) \cdot n_2(A_2) \cdot \dots \cdot n_n(A_n) \quad (7)$$

式中: s 为根据数据的不同特征实现数据融合的系数。

4 实验结果与分析

为验证文中所提的能源互联网多源数据处理和融合技术的有效性与可靠性,以电力系统实测数据为例,选取某省电力系统实际运行数据、终端监测数据、环境参数数据。利用多源数据处理结果,围绕能源互联网故障信息诊断与挖掘这一实际应用场景需求,进行数据融合。

多源数据处理实验结果值如表1所示。

表1 多源数据处理实验结果值

Tab.1 Multi-source data processing experimental results

编号	电流 (标么值)	电压 (标么值)	能量 (标么值)	功率因数 (标么值)	温度/ ℃	风速 (标么值)
1	0.44	0.03	-1.00	0.25	32	0.80
2	0.43	0.37	0.72	-0.49	33	0.40
3	0.49	-0.07	-0.75	-0.37	35	0.18
4	-0.70	-0.25	0.01	0.22	31	0.60
5	-0.81	-0.18	0.56	-0.80	30	0.90
6	0.81	-0.55	-0.32	0.07	32	0.70
7	-0.02	0.64	0.14	-0.72	31	0.37

根据能源互联网多源异构数据特征,寻找最佳数据融合点,变化趋势如图6所示。

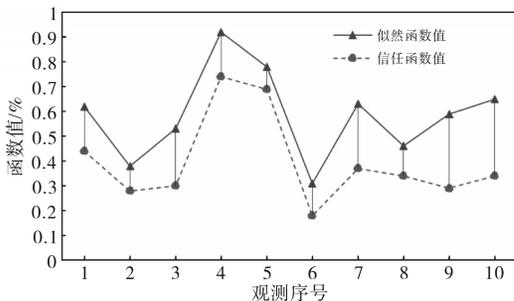


图6 数据融合过程中的信任区间变化

Fig.6 Change of confidence interval during data fusion

对电力系统内经云计算处理后的多源数据融合实验结果分别采用DS推理法、BP神经网络

(back propagation neural network, BPNN)与DS推理法结合(BPNN-DS)算法及文中多主成分分析(principal component analysis, PCA)与DS推理法结合(PCA-DS)算法进行信息处理和数据挖掘的有效性的实验对比,三种实验融合结果的算法对比结果如图7所示。

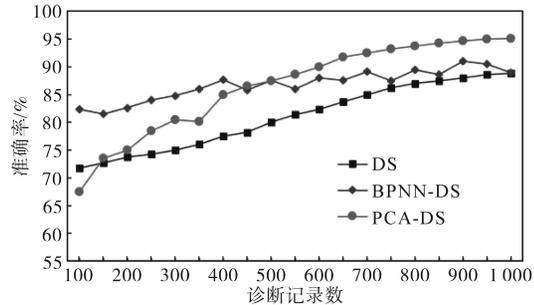


图7 算法对比图

Fig.7 Algorithm comparison diagram

实验表明,基于主成分分析的DS能源互联网数据融合模型从不同角度出发,结合云计算对能源互联网异构数据进行融合处理,提高了数据融合的水平。

5 结论

文中提出了能源互联网多源数据处理与融合技术,该技术的提出在一定程度上能源互联网的智能化水平,利用云计算和边缘计算为能源互联网数据分析提供了数据支撑和决策支持。通过深入分析能源互联网建设过程中出现的数据处理和储存问题,设计了基于边缘计算的能源互联网云计算处理与融合架构,利用边缘计算和云计算的整合储存和处理能力,有效地提高了电网数据的处理速度,并实现了能源互联网多源用电数据融合,提高了能源互联网在状态评估、辅助决策等等应用中的数据可靠性。

参考文献

[1] 刘凯诚,钟鸣,曾平良,等.考虑分布式可再生能源和储能智能配电网可靠性评估综述[J].电测与仪表,2021,58(7):1-11.
Liu Kaicheng, Zhong Ming, Zeng Pingliang, et al. A review of reliability evaluation of smart distribution network considering distributed renewable power and energy storage[J]. Electrical Measurement and Instrumentation, 2021, 58(7):1-11.

[2] Mk A, Rs B, Ro A, et al. On the trade-off between timeliness and accuracy for low voltage distribution system grid monitoring utilizing smart meter data[J]. International Journal of Electrical Power & Energy Systems, 2020, 121:106090.

- [3] 毛先胤,文屹,马晓红,等.基于电力大数据的多源异构数据融合技术研究与应用[J].电力大数据,2020,23(8):33-39. Mao Xianyin, Wen Yi, Ma Xiaohong, *et al.* Research and application of multi-source heterogeneous data fusion technology based on power big data[J]. Power System and Big Data, 2020, 23(8):33-39.
- [4] 威尔江,彭道刚,关欣蕾,等.智能配电自动化远方终端数据采集系统设计[J].电气传动,2018,48(7):60-65. Qi Erjiang, Peng Daogang, Guan Xinlei, *et al.* Design of remote terminal data acquisition system for intelligent distribution automation[J]. Electric Drive, 2018, 48(7):60-65.
- [5] 田鹏,秦贺,侯乃乾,等.基于边缘计算和广域同步测量技术的配电网故障定位方法[J].电气工程学报,2020,15(1):83-88. Tian Peng, Qin He, Hou Naigan, *et al.* Fault location method of distribution network based on edge computing and wide area synchronous measurement technology[J]. Journal of Electrical Engineering, 2020, 15(1):83-88.
- [6] 冯成.多源异构数据融合关键技术研究[D].北京:北京邮电大学,2020. Feng Cheng. Research on key technologies of multi-source heterogeneous data fusion[D]. Beijing: Beijing University of Posts and Telecommunications, 2020.
- [7] 盛万兴,方恒福,沈玉兰,等.考量量测时延时基于3种数据融合的配网状态估计[J].电力系统及其自动化学报,2019,31(12):108-115. Sheng Wanxing, Fang Hengfu, Shen Yulan, *et al.* Distribution network state estimation considering measurement delay based on three kinds of data fusion[J]. Proceedings of the CSU-EPSA, 2019, 31(12):108-115.
- [8] 周杨珺,梁朔,俞小勇,等.基于多源异构数据的配电网运行分析平台:系统架构与技术实现[J].南方电网技术,2018,12(8):59-64. Zhou Yangjun, Liang Shuo, Yu Xiaoyong, *et al.* Distribution network operation analysis platform based on multi-source heterogeneous data: system architecture and technical implementation[J]. Southern Power System Technology, 2018, 12(8):59-64.
- [9] 王伏亮,李澄,李春鹏,等.基于边缘计算的中低压配电网多类型传感器接入技术[J].自动化与仪表,2020,35(7):11-15. Wang Fuliang, Li Cheng, Li Chunpeng, *et al.* Multi-type sensor access technology for medium and low voltage distribution network based on edge computing [J]. Automation & Instrumentation, 2020, 35(7):11-15.
- [10] Okay F Y, Ozdemir S. A fog computing based smart grid model[C]//2016 International Symposium on Networks, Computers and Communications (ISNCC). IEEE, 2016.
- [11] 马洲俊,樊飞龙,王勇,等.基于多源异构数据的配电网故障信息挖掘与诊断[J].供用电,2018,35(8):31-39. Ma Zhoujun, Fan Feilong, Wang Yong, *et al.* Distribution network fault information mining and diagnosis based on multi-source heterogeneous data[J]. Distribution & Utilization, 2018, 35(8):31-39.
- [12] Barik R K, Gudey S K, Reddy G G, *et al.* Fog grid: leveraging fog computing for enhanced smart grid network[C]//2017 14th IEEE India Council International Conference (INDICON), 2017.
- [13] 常利伟,田晓雄,张宇青,等.基于多源异构数据融合的网络安全态势评估体系[J].智能系统学报,2021,16(1):38-47. Chang Liwei, Tian Xiaoxiong, Zhang Yuqing, *et al.* Based on data fusion of multi-source heterogeneous network security situation assessment system[J]. CAAI Transactions on Intelligent System, 2021, 16(1):38-47.
- [14] 黄欣,吴伟杰,张伊宁,等.电力系统多源异构数据混合式存储方式研究[J].机电工程技术,2021,50(1):77-79. Huang Xin, Wu Weijie, Zhang Yining, *et al.* Research on hybrid storage mode of multi-source heterogeneous data in power system[J]. Mechanical and Electrical Engineering Technology, 2021, 50(1):77-79.
- [15] 彭道刚,赵晨洋,关欣蕾,等.基于IEC61850的智能配电网主站级通信系统研究[J].电气传动,2018,48(12):44-48. Peng Daogang, Zhao Chenyang, Guan Xinlei, *et al.* Research on intelligent distribution network master station-level communication system based on IEC61850[J]. Electric Drive, 2018, 48(12):44-48.
- [16] 马立红,王海生,宁光涛,等.基于多源数据融合与分析的配电网可视化诊断系统的构建[J].电子技术与软件工程,2020(2):204-207. Ma Lihong, Wang Haisheng, Ning Guangtao, *et al.* Construction of visual diagnosis system of distribution network based on multi-source data fusion and analysis[J]. Electronic Technology and Software Engineering, 2020(2):204-207.

收稿日期:2021-02-03

修改稿日期:2021-04-14