

供电企业基于三层分析模型的线损异常分析及处理研究

邵 丹,石立彬,史静远,郭晓松,秦晓丹

(国网邢台供电公司,河北 邢台 054001)

摘 要:线损精益化管理一直都是国家电网公司一项重点工作,近几年线损数据获取技术日趋成熟,但是分析方法依旧匮乏,严重影响了高损治理的效果。为解决电网运行中高损治理不佳的问题,我们提出了用大数据分析方法来传统线损管理问题——供电企业基于三层分析模型的线损异常分析及处理研究。首先,搭建了一个由 Hampel 抗差算法、加权皮尔逊算法和随机森林算法三种不同算法构成的三层线损异常分析模型;通过该模型,可以结合不同用户用电量大小不一的特点,自上而下的分析大电量异常用户和小电量异常用户。然后将该模型嵌入公司已有的线损监控平台中,可以实现对全省线损数据的实时监测、有效挖掘、深度分析、精准定位和工单管控,形成基于三层分析模型的线损异常分析及处理新方法。该方法对分析处理 10 kV 高损线路和 0.4 kV 高损台区中效果显著。

关键词:汉佩尔抗差算法;加权皮尔逊算法;随机森林算法;三层线损分析模型;线损监控平台

文章编号:2096-4633(2019)10-0078-06 **中图分类号:**TM744 **文献标志码:**B

电能是关系着国家经济命脉的核心资源,也是国家电网公司经营、发展的重要基础。然而,近几年来随着用电量的快速增长、配网结构复杂多样化,电能损耗(又称线损)问题日益突出,阻碍了我国电网持续、稳定、健康发展。

线损是供电企业一项综合性技术经济指标,它可以反映出供电企业的电网规划、生产技术和运行管理水平,长期以来受到各级供电企业的高度重视。线损可分为理论线损和管理线损,其中理论线损指电网各元件的电能损耗,管理线损指因管理因素造成的电能损耗。线损作为核定输配电成本的重要因素,增损风险与降损受益均由电网企业承担。

随着智能电表的推广应用,用电信息采集系统建设的有力推进,以及配用电信息融合集成技术、互动技术、大数据技术的日益成熟,为线损管理向智能化、精益化、互动化方向发展提供了坚强支撑。当发生采集装置故障、采集数据传输失败、线变关系错误、偷电漏电等情况都会引发线损异常^[1-3]。如何根据各计量装置采集的数据,进一步分析造成高损异常的原因、定位电量异常用户,是各级供电企业面临的实际问题。为减少电网运行成本,实现降损增效的工作目标,国网公司对电网节能降损工作提出了更高的要求。但是影响线损的因素较多,实现

“对症下药”难度极高,公司面临的实际问题有以下几点:

(1)线损数据统计由月到日,数据量剧增,无法筛选有价值信息;

(2)传统线损异常分析方法落后,难以定位异常用户和原因;

(3)传统分析方法查全率、准确度低(50% - 65%),现场排查工作难度较大、效率低。

针对以上问题和痛点,通过广泛调研行业现状、收集各专业数据信息、深入应用大数据分析方法,构建了线损异常分析模型,搭建了线损监控平台,形成了基于三成分析模型的线损异常分析及处理新方法,实现了对线损数据的实时监测、有效挖掘、深度分析、精准定位和工单管控。

1 线损异常分析及处理的核心

该方法的核心是基于三层数据过滤分析的数学模型,主要包括:第一层数据抗差修复,通过 Hampel 抗差算法对系统提取的数据进行坏数据辨识和抗差修复;第二层大电量用户异常筛查,通过加权皮尔逊算法对造成线损曲线波动的大电量用户进行筛查分析;第三层是小电量用户异常筛查,通过随机森林算法对小电量用户潜在的异常用电行为进行研判。

1.1 数据特征变量提取

引起线损异常的原因主要有以下几种情况:电量采集异常、接线/设备故障、窃电、功率因数异常、运行方式问题、供带关系错误等。为精准识别上述异常,需要在各系统中广泛提取关联性较强的特征变量,并结合关联规则算法进行分类和降维。考虑到系统现有的数据,从以下 3 个维度进行分析:表计电量变化特征、计量异常事件、历史异常用电行为^[4-5]。

用电量变化特征:首先,通过线损一体化系统、用电信息采集系统、电能量采集系统等信息系统,导出一定周期内电力用户的用电量数据。然后,结合对应的特征参数算法得出电量曲线变化特征,包括:日均用电量占比、用电量变异系数、用电量变化趋势度等。

计量异常事件:当发生计量异常时,计量装置会自动记录,并上传异常事件报文至用电信息采集系统。包括:电能表飞走、电能表开盖记录、电能表停走、电流失流、失压断相、潮流反向、负荷超容等。

历史异常用电行为:结合历史数据评估用户当前的异常用电概率,调取近一年的异常用电工单记录,匹配相关用户数据,统计发生过电量采集异常、接线/设备故障、窃电行为、功率因数异常、运行方式问题、供带关系错误等异常的次数^[6-8]。

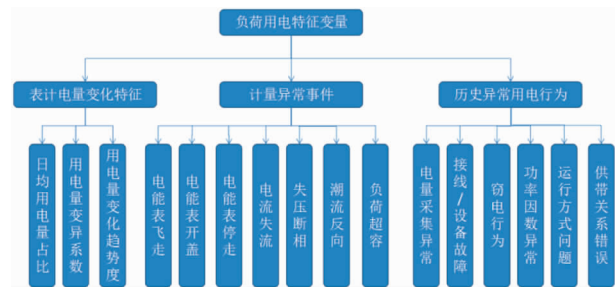


图 1 用电特征变量组成

Fig. 1 Composition of electricity characteristic variables

为生成最终数学模型,调取河北南网 2017 年 4 月至 2018 年 4 月期间 11 308 条 10 kV 线路、196 735 个台区及所带全部高低压用户的用电特征数据。总数据量约为 2.25 亿条。建立该模型后进行测试应用,整个运行时间约为 45 分钟,准确率约为 87.51%,证明该数据量足以保证后续分析结果的可靠性,同时不影响程序运算的效率。

1.2 数据预处理

基于 Hampel 算法的异常电量数据识别修复。

由于导出的电量数据可能受到因设备、系统、通信等影响,而产生的较大偏差,即异常电量数据。所以在计算前需要进行异常数据识别和修复。通过 Hampel 抗差算法构建三段截尾 ρ 函数将电量数据为四个区,记为 I、II、III、IV,如图 2 所示。

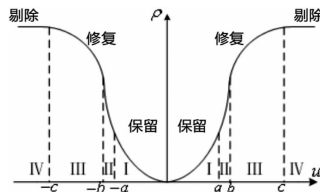


图 2 Hampel 估计法的 ρ 函数

Fig. 2 ρ function of hamper estimation method

对 I 区可靠电量数据赋予较大权重进行保留; II 区和 III 区对于偏差明显的电量数据赋予较小权重进行抗差限制; IV 区对于偏差较大的电量数据取权重因子为 0,予以淘汰。

1.3 数据分析

1.3.1 基于加权皮尔逊算法的大电量用户分析

在线损监测过程中,可以发现:一是部分 10 kV 线路线损率曲线与所带专公变电量曲线存在着线性关系;二是部分 0.4 kV 台区线损率曲线与所带低压用户电量曲线存在着线性关系。这种线性关系可以通过皮尔逊相关系数来描述^[9-10]。

考虑传统皮尔逊算法在运算过程中,不能重点体现关键节点对计算结果的影响程度。如某条线路在一段时间内的 i 日出现大幅线损波动,需要对 i 日所有专公变的用电情况进行重点分析,可以加大 i 日的用电相关度对于结果 r 值的影响,即对 i 日的计算结果进行加权,如图 3 所示。

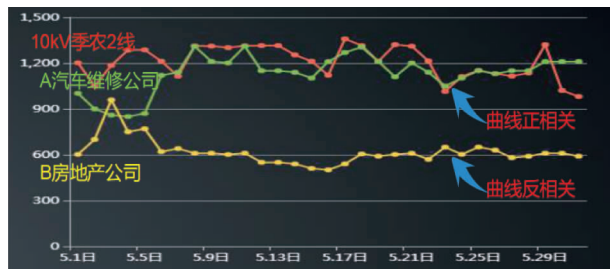


图 3 皮尔逊相关系数示意图

Fig. 3 Schematic diagram of Pearson correlation coefficient

结合加权皮尔逊算法,分析统计周期内两者的相关性,在正常情况下,每个用户的用电量曲线应该与线路损耗电量曲线相关性不是很强,即计算结果 r 值较接近 0(针对多个样本进行计算,发现 85% 以上变台的相关性 r 值在 -0.4 至 0.4 以内),如果 r

值接近于 -1 或者 1,说明存在计量异常,用户用电量会对线损产生一定影响。可以根据计算结果,有针对性的进行现场用电检查。如表 1 所示。

表 1 皮尔逊相关系数反应高损原因

Tab. 1 High-loss reason of of Pearson correlation coefficient response

r 的范围	反应的高损原因
$0.6 < r \leq 1.0$	表计倍率过小、超负荷用电、台区远端用电量过大
$0.3 < r \leq 0.6$	如负荷占比较小,难以确定是否存在异常用电行为
$-0.2 \leq r \leq 0.3$	
$-0.4 \leq r < -0.2$	
$-1.0 \leq r < -0.4$	窃电、漏电、表计倍率过大、台区轻载、频发采集失败

在实践应用的过程中,一些用户日均用电量占比较小,在计算过程中相似度 r 值在 -0.4 至 0.4 间,即使存在异常用电行为也不能对线路或台区线损曲线产生明显影响,也就是产生了特征“淹没”。如何能够挖掘这部分用户的异常用电情况,需要考虑用户用电曲线的其他特征进一步分类分析。

1.3.2 基于随机森林算法的小电量用户分析

对于加权皮尔逊相关性分析无法识别的小电量用户,需要综合考虑其它线损潜在关联因素,建立多维评价指标对用户用电特性进行分析。在相关信息系统中调取各用户用电量变化特征、计量异常事件、历史异常用电行为等 3 个维度的 16 个特征变量。选择随机森林算法建立了多个决策树,可以获得更准确和稳定的预测。

构建基于随机森林算法的线损异常分析模型,大致过程如下。

使用“Bootstrap”方法从用电特征样本库中选出 m 个样本,构成训练集;

从所有用电特征中随机选择 k 个特征,对选出的样本利用这些特征建立决策树;

重复以上两个步骤 n 次,即生成 n 棵决策树,形成随机森林模型;

对于测试集中的高损对象,经过每棵树决策,最后投票确认为存在哪一种异常用电原因。

基于随机森林算法的线损异常分析模型,利用 Bootstrap 方法对原始训练集进行随机采样,同时对 16 个特征变量进行随机采样,这两个过程都是独立进行的,进而使用并行计算来大幅提高数

据分析的效率,因此具有适应海量线损数据处理的优势^[11-17]。

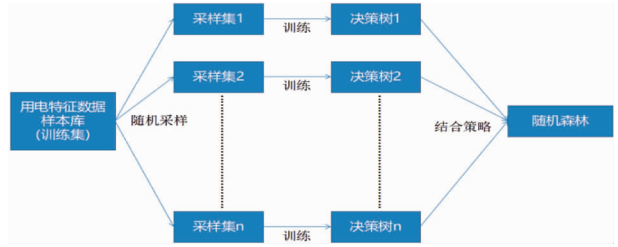


图 4 随机森林形成示意图

Fig. 4 Schematic diagram of random forest formation

1.4 模型评价

在评价模型检索能力时,查全率 (Recall) 与查准率 (Precision) 是两个最常用的指标。它们表示模型的“过滤能力”,即让有效信息“通过”,“阻止”无关信息,这里表示筛查到所有异常用户的能力。

选取存在异常用电行为的用户 500 户,不存在异常用电行为的用户 2 000 户,组合而成验证集进行模型计算。分析得出异常用户 445 户,正常用户 2 055 户,如表 2 所示。

表 2 查全率和查准率验证

Tab. 2 Verification form of recall and precision

	真异常用电用户 (实际的)	假异常用电用户 (实际的)
真异常用电用户 (预测的)	431	14
假异常用电用户 (预测的)	69	1 986

因此,经验证集测试该模型的查全率为 86.20%,查准率为 95.78%。本模型能够满足线损治理工作中异常用电分析的精准度要求,为监测人员提供更可靠的分析结果,辅助制定现场排查方案。

2 结果验证

训练集为河北南网 2017 年 4 月至 2018 年 4 月期间 11 308 条 10 kV 线路、196 735 个台区及所带所有高低压用户的用电特征数据。测试集为河北省沧州市 2018 年 5 月期间的 10 条高损线路及其所带高低压用户数据。

首先,对测试集 10 条高损线路所带专公变进行加权皮尔逊相似度分析,分别计算与线路线损率相关度、同线路下其他用户曲线相关度。结果如表 3 所示。

表 3 专公变电量曲线皮尔逊系数计算结果

Tab. 2 Pearson coefficient calculation of voltage curve of special public transformer

序号	专公变名称	电量曲线皮尔逊相关度	所属线路
1	综全中隔 005	0.893	10 kV 凤南 I 5716 线
2	混油房李变压器	0.877	徐梁东线 5314
3	运西同心电焊加工部	0.871	10 kV 南纸 IV 线(5724)
4	纸坊扬水站变压器柱上变压器	0.855	韩纸线 5613
5	综全大合庄 9 号	0.812	10 kV 华莘 II 线(56A3)
⋮	⋮	⋮	⋮
525	农马庄农用 1 变压器	-0.407	徐梁东线 5314
526	农全郑龙洼 2010046	-0.414	海尤线 558
527	原一电子有限公司	-0.499	海尤线 558
528	张窑机窑 049	-0.530	韩纸线 5613
529	市政设施管理处	-0.606	10 kV 南纸 IV 线(5724)

其中,有 31 个用户的用电量曲线与线损率曲线皮尔逊相关系数大于 0.6,即极强正相关;有 5 个用户小于 -0.4,即极强反相关。极强正相关和极强反相关电量曲线例如图 5 所示。

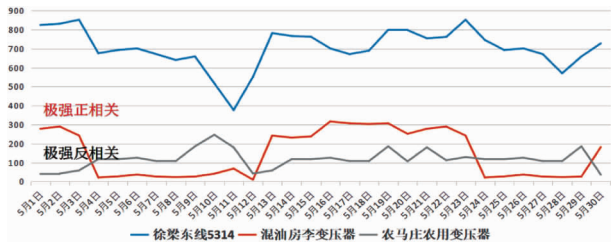


图 5 皮尔逊系数极强正相关和极强反相关示例

Fig. 5 Examples of strong positive correlation and strong negative correlation of Pearson coefficients

同理,对测试集中 6383 个低压用户用电量曲线与所属 315 个台区线损率曲线进行相关度计算。其中,有 6% 的用户电量曲线相关度大于 0.6,即极强正相关;有 1.4% 的用户电量曲线相关度小于 -0.4,即极强反相关。

然后,结合表 1 皮尔逊相关系数反应的高损原因,可以得出极强正相关和极强反相关的表计用电异常概率,进而辅助线损工作人员更具针对性的开展现场用电检查工作。

下面需要对皮尔逊相关系数在 -0.4 至 0.6 之间的用户进一步分析,以 16 种强关联因素数据作为输入,以 6 种异常用电原因作为输出,建立随机森林分析模型。利用 Bootstrap 在样本库随机抽取 W 个样本子集,用于生成 W 棵决策树。图 6 显示了当 w 取不同值

时,基于随机森林算法预测模型的预测值与实际值之间的平均绝对百分误差。在综合考虑建模速度和预测误差的情况下,取 W = 350 为最佳决策树棵数。

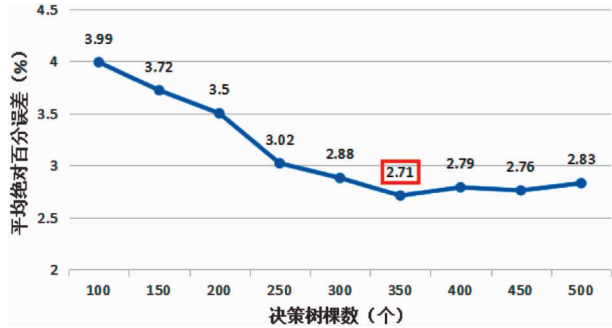


图 6 不同森林规模下的分析误差

Fig. 6 Analysis error under different forest scales

进而搭建随机森林分析模型如图 7 所示。

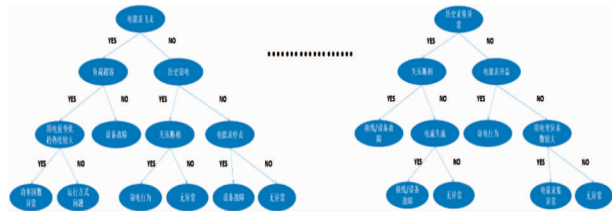


图 7 随机森林分析模型

Fig. 7 Stochastic forest analysis model

将 10 条线路下的小电量用户数据输入随机森林模型,得出异常用电行为分布如图 8 所示。

可见,大部分线路高损主要原因在于电量采集异常,同时接线和设备故障问题也是重要影响因素之一。以此分析结果,可以得出各条线路的高低电压用户异常用电情况,结合配网 GIS 地图进一步锁定

异常用电位置。

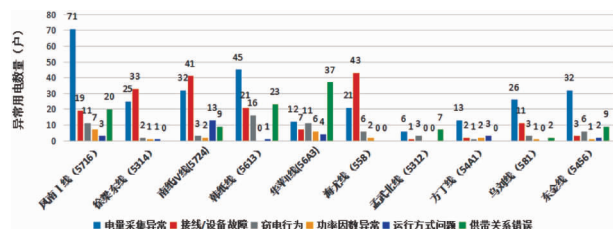


图8 用电异常随机森林分析模型

Fig. 8 Stochastic forest analysis model for electricity abnormality

基于以上三层线损异常分析模型建立了线损监控平台,具有线损指标监测、高损对象筛查、异常用电分析、线损工单管理等功能,效果如图9所示。



图9 线损监控平台可视化界面

Fig. 9 Visualization interface of line loss monitoring platform

3 结论

3.1 创新数据提取方法,有效挖价值数据

该方法在各线损管理系统中广泛提取数据信息,结合关联规则算法进行线损异常关联性分类和PCA降维,进一步提取3个类别16种强相关特征变量进行分析,降低数据提取的难度,提升了分析模型的精度;

3.2 创新线损异常分析模型,精准定位异常用户和原因

该方法通过有效融合Hampel抗差算法、加权皮尔逊算法、随机森林算法建立线损异常分析模型。首先,实现数据清洗、过滤和修复,保留有效数据;然后利用高压用户电量大的特点,通过用电量与线损率的相关性分析,实现高压异常用户的分析;最后,对剩余低压用户用电量小的特点,选取用户档案信息、计量装置信息、采集信息、异常事件信息等更多用户数据开展分析,实现低压异常用户的分析。模型的三种不同算法,是根据数据特点、用户用电量大小而对应选取的,三层分析,自上而下,逐层开展,实现异常用户全覆盖,异常问题准确定位。

3 独创搭建线损监控平台,优化线损治理过程

以线损异常分析模型为核心,搭建线损监控平

台,与国网推广的线损系统建立数据接口,实现了线损指标实时监测、高损对象筛查、异常用电分析、线损工单管理等,并制定降损方案,提升现场排查工作效率,促进线损治理效果。

4 应用成效

该方法在河北南网各城网及各县公司进行测试应用。目前已经累计应用一年。分析治理高损用户13 290余户,成功降损10 kV线路1 031条、0.4 kV台区9 632个,避免电量流失2 000万度,挽回经济损失约1 000万元,大幅提升了高损异常分析、治理效率和效果。该方法从供电公司重点关注的线损管理问题出发,关注高损对象的数据变化情况,对影响线损波动的因素进行了深入发掘,能够有效发现线损管理的薄弱环节,实现全方位的线损数据监测和分析,为公司挽回电能损耗和经济损失,实现企业效益最大化,促进供电企业健康、稳定的发展。

参考文献:

- [1] 耿俊成,张小斐,袁少光,等. 基于大数据分析的电网设备质量评价[J]. 电力大数据,2018,21(05):36-40.
GENG Juncheng, ZHANG Xiaofei, YUAN Shaoguang, et al. Quality evaluation of power grid equipment based on big data analysis [J]. Power Systems and Big Data, 2018, 21(05): 36-40.
- [2] 辛苗苗,张延迟,解大. 基于电力大数据的用户用电行为分析研究综述[J]. 电气自动化,2019,41(01):1-4+27.
XIN Miaomiao, ZHANG Yanchi, XIE Da. Summary of researches on consumer behavior analysis based on big power data [J]. Electrical Automation, 2019, 41(01): 1-4+27.
- [3] 孙利雄,秦锐,崔大铭. 基于运行状态和寿命的设备全寿命期评价分析研究[J]. 电力大数据,2017,20(09):10-13.
SUN Lixiong, QIN Kun, CUI Daming, et al. Evaluation and analysis of equipment life cycle based on running state and life [J]. Power Systems and Big Data, 2017, 20(09): 10-13.
- [4] 牟婷婷,陆微,王兰君,辛洁晴. 基于主成分分析的用电模式稳定性分析[J]. 电力系统自动化,2017,41(19):59-65.
MOU Tingting, LU Wei, WANG Lanjun, XIN Jieqing. Stability Analysis of Consumption Mode Based on Principal Component Analysis [J]. Power System Automation, 2017, 41(19): 59-65.
- [5] 申庆斌,武志宏,张媛,等. 基于电力大数据的用户能效服务研究[J]. 电力需求侧管理,2017,19(04):29-31.
SHEN Qingbin, WU Zhihong, ZHANG Yuan, et al. Research on energy efficiency service of based on big data of electric power [J]. Power Demand Side Management, 2017, 19(04): 29-31.
- [6] 陈俐冰,何容,邱林,等. 电力客服中心用户行为分析研究与实现[J]. 计算机技术与发展,2017,27(02):116-119.

- + 124.
- CHEN Libing, HE Rong, QIU Lin, et al. Research and implementation of user behaviors analysis of electric power customerservice center[J]. Computer Technology and Development, 2017,27(02):116-119+124.
- [7] 沈玉玲,吕燕,陈瑞峰. 基于大数据技术的电力用户行为分析及应用现状[J]. 电气自动化,2016,38(03):50-52.
- SHEN Yuling, LU Yan, CHEN Ruifeng. Power user behavior analysis and application status based on big data technology [J]. Electrical Automation,2016,38(03):50-52.
- [8] JUN ZHU, ERIC ZHUANG, JIAN FU, et al. A framework-based approach to utility big data analytics[J]. IEEE Transactions on Power Systems,2016,31(03):2455-2462.
- [9] 刘世成,张东霞,朱朝阳. 能源互联网中大数据技术思考[J]. 电力系统自动化,2016,40(08):14-21+56.
- LIU Shicheng, ZHANG Dongxia, ZHU Chaoyang. A view on big data in energy internet [J]. Power System Automation,2016,40(08):14-21+56.
- [10] 彭小圣,邓迪元,程时杰,等. 面向智能电网应用的电力大数据关键技术[J]. 中国电机工程学报,2015,35(03):503-511.
- PENG Xiaosheng, DENG Diyuan, CHENG Shijie, et al. Key technologies of electric power big data and its application prospects in smart grid [J]. Proceedings of the CSEE,2015,35(03):503-511.
- [11] 张东霞,苗新,刘丽萍,等. 智能电网大数据技术发展研究[J]. 中国电机工程学报,2015,35(01):2-12.
- ZHANG Dongxia, MIAO Xin, LIU Liping, et al. Research on development strategy for smart grid big data [J]. Proceedings of the CSEE,2015,35(01):2-12.
- [12] 杨旭英,周明,李庚银. 智能电网下需求响应机理分析与建模综述[J]. 电网技术,2016,40(01):220-226.
- YANG Xuying, ZHOU Ming, LI Gengyin. Survey on demand response mechanism and modeling in smart grid [J]. Power System Technology,2016,40(01):220-226.
- [13] 赵刚. 大数据技术与应用实践指南[M]. 北京:电子工业出版社,2013:56-58.
- [14] BOYANG LI, MITHAT C. KISACIKOGLU, CHEN LIU, et al. Big data analytics for electric vehicle integration in green smart cities [J]. IEEE Communications Magazine,2017,55(11):19-25.
- [15] BOYANG LI, MITHAT C. KISACIKOGLU, CHEN LIU, et al. Big data analytics for electric vehicle integration in green smart cities [J]. IEEE Communications Magazine,2017,55(11):19-25.
- [16] ZHENYU ZHOU, HOUJIAN YU, CHEN XU, et al. Dependable content distribution in D2D-based cooperative vehicular networks: a big data integrated coalition game approach [J]. IEEE Transactions on Intelligent Transportation Systems,2018,19(03):953-964.
- [17] 吴克河,姚鹏. 基于 AHP-CRITIC 算法的线损四分管理后评估方法[J]. 广东电力,2018,31(02):82-87.
- WU Kehe, YAO Peng. Post-evaluation method for four-division line loss management based on AHP-CRITIC algorithm [J]. Guangdong Electric Power,2018,31(02):82-87.

收稿日期:2019-09-17

作者简介:



邵 丹(1985),女,硕士,高级工程师,主要从事电力企业运营监测、线损管理及互联网技术工作。

(本文责任编辑:范 斌)

Analysis and treatment of line loss anomalies based on three-layer analysis model in power supply enterprises

SHAO Dan, SHI Libin, SHI Jingyuan, GUO Xiaosong, QIN Xiaodan

(State Grid Xingtai Power Supply Company, Xingtai 054001 Heibei, China)

Abstract: Line loss lean management has always been a key task of the State Grid Corporation. In recent years, line loss data acquisition technology has become increasingly mature, but the analysis methods are still scarce, which seriously affects the effect of high-loss management. In order to solve the problem of poor high-loss management in power grid operation, a big data analysis method is proposed to solve the traditional line loss management problem—the line loss anomaly analysis and processing based on the three-layer analysis model of power supply enterprises. Firstly, a three-layer linear damage anomaly analysis model composed of Hampel resistance algorithm, weighted Pearson algorithm and random forest algorithm is built. this model can combine the characteristics of different users (different power consumption, and analyze the users of large power anomaly and small power anomaly from top to bottom. Then the model is embedded in the existing line loss monitoring platform of the company, which can realize real-time monitoring, effective mining, depth analysis, precise positioning and work order control of the line loss data in the province, and form a new method of line loss anomaly analysis and processing based on the three-layer analysis model. This method is effective in analyzing and treating 10 kV high loss lines and 0.4 kV high loss platform areas.

Key words: Hampel robust algorithm; weighted Pearson algorithm; stochastic forest algorithm; three-layer line loss analysis model; line loss monitoring platform