

DOI:10.12158/j.2096-3203.2020.04.023

基于预测模型的发电厂异常数据辨识方法

高骞¹, 张浩天², 汤奕³

(1. 国网江苏省电力有限公司, 江苏 南京 210024; 2. 东南大学网络空间安全学院, 江苏 南京 210096; 3. 东南大学电气工程学院, 江苏 南京 210096)

摘要:发电统计数据是我国实施电力监管的重要依据, 其中的发电厂用电率作为反映电厂生产效能、论证节能降耗情况的关键指标, 验证其上报数据的真实性和准确性十分重要。为此, 提出一种基于预测模型的发电厂异常数据辨识方法。该方法首先利用 Adaboost 改进非线性有源自回归模型 (NARX) 神经网络构建发电厂用电率预测模型, 通过不断引入上报值以动态的方式对当前时刻发电厂用电率进行预测。当发电厂用电率时间序列出现突变时, 残差时间序列会出现明显的增大或减小, 进而利用孤立森林算法得到各残差向量组的异常分值从而辨识出异常点。最后, 利用该方法对注入了虚假数据的实际发电数据进行辨识, 验证了所提方法的有效性。

关键词:发电厂用电率; 异常数据; 非线性有源自回归模型 (NARX) 神经网络; Adaboost 算法

中图分类号: TM76

文献标志码: A

文章编号: 2096-3203(2020)04-0164-07

0 引言

电力统计数据是政府实施宏观调控给、加强专业监管、制定国民经济规划以及能源、电力等专项规划的重要依据, 是研判经济和能源形势、制定产业政策的重要参考^[1]。高度重视电力监管统计工作, 不仅要系统地收集、整理、共享电力统计信息, 还要利用统计信息和数据, 深入开展综合分析和专题研究, 更好地为能源和经济决策、行业管理和监管以及“放管服”改革服务。

随着电力企业改革的不断深化和发展, 高耗能行业的节能降耗工作对解决我国能源问题至关重要, 也是实现我国“十三五”节能降耗目标的根本保证, 国家能源局和政府部门进行电力监管统计工作时十分重视节能降耗相关指标的真实有效性。但现行的发电统计模式缺乏有效的数据质量管控体系, 甚至存在部分电厂为满足节能降耗和碳排放等要求上报虚假数值, 形成异常数据的可能, 影响相关部门和企业对发电情况的认知。因此, 需针对关键指标开展异常数据辨识工作, 以监管电厂实际运行水平。

发电厂用电率是论证节能降耗情况的重要依据, 是衡量发电机组的经济性能的重要技术指标, 且与其他经济指标如供电煤耗、负荷率之间是相互联系相互影响的。如厂用电率每变化 1% 对供电煤耗的影响系数为 3.499%, 负荷率每下降 1 个百分点影响厂用电率升高 0.06 个百分点^[2]。因此, 将发电

厂用电率作为关键指标, 利用历史数据分析发电厂用电率潜在规律, 辨识上报值是否异常。

受电力负荷和调度、季节等因素的综合影响, 发电厂用电率呈现出明显时间序列特征。针对时间序列的异常检测方法包括基于频率^[3]的方法和基于机器学习的方法^[4]。前者优点是处理高效, 可以在线性的时间和空间内检测新颖模式, 但由于采用了时间序列离散化技术, 有意义的数据可能在转换的过程中丢失; 后者的基本思想是对序列进行预测, 通过预测值与实际值的差值确定异常值, 主要包括两大类: 支持向量机法^[5-6]和神经网络法^[7-9], 其中非线性有源自回归模型 (nonlinear auto regressive models with external input, NARX) 作为动态神经网络的一种, 对历史信息有记忆和联想功能, 对时间序列的时变规律特性适应能力强, 目前已经被用于风速、光伏发电功率预测等领域^[10-11]。

基于预测的异常检测方法对时间序列预测精度有很高的要求, 然而大量实验测试结果表明, 单一预测模型可能存在对序列信息挖掘不充分, 从而导致突变点处预测精度不佳以及预测滞后的现象^[12-13]。

针对以上问题, 文中引入 Adaboost 集成学习算法^[14-15], 结合 NARX 神经网络构造发电厂用电率预测模型, 以获得更加精准的发电厂用电率预测值。利用预测值和上报值的残差序列, 结合与发电厂用电率强相关的发电指标, 构造向量组, 通过孤立森林算法^[16-19]给定异常分值, 由此提出基于预测模型的异常数据辨识方法。最后通过某省网的数据开展实例分析, 验证所提方法的有效性。

收稿日期: 2020-02-25; 修回日期: 2020-03-19

基金项目: 国家自然科学基金资助项目 (51577030)

1 发电厂统计数据指标分析

发电厂在启动、运转、停机、检修过程中有大量电动机拖动的机械设备,用以保证机组本身的主要设备和输煤、除灰、除尘及水处理等辅助设备的正常运行,其总的耗电量统称为厂用电。厂用电耗电量与电厂类型、燃料种类及燃烧方式、蒸汽参数等因素有关。厂用电耗电量占发电厂全部发电量的百分数称为厂用电率。

厂用电率是电厂能耗监督的主要综合指标,是当前衡量发电机组经济运行情况、节能降耗情况的重要依据。厂用电率与其他反映电厂实际运行情况的指标之间又有着相互的影响和关联。例如,在火电厂同容量机组中,煤种发热值高低直接影响燃煤量。燃煤发热值低时燃煤量增加,直接影响磨组合及磨煤电耗。燃用较低热值煤质较中等热值时厂用电率将增加 0.2% 以上^[20]。在 70% 以上负荷时负荷率每增加 1%,厂用电率减少 0.03%;在 60%~70% 负荷范围内,负荷率每减少 1%,厂用电率增加 0.05%。

因此,将发电厂用电率作为反映电厂生产效能的研究目标,利用历史数据挖掘发电厂用电率时间序列潜在规律,通过不断引入新的上报数据来预测下一个时间点。当发电厂用电率时间序列出现突变时,残差时间序列会出现明显的增大或减小从而识别出异常值,达到监管统计的目的。

2 发电厂用电率预测模型

2.1 预测模型输入数据处理

2.1.1 特征选择

发电统计数据存在特征维度高、耦合性强等特点,且不同的机组类型存在特定的发电指标,如抽水蓄能机组包含抽水综合效率、抽水蓄能耗电量等特有指标,燃煤机组包含发电用原煤量、煤折标系数等特有指标。若将无关特征输入网络则会消耗计算性能,增加学习任务时间。因此针对不同的机组类型,首先利用相关性分析筛选与发电厂用电率关联性强的指标,减轻学习过程的负担。

2.1.2 特征降维

经特征选择剔除与结果相关性小的数据后,高维度的发电统计数据仍存在引发维度灾难的隐患。主成分分析法(principal components analysis, PCA)是一种数据压缩和特征降维的多变量统计分析技术^[21-22],对 q 项发电指标 $I(t)$ 进行 PCA 分析,计算各项发电指标的特征值与贡献率。选取预测误差

最小时的累计贡献率 R 所对应的 p 维主成分变量 $I'(t)$ 作为网络的外部输入,从而将网络外部输入维数由 q 降至 p 。

2.1.3 缺失数据处理

发电数据具有时序连续性的特点,正常情况下不会发生急剧的跳变。因此,针对部分月份缺失的发电数据,采用线性插值法对其进行修复,即:

$$X_{k+i} = X_k + i(X_{k+j} - X_k)/j \quad 0 < i < j \quad (1)$$

式中: X_k, X_{k+j}, X_{k+i} 分别为 k 时刻, $k+j$ 时刻和 $k+i$ 时刻的参数。

2.2 基于 Adaboost 的发电厂用电率预测模型

构建基于 Adaboost 改进的 NARX 主要分为 5 个步骤。

步骤 1:对外部输入时间序列 $I(t)$ 进行 PCA 降维处理,得到主成分变量 $I'(t)$,与发电厂用电率 $O(t)$ 合并为样本数据集输入 NARX 神经网络进行训练,得到预测结果 $O'(t)$ 。

步骤 2:将 $O'(t)$ 作为预测模型的输入,实际发电厂用电率 $O(t)$ 作为输出,构成样本集 $T = \{(O'(1), O(1)), (O'(2), O(2)), \dots, (O'(m), O(m))\}$ 。

步骤 3:初始化样本权重。

$$\begin{cases} D_k = (w_{k_1}, w_{k_2}, \dots, w_{k_m}) \\ w_{k_i} = \frac{1}{m} \quad i = 1, 2, \dots, m \end{cases} \quad (2)$$

式中: D_k 为第 k 次训练中 T_1 的权重分布; w_{k_i} 为第 i 个训练样本的权重。

步骤 4:网络训练包括以下 4 块内容。

(1) 采用具有权重分布 D_k 的训练样本集 T_1 训练 NARX,得到弱回归预测器 $G_k(O')$ 。

(2) 计算 $G_k(O')$ 在训练样本集 T_1 上的最大误差为:

$$E_k = \max |O(i) - G_{k_i}(O'(i))| \quad i = 1, 2, \dots, m \quad (3)$$

式中: $O'(i)$ 为第 i 个训练样本对应的实际发电厂用电率; $G_{k_i}(O'(i))$ 为 $G_k(O')$ 对第 i 个训练样本的预测结果。各训练样本线性相对误差为:

$$e_{k_i} = \frac{|O(i) - G_k(O'(i))|}{E_k} \quad (4)$$

则 $G_k(O')$ 在训练样本集 T_1 上的误差为:

$$e_k = \sum_{i=1}^m w_{k_i} e_{k_i} \quad (5)$$

式中: w_{k_i} 为第 k 次训练中第 i 个训练样本的权重。

(3) 计算 $G_k(O')$ 的系数 α_k :

$$\alpha_k = \frac{e_k}{1 - e_k} \quad (6)$$

(4) 更新样本权重 D_{k+1} , 第 $k+1$ 个弱学习器的样本集权重系数为:

$$w_{k+1,i} = \frac{w_{k,i}}{Z_k} \alpha_k^{1-e_{k,i}} \quad (7)$$

步骤 5: 将 K 个 NARX 弱学习器组合成强学习器, 并用该模型预测测试样本集 T_2 。

$$f(O') = \sum_{k=1}^K \frac{\alpha_k}{\sum_{k=1}^K \alpha_k} G_k(O') \quad (8)$$

对 Adaboost 算法每一次循环所得基本学习器 $G_k(O')$ 采取加权累加方式, 得到最终强学习器 $f(O')$ 。计算 $f(O')$ 预测误差, 当预测误差开始增加时停止循环, 将对应的循环次数作为最终 K 值。

3 基于预测模型的异常数据辨识

3.1 异常分值计算

异常分值计算流程如图 1 所示, $O(t)$ 为发电厂用电率上报值, $\hat{O}(t)$ 为发电厂用电率预测值, $U(t)$ 为二元值, 用 +1, -1 表示。

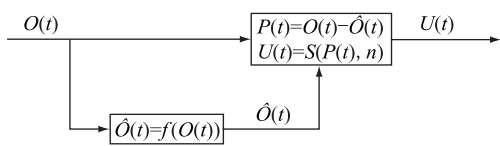


图 1 异常分值计算流程

Fig.1 Calculation process of abnormal score

利用 1.2 节所得强学习器 $f(O')$ 进行预测, 将发电厂用电率预测值 $\hat{O}(t)$ 与发电厂用电率上报值 $O(t)$ 做差得到残差时间序列 $P(t)$, 组合相关性高的其余发电指标, 利用孤立森林算法计算各向量组的异常分值 $S(P(t), n)$ 。其定义如下:

$$c(n) = 2H(n-1) - [2(n-1)/n] \quad (9)$$

$$S(P(t), n) = 0.5 - 2^{-\frac{E(h(P(t)))}{c(n)}} \quad (10)$$

式中: n 为残差向量组中样本的个数; $H(i)$ 为谐波次数; $c(n)$ 为二叉搜索树的平均路径长度, 用于 $h(P(t))$ 的标准化; $E(h(P(t)))$ 为样本点 $P(t)$ 在孤立森林中所有孤立树的路径长度的平均值。若异常分值大于阈值, 则 $U(t)$ 的值为 -1, 该点判定为异常值。

3.2 异常检测结果评价

文中以误报率 (false positive rate, FPR) 和检出率 (true positive rate, TPR) 来量化评价异常数据分析方法, 检出率和误报率的计算公式如下:

$$\mu_{\text{TPR}} = \theta_{\text{TP}} / (\theta_{\text{TP}} + \theta_{\text{FN}}) \quad (11)$$

$$\mu_{\text{FPR}} = \theta_{\text{FP}} / (\theta_{\text{FP}} + \theta_{\text{TN}}) \quad (12)$$

式中: θ_{TP} 为真阳性数量 (实际发电厂用电率为正常

检测结果为正常); θ_{FN} 为假阴性数量 (实际发电厂用电率为正常检测结果为异常); θ_{FP} 为假阳性数量 (实际发电厂用电率为异常检测结果为正常); θ_{TN} 为真阴性数量 (实际发电厂用电率为异常检测结果为异常)。具体流程如图 2 所示。

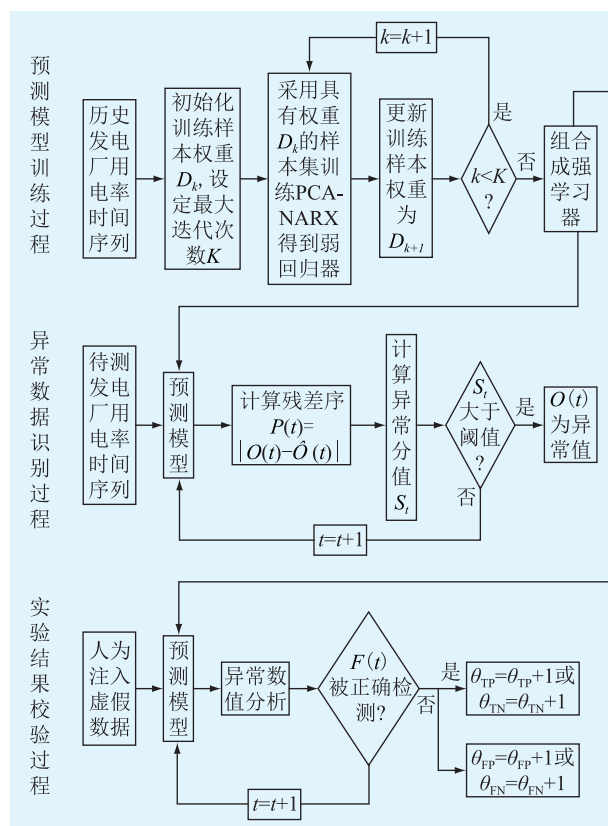


图 2 异常检测流程

Fig.2 Flowchart of anomaly detection

4 实例分析

4.1 实验数据来源

为验证所提方法的科学性和可靠性, 使用某省 2010 年 1 月至 2015 年 10 月发电厂生产情况月报表作为训练集, 筛选其中 50 家电厂的发电数据作为训练样本。以某一家燃煤电厂为例, 该电网上月报表中有 52 项发电指标, 计算发电厂用电率与各指标之间的 Pearson 相关系数 r , 如表 1 所示。

表 1 指标间相关系数

Table 1 Correlation coefficient between indicators

指标	相关系数 r
原煤量	0.972
发电标准煤耗	0.914
供电标准煤耗	0.913
⋮	⋮
热电比	0.260
供热原煤耗	0.258
入厂煤热值差	0.240

剔除 $|r| \leq 0.3$, 即不存在线性相关的指标, 筛选过的发电统计指标包括发电用原煤量、发电标准煤耗、运行小时、运行负荷率等 40 项。

4.2 预测模型参数确定

根据 1.1.2 节描述的 PCA 提取主成分的方法, 对除发电厂用电率之外的 39 项发电指标进行 PCA 主成分提取, 计算贡献率。

经过 PCA 提取的前 4 个主成分贡献率分别为 59.74%, 22.18%, 14.70% 和 3.27%, 经实验可知, 当累计贡献率 R 为 96.6% 时能反映原始数据提供的绝大部分信息。因此主成分个数 p 确定为 3, 将网络的外部输入由 40 维降低为 3 维, 优化网络的输入。

为进一步提高预测精度, 采用参数寻优方法对 NARX 网络模型进行逐步调优。在固定输入不变的情况下, 通过不断增加 NARX 网络的隐藏层神经元个数来测试不同神经元数量所起到的预测效果。不同神经元数量对预测结果精度的影响如图 3 所示。实验结果表明隐藏层神经元数量为 7 时预测效果最好, 当神经元数量增加到 8 时训练误差率上升, 表示模型过度学习, 故确定最终隐藏层神经元数量为 7。

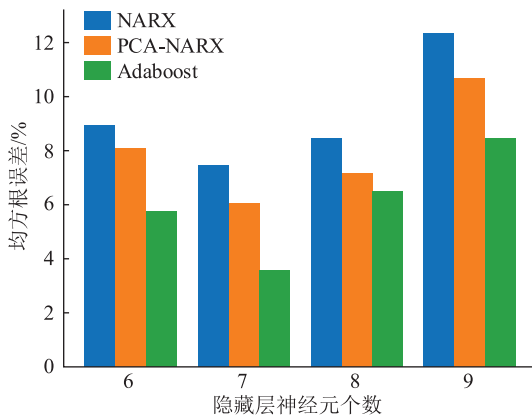


图 3 神经元个数对预测精度的影响

Fig.3 The effects of the number of neurons to the prediction accuracy

通过预测结果的根均方误差 RMSE 来确定最佳循环次数 K 。不同循环次数对预测结果精度的影响如图 4 所示。实验可知, 当循环次数为 4 时, 模型预测结果的误差最低, 因此循环次数确定为 4 次。

4.3 异常判断结果分析

分别采用训练好的单一 NARX 网络模型和经过 Adaboost 集成学习改善的模型对 2015 年 11 月至 2018 年 10 月的发电厂用电率进行预测, 2 种模型的预测结果如图 5 所示, 各模型预测结果的误差评价指标如表 2 所示。

由图 5 和表 2 可知, 使用文中方法能准确预测

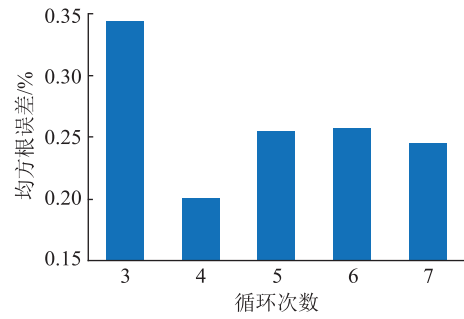


图 4 循环次数对预测精度的影响

Fig.4 The effects of loop count to the prediction accuracy

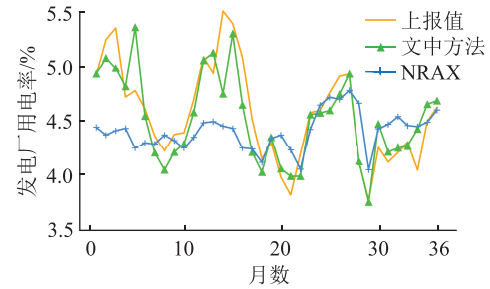


图 5 2 种方法预测结果对比

Fig.5 Comparisons of two methods for predicting results

表 2 2 种方法预测精度对比

Table 2 Comparison of prediction accuracy

模型	均方根误差 RMSE	平均绝对百分误差 MAPE/%
文中方法	0.201	3.64
NARX	0.352	7.61

发电厂用电率的变化趋势, 预测滞后现象得到改善, 且缩减了发电厂用电率突变点处的预测误差。相比于 NARX 预测模型, Adaboost 组合预测模型的均方根误差 RMSE 减小 0.151, 平均绝对百分误差 MAPE 减小 3.97%。

将发电厂用电率预测值与电厂上报值做差得到时序残差序列, 组合残差序列与相关性最高的 3 个发电指标原煤量、发电标准煤耗和供电标准煤耗构成残差向量组, 利用孤立森林算法对残差向量组进行分析, 得到各个样本点的异常分值 S 。图 6、图 7 和图 8 为发电厂用电率异常分值等高线, 样本点所处区域颜色越浅表示该点异常分值越小, 所处区域颜色越深表示该点异常分值越大。其中红点表示该点 $U(t)$ 值为 -1, 判定为异常数据; 黑点表示该点 $U(t)$ 为 +1, 判定为正常数据。

根据不同发电指标组成的残差向量组进行异常数值分析可能得到不同异常判别结果。当以残差-原煤量构成向量组时, 判别为异常数值的发电厂

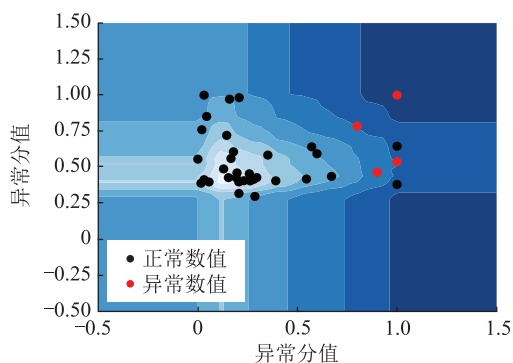


图6 残差-原煤量向量组发电厂
用电率异常分值等高线

Fig.6 Anomaly score contour based on residual and heating capacity

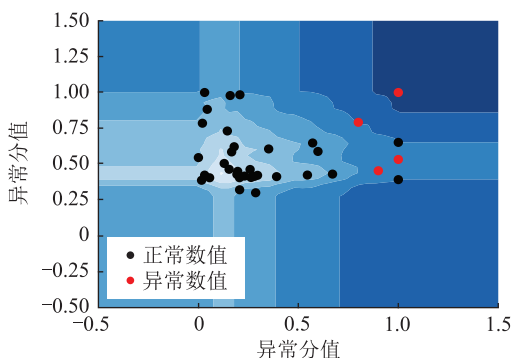


图7 残差-发电标准煤耗向量组发电厂
用电率异常分值等高线

Fig.7 Anomaly score contour based on residual and standard coal consumption rate

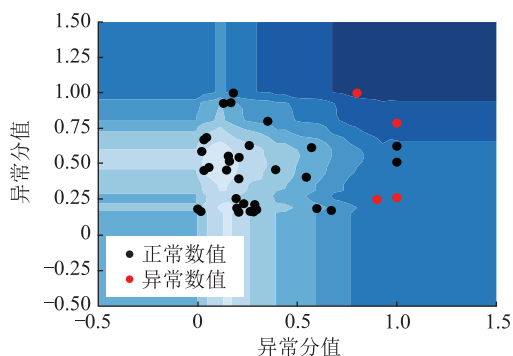


图8 残差-供电标准煤耗向量组发电厂
用电率异常分值等高线

Fig.8 Anomaly score contour based on residual and standard coal consumption rate of power supply

用电率样本点为2016年7月、2017年5月、2017年12月和2018年1月;当以残差-发电标准煤耗构成向量组时,判别为异常数据的发电厂用电率样本点为2016年7月、2016年9月、2017年10月和2017年12月;当以残差-供电标准煤耗构成向量组时,判别为异常数值的发电厂用电率样本点为2016年7月、2016年9月、2017年10月和2017年12月。

为了验证基于预测的发电厂用电率异常数据分析方法,人为的注入异常样本代替正常样本,利用50家电厂2010年1月至2015年10月的数据进行模型训练,预测2015年11月至2018年10月的异常情况(其中正常样本1600个,异常样本200个)。利用文中方法对带有异常数据的测试样本进行异常检测,表3为以3种不同发电指标构成向量组进行检测的结果。

表3 3种向量组的异常检测结果对比

Table 3 Comparison of abnormal detection results of three vector groups

项目	残差-原煤量	残差-发电标准煤耗	残差-供电标准煤耗
真阳性	1 578	1 580	1 585
假阳性	18	17	13
真阴性	182	183	187
假阴性	22	20	15
误报率/%	9.0	8.5	6.5
检出率/%	98.6	98.75	99.0

由表3可知,以残差-原煤量组成向量组的异常检测误报率为9%,检出率为98.6%;以残差-发电标准煤耗组成向量组的异常检测误报率为8.5%,检出率为98.75%;以残差-供电标准煤耗组成向量组的异常检测误报率为6.5%,检出率为99.0%。3种由残差和相关发电指标构成的向量组均有较好的异常数据识别能力,验证了文中方法的有效性。

5 结语

针对发电厂用电率数据异常检测的实际需求,提出一种基于预测模型的发电厂用电率异常数据分析方法,解决因单一模型受内在随机性影响导致整体预测精度下降,进而阻碍异常检测的问题。

经过实测数据验证,在数据预测阶段,使用文中改进的模型能有效减小预测误差。在异常数据检测阶段,针对电厂特性筛选与发电厂用电率相关性高的发电指标(如原煤量、发电标准煤耗、供电标准煤耗等),将残差与发电指标组合构成向量组进行异常数值分析,具有良好的识别率,为实现准确的不良数据定位提供了可能。后续研究工作中,针对发电厂用电率数据无标签、样本不均衡的现状,尝试深度学习等方法进一步提高该方法的综合性能及适用性也是重点关注的研究点之一。

参考文献:

[1] 国家能源局综合司. 国家能源局综合司关于进一步做好电力监管统计信息报送工作的通知: 国能局发[2017] 90号 [A/OL]. (2017-10-18) [2019-10-30]. <http://fjb.nea.gov.cn/>

- news_view.aspx? id=28398.
National Energy Administration. Notice of the National Energy Administration on further improving the reporting of statistical information on electricity regulation; National Energy Administration issued No.90[A/OL].(2017-10-18)[2019-10-30]. http://fjb.nea.gov.cn/news_view.aspx? id=28398.
- [2] 邢希东. 600 MW 火电机组降低厂用电率措施[J]. 中国电力,2007(9):60-64.
XING Xidong. Countermeasures to reduce the auxiliary power consumption rate of 600 MW thermal power units[J]. Electric Power,2007(9):60-64.
- [3] 陈波,刘厚泉,赵志凯. 时间序列多尺度异常检测方法[J]. 计算机工程与应用,2018,54(20):122-127.
CHEN Bo,LIU Houquan,ZHAO Zhikai. Method of multi-scale anomaly detection in time series[J]. Computer Engineering and Applications,2018,54(20):122-127.
- [4] LAURA B,KAUSLER B X,MARTIN S,et al. Time series anomaly detection based on shapelet learning[J]. Computational Statistics,2018.
- [5] CHEN Y T,QIAN J,SALIGRAMA V. A new one-class SVM for anomaly detection[C]//IEEE International Conference on Acoustics,Speech and Signal Processing. 2013:3567-3571.
- [6] 刘亚丽,孟令愚,丁云峰. 电网工控系统流量异常检测的应用与算法改进[J]. 计算机系统应用,2018,27(3):173-178.
LIU Yali,MENG Lingyu,DING Yunfeng. Application and algorithm improvement of abnormal traffic detection in smart grid industrial control system[J]. Computer System & Applications,2018,27(3):173-178.
- [7] 李鑫,孟亨,杨桢,等. 基于 BP 神经网络递推积分 PI-重复控制在微电网 APF 中的研究[J]. 电力系统保护与控制,2019,47(6):132-140.
LI Xin,MENG Heng,YANG Zhen,et al. Research on recursive integral PI-repetitive control based on BP neural network in micro-grid APF[J]. Power System Protection and Control,2019,47(6):132-140.
- [8] 杨光. 基于神经网络的异常用电检测算法模型研究[J]. 供用电,2016,33(10):56-59.
YANG Guang. Research and analysis on the algorithm model of anomaly detection based on neural network[J]. Power Utilization,2016,33(10):56-59.
- [9] 林峻,严英杰,盛戈峰,等. 考虑时间序列关联的变压器在线监测数据清洗[J]. 电网技术,2017,41(11):3733-3740.
LIN Jun,YAN Yingjie,SHENG Gehao,et al. Online monitoring data cleaning of transformer considering time series correlation[J]. Power System Technology,2017,41(11):3733-3740.
- [10] FENG Huo. NARX-network based wind speed estimation for wind turbines[C]//Proceedings of 2016 International Conference on Power,Energy Engineering and Management. 2016:5.
- [11] 付青,单英浩,朱昌亚. 基于 NARX 神经网络的光伏发电功率预测研究[J]. 电气传动,2016,46(4):42-45.
FU Qing,SHAN Yinghao,ZHU Changya. Generated power forecasting for photovoltaic system based on NARX neural network[J]. Electric Drive,2016,46(4):42-45.
- [12] 胡倩,陈红坤,孙志达. 风电场短期风速的集成学习预测模型[J]. 电力系统及其自动化学报,2016,28(1):1-4.
HU Qian,CHEN Hongkun,SUN Zhida. An ensemble learning predictive model for short-term wind speed forecasting[J]. Proceedings of the CSU-EPSA,2016,28(1):1-4.
- [13] 芦婧,曾明. 短期风速的 Adaboost_GRNN 组合预测模型[J]. 电力系统及其自动化学报,2019,31(4):70-76.
LU Jing,ZENG Ming. Adaboost_GRNN combination forecasting model for short-term wind speed[J]. Proceedings of the CSU-EPSA,2019,31(4):70-76.
- [14] 谭津,邓长虹,杨威,等. 微电网光伏发电的 Adaboost 天气聚类超短期预测方法[J]. 电力系统自动化,2017,41(21):33-39.
TAN Jin,DENG Changhong,YANG Wei,et al. Ultra-short-term photovoltaic power forecasting in microgrid based on Adaboost clustering[J]. Automation of Electric Power Systems,2017,41(21):33-39.
- [15] 吴俊利,张步涵,王魁. 基于 Adaboost 的 BP 神经网络改进算法在短期风速预测中的应用[J]. 电网技术,2012,36(9):221-225.
WU Junli,ZHANG Buhan,WANG Kui. Application of Adaboost-based BP neural network for short-term wind speed forecast[J]. Power System Technology,2012,36(9):221-225.
- [16] 李新鹏,高欣,阎博,等. 基于孤立森林算法的电力调度流数据异常检测方法[J]. 电网技术,2019,43(4):1447-1456.
LI Xinpeng,GAO Xin,YAN Bo,et al. An approach of data anomaly detection in power dispatching streaming data based on isolation forest algorithm[J]. Power System Technology,2019,43(4):1447-1456.
- [17] 刘佳翰,陈克绪,马建,等. 基于卷积神经网络和随机森林的三相电压暂降分类[J]. 电力系统保护与控制,2019,47(20):112-118.
LIU Jiahao,CHEN Kexu,MA Jian,et al. Classification of three-phase voltage dips based on CNN and random forest[J]. Power System Protection and Control,2019,47(20):112-118.
- [18] 王继业,朱欣焰,赵光,等. 基于深度神经网络和 GIS 数据的最优停电模型研究[J]. 电力系统保护与控制,2019,47(16):58-63.
WANG Jiye,ZHU Xinyan,ZHAO Guang,et al. Research on optimal outage model based on deep artificial neural network and GIS data[J]. Power System Protection and Control,2019,47(16):58-63.
- [19] LIU F T,KAI M T,ZHOU Z H. Isolation-based anomaly detection[J]. ACM Transactions on Knowledge Discovery from Data,2012,6(1):1-39.
- [20] 贾育康,赵力芬. 300 MW 火电机组厂用电分析及节电措施[J]. 山西电力,2009(2):49-51.
JIA Yukang,ZHAO Lifen. Analysis of service power and electricity saving measures of 300 MW thermal power unit[J]. Shanxi Electric Power,2009(2):49-51.
- [21] 田野,张程,毛昕儒,等. 运用 PCA 改进 BP 神经网络的应用

电异常行为检测[J]. 重庆理工大学学报(自然科学版), 2017,31(8):125-133.

TIAN Ye,ZHANG Cheng,MAO Xinru,et al. Research on abnormal behavior of power consumption based in BP neural network with PCA[J]. Journal of Chongqing University of Technology(Natural Science),2017,31(8):125-133.

[22] 李磊,严正,冯冬涵,等. 结合主成分分析及生产函数的电网智能技术评价探讨[J]. 电力系统自动化,2014,38(11):56-61.

LI Lei,YAN Zheng,FENG Donghan,et al. Discussion on intelligent technology evaluation of electrical power grid based on

principal component analysis and production function[J]. Automation of Electric Power Systems,2014,38(11):56-61.

作者简介:



高骞

高骞(1980),男,硕士,高级工程师,从事电力统计分析等工作(E-mail:gaoqian_1980@163.com);

张浩天(1995),男,硕士在读,研究方向为网络空间安全;

汤奕(1977),男,博士,教授,研究方向为电力系统稳定分析。

Abnormal data analysis method of power plant data based on forecast model

GAO Qian¹, ZHANG Haotian², TANG Yi³

(1. State Grid Jiangsu Electric Power Co.,Ltd.,Nanjing 210024,China;

2. School of Cyber Science and Engineering,Southeast University,Nanjing 210096,China;

3. School of Electrical Engineering,Southeast University,Nanjing 210096,China)

Abstract: As an important index reflecting the production efficiency and energy saving of power plants, the authenticity and accuracy of the reported data of power plant power consumption rate are very important. Therefore, a method of abnormal data identification based on predictive model is proposed. Firstly, the NARX neural network is improved by Adaboost to build the forecasting model of the power consumption rate. Then, the present power consumption rate is forecasted in a dynamic way by continuously introducing the reported value. When the time series of the power consumption rate suddenly changes, the residual time series obviously increase or decrease. Then the outliers of each residual vector group are obtained by using the isolation forest algorithm. Finally, the method is used to identify the actual generation data injected with false data, and the effectiveness of the proposed method is verified.

Keywords: power plant power consumption rate; abnormal data; nonlinear auto regressive models with external input(NARX) neural network; Adaboost algorithm

(编辑 钱悦)