

基于数据分析的配电网故障数据特征变量提取

杨凤生¹,熊波¹,蔡广林²,杨琦岑²

(1. 贵阳供电局, 贵州 贵阳, 550004; 2. 广州思泰信息技术有限公司, 广东 广州, 511493)

摘要:配电网线路故障预测是提升配电网可靠性指标的重要手段,为了构建性能稳定、预测能力强的线路故障预测模型,需要保证模型输入特征变量的有效性、强相关性和无冗余性。为合理确定线路故障预测模型的输入特征变量,本文采用数据探索和挖掘的分析方法对馈线故障及其影响因素之间的关系进行了分析研究,以皮尔森相关系数为计算指标,对大量实际馈线故障数据与其影响因素进行相关性统计,从馈线故障的时间-地域特性、外部影响因素、自相关特性、运行影响因素等四个维度筛选出了馈线故障影响因素特征变量作为馈线故障预测模型的输入变量,直观有效地剔除无关故障特征变量。因此,所提出方法可用于配电网大数据的预处理分析和提取,为配电网故障预测提供重要方法和数据基础。

关键词:故障预测;特征变量;可靠性

文章编号:2096-4633(2018)03-0022-05 **中图分类号:**TM7 **文献标志码:**B

配电网运行可靠性是配电网运行维护工作的一项重要指标,直接关系到家家户户的用电可靠性^[1-3]。现有研究工作中,对配电网可靠性的研究主要有配电网可靠性评估、配电网可靠性预测两种。配电网可靠性评估基于配电网结构和多年的元件可靠性参数历史数据为基础,对配电系统建立可靠性模型,利用评估算法求得配电网可靠性指标^[4-5]。但是设备可靠性参数往往难以得到,且存在建模困难、计算量大的问题。配电网可靠性预测根据配电网历史可靠性指标的变化趋势应用灰色预测、回归预测和趋势预测等方法进行可靠性指标预测^[6]。但其不能反映可靠性指标受相关因素的影响,有学者进一步研究了考虑相关因素为自变量、可靠性指标为因变量的预测的方法^[7-8]。然而,可靠性指标不能够直观地感受实际的系统表现。在实际运行中,配电网运行人员希望采取更具有前瞻性的方法,能够预先判断配电网是否可能发生故障,从而在故障发生前采取有效措施。在此背景下,配电网故障预测应运而生。

馈线故障预测首先对所需的数据进行收集及预处理,进而获取完善有效的数据并基于所获得的数据进行预测模型的构建。其中,构建性能稳定、预测能力强的故障预测模型必须保证模型的输入特征变量是有效、强相关和无冗余的,目前研究工作中对输入变量的特征分析并无过多研究,只是笼统的将认为相关的变量作为输入变量,该做法可能造成信息冗余而

影响预测模型。为此,基于配电网故障预测的科学性原则,本文首先对馈线故障影响因素的数据进行探索分析,发掘馈线故障及其影响因素之间的相关性,进一步地,将故障预测模型的输入特征变量进行分类,最终提出配电网馈线故障预测的特征变量集,以期提高配电网馈线故障预测的科学性。

1 馈线故障影响因素数据分析方法

本文通过数据探索的方法对馈线故障及其影响因素之间的关系进行分析,初步筛选出馈线故障影响因素特征变量(以下简称“故障特征变量”),为进一步采用特征选择算法筛选出最优故障特征变量奠定基础。

数据探索分析是通过绘制图表、统计、相关性分析、数据挖掘、大数据技术等方法探索数据规律的一种方法^[9-11]。数据探索有三个特点:一是让数据“说话”,强调从数据出发,探索数据之间的内在规律,而不是传统的假设后验证结论。二是不拘泥与传统的统计方法,分析方法完全根据数据进行选择。三是数据分析结果倾向与采用数据可视化方法,直观地发现数据的内在价值。

在相关性分析方法上,本文采用皮尔森相关系数进行相关性定量分析^[12-15]。假设变量 X ,其对应的另一变量为 Y ,则皮尔森相关系数计算方式如下。

$$r_{XY} = \frac{COV(X,Y)}{\sqrt{Var(X) \times Var(Y)}} \quad (1)$$

其中,变量 X 和 Y 的相关程度的划分为: $|r| < 0.4$ 为低度线性相关, $0.4 \leq |r| \leq 0.7$ 为显著线性相关, $0.7 \leq |r| < 1$ 为高度线性相关。

2 馈线故障的时间-地域特性

为进行馈线故障的特性分析,本文调研某地区120条馈线数据作为研究对象,基于这些馈线的数据进行统计、相关性分析。本节通过时间和空间两个维度,分析馈线故障在不同月份、不同地域上的差异性,并进行相关性分析,具体如下。

2.1 时间特性分析

时间特性分析目的一方面在于探索馈线故障的发生在不同月份是否存在差异性,另一方面,在于分析馈线故障的历史相关性(即馈线故障和之前月份发生故障的关联性)。

图1和图2分别给出了2014年1月至12月的月平均故障数直方图、2014年4月至2015年6月这14个月中120条馈线的第 n 月故障数与其 $n-i$ ($i=1,2,3$) 月故障数之间的散点关系图。

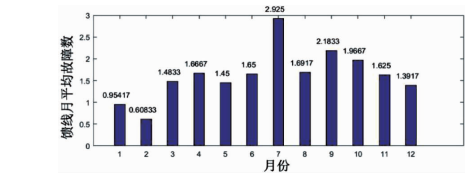


图1 馈线故障时间特性分析

Fig.1 Analysis of time feature for feeder outage

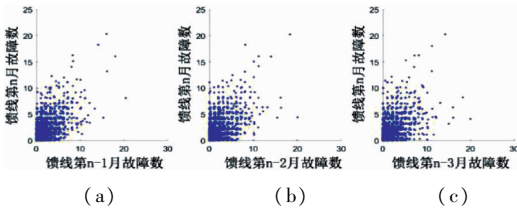


图2 馈线故障自相关性分析

Fig.2 Self correlation analysis of feeder outage

从图1可知,馈线月平均故障数在7和9月最高、1和2月最低,在不同月份馈线月平均故障数具有差异性,馈线月平均故障次数的大小和时间存在关系。因此,将“故障时间”以月为单位作为一个故障特征变量,并依据不同月份对应的馈线月平均故障数的大小对月份划分等级,结果如表1。

从图2可直观看出来馈线第 n 月故障数与第 $n-1$ 、 $n-2$ 、 $n-3$ 月故障数存在相关关系,基于式(1)可得,图2(c)中馈线第 n 月故障数与其第 $n-i$ ($i=1,2,3$) 月馈线故障数的线性相关系数 $|r|$ 分别为:

0.47、0.53、0.59,都为显著线性相关。因此,可初步将馈线第 $n-1$ 、 $n-2$ 、 $n-3$ 月的故障数作为故障特征变量。特别说明的是,此处不进行更早之前的故障自相关特性分析,是因为这类故障自相关特性的特征变量之间存在互相关性,分析提取太多的特征变量会增加不必要的冗余特征。

表1 故障发生时间分级结果

Tab.1 Classification results of outage occurrence time

月 份	馈线月平均故障次数范围	分级
1、2	[0,1)	1
3、4、5、6、8、11、2	[1,1.75)	2
9、10	[1.75,2.5)	3
7	[2.5,+∞)	4

2.2 地域特征性分析

因不同馈线所属变电站对应的地域条件存在较大的差异,所以本文采用特征变量“馈线所属变电站”使得馈线地理特性特征化,并绘制已收集数据中的19个变电站的馈线月平均故障数直方图如图3所示。

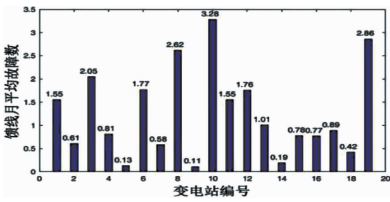


图2 馈线故障地理特性分析

Fig.2 Analysis of the geographical feature for feeder outage

由图2可知,不同的变电站所对应的地域条件下馈线月平均故障数具有较大的差异,馈线故障发生情况存在地域差异特性。因此,将“馈线所属变电站”作为故障特征变量,并依据变电站所对应的馈线月平均故障数的大小对“馈线所属变电站”划分等级,结果如表2所示。

表2 馈线故障地理特性分级结果

Tab.2 Classification results of geographical feature for feeder outage

变电站编号	馈线月平均故障次数范围	分级
5、9、14、18	[0,0.5)	1
2、4、7、15、16、17	[0.5,1)	2
1、6、11、12、13	[1,2)	3
3、8、10、19	[2,+∞)	4

3 馈线故障的外部环境影响因素

在馈线故障与外部影响因素的关系分析中,外部影响因素主要指的是外界气温及恶劣天气(如降水、雷暴、大风等)的影响。分别分析如下。

3.1 外界气温与馈线故障的相关性分析

绘制在不同的温度特征变量和 120 馈线 2014 年 1 月至 2015 年 6 月时间段中馈线月故障数变量之间的散点关系图,如图 4 所示。

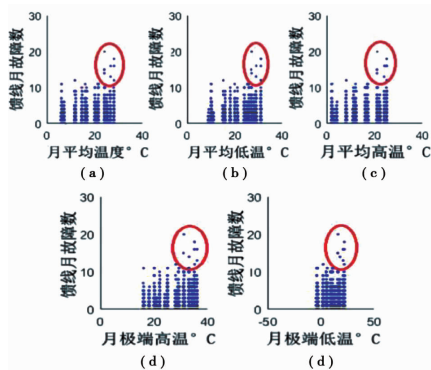


图 4 馈线故障与温度特征变量的关系分析

Fig. 4 Analysis of relationship between feeder outage and temperature feature variables

从图 4 中可发现只有当各类温度特征变量的值较高的情况下,才出现有馈线处于较高的月故障次数。如图 4(a)中当月平均温度大于 24°C 时,部分馈线才开始出现月故障数大于 10 次/月的情况。证明温度的升高对馈线的故障有一定的影响,因此将此六类温度特征变量作为故障特征变量。

3.2 恶劣天气与馈线故障的相关性分析

统计降水量、雷暴日、大风日三类恶劣天气特征不同等级下 120 条馈线 2014 年 1 月至 2015 年 6 月的月平均故障数,绘制直方图如图 5。

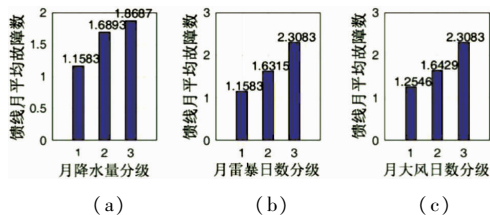


图 5 三类恶劣天气下馈线故障情况分析

Fig. 5 Analysis of feeder outage in three kinds of bad weather

由图 5 发现随着月降水量上升、月雷暴日和大风日数增多,馈线月平均故障数也逐渐增大,呈正相关关系。因此,将此“月降水量分级”、“月雷暴日数分级”、“月大风日数分级”作为故障特征变量。

4 馈线故障的自相关特性

为分析馈线故障与自身影响因素的关系,本节开展了馈线故障与设备投运时间、设备数量及分段绝缘导线/电缆总长度的相关性分析。

4.1 馈线故障与设备投运时间的关系

已有的研究表明设备故障率与投运时间的关系呈现浴盆曲线。设备在投运初期 I 阶段故障率较高,然后逐渐下降到了 II 阶段其故障率较为平稳,最后随着投运时间的增长进入 III 阶段故障率开始快速上升。依据馈线上各类设备的平均投运时间,以月为单位绘制馈线上各类设备和馈线月故障次数的散点关系图如图 6。

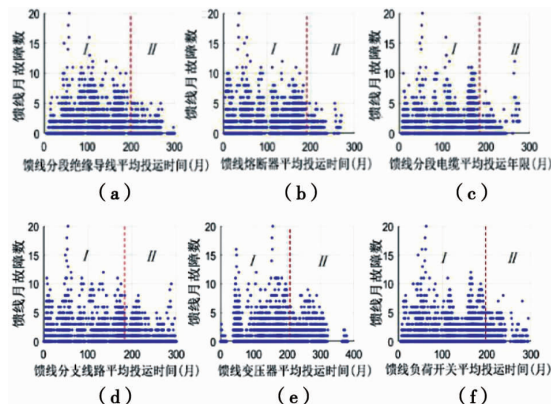


图 6 馈线故障与设备投运时间的关系分析

Fig. 6 Analysis of relationship between feeder outage and equipment operation time

由图 6 可知,各类设备的平均投运时间与馈线月故障数的关系基本符合设备故障率的“浴盆曲线”规律。在设备的投运初期即 I 阶段,馈线月故障发生次数较大的情况发生较多;随后进入 II 阶段后馈线月故障发生次数随着设备投运时间逐渐增大后保持平稳。本文所采集数据中未呈现 III 阶段的故障规律,可能是因为设备还未进入第 III 阶段时电力公司已更换设备。鉴于馈线故障与设备投运时间存在的非线性关系,将各类设备的投运时间作为故障特征变量。

4.2 馈线故障与设备数量的关系

基于设备数量越多馈线累加的故障率越高的假设,本文进行实际的数据分析。绘制 120 条馈线 2014 年 1 月至 2015 年 6 月的月故障数与设备数量的散点关系图如图 7。

分别计算图 7 中馈线月故障数与馈线分支线路数、变压器数、分段绝缘导线数、熔断器数的相关系数 $|r|$ 为 0.52、0.63、0.63、0.62,都为显著线性相关。而馈线月故障数与负荷开关、分段电缆数量的相关系数 $|r|$ 分别为 0.26、-0.12,为弱线性相关。因此仅将馈线分支线路数、变压器数、分段绝缘导线数、熔断器数初步选定作为故障特征变量,而弱线性相

关的故障特征变量进行剔除。

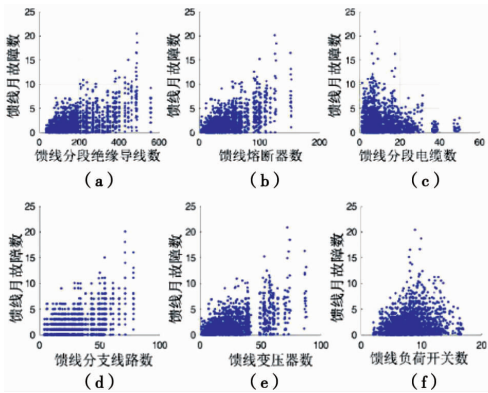


图7 馈线故障与设备数量的关系分析

Fig. 7 Analysis of relationship between feeder outage and equipment number

4.3 馈线故障与绝缘导线/电缆长度的关系

本小节分析“馈线绝缘导线长度”和“馈线电缆长度”特征变量与馈线月故障发生次数的关系。绘制120条馈线在2014年1月至2015年6月的故障发生次数与这两个特征变量的三维散点关系图如图8所示。

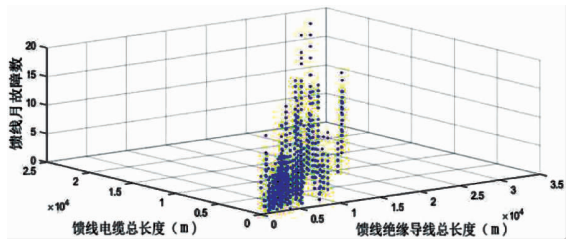


图8 馈线故障与绝缘导线/电缆长度的关系分析

Fig. 8 Analysis of relationship between feeder outage and insulated wire/cable length

从图8中可直观看出随着电缆长度和绝缘导线长度增加,对应馈线的月故障发生次数有上升趋势。然后,分别计算图8中馈线月故障数和馈线绝缘导线长度、电缆长度的相关系数分别为0.62、0.59,证明是显著线性相关。因此选取“馈线绝缘导线长度”和“馈线电缆长度”为故障特征变量。

5 馈线运况对馈线故障的影响

馈线故障的运行影响因素主要是提取了“馈线月平均负荷”和“馈线月最大负荷”为特征变量。绘制120条馈线2014年1月至2015年6月的两个特征变量与月故障数的三维散点关系图进行分析。

从图9可发现随着馈线月平均负荷和馈线月最大负荷增大,对应的大部分馈线月故障发生次数在上升,且在月最大负荷较大时发生的才存在的馈线月故

障发生次数较大的情况;进一步计算馈线月故障数和月平均负荷、月最大负荷的相关系数 $|r|$ 分别为0.47、0.58,证明是显著线性相关。因此选取“馈线月平均负荷”和“馈线月最大负荷”为故障特征变量。

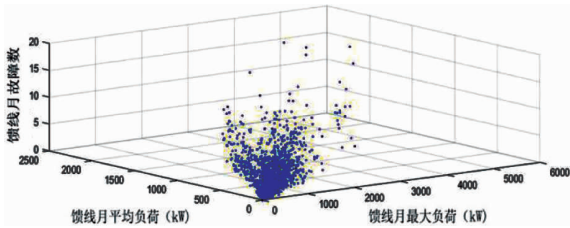


图9 馈线故障与运行影响因素的关系分析

Fig. 9 Analysis of relationship between feeder outage and operational influential factors

6 结语

本文通过数据探索分析对数据预处理所得的故障特征变量进行初步筛选,通过统计、制图、相关性分析等方法直观有效地剔除无关故障特征变量。提出了四个维度的故障特征变量:馈线故障时间-地域特性(故障时间、 $n-1$ 、 $n-2$ 、 $n-3$ 月故障数、馈线所属变电站);外部环境影响因素(月平均气温、月平均高/低温、极端高/低温数据、月雷暴日分级、月大风日分级);自身影响因素(熔断器平均投运时间、分段电缆平均投运时间、负荷开关平均投运时间、变压器数、熔断器数、分段绝缘导线长度);运行影响因素(馈线月平均负荷、馈线月最大负荷。结果为构建馈线故障预测模型提供了基础。

参考文献:

[1] BIE Z, ZHANG P, LI G, et al. Reliability evaluation of active distribution systems including microgrids[J]. IEEE Transactions on Power Systems, 2012, 27(04):2342-2350.

[2] XU X, MITRA J, WANG T, et al. Evaluation of operational reliability of a microgrid using a short-term outage model[J]. IEEE Transactions on Power Systems, 2014, 29(05):2238-2247.

[3] 赵洪山,王莹莹,陈松.需求响应对配电网供电可靠性的影响[J].电力系统自动化,2015,39(17):49-55.

ZHAO Hongshan, WANG Yingying, CHEN Song. Impact of demand response on distribution system reliability[J]. Automation of Electric Power Systems, 2015, 39(17):49-55.

[4] CONTI S, NICOLOSI R, RIZZO S A. Generalized systematic approach to assess distribution system reliability with renewable distributed generators and microgrids[J]. IEEE Transactions on Power Delivery, 2011, 27(01):261-270.

[5] 赵腾,张焰,张东霞.智能配电网大数据应用技术与前景分

- 析[J]. 电网技术, 2014, 38(12): 3305 – 3312.
- ZHAO Teng, ZHANG Yan, ZHANG Dongxia. application technology of big data in smart distribution grid and its prospect analysis [J]. Power System Technology, 2014, 38(12): 3305 – 3312.
- [6] 程鹏飞, 何江, 马岩. 10 kV 城市配电网可靠性基础参数预测方法分析[J]. 广东电力, 2014, 7(08): 94 – 97 + 101.
- CHENG Pengfei, HE Jiang, MA Yan. Prediction method for reliability underlying parameters of 10 kV city distribution network [J]. Guangdong Electric Power, 2014, 7(08): 94 – 97 + 101.
- [7] 刘亚南, 卫志农, 钟淋涓, 等. 基于 PCA 和 RVM 的电网供电可靠性预测模型研究[J]. 电力系统保护与控制, 2012, 40(20): 101 – 105.
- LIU Yanan, WEI Zhinong, ZHONG Linjuan, et al. Study on the forecasting model of power supply reliability based on PCA and RVM [J]. Power System Protection & Control, 2012, 40(20): 101 – 105.
- [8] 宋云亭, 吴俊玲, 彭冬, 等. 基于 BP 神经网络的城网供电可靠性预测方法[J]. 电网技术, 2008, 32(20): 56 – 59.
- SONG Yunting, WU Junling, PENG Dong, et al. A BP neural network based method to predict power supply reliability of urban power network[J]. Power System Technology, 2008, 32(20): 56 – 59.
- [9] 袁捷. 贵州电网大数据应用探讨[J]. 电力大数据, 2017, 20(12): 4 – 7.
- YUAN jie. Discussion on application of big data in Guizhou power grid[J]. Power Systems And big Data, 2017, 20(12): 4 – 7.
- [10] 彭晓平, 段琳, 程旭明, 等. 基于电力大数据的实时流数据处理技术研究[J]. 电力大数据, 2017, 20(08): 58 – 60.
- PENG Xiaoping, DUAN Lin, CHENG Xuming. Research on real time flow data processing technology based on large power data [J]. Power Systems And big Data, 2017, 20(08): 58 – 60.
- [11] 王守相, 梁栋, 葛磊蛟. 智能配电网态势感知和态势利导关键技术[J]. 电力系统自动化, 2016, 40(12): 2 – 8.
- WANG Shuoxiang, LIANG Dong, GE Leijiao. Key technologies of situation awareness and orientation for smart distribution systems [J]. Automation of Electric Power Systems, 2016, 40(12): 2 – 8.
- [12] 刘巍, 黄翌, 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(S1): 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.
- [13] 刘科研, 盛万兴, 张东霞, 等. 智能配电网大数据应用需求和场景分析研究[J]. 中国电机工程学报, 2015, 35(02): 287 – 293.
- LIU Keyan, SHENG Wanxing, ZHANG Dongxia, et al. Big data application requirements and scenario analysis in smart distribution network [J]. Proceedings of the CSEE, 2015, 35(02): 287 – 293.
- [14] 苗新, 张东霞, 孙德栋. 在配电网中应用大数据的机遇与挑战[J]. 电网技术, 2015, 39(11): 3122 – 3127.
- MIAO Xin, ZHANG Dongxia, SUN Dedong. The opportunity and challenge of big data's application in power distribution networks [J]. Power System Technology, 2015, 39(11): 3122 – 3127.
- [15] HUGHES G. On the mean accuracy of statistical pattern recognizers [J]. IEEE Transactions on Information Theory, 1968, 14(01): 55 – 63.

收稿日期: 2018 – 02 – 17

作者简介:



杨凤生(1983), 男, 汉族, 硕士, 高级工程师。主要从事配电网运维与监测研究工作。

(本文责任编辑: 范 斌)

Characteristic variable extraction of fault data in distribution feeder based on data analysis

YANG Fengsheng, XIONG Bo, CAI Guanglin, YANG Qichen

(1. Guiyang Power Supply Bureau of Guizhou Power Grid Co., Ltd., Guiyang 550004 Guizhou, China;

2. Guangzhou Si Tai Information Technology Co., Ltd., Guangzhou 511493 Guangdong, China)

Abstract: Distribution feeder fault prediction is an important way to improve the distribution network reliability, in order to construct a feeder fault prediction model with stable performance and strong forecasting ability, it is necessary to ensure the validity, strong correlation and non – redundancy of the model input feature variables. For reasonable determine the characteristics of input variables of the feeder fault prediction model, this article adopts the analysis method of data exploration and mining of the relationship between the feeder fault and its influencing factors are analyzed, Specifically, through correlation analysis of a large number of actual feeder fault data based on Pearson Correlation Coefficient, The correlation statistics of large number of actual feeder fault data and its influencing factors were carried out. From four dimensions of feeder fault such as time – regional characteristics, external influencing factors, self – influencing factors, and operation influencing factors, the characteristic variables of feeder fault influence factors are selected as input variables of feeder fault prediction model, The independent fault feature variables are eliminated intuitively and effectively. Therefore, the proposed method can be used for pretreatment analysis and extraction of big data of distribution network, which provides important method and data base for distribution network fault prediction.

Key words: fault prediction; characteristic variable; reliability