

10-30-2023

Learning and Analysis of Dynamic Models for Grid Discrete Events Based on Log Information

Danlong Zhu

Shenzhen Power Grid Utility Co. LTD, Shenzhen 518048, China, zhudanlong@sz.csg.cn

Yunqi Yan

Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

Ying Chen

Department of Electrical Engineering, Tsinghua University, Beijing 100084, China, chen_ying@tsinghua.edu.cn

Jiaqi Zhang

Department of Electrical Engineering, Tsinghua University, Beijing 100084, China

See next page for additional authors

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact xtfzxb@126.com.

Learning and Analysis of Dynamic Models for Grid Discrete Events Based on Log Information

Abstract

Abstract: With the increasing scale of power grid, the massive amount of log information generated by devices in the power grid poses a challenge to the manual analysis of abnormal grid conditions. The log information generated during the operation of the power grid has the typical discrete sequential characteristics. By analyzing the log information of grid alarm messages, a station event transition probability model and an event sequence risk calculation method are proposed to effectively model and analyze the abnormal operation level of primary and secondary systems in substations. The proposed method not only successfully identifies the event sequences corresponding to grid failures that have been manually recorded by dispatchers but also identifies the abnormal alarm information sequences that have not been recorded by dispatchers. This helps analyze and assess the operational situation of power grid equipment, discover potential risks, and improve the efficiency of substation maintenance.

Keywords

power system, alarm information, dynamic feature learning, anomaly detection, risk assessment

Authors

Danlong Zhu, Yunqi Yan, Ying Chen, Jiaqi Zhang, Longxing Jin, and Wei Fu

Recommended Citation

Zhu Danlong, Yan Yunqi, Chen Ying, et al. Learning and Analysis of Dynamic Models for Grid Discrete Events Based on Log Information[J]. Journal of System Simulation, 2023, 35(10): 2193-2201.

基于日志信息的电网离散事件动态模型学习和分析

朱丹龙¹, 闫云琦², 陈颖^{2*}, 张家琦², 晋龙兴¹, 付威¹

(1. 深圳供电局有限公司, 广东 深圳 518048; 2. 清华大学 电机系, 北京 100084)

摘要: 随着电网规模的增加, 电网设备产生的海量日志信息为人工分析电网异常状态带来了挑战。电网运行中的日志信息具有典型的离散序列特征, 通过分析电网告警信息日志, 提出了厂站事件转移概率模型和事件序列风险计算方法, 可以有效建模和分析厂站一二次系统运行异常水平。该方法不仅可以成功地识别出已被调度员人工标记的电网故障对应事件序列, 还能识别出未曾被调度员记录的异常告警信息序列, 有助于分析研判电网设备运行态势, 发现潜在风险, 提升变电站运维效率。

关键词: 电力系统; 告警信息; 动态特征学习; 异常识别; 风险评估

中图分类号: TM769; TP391 文献标志码: A 文章编号: 1004-731X(2023)10-2193-09

DOI: 10.16182/j.issn1004731x.joss.23-FZ0800

引用格式: 朱丹龙, 闫云琦, 陈颖, 等. 基于日志信息的电网离散事件动态模型学习和分析[J]. 系统仿真学报, 2023, 35(10): 2193-2201.

Reference format: Zhu Danlong, Yan Yunqi, Chen Ying, et al. Learning and Analysis of Dynamic Models for Grid Discrete Events Based on Log Information[J]. Journal of System Simulation, 2023, 35(10): 2193-2201.

Learning and Analysis of Dynamic Models for Grid Discrete Events Based on Log Information

Zhu Danlong¹, Yan Yunqi², Chen Ying^{2*}, Zhang Jiaqi², Jin Longxing¹, Fu Wei¹

(1. Shenzhen Power Grid Utility Co. LTD, Shenzhen 518048, China; 2. Department of Electrical Engineering, Tsinghua University, Beijing 100084, China)

Abstract: With the increasing scale of power grid, the massive amount of log information generated by devices in the power grid poses a challenge to the manual analysis of abnormal grid conditions. The log information generated during the operation of the power grid has the typical discrete sequential characteristics. By analyzing the log information of grid alarm messages, a station event transition probability model and an event sequence risk calculation method are proposed to effectively model and analyze the abnormal operation level of primary and secondary systems in substations. The proposed method not only successfully identifies the event sequences corresponding to grid failures that have been manually recorded by dispatchers but also identifies the abnormal alarm information sequences that have not been recorded by dispatchers. This helps analyze and assess the operational situation of power grid equipment, discover potential risks, and improve the efficiency of substation maintenance.

Keywords: power system; alarm information; dynamic feature learning; anomaly detection; risk assessment

收稿日期: 2023-07-02 修回日期: 2023-09-12

基金项目: 国家自然科学基金企业创新发展联合基金重点资助项目(U22B2096)

第一作者: 朱丹龙(1988-), 男, 工程师, 本科, 研究方向为电力系统自动化。E-mail: zhudanlong@sz.csg.cn

通讯作者: 陈颖(1979-), 男, 研究员, 博士, 研究方向为电力系统高性能计算、信息物理系统、电网韧性。

E-mail: chen_ying@tsinghua.edu.cn

0 引言

电网中存在大量量测数据采集与监控设备, 电网日志信息近乎实时地反映了电网设备的运行状态。电网正常运行、方式变更、倒闸检修、异常故障等不同状态的特征, 都可以从告警信息中分析获取。已有的告警信息分析方式主要依赖调度员的工作经验, 需要高度集中注意力。随着经济社会的发展, 电网规模日益庞大, 接入设备数量增多且种类繁多, 主站告警信息内容更加复杂且上传频率较高, 调度员分析处理告警信息的压力愈发增大。因此, 有必要研究电网告警信息的智能化分析技术, 实现对关键事件序列特征的学习和识别, 辅助调度员分析处理海量上传告警信息, 提升电网运行可靠性的水平。

目前, 电网告警信息的处理面临着“数据过剩, 信息缺少”的问题。电网监测装置对电网设备的运行状态进行离散采样, 并上传告警信息进行状态描述。由于电网具有动态性, 单独一条告警信息不足以完整表达系统的故障传递特性, 因此, 需要对系统异常进行动态的检测和预报^[1-2]。当设备运行在异常状态或发生故障的前后, 其关联的监测装置上传的告警信息序列包含着相应的异常或故障信息。通过挖掘和分析这些异常或者告警信息, 可以得到两方面的重要特征信息: ①表征电网故障的特征信息, 即挖掘是否存在特定的告警信息组合, 能够表征着特定种类的电网故障, 所得到的特征信息序列可用于构建和维护电网故障诊断中专家系统的规则库。②电网设备故障发生前的特征信息, 或者说设备处于异常运行状态下的特征信息, 而此时故障和事故尚未发生。所得到的特征信息可用于深层次电网态势感知, 对可能发生的故障或正在发生的事故向调度员进行预警。

已有文献针对电网告警信息组成的事件序列的研究主要以故障诊断为目的, 具体方法可以归纳为知识驱动和数据驱动两类^[3]。知识驱动的研究

从电网一二次设备配合逻辑出发, 构建故障过程事件序列推演模型, 如贝叶斯网络^[4]、Petri网^[5]和专家系统^[6]等。知识驱动需由专业人员提供支持, 模型往往是固定的, 受限于特定的研究目标, 泛化能力差, 难以适应新的故障情景。数据驱动主要采用数据挖掘的方法, 从历史事件日志挖掘故障对应模式, 如应用频繁项挖掘、序列模式挖掘等, 得到故障相关事件关联模式或者事件序列模式。为了提升模式挖掘和识别效率, 有研究通过对告警信息聚类实现预处理^[7-8]。也可通过标注正常和故障事件序列样本, 采用机器学习算法得到识别故障事件序列的分类器^[3,9]。数据驱动的方法不依赖于专家经验, 可以从数据中学习获取新的知识, 从而支持相关决策的应用。数据驱动的效果也取决于数据的规模、质量和获取的难易程度^[10]。以故障诊断为例, 采用机器学习的算法需要标注正常和故障事件序列数据。但是现实中电网故障是偶发事件, 发生概率较低, 本文所研究的某市电网5年内告警信息规模在5千万条以上, 但仅有约400起故障案例。常见的有监督机器学习模型难以适用于这类高度不均衡的数据集^[11]。此外, 也有学者提出针对告警信息文本语义分析的挖掘方法。文献[12]基于双向LSTM神经网络对电网故障案例文本数据展开因果挖掘。文献[13]结合有限状态机和告警信息对继电保护装置的拒动情况进行故障追踪, 能够对海量告警信息进行高效解析, 并识别出拒动原因。然而, 上述工作只是对故障进行溯因, 无法实时监测系统可能面临的运行状态异常与风险。

本文所研究的故障事件序列特征学习与自主识别技术, 旨在从时间上连续、空间上多点并发上传的电网日志信息中发现、学习电网故障事件序列的各类特征模式, 实现电网异常及故障的预警, 从而避免故障事件进一步发展升级, 辅助调度员高效分析处理海量告警信息, 提升电网故障处置效率。

1 电网日志信息事件序列建模

电网日志信息中包含着电网运行态势的关键信息, 因此, 无论是电网故障的预警、诊断和恢复, 还是变电站自动化运维, 都离不开对日志信息的分析。已有的关于电网日志信息的研究, 可以分为2类: ①对日志信息的语义解析, 一般以单条日志信息为研究对象; ②将每条日志信息作为一个事件, 关注的是连续的日志信息所形成的事件序列, 研究事件序列所反映的电网运行状态和变化。由于电网发生的故障往往存在因果性、时序依赖性。因此, 需要对电网事件序列进行建模, 本文对电网日志信息组成的事件序列做如下形式化定义:

$$S = \{(t_i, e_i)\}_{1 \leq i \leq n} \quad (1)$$

式中: S 为长度为 n 的事件序列; e_i 为序列中第 i 个发生的事件; t_i 为 e_i 发生的时间戳。每个事件 e_i 表示一种告警信息的实例。

电网告警信息事件序列具有独特的性质。首先, 电网告警信息事件序列包含多维度信息, 如序列中的每对二元组 (t_i, e_i) 都描述了一个事件的时间戳、产生厂站、一二次设备、装置属性和动作行为。其次, 与时间序列不同的是, 事件序列的时间戳处于连续的时间域, 而时间序列则通常由等间隔的采样形成。电网告警信息事件序列中的事件并非相互独立的, 如设备故障导致保护动作后会引发一系列上送信息, 其反映的是设备运行状态的变化和量测控制装置动作之间复杂的耦合作用。事件之间相互影响, 进而在连续的时间域上展现出复杂的变化动态。此外, 无论是以地区电网还是单个厂站的告警信息事件序列为研究目标, 事件序列数据都是并发产生的。以地区电网而言, 同一时间段内存在多个厂站同时上传告警信息, 但不同厂站的告警信息事件序列之间可能并不具备关联关系, 这一点对于同一厂站的事件序列亦然。

2 电网日志信息序列特征学习方法

2.1 事件序列划分

调度中心收到的告警信息由变电站监测系统采集并上传。可将变电站告警信息事件日志表示为 $D = ((t_1^D, e_1^D), (t_2^D, e_2^D), \dots, (t_n^D, e_n^D))$, 其中, t_i^D 代表告警信息事件发生的时间戳, e_i^D 代表告警信息事件, $t_i^D \leq t_{i+1}^D, 1 \leq i \leq n$ 。

表1为基于某市电网5年告警信息日志的变电站告警事件统计结果。图1展示了某变电站1个月内告警信息数量在时间上的分布情况, 以小时为最小统计单位(组数720)。可以看出, 电网事件在时间上的分布具有稀疏性, 而这一特点则对应于电网异常或故障的偶发性。为进一步说明该稀疏特性, 考虑对变电站告警信息事件日志统计得到时间分布向量 V_T 。以一天24h为例, $V_T(i)$ 是统计第 i 小时内变电站发生的事件数量。变电站对应的时间分布向量 V_T 的非零元素占比越低, 则说明该变电站的事件分布稀疏程度越高。

表1 变电站告警事件统计结果

变电站电压等级/kV	变电站数量	日志长度中位数	事件种类中位数	采集模板事件种类
110	241	34 216	1 050	527
220	69	92 587	2 312	962
500	11	94 567	2 520	934

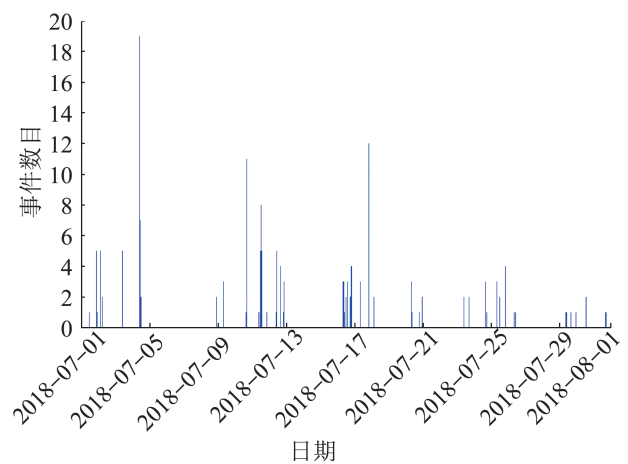


图1 告警信息事件数量分布

Fig. 1 Distribution of number of alarms information

为了能准确地刻画偶发的电网事件演变的动态过程，电网事件序列应当具有两方面的特征。首先，同一序列内的事件之间具有紧密的联系，包括电气、通信、控制等多种关联逻辑。其次，每一个序列内部的事件在时间上分布相对集中。为了得到此类事件序列，本文设计了电网事件序列的划分算法，其中，变电站告警信息事件日志记为 $D = ((t_1^D, e_1^D), (t_2^D, e_2^D), \dots, (t_n^D, e_n^D))$ ，每条告警信息事件对应的事件序列的编号记为 $T = (trace_id_1, trace_id_2, \dots, trace_id_n)$ 。

算法1 变电站日志信息事件序列划分方法

Input: 变电站告警信息事件日志 D ，事件序列划分时间间隔 ΔT

Output: 告警信息事件序列编号 $T = (trace_id_i)_{1 \leq i \leq n}$

$trace_id_i \leftarrow 0, 1 \leq i \leq n$

for $k \leftarrow 2$ to n do

if $t_k^D - t_{k-1}^D < \Delta T$ then

$trace_id_k \leftarrow trace_id_{k-1}$

else

$trace_id_k \leftarrow trace_id_{k-1} + 1$

End

算法1所描述的对事件序列的划分策略，事实上等价于机器学习中基于密度的聚类方法^[14]。在本研究中，样本分布的紧密程度由日志信息上传时刻的间隔长短来衡量，相邻时间间隔较短的事件之间具备可连接性，在可连接事件的基础上不断拓展即可得到事件序列，每一个划分出的事件序列对应于一个样本簇。

2.2 变电站事件转移概率模型

将变电站不同类型的事件集合表示为 $\Omega = \{E_1, E_2, \dots, E_M\}$ ，考虑长度为 k 的事件序列 $s = ((t_1, e_1), (t_2, e_2), \dots, (t_k, e_k))$ ， $e_i \in \Omega, 1 \leq i \leq k$ 。为了刻画事件序列产生和演变的机制，假设变电站事件的转移具有一阶马尔可夫性：

$$P(e_{i+1}|e_i, e_{i-1}, \dots, e_1) = P(e_{i+1}|e_i) = P(e_i \rightarrow e_{i+1}) \quad (2)$$

则定义在该状态空间上的马尔可夫过程转移矩阵为 $P_{ij} = P(E_i \rightarrow E_j) = P(E_j|E_i), 1 \leq i, j \leq M$ 。因此，可以将变电站事件日志看作马尔可夫过程的采样样本集合，按照2.1节中方法划分得到的每条事件序列即为一条单独的样本。记获得的事件序列集合为 S ，集合中的每一条事件序列为 s ，序列 s 长度为 l_s 。根据该样本数据可估计转移矩阵 P 的参数：

$$\hat{P}_{ij} = \hat{P}(E_j|E_i) = \frac{\sum_{s \in S} \sum_{1 \leq k \leq l_s - 1} 1(e_k = E_i, e_{k+1} = E_j)}{\sum_{s \in S} \sum_{1 \leq k \leq l_s - 1} 1(e_k = E_i)} \quad (3)$$

式中： \hat{P} 为估计的转移矩阵； $\hat{P}(E_j|E_i)$ 为估计的事件 E_i 到 E_j 的转移概率。

同样地，可以估计初始状态分布 $\pi^0 = (\pi_1^0, \pi_2^0, \dots, \pi_M^0)$ ，估计方法为

$$\hat{\pi}_i^0 = \frac{\sum_{s \in S} 1(e_0 = E_i)}{|S|} \quad (4)$$

根据式(3)的参数估计方法，遍历变电站的事件日志，统计得到刻画该变电站告警信息事件转移概率分布的 bi-event 矩阵。bi-event 矩阵可对应一个有向概率图模型，其中，节点代表不同事件，边权代表事件之间的转移概率。

由表1可知，得到的事件转移 bi-event 矩阵的维度等于变电站的事件种类数目 $|\Omega|$ 。而该转移矩阵是高度稀疏的，大量的事件对组合并没有在历史日志中出现过，对应的事件转移概率图连通性较弱。同时，存在大量的“多词同义”的现象，即多个告警事件实际上描述的是同一个过程内容。此外，不同的变电站，即便是属于同一电压等级，电气接线结构和一二次设备配合设置相似，所得到的转移矩阵维度仍然大不相同，不便于分析。为此，考虑采用特征选择方法，获取同一电压等级变电站事件空间的统一低维表示。

按照电网公司制定的告警信息模板, 可以将告警信息进行标准化处理, 提取告警时间、对应电压等级、告警设备位置等信息特征。本文主要关注系统风险在设备之间的转移扩散特性, 因此, 以告警信息对应的设备位置作为目标特征, 将原始的样本空间降维至一个较小的事件空间。由于告警信息标准化处理后可以直接提取“设备位置”这一信息特征, 因此, 该特征提取模型无需训练, 降维方法整体的计算量较小。

对变电站的上千种事件类型做聚合, 得到的降维后的事件空间记为 Ω 。考虑告警信息事件的设备位置对应于采集遥信信息模板中的设备名称, 表 2 为某电网公司 2019 年制定的 220 kV 变电站采集遥信信息模板示例。

表 2 变电站采集遥信信息模板

设备名称	规范信息描述
220 kV 母线保护	220 kV×M×M 母线保护 I 动作跳 220 kV×M
	220 kV×M×M 母线保护 I 动作跳 220 kV 母联(分段)20××
	220 kV×M×M 失灵保护 I 动作跳 220 kV×M
	220 kV×M×M 母线保护 I 动作线路跟跳
.....	
220 kV 线路间隔	220 kV××线××主一保护动作
	220 kV××线××主一保护重合闸动作
	220 kV××线××保护第一组出口跳闸
	220 kV××线××保护第二组出口跳闸
.....	

2.3 事件序列风险评估方法

设待评估风险的告警信息事件序列为 $s = ((t_1, e_1), (t_2, e_2), \dots, (t_{l_s}, e_{l_s}))$, 基于事件转移概率模型, 可根据式(5)计算该事件序列的概率:

$$\begin{aligned}
 P(e_1 e_2 \dots e_{l_s}) &= \\
 P(e_1) P(e_2 | e_1) \dots P(e_{l_s} | e_1 e_2 \dots e_{l_s-1}) &= \\
 P(e_1) P(e_2 | e_1) P(e_3 | e_2) \dots P(e_{l_s} | e_{l_s-1}) &= \\
 P(e_1) \prod_{1 \leq i \leq l_s-1} P(e_{i+1} | e_i) & \quad (5)
 \end{aligned}$$

本文将事件序列产生概率的负对数定义为事件序列累积风险:

$$\begin{aligned}
 R_s = -\lg P(s_j = ((t_1, e_1), (t_2, e_2), \dots, (t_j, e_j))) &= \\
 -\lg \left(P(e_1) \prod_{1 \leq i \leq l_s-1} P(e_{i+1} | e_i) \right) &= \\
 -\lg P(e_1) + \sum_{1 \leq i \leq l_s-1} -\lg(P(e_{i+1} | e_i)) &= \\
 -\lg P(e_1) + \sum_{1 \leq i \leq l_s-1} -\lg(P(E_n = e_{i+1} | E_m = e_i)) & \quad (6)
 \end{aligned}$$

整理得到

$$R_s = \begin{cases} -\lg P(e_1), & l_s = 1 \\ -\lg P(e_1) + \sum_{1 \leq i \leq l_s-1} -\lg(P(E_n = e_{i+1} | E_m = e_i)), & \\ l_s > 1 \end{cases} \quad (7)$$

事件序列的累积风险值 R_s 的意义可结合信息论中自信息概念阐释。信息论中的自信息用于度量一个随机事件所包含的信息量, 其取值只与事件发生的概率相关, 假设事件 ω_n 发生的概率为 $P(\omega_n)$, 则自信息定义为

$$I(\omega_n) = -\log(P(\omega_n)) = \log\left(\frac{1}{P(\omega_n)}\right) \quad (8)$$

根据式(8)可以看出, 事件 ω_n 发生的概率越小, 则它发生后的自信息量越大, 也就是说该事件发生带来的“意外”程度越高。信息论中用熵来度量不确定性, 其计算方式为随机变量的自信息的期望。为刻画事件序列的异常水平, 采取类似自信息的定义来描述事件转移过程的异常程度。式(7)所定义的事件序列累积风险 R_s 可以对一个完整序列异常程度进行整体性的描述, 具备积分的性质, 但对于事件序列风险累积过程的动态刻画能力不足。为了增强所构建指标的表达力, 考虑采用滑动时间窗来计算事件序列的风险向量 V_R :

给定时间窗口宽度 w 和窗口滑动时间步长 d , 事件序列 $s = ((t_1, e_1), (t_2, e_2), \dots, (t_{l_s}, e_{l_s}))$, 风险向量 $V_R = ((t_1^w, r_1^w), (t_2^w, r_2^w), \dots, (t_l^w, r_l^w))$, l_s, l 分别为 s 和 V_R 的长度。滑动窗 W 里包含的事件序列记为 s_w 。随着时间窗的不断滑动, 记当前窗口 W 的时间起始点为 t_k^w , 新的事件从右端加入当前窗口对应的序列 s_w 中, 同时左端不再属于当前窗口的事

件从 s_w 中剔除。每一次滑动后，计算当前窗口对应的事件序列 s_w 的累积风险 r_k^W 。由滑动窗的工作原理可知，风险向量长度为

$$l = \frac{w + t_i - t_1}{d} \quad (9)$$

采用滑动时间窗计算得到风险向量 V_R ，用来刻画事件序列异常水平，具有以下优点：

(1) 原本采样自事件空间，时间上非均匀分布的告警信息事件序列的异常水平，通过一个均匀采样的一维时间序列来描述；

(2) 与整个序列的累积风险值 R_s 相比，风险向量 V_R 可以反映序列风险值累积变化的动态过程；

(3) 通过调整滑动窗的参数，即窗口宽度 w 和滑动步长 d ，可以控制风险值采样计算的精度，从而关注特定的风险变化特征。

本文提出一种基于滑动时间窗的事件序列风险向量计算方法，如算法2所示。算法2的时间复杂度为 $O(n)$ ，其中， n 为序列中事件数目。在大多数情况下，由于变电站事件日志的稀疏性，能够保证 $n \ll l$ ，进而确保算法的高效性。

算法2 基于滑动时间窗的事件序列风险向量计算方法

Input: 变电站告警信息事件日志 $D = ((t_1^D, e_1^D), (t_2^D, e_2^D), \dots, (t_n^D, e_n^D))$ ，变电站事件转移矩阵 P ，滑动窗时间宽度 w 和滑动时间步长 d

Output: 风险向量 $V_R = ((t_1^W, r_1^W), (t_2^W, r_2^W), \dots, (t_l^W, r_l^W))$

初始化： $t_0^W \leftarrow t_1^D - w - d$ ； $r_0^W \leftarrow 0$ ； $k, i \leftarrow 1$ ； $s_w \leftarrow \text{FIFO对象}$

while $i \leq n$ **do** //遍历数据集中的事件并加入时间窗

$$t_k^W \leftarrow t_{k-1}^W + d$$

$$r_k^W \leftarrow r_{k-1}^W$$

$$r_k^W \leftarrow \text{DeleteEventUpdate}(t_k^W, r_k^W, s_w)$$

$$p \leftarrow i$$

$$r_k^W, p \leftarrow \text{AddEventUpdate}(t_k^W, r_k^W, s_w)$$

if $s_w.\text{length} = 0$ **then**

$$t_{k+1}^W \leftarrow t_p^D - w - d$$

$$r_{k+1}^W \leftarrow 0$$

$$k \leftarrow k + 1$$

end

$$i \leftarrow p$$

$$k \leftarrow k + 1$$

end

while $s_w.\text{length} > 0$ **do**

$$t_k^W \leftarrow t_{k-1}^W + d$$

$$r_k^W \leftarrow r_{k-1}^W$$

$$\text{DeleteEventUpdate}(t_k^W, r_k^W, s_w, p)$$

$$k \leftarrow k + 1$$

End

3 算例测试

3.1 数据准备

以某变电站5年告警信息事件日志为测试数据。由于现实中变电自动化系统的建设与应用具有差异性，导致多厂家、多版本的信息采集系统并存，而其采用的信息采集和描述规范也不尽相同，导致得到的告警信息数据很多并不符合电网公司现行的信息采集规定中的模板描述。因此，需要先对告警信息进行标准化建模和处理，为后续的分析提供规范的数据。根据命名规范^[15]，一条告警信息文本可以分为3部分内容。

(1) 时间：本文所研究的是调度中心接收到的告警信息，所记录的时间为主站监控系统收到告警信息报文的时间。

(2) 告警信息内容：包含厂站、电压、间隔设备、信号行为等信息。

(3) 告警信息动作：描述信号的状态，包含“动作”和“复归”2种状态；对于开关分合闸告警而言，则没有该部分信息。

根据上述信息分类，可以结合告警信息专业词典进行分词处理。告警信息专业词典中归纳总结了电网设备与行为的多种表述。获得专业词典

后, 即可采取成熟的分词方法(如双向最大匹配算法)进行分词处理, 实现告警信息的自动解析。分词结果如表 3 所示。

表 3 电网告警信息分词示例

Table 3 Example of tokenization for grid alarm information

告警信息	
处理前	2018-09-15 14:17:25, 秀丽站 #1、#2 主变故障录波屏启动录波 复归
	2018-09-26 17:00:00, 塘朗站 220 kV 翡朗甲线 9517 保护第二组出口跳闸 动作
	2018-09-28 15:52:00, 玉翠站 110 kV 民翠线 1679 保护通信中断 动作
处理后	2018-09-15 14: 17: 25/TIME 秀丽站/SUB #1/EQP #2 主变/EQP 故障录波屏/CM 启动录波/BEH 复归/ACT
	2018-09-26 17:00:00/TIME 塘朗站/SUB 220 kV/VOLT 翡朗甲线 9517/EQP 保护/CM 第二组/OTH 出口跳闸/BEH 动作/ACT
	2018-09-28 15:52:00/TIME 玉翠站/SUB 110 kV/VOLT 民翠线1679/EQP 保护/CM 通信中断/BEH 动作/ACT

根据分词处理后的告警信息, 提取告警信息对应设备位置特征, 形成事件聚合后的告警信息序列, 供后续风险分析使用。

3.2 风险分析结果

取滑动窗窗口宽度 $w=5 \text{ min}$, 滑动步长 $d=1 \text{ min}$ 。图 2 为该变电站 5 年告警信息事件日志时间分布。图 3 为该变电站 5 年告警信息事件日志对应的风险时间序列。

事件序列划分时间间隔阈值 $\Delta T=1 \text{ h}$, 将该变电站的告警信息事件日志划分得到 2 432 条事件序列, 事件序列的风险向量最大值的经验累积分布如图 4 所示。计算结果显示, 97.6% 的事件序列的风险最大值不大于 10。

每个事件序列都可以计算得到对应的风险向量 V_R , 即风险的时间序列。根据定义, 事件序列的风险反映了序列的异常水平, 可以通过该指标区分正常和异常序列。此处考虑设定异常水平阈值 c , 当序列的风险超过该阈值, 将事件序列识别

为异常, 反之则判断为正常序列。上述异常检测方法的效果与风险阈值 c 的取值紧密相关。

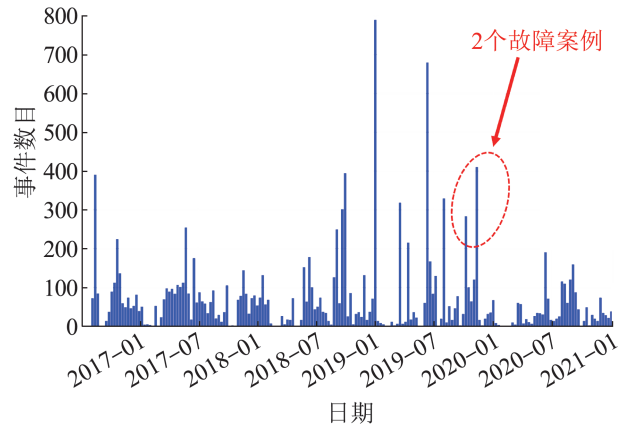


图 2 告警信息事件日志时间分布
Fig. 2 Time distribution of alarm information event logs

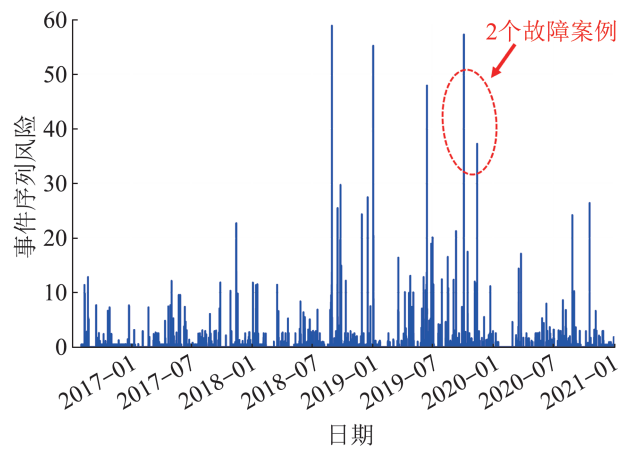


图 3 告警信息事件风险时间序列
Fig. 3 Risk time series of alarm information events

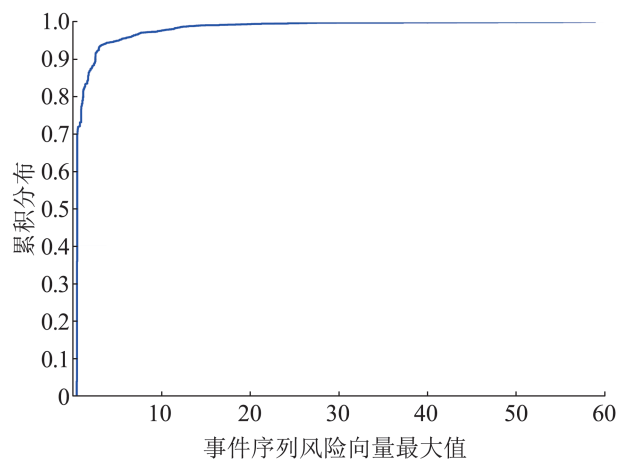


图 4 经验累积分布
Fig. 4 Empirical cumulative distribution

在图3中存在其他事件序列风险较高,但未记录为故障案例的告警信息。通过查看对应的告警信息序列,发现线路、电容器、变压器等设备对应的开关均发生了动作,保护也产生了异常。通过交叉对比这一时段内的检修记录,发现该序列对应了电网检修,并非为故障。但由于检修状态相对于电网的正常运行状态而言更加异常,因此,也会被本文方法检出。由于本文方法并不片面地考虑准确性,而是保证在不漏报异常的前提下提升运维效率。因此,在实际运行中,可以根据检修计划等信息综合研判异常序列识别结果。

为了进一步说明本文方法的有效性,考虑采用某市电网5年所有110 kV和220 kV电压等级的输变电故障记录和告警信息日志测试验证。首先,对变电站的告警信息日志按照时间间隔 $\Delta T=1$ h划分事件序列,得到共计1 372 307条事件序列,分别计算其风险时间序列。其次,对照电网调度部门故障记录日志,查找故障对应的事件序列,得到共计316条故障事件序列。采用风险阈值 $c=5$,得到召回率为 $R=0.89$,假正例率 $FPR=0.04$ 。不同的阈值 c 取值下,识别电网异常事件的召回率和假正例率具体数据如表4所示。通常,异常识别应用对于召回率更加看重,所以在人工校核成本可接受范围内,应当选取召回率较高的阈值 c 。

表4 异常序列识别模型在测试集上的性能

Table 4 Performance of anomaly sequence detection model on test set

风险阈值	召回率	假正例率
1	0.99	0.43
2	0.95	0.15
4	0.90	0.06
5	0.89	0.04
10	0.74	0.02
20	0.53	0.01

为验证风险阈值 c 的影响,采用受试者工作特性(receiver operating characteristic, ROC)曲线检验本文模型的性能,如图5所示。图5中的每个点都是一个可能的风险阈值 c 的取值,ROC曲线同

样可以指导阈值 c 的选取。如要求模型的召回率尽可能高的同时,希望假正例率尽可能低,为了找到符合要求的阈值取值,则应在图5所示的ROC曲线的左上段选取。但模型性能的权衡始终取决于具体业务的需求和成本的考量,如故障识别的场景对于召回率更加看重,因此,在人工校核成本可接受范围内,应当选取召回率较高的阈值 c 。

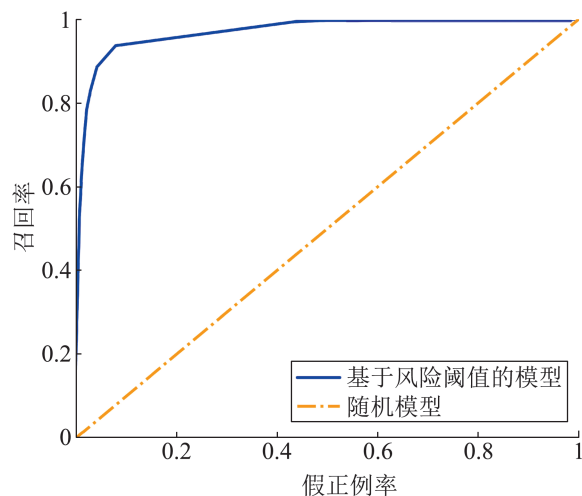


图5 不同模型在测试数据集上的ROC曲线

Fig. 5 ROC curve of different model on test data set

4 结论

本文对电网日志信息进行了分析,建立了基于告警信息的电网离散事件分析模型。事件转移概率模型和事件序列风险计算方法以事件序列产生概率为异常性指标,可以在不考虑系统量测数据的前提下,通过统计分析系统中各类设备上传到主站的告警信息,有效地刻画变电站运行异常水平。本文基于某变电站实际告警信息及对应的故障案例验证了本文方法的有效性。本文方法能识别出未曾被调度员记录为故障的异常告警信息序列,保证了异常告警信息序列不漏报,同时能够发掘海量事件序列中的少数异常序列,有助于专业人员分析研判电网设备运行态势,发现潜在风险,提升设备运维效率。本文模型和方法,实现方式简单,计算复杂度低,可在线更新。

在后续的研究中, 需要考虑如何结合关键异常类型特点构建异常事件序列识别器, 灵活处理不同故障序列之间的事件链长度差异, 提升风险分析的准确性。可以基于本文所述的风险分析结果研究包含多个变电站的大型电网异常状态识别或连锁故障诊断, 提升大电网运维效率与安全水平。

参考文献:

- [1] 魏利胜, 张倩. 基于改进的UKF智能电网虚假数据攻击检测[J]. 系统仿真学报, 2023, 35(7): 1508-1516.
Wei Lisheng, Zhang Qian. Detection of False Data Injection Attack in Smart Grid Based on Improved UKF [J]. Journal of System Simulation, 2023, 35(7): 1508-1516.
- [2] 刘兴华, 董思文, 田佳强. 基于卡尔曼滤波分解的配电网安全状态估计[J]. 系统仿真学报, 2023, 35(6): 1191-1202.
Liu Xinghua, Dong Siwen, Tian Jiaqiang. Secure State Estimation of Distribution Network Based on Kalman Filter Decomposition[J]. Journal of System Simulation, 2023, 35(6): 1191-1202.
- [3] 汪崔洋. 基于告警信号文本挖掘的电力调控智能分析与故障诊断[D]. 杭州: 浙江大学, 2019.
Wang Cuiyang. Intelligent Analysis and Fault Diagnosis of Power Dispatching Based on Alarm Signal Text Mining[D]. Hangzhou: Zhejiang University, 2019.
- [4] Zhu Yongli, Huo Limin, Lu Jinling. Bayesian Networks-based Approach for Power Systems Fault Diagnosis[J]. IEEE Transactions on Power Delivery, 2006, 21(2): 634-639.
- [5] Lo K L, Ng H S, Trecat J. Extended Petri Net Models for Fault Diagnosis for Substation Automation[J]. IEE Proceedings-generation, Transmission and Distribution, 1999, 146(3): 229-234.
- [6] Fukui C, Kawakami J. An Expert System for Fault Section Estimation Using Information From Protective Relays and Circuit Breakers[J]. IEEE Power Engineering Review, 1986, PER-6(10): 29-30.
- [7] 邓甜甜, 熊荫乔, 何贤浩. 一种基于时序性告警的新型聚类算法[J]. 计算机科学, 2020, 47(增1): 440-443, 473.
Deng Tiantian, Xiong Yinqiao, He Xianhao. Novel Clustering Algorithm Based on Timing-featured Alarms [J]. Computer Science, 2020, 47(S1): 440-443, 473.
- [8] 林斌, 曾瑛, 李星南, 等. 基于近邻传播聚类的电力通信告警分析方法[J]. 电子设计工程, 2016, 24(16): 142-145.
Lin Bin, Zeng Ying, Li Xingnan, et al. An Analysis Method of Power Communication Alarm Information Based on Affinity Propagation Clustering[J]. Electronic Design Engineering, 2016, 24(16): 142-145.
- [9] 郭云. 基于深度学习的电网故障诊断[D]. 北京: 华北电力大学(北京), 2020.
Guo Yun. Power System Fault Diagnosis Based on Deep Learning[D]. Beijing: North China Electric Power University(Beijing), 2020.
- [10] Andrade J R, Rocha C, Silva R, et al. Data-driven Anomaly Detection and Event Log Profiling of SCADA Alarms[J]. IEEE Access, 2022, 10: 73758-73773.
- [11] 刘家瑞, 杨国田, 王孝伟. 基于孪生深度神经网络的风电机组故障诊断方法[J]. 系统仿真学报, 2022, 34(11): 2348-2358.
Liu Jiarui, Yang Guotian, Wang Xiaowei. A Wind Turbine Fault Diagnosis Method Based on Siamese Deep Neural Network[J]. Journal of System Simulation, 2022, 34(11): 2348-2358.
- [12] 杜修明, 秦佳峰, 郭诗瑶, 等. 电力设备典型故障案例的文本挖掘[J]. 高电压技术, 2018, 44(4): 1078-1084.
Du Xiuming, Qin Jiafeng, Guo Shiyao, et al. Text Mining of Typical Defects in Power Equipment[J]. High Voltage Engineering, 2018, 44(4): 1078-1084.
- [13] 王哲. 基于文本挖掘的电网告警信息智能识别及应用研究[D]. 济南: 山东大学, 2021.
Wang Zhe. Research on Intelligent Recognition and Application of Alarm Information in Power Grid Based on Text Mining[D]. Ji'nan: Shandong University, 2021.
- [14] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
Zhou Zhihua. Machine Learning[M]. Beijing: Tsinghua University Press, 2016.
- [15] 深圳供电局有限公司. 深圳电网调度及厂站监控系统信息采集技术管理细则[R]. [出版地不详]: 深圳供电局有限公司, 2019.