

基于 K-邻近法的电网关键断面在线分布式发现方法

王彬¹, 郭文鑫¹, 刘文涛¹, 卢建刚¹, 向德军¹, 周哲民², 余志文¹

(1. 广东电网有限责任公司电力调度控制中心, 广东 广州 510600;

2. 北京清大高科系统控制有限公司, 北京 100084)

摘要: 随着可再生能源大规模接入电网, 电力系统正面临着越来越复杂的运行环境, 从而对电网在线操作的时间粒度提出了更高的要求。关键断面在线发现以及其极限传输容量计算是保证大电网在线安全运行的重要手段, 调度员通过调控关键断面控制电网安全、稳定运行。从数据驱动的角度出发, 对电网在线运行状态建立特征集合, 运用层次聚类法和分布式特征选择筛选出影响断面出现与否的关键特征, 随后运用二分类分布式的 K-邻近法(KNN)建立特征集合与断面出现与否的映射。算例分析表明, 相比于传统方法, 所提机器学习方法可以大大减少关键断面在线发现所需时间, 且关键断面预测精度达到工程应用需求。

关键词: 关键断面; K-邻近法; 分布式; 机器学习; 数据挖掘

Power system distributed key section detection online based on K nearest neighbor algorithm

WANG Bin¹, GUO Wenxin¹, LIU Wentao¹, LU Jiangang¹, XIANG Dejun¹, ZHOU Zhemin², YU Zhiwen¹

(1. Guangdong Power Grid Power Dispatching Control Center, Guangzhou 510600, China;

2. Qing Da Gao Ke System Control Company, Beijing 100084, China)

Abstract: The renewable energy such like wind energy and solar energy is rather uncertain and intermittent, thus it has raised a more strict requirement for the efficiency of operating. Key section automatic detection and total transfer capability calculation online is the key way to guarantee security of large power grid by operators, thus it has been put much more attention. From the perspective of data-driven, in this paper, feature set is built to describe the state, and then machine learning method is utilized to map the feature set to whether the section exists or not, where KNN method is to play as a classifier. Numerical tests show that the proposed machine learning method can reduce the time needed to discover key sections and have high performance in accuracy compared with traditional methods.

This work is supported by the Science and Technology Program of China Southern Power Grid: the Key Technology and Demonstration Application for Security Feature Selection and Knowledge Discovery in Complex Large-scale Power System Based on Big Data (No. GDKJ00000058).

Key words: key section; K nearest neighbor algorithm (KNN); distributed; machine learning; data mining

0 引言

近年来可再生能源产业迅猛发展, 其在电网能源中的占比逐渐加大。在可预见的将来, 可再生能源作为传统化石能源的替代品将在人类一次能源中占有更大的比重。然而, 以风能和太阳能为代表的可再生能源在环境友好的同时也给电网安全运行带

来了严峻的挑战^[1-2]。究其原因, 是因为可再生能源的出力不受人控制也较难预测, 具有很强的间歇性和不确定性。因此, 相比于传统以火电为一次能源的电网, 如今电网正面临更大的复杂性和不确定性。因此, 智能电网对电力系统薄弱环节发现、安全预警以及相应预警策略制定的时间粒度均提出了更高的要求^[3-5]。

电力系统在线关键断面自动发现是保证电网安全运行的重要手段^[6-7]。计算和控制关键断面极限传输容量, 一方面可以及时发现电网中存在的传输容量隐患, 另一方面可以在事故发生之前制定

基金项目: 南方电网科技项目(GDKJ00000058)“面向大数据的复杂大电网安全特征选择和知识发现的关键技术与示范应用”

相应的措施规避风险^[8]。传统的关键断面在线自动发现方法通常以建模穷举扫描或者简单的启发式方法来搜索关键断面^[6-7,9]，一方面建模穷举扫描或者简单的启发式搜索具有较大的算法复杂度，其对于大电网而言需要较长的计算时间，从而增加了在线运行的时间粒度；另一方面，由于电网运行方式具有一定的周期性，关键断面通常会重复出现，所以在运用建模穷举扫描或者启发式搜索时通常会对某些不常出现的关键断面进行重复搜索。为了提升关键断面的检测效率，本文运用一种基于数据的机器学习方法。

机器学习作为人工智能的重要分支，其价值正在从学术层面逐渐转向到应用层面。传统的机器学习是将模型与数据有机结合的方法，具体来说，机器学习的数学模型并不复杂，但是需要大量的数据作为支撑，从而从数据中发掘出最多的信息，更具有面向对象的特点^[10-13]，数据挖掘技术已经在电力系统的多个领域内得到了尝试与应用^[14-17]。电网的实时运行数据本身具有大量的数据量，而基于实际数据的仿真数据具有更大的体量，从而为数据驱动算法奠定了基础条件。本文尝试用 K-邻近法(KNN)预测关键断面是否出现。这一方法在包括电力系统在内的许多领域得到了广泛应用^[18-20]。首先，运用分层聚类法和分布式特征选择来选择描述电网运行状态的关键特征；接着，建立关键特征到各个断面的映射，用以判断各个断面是否出现。特征选择以及关键断面判断均以分布式方式进行。算例分析表明本文所提机器学习方法可以大大减少关键断面在线发现所需时间，且关键断面预测精度达到工程应用需求。

1 方法框架

本文所提基于 K-邻近法(KNN)的电网关键断面在线分布式发现方法的主要框架主要由以下几个部分组成：电网在线运行状态获取与蒙特卡洛仿真、特征聚类与分布式特征选择、基于分布式 KNN 的关键断面分类器，其整体架构示意图如图 1 所示。现将该算法的几个部分分别阐述如下。

1) 历史数据与运行状态获取

首先本计算系统从能量管理系统(Energy Management System, EMS)中获取电网在线实际运行状态，通过传统的关键断面在线发现方法可以发现一系列该运行状态下的关键断面^[6-7]。同时，在获取电力系统实际运行状态以后，运用蒙特卡洛模拟产生该状态邻域内海量的仿真状态。通过蒙特卡洛法，微调系统中的可控量，如发电机的节点电压和

有功出力、可控负荷节点电压等，重新结算潮流和关键断面极限传输容量，由此得到海量的仿真状态，这些数据将作为历史数据用于后续训练^[21-23]。

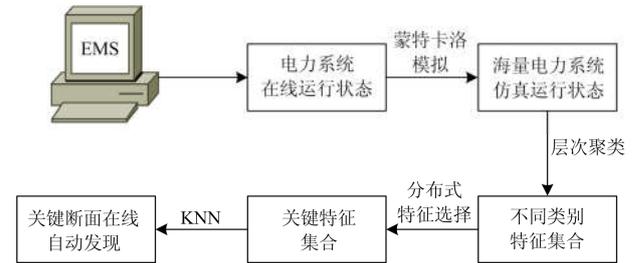


图 1 基于 K-邻近法(KNN)的电网关键断面在线分布式发现方法功能架构图

Fig. 1 Function framework of power system distributed key section detection online based on K nearest neighbor algorithm

2) 基于层次聚类的分布式特征选择

在得到海量仿真数据以后，将仿真数据与实际状态数据合并为电力系统仿真大数据，随后对每个状态建立特征集合。特征集合由系统可控潮流量构成(如发电机电压、发电机有功、500 kV 变电站电压等)。由于电网规模庞大，特征量的规模也十分巨大，给后续的机器学习造成瓶颈。然而，各个特征之间并不是独立的，有些特征因为电气距离比较近往往具有较强的相关性，因此有必要对特征个数进行削减，进行降维处理。

数据科学中主要的降维方法有两种：特征选择与特征提取。特征选择的主要目标是在海量待选择的特征中通过某种指标筛选出最有代表性的若干特征，筛选出的特征个数会远远小于未被筛选的特征个数，从而达到降维的目的。而特征提取的主要思想是将原有特征映射到另一个空间中，该空间可以是另一个线性空间或者是非线性空间，根据映射的规则不同，可以有不同的映射方法，如主成分分析等。

由于电力系统在线运行特征量具有十分明确的物理意义，若使用特征提取将使其物理含义不再清晰。本文采取文献[23]中的特征选择方法，首先采用层次聚类所有待选特征按照分布式计算节点的个数进行聚类分析，将相似的特征归纳到同一个类中并由同一个计算节点处理；接着，在每个计算节点上进行分布式特征选择，将最能代表该组特征集合的若干个特征遴选出来；最后，将所有节点上遴选出的特征进行合并得到关键特征集合。

3) 基于 K-邻近法(KNN)的关键断面在线发现

KNN 是一种监督学习算法，主要用于分类和回归之中，其思想为建立新样本与已有样本之间的远

近关系并将此关系用来评判回归量的特性^[18-20]。海量的训练样本是 KNN 发挥重要能力的基础,而电力仿真大数据每天的数据量超过 24 G,符合 KNN 计算的数据量需求。本文首先对每个断面单独建立 K-邻近法(KNN)二分类模型并进行分布式的训练;接着用 KNN 模型分布式地对每个关键断面是否出现做出预测。利用 KNN 分布式预测模型可以大大减少关键断面在线发现所需时间,且关键断面预测精度达到工程应用需求。

2 关键技术细节

2.1 关键断面分布式发现

基于 K-邻近法(KNN)的电网关键断面在线分布式预测模型的架构如图 2 所示。

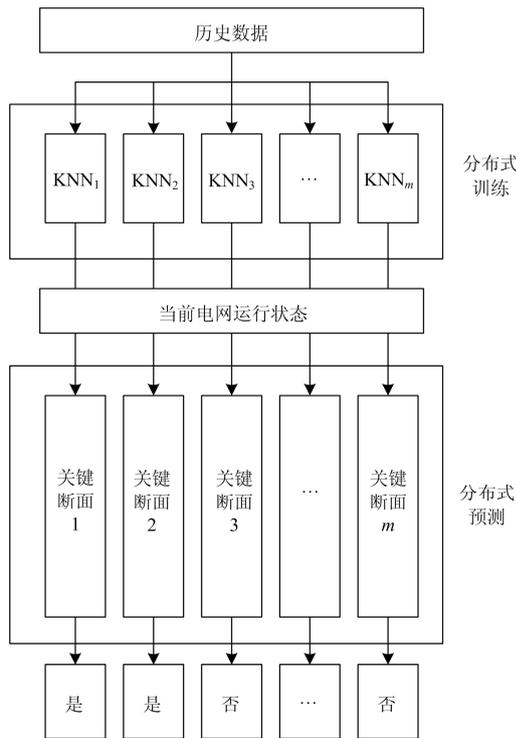


图 2 基于 K-邻近法(KNN)的电网关键断面在线分布式预测模型的架构图

Fig. 2 Function framework of power system distributed key section prediction online based on K nearest neighbor algorithm

基于历史实际和仿真大数据,针对每个断面独立地建立 K-邻近法(KNN)二分类模型,这些模型之间可以进行分布式的训练,节约训练时间;接着,结合当前电网运行状态,用 KNN 模型分布式地对每个关键断面是否出现做出快速预测。

2.2 KNN 分类器

基于 K-邻近法(KNN)的主要思想是将待处理

的对象映射到某个空间中,这个空间既可以是由放射变换和欧氏距离定义的线性空间,也可以是由非线性核函数定义的非线性空间,甚至可以是哈密特空间。所以 KNN 拥有相当多的变种,本文运用基于放射变换和欧氏距离的 KNN 算法^[11-12]。

1) 样本空间的建立以及距离定义

定义某一时刻电网的状态可以用一个一维向量表示。

$$\mathbf{X} = [x_1, x_2, x_3, \dots, x_n] \quad (1)$$

式中: \mathbf{X} 为该状态; x_i 为第 i 个特征,共 n 个特征,而该状态对应的某个特定的关键断面出现与否用一个 0-1 变量表示,则该状态完整表示为

$$\hat{\mathbf{X}} = [x_1, x_2, x_3, \dots, x_n; b] \quad (2)$$

式中, b 代表该关键断面出现与否的 0-1 变量。定义 n 维放射空间

$$\mathbf{S} = [s_1, s_2, s_3, \dots, s_n] \quad (3)$$

定义空间中 \mathbf{X} 与 \mathbf{Y} 两点距离为欧式距离,即

$$D(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

定义两个对象之间的相似度即为两者之间的欧式距离,即

$$\text{Simi}(\mathbf{X}, \mathbf{Y}) = D(\mathbf{X}, \mathbf{Y}) \quad (5)$$

至此,将所有的训练数据集映射至该空间中,每个对象均对应着某一关键断面是否出现的状态。

2) 相似对象的查找

对于一个新的对象 \mathbf{Y} :

$$\mathbf{Y} = [y_1, y_2, y_3, \dots, y_n] \quad (6)$$

首先将其映射到样本空间,接着根据样本空间中个体的相似度定义,查找出距离该新样本最近的 K 个历史样本。定义历史数据的全集为

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \dots, \mathbf{X}_N] \quad (7)$$

对于新对象 \mathbf{Y} ,寻找和其距离最近的 K 个对象,定义该集合为 \mathbf{H}_K :

$$\mathbf{H}_K = [\mathbf{X}_1^K, \mathbf{X}_2^K, \dots, \mathbf{X}_K^K] \quad (8)$$

对于 \mathbf{H}_K 中的所有对象,统计其关键断面出现频率,若该关键断面出现频率大于某一阈值 ε ,则认为该关键断面会出现,否则该关键断面不会出现:

$$P(b=1) \geq 0.5 \quad (9)$$

对 KNN 模型,可以总结出具有以下优点:

1) 对历史数据的处理十分简单。由于本文采用放射空间和欧式距离的组合方法,原始数据不加转换地转换为放射空间中的坐标,故十分节省有限的计算资源。

2) 发挥共享内存与并行计算的优势。由于对于

所有断面均未进行空间映射，则众多的历史数据可以在内存中共享，即虽然不同断面的计算过程在不同计算节点上进行，但是所有计算可以共享一个内存，内存的大大缩减有利于进行内存计算，从而可以最大程度地发挥 Spark 大数据构架的优点。

3 算例测试

本文算例测试在广东电网进行。广东电网网架结构复杂，元件众多，是我国最大的电网之一。广东电网的电气分区图如图 3 所示。同时广东电网的发电负荷水平较高，存在许多需要负载率高、需要调度员重点调控的关键断面。因此，广东电网适合于做基于 K-邻近法(KNN)的电网关键断面在线分布式发现方法的算例分析。

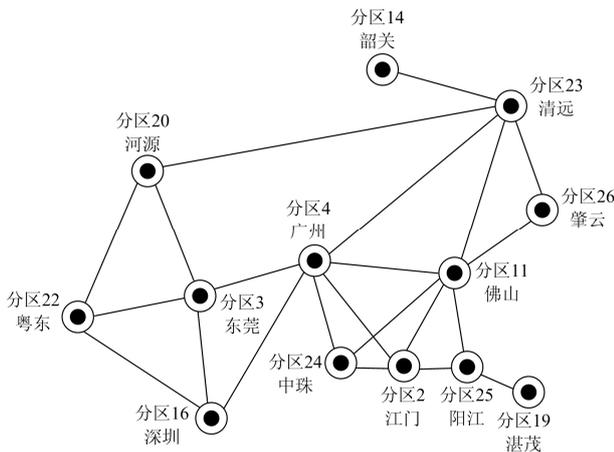


图 3 广东电网电气分区图

Fig. 3 Electric zone of Guangdong Power Grid

算例运行环境为：主节点为双核 Intel Xeon 2630 v3.0 2.4 GHz, RAM 64 G; 从节点为 Intel Xeon 2630 3.0 2.4 GHz, RAM12 G。

算例运行数据为：将 2016 年 5 月 1 日至 2016 年 6 年 30 日期间，每隔 15 min 实际运行数据以及部分仿真数据作为历史数据，用于分布式地训练 KNN 模型。以 2016 年 7 月 1 日至 2016 年 7 月 7 日期间，每隔 15 min 实际运行数据作为测试数据，预测断面出现的情况，并与传统建模方法结果^[6-7]进行比较分析。

为了量化关键断面预测准确程度，本文定义准确率 P_R 、漏警率 P_L 和虚警率 P_X ，那么对于正确预测次数为 R 、漏警次数为 L 、虚警次数为 X 的某一断面准确率 P_R 为

$$P_R = \frac{R}{R + X + L} \quad (10)$$

漏警率 P_L 为

$$P_L = \frac{L}{R + X + L} \quad (11)$$

虚警率 P_X 为

$$P_X = \frac{X}{R + X + L} \quad (11)$$

设定阈值 ε 为

$$\varepsilon = 0.5 \quad (12)$$

通过两个月历史仿真大数据的训练，共发现 139 个关键断面，将这些关键断面对应的在线分布式预测模型在 1 周的测试数据集上进行测试，KNN 模型的预测结果统计如表 1 所示。

表 1 KNN 模型的预测结果统计

Table 1 Forecast results of KNN model

平均准确率	平均虚警率	平均漏警率
93.59%	4.38%	2.03%

从表 1 的预测结果可以看出，本文所提方法具有较高的准确率，且错误预测中占比较大的是虚警，对电网安全运行影响较小，关键断面预测精度达到工程应用需求。KNN 模型与传统模型的预测时间比较如表 2 所示。

表 2 KNN 模型与传统模型的预测时间比较

Table 2 Comparison of KNN model and traditional model for prediction time

方法	预测时间
KNN	<0.1 ms
传统方法	0.8 s

结合表 1 和表 2 可以发现，在关键断面预测精度达到工程应用需求的条件下，基于 K-邻近法的电网关键断面在线分布式发现方法能够大大减少关键断面预测时间，从而极大地提升在线运行和控制的时间粒度。

4 结论

本文首先根据电力系统实时运行状况建立了特征集合，接着运用层次聚类和分布式特征选择对特征降维处理，最后利用 KNN 方法建立降维后的特征与关键断面出现与否的映射，形成预测模型。算例分析表明本文所提方法不仅对关键断面的预测有着很高的准确度，预测精度达到工程应用需求，而且大大缩减了传统算法计算关键断面所需时间，从而提高了在线关键断面辨识的时间粒度。

参考文献

[1] US-Canada Power System Outage Task Force. Final report on the August 14th blackout in the United States and Canada: causes and recommendations[EB/OL].

- [2003-10-15]. <http://www.nerc.com>.
- [2] LIU Yanhua, LI Yan, CHI Yongning, et al. Analysis on a large scale wind turbines cascading trip-off accident in North China[C] // 2013 IEEE Grenoble Conference PowerTech (POWERTECH), June 16-20, 2013, Grenoble, France:1-6.
- [3] 孙宏斌, 胡江溢, 刘映尚, 等. 调度控制中心功能的发展——电网实时安全预警系统[J]. 电力系统自动化, 2004, 28(15): 1-6.
- SUN Hongbin, HU Jiangyi, LIU Yingshang, et al. Development of the power dispatching control center-real time power security early warning system[J]. Automation of Electric Power Systems, 2004, 28(15): 1-6.
- [4] 严剑峰, 于之虹, 田芳, 等. 电力系统在线动态安全评估和预警系统[J]. 中国电机工程学报, 2008, 28(34): 87-93.
- YAN Jianfeng, YU Zhihong, TIAN Fang, et al. Dynamic security assessment and early warning system of power system[J]. Proceedings of the CSEE, 2008, 28(34): 87-93.
- [5] 黄天恩, 孙宏斌, 郭庆来, 等. 基于电网运行仿真大数据的知识管理和超前安全预警[J]. 电网技术, 2015, 39(11): 3080-3087.
- HUANG Tianen, SUN Hongbin, GUO Qinglai, et al. Knowledge management and security early warning based on big simulation data in power grid operation[J]. Power System Technology, 2015, 39(11): 3080-3087.
- [6] 赵峰, 孙宏斌, 张伯明. 基于电气分区的输电断面自动发现[J]. 电力系统自动化, 2011, 35(5): 42-46.
- ZHAO Feng, SUN Hongbin, ZHANG Boming. Zone division based automatic discovery of flowgate[J]. Automation of Electric Power Systems, 2011, 35(5): 42-46.
- [7] 赵峰, 孙宏斌, 谭嫣, 等. 综合考虑多种电网安全主题的关键断面自动发现方法[J]. 电网技术, 2014, 38(5): 1169-1174.
- ZHAO Feng, SUN Hongbin, TAN Yan, et al. Automatic discovery method of power system critical flowgates considering multiple security themes in power grid[J]. Power System Technology, 2014, 38(5): 1169-1174.
- [8] 蒋维勇, 张伯明, 吴文传, 等. 应用于在线调度决策的极限传输容量计算方法[J]. 电力系统自动化, 2008, 32(10): 12-17.
- JIANG Weiyong, ZHANG Boming, WU Wenchuan, et al. A total transfer capability calculation method for power system operation and decision[J]. Automation of Electric Power Systems, 2008, 32(10): 12-17.
- [9] 赵峰, 孙宏斌, 黄天恩, 等. 电网关键断面及安全运行规则自动发现系统设计与工程实现[J]. 电力系统自动化, 2015, 39(1): 117-123.
- ZHAO Feng, SUN Hongbin, HUANG Tianen, et al. Design and engineering application of automatic discovery system for critical flowgates and security operation rules in power grids[J]. Automation of Electric Power Systems, 2015, 39(1): 117-123.
- [10] HAN Jiawei, KAMBER M. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2001.
- [11] 边肇祺, 张学工. 模式识别[M]. 北京: 清华大学出版社, 2000.
- [12] 周志华, 杨强. 机器学习及其应用[M]. 北京: 清华大学出版社, 2011.
- [13] 夏定纯, 徐涛. 人工智能技术与方法[M]. 武汉: 华中科技大学出版社, 2004.
- [14] 向德军, 王彬, 郭文鑫, 等. 基于人工神经网络的电力系统精细化安全运行规则[J]. 电力系统保护与控制, 2017, 45(18): 38-43.
- XIANG Dejun, WANG Bin, GUO Wenxin, et al. Fine security rule for power system operation based on artificial neural network[J]. Power System Protection and Control, 2017, 45(18): 38-43.
- [15] 周勳, 张波, 刘日亮, 等. 基于数据驱动的线路保护定值自适应整定方法研究[J]. 电力系统保护与控制, 2017, 45(24): 50-56.
- ZHOU Meng, ZHANG Bo, LIU Riliang, et al. An adaptive adjustment method of line protection setting based on data-driven[J]. Power System Protection and Control, 2017, 45(24): 50-56.
- [16] 崔明建. 基于ANN概率生成模型的风电功率爬坡事件大数据场景预测[J]. 陕西电力, 2017, 45(7): 83-91.
- CUI Mingjian. Big data scenarios prediction of wind power ramp events based on ANN probability generation model[J]. Shaanxi Electric Power, 2017, 45(7): 83-91.
- [17] 赵建平, 李刚. 基于GA-Elman神经网络的短期风电功率预测[J]. 陕西电力, 2017, 45(3): 23-26.
- ZHAO Jianping, LI Gang. Short term wind power prediction based on GA-Elman neural network[J]. Shaanxi Electric, 2017, 45(3): 23-26.

- [18] 钱海峰, 陈阳舟, 李振龙, 等. 核函数法与最邻近法在短时交通流预测应用中的对比研究[J]. 交通与计算机, 2008, 26(6): 18-21.
QIAN Haifeng, CHEN Yangzhou, LI Zhenlong, et al. Comparison of kernel approach in short term forecasting of traffic flow[J]. Computer and Communications, 2008, 26(6): 18-21.
- [19] 赖志柱, 戈冬梅, 张云艳. 求解 TSP 问题的改进最邻近法[J]. 贵州工程应用技术学院学报, 2016, 34(1): 139-142.
LAI Zhizhu, GE Dongmei, ZHANG Yunyan. Improved nearest neighbor method for solving TSP[J]. Journal of Guizhou University of Engineering Science, 2016, 34(1): 139-142.
- [20] 张慕宇, 杨智春, 丁燕, 等. 采用主成分分析与最邻近法的复合材料板损伤检测实验[J]. 西北工业大学学报, 2010, 28(5): 786-791.
ZHANG Muyu, YANG Zhichun, DING Yan, et al. A better experimental method for damage detection of composite plate using principal component analysis (PCA) and nearest neighbor principle[J]. Journal of Northwestern Polytechnical University, 2010, 28(5): 786-791.
- [21] 蒋维勇, 孙宏斌, 张伯明, 等. 电力系统精细规则的研究[J]. 中国电机工程学报, 2009, 29(4): 1-7.
JIANG Weiyong, SUN Hongbin, ZHANG Boming, et al. Fine operational rule of power system[J]. Proceedings of the CSEE, 2009, 29(4): 1-7.
- [22] 孙宏斌, 赵峰, 蒋维勇, 等. 电网精细规则在线自动发现系统架构与功能设计[J]. 电力系统自动化, 2011, 35(18): 81-86.
SUN Hongbin, ZHAO Feng, JIANG Weiyong, et al. Framework and functions of fine operational rules online automatic discovery system for power grid[J]. Automation of Electric Power Systems, 2011, 35(18): 81-86.
- [23] 黄天恩, 孙宏斌, 郭庆来, 等. 基于电网运行大数据的在线分布式安全特征选择[J]. 电力系统自动化, 2016, 40(4): 32-40.
HUANG Tianen, SUN Hongbin, GUO Qinglai, et al. Distributed security feature selection online based on big data in power system operation[J]. Automation of Electric Power Systems, 2016, 40(4): 32-40.

收稿日期: 2018-11-06; 修回日期: 2019-01-15

作者简介:

王彬(1987—), 男, 博士, 工程师, 主要研究方向为电力系统调度自动化; E-mail: wangbin_gd@qq.com

郭文鑫(1985—), 男, 硕士, 高级工程师, 主要研究方向为电力系统调度自动化。E-mail: guowenxin1985@126.com

(编辑 魏小丽)