

DOI: 10.19650/j.cnki.cjsi.J2210627

考虑行业关联度的工业用户用电异常识别研究*

陈静,郑垂锭,李桂敏,江灏,缪希仁
(福州大学电气工程与自动化学院 福州 350108)

摘要:针对工业用户的行业属性对其用电模式的影响,本文提出一种考虑行业关联度的工业用户用电异常识别方法。基于真实工业用户用电负荷数据生成多个行业类别的典型负荷特征曲线。运用改进灰色关联度算法计算电力用户用电特征与各个行业典型用电特征之间的关联性,生成用户的行业关联特征;利用多头注意力机制(MHA)提取用户负荷序列特征,与行业关联特征相结合,采用变分自动编码器(VAE)所提供的重构误差作为异常判定度量,建立MHA-VAE深度异常检测模型,实现对多种类型工业用户用电异常的识别。结果表明,引入用户的行业关联特征后异常检测的准确率、检出率和误检率分别为96.84%、98.02%、4.35%,与仅考虑用户负荷特征相比准确率提高1.06%,误检率降低2.24%。

关键词:用电异常识别;多头注意力机制;变分自动编码器;行业关联度

中图分类号: TM 714 TH 86 文献标识码: A 国家标准学科分类代码: 470.40

Research on power consumption anomaly identification of industrial users considering industry relevance

Chen Jing, Zheng Chuiding, Li Guimin, Jiang Hao, Miao Xiren

(College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China)

Abstract: In view of the influence of industrial users' industry attributes on their power consumption patterns, a power consumption anomaly identification method considering industry relevance is proposed in this article. Based on the real industrial consumer power consumption data, the typical load characteristic curves of each industry are generated, and the improved grey correlation degree algorithm is used to calculate the relevance between the power consumption characteristics of power users and the typical power consumption characteristics of the industry. In this way, the industry relevance characteristics of users are achieved. The multi-head attention (MHA) is used to extract the features contained in load sequences. Combined with the industry relevance features, the reconstruction error provided by the variational autoencoder (VAE) is used as the anomaly decision metric to formulate the MHA-VAE depth anomaly detection model to identify various types of industrial users' power consumption anomalies. Results show that, the accuracy, detection rate and false detection rate after introducing users' industry relevance are 96.84%, 98.02%, and 4.35%, respectively. Compared with only considering the load characteristics of users, the accuracy is increased by 1.06% and the error detection rate is reduced by 2.24%.

Keywords: power consumption anomaly identification; multi-head attention; variational auto-encoder; industry relevance

0 引言

电力市场的售电侧改革使得越来越多售电公司作为连接电力批发侧和电力零售侧的市场主体出现。深入掌握客户用电行为的变化规律,将是实现电力营销“精细化决策”的重要途径^[1],与售电公司的利润密切相关。而负

荷采集设备故障、窃电等非法行为以及用户的非典型用电行为,都会造成用电负荷数据出现异常^[2]。采集设备故障与窃电等欺诈行为会导致售电公司的直接经济损失^[3];未能及时查悉用户的非典型用电行为则可能会导致售电公司的购售电策略和负荷管理工作与实际需求产生偏差^[4]。

收稿日期:2022-10-26 Received Date: 2022-10-26

* 基金项目:福建省自然科学基金(2022J01566)项目资助

随着智能电网的发展,电力负荷采集边缘基础设施不断推广及普及^[5],大量蕴藏用户用电内在规律的信息数据被广泛采集,使得通过用户侧海量负荷数据对电力用户异常用电行为进行识别分析成为现实^[6]。通过基于数据驱动模型,快速有效地察觉电力用户负荷异常变动情况^[7],及时对其异常用电行为做出反应,能够为电网侧的精细化管理和售电公司的营销服务、风险管理提供辅助信息,对提升电网企业、售电商、用户的综合利益具有重大意义。

目前,面对日趋隐蔽、多样、快速的异常用电行为特点^[8],学者们从用电负荷特征着手,提出了基于回归^[9]、分类^[10]、聚类^[11-14]等方法的异常用电识别分析方法,通过对比正、异常负荷序列特征的相似性判定是否存在异常。其中,基于聚类算法的用电异常检测是近年来研究的热点。文献[12]结合负荷预测结果,对预测值与实际值的残差项进行聚类,将离群的用户认定为异常。文献[13]基于空间密度聚类结果构建异常用电数据域,通过对比用户负荷曲线与异常数据域上下界的距离判断用电是否异常。文献[14]为了能更加准确地衡量电力用户之间负荷曲线的差异,采用快速动态时间扭曲距离作为聚类过程中的相似量度,并采用密度异常指数量化负荷曲线的异常程度。但电力用户用电负荷受到多种因素影响^[15],如用户家庭人数、从事行业等自身因素,GDP的增长、产业结构的调整、国民消费水平等社会因素,以及天气气温、地理位置等自然因素。这些因素从各个维度影响着电力用户的用电规律,仅以负荷特征为导向的异常检测方法识别出的异常用户中往往有一部分为误报,影响到了实际工程应用。

深入挖掘历史负荷特征与其影响因素之间的关联关系,将其融入异常用电识别模型,对提高异常用电检测精度具有显著的现实意义^[16]。文献[17]利用多元逻辑回归模型,分析居民用户的用电行为与家庭成员人数、从事行业等家庭特征间的潜在联系。文献[18]将电力用户的台区号引入异常检测模型,显著提升了检测效果。文献[19]综合用户消费历史统计信息和电性参数生成11个特征,用于窃电检测系统。研究者们基于数据驱动的方法,分析用电负荷特征与各种影响因素的相关性,引入负荷关联参考因素,以更多维度判定用户负荷特征的正常与否。

对于工业用户而言,不同行业的用户由于生产规模、经营方式的差异,往往具有不同的用电模式。当某一电力用户的异常负荷特征恰好契合其他行业用户的正常负荷特征时,检测算法会将其认定为正常负荷特征,导致异常检测出现误检的情况。针对此问题,本文引入用户的行业关联特征,解决因行业差异导致的误判问题。

本文以中国南方某省工业电力用户为研究对象,借助现有历史负荷数据,以行业关联参考因素作为影响因素,开展工业用户异常用电识别研究,识别的异常用电行为类型包括负荷采集设备故障、窃电行为以及工业用户因生产过程变化所造成的非典型用电行为。首先,运用改进灰色关联度算法对电力用户与行业典型负荷进行关联度分析,为不同电力用户获取相对应的行业关联参考度。其次,通过多头注意力机制对负荷序列进行变换,着重负荷特征信息。最后,将负荷特征与行业关联特征相融合,以变分自动编码器提供的重构误差判断电力用户是否存在异常用电的情况。试验结果表明,与传统异常检测算法相比,本文使用的考虑行业关联度的深度异常检测模型识别效果更加优秀,能有效检测出多种类型的用电异常用户,同时降低误检率。

1 用户特征提取

1.1 行业关联度构建

工业电力用户的行业属性,一定程度上决定了用户的生产经营规律,从而对用户的用电模式造成影响。除此之外,行业间上下游企业间的供应、需求关系也相互影响着彼此的负荷特征。而灰色关联分析算法理论可以通过研判各类电力用户负荷序列曲线间相似程度、紧密与否等因素,给定一个关联度值^[20-21]。为探究行业类别间负荷关联性,挖掘行业典型负荷与用户负荷之间的相关性,采用改进灰色关联度算法构建工业电力用户的行业关联度,主要步骤如下:

1) 将用户典型日负荷序列 $\mathbf{X}$ 作为待测序列,与 M 个不同行业的典型日负荷序列 $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_M$ 构成序列矩阵:

$$\mathbf{O} = (\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_M) = \begin{pmatrix} x_1 & y_1^1 & \cdots & y_1^M \\ x_2 & y_2^1 & \cdots & y_2^M \\ \vdots & \vdots & \ddots & \vdots \\ x_T & y_T^1 & \cdots & y_T^M \end{pmatrix} \quad (1)$$

式中: $x_i (i = 1, 2, \dots, T)$ 为用户典型日负荷序列 $\mathbf{X}$ 中第 i 个时间节点的负荷值, $y_i^m (m = 1, 2, \dots, M)$ 为第 m 个行业日负荷序列 $\mathbf{Y}_m$ 中第 i 个时间节点的负荷值。

2) 构建待测负荷序列 $\mathbf{X}$ 与各个行业典型日负荷序列 $\mathbf{Y}_m (m = 1, 2, \dots, M)$ 之间的差值矩阵 $\mathbf{I}$ 。

$$\mathbf{I} = \begin{pmatrix} u_1^1 & u_1^2 & \cdots & u_1^M \\ u_2^1 & u_2^2 & \cdots & u_2^M \\ \vdots & \vdots & \ddots & \vdots \\ u_T^1 & u_T^2 & \cdots & u_T^M \end{pmatrix} \quad (2)$$

矩阵中的元素 u_i^m 表示序列 $\mathbf{X}$ 中第 i 个元素 x_i 与第 m 个行业负荷序列 $\mathbf{Y}_m$ 中对应元素 y_i^m 之间差值的绝对值。

3) 计算两级最小差 $S_{\min}$ 与两极最大差 $S_{\max}$ 。

$$\begin{cases} S_{\min} = \min_m \min_t u_t^m \\ S_{\max} = \max_m \max_t u_t^m \end{cases} \quad (3)$$

4) 计算待测负荷序列与各个行业序列之间的关联系数,进而计算出用户与各个行业的关联度值。

$$\begin{cases} \lambda_t^m = \frac{S_{\min} + \rho S_{\max}}{u_t^m + \rho S_{\max}} \\ r_m = \frac{1}{T} \sum_{t=1}^T \lambda_t^m \end{cases} \quad (4)$$

式中: λ_t^m 表示待测序列 $\mathbf{X}$ 中第 t 个元素与第 m 个行业序列对应元素的关联系数; ρ 为分辨系数,用于解决两极最大差过大时引起的关联度失真的问题,本文取 $\rho = 0.5$; r_m 表示待测用户负荷序列 $\mathbf{X}$ 与第 m 个行业负荷序列的关联度值。

传统灰色关联度算法由于式(2)中差值矩阵数值取绝对值,最后关联度计算结果总是大于 0,与负荷序列实际存在负关联度的情况相矛盾。因此,引入关联符号函数 $\text{sgn}(\cdot)$ 确定关联度的正负性,则考虑正负关联度的改进灰色关联度计算如式(5)所示。

$$\begin{cases} \hat{\lambda}_t^m = \lambda_t^m \cdot \text{sgn}(\Delta x_t \cdot \Delta y_t^m) \\ \hat{r}_m = \frac{1}{T} \sum_{t=1}^T \hat{\lambda}_t^m \end{cases} \quad (5)$$

式中: $\hat{\lambda}_t^m$ 表示考虑负关联度后待测用户负荷序列 $\mathbf{X}$ 中第 t 个元素与第 m 个行业负荷序列对应元素的关联系数; $\hat{r}_m$ 表示 $\mathbf{X}$ 与第 m 个行业序列的关联度值; Δx_t 是用户序列 $\mathbf{X}$ 中第 $t+1$ 时间节点相对于第 t 个时间节点负荷值的增量; Δy_t^m 是第 m 个行业负荷序列中第 $t+1$ 时间节点相对于第 t 个时间节点负荷值的增量。

1.2 多头注意力机制

用户的用电负荷序列中,某个时间节点的负荷数据与整体负荷序列关联性的高低决定了其变动对整体的影响力。若忽略这一特性,直接采用原始数据进行模型的训练及检测容易造成过拟合,影响检测效果。而注意力机制^[22]对于序列数据具有过滤无用数据,关注重点数据的特点,对与整体负荷序列关联性较高的负荷特征数据有着更高的关注度,起到提炼负荷特征的作用。

如图 1 所示,多头注意力机制由自注意力层、拼接层和线性变换层组成^[23]。其中,自注意力层包括点积模块和 Softmax 函数模块。多头注意力机制 (multi-head attention, MHA) 通过集成多个参数不同的自注意力网络,从不同的角度对负荷数据的依赖关系进行挖掘,相较于传统注意力机制能够更加准确地表示数据特征。

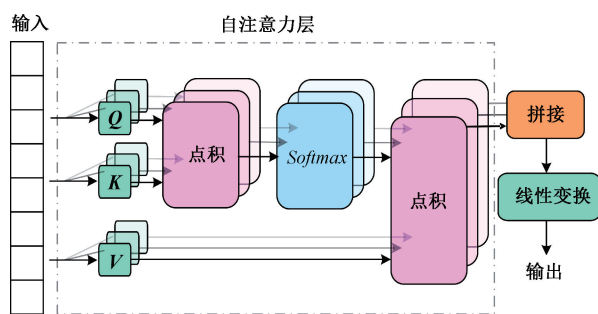


图 1 多头注意力机制模型结构

Fig. 1 The model network architecture of MHA

首先,将用户日负荷序列 $\mathbf{X} = (x_1, x_2, \dots, x_T)$ 输入多个自注意力层中,计算出每层的注意力值矩阵。对第 i 个自注意力层初始化 3 个不同的线性投影矩阵 $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V$, 将输入负荷样本 $\mathbf{X}$ 映射到询问 $\mathbf{Q}_i$, 键 $\mathbf{K}_i$ 与值 $\mathbf{V}_i$ 矩阵,而后计算出每层的结果。计算过程如式(6)所示。

$$\begin{cases} \mathbf{Q}_i = \mathbf{X} \mathbf{W}_i^Q \\ \mathbf{K}_i = \mathbf{X} \mathbf{W}_i^K \\ \mathbf{V}_i = \mathbf{X} \mathbf{W}_i^V \\ \mathbf{H}_i = \text{softmax} \left(\frac{\mathbf{Q}_i \mathbf{K}_i^T}{\sqrt{d}} \right) \mathbf{V}_i \end{cases} \quad (6)$$

式中: $\mathbf{H}_i$ 为第 i 个自注意力层的输出结果; d 为 3 个线性投影矩阵的维数。 $\text{softmax}(\cdot)$ 函数用于限制矩阵数值范围。

其次,将多个自注意力层的计算结果进行拼接。

最后,将拼接结果进行线性变化,输出注意力值矩阵 $\mathbf{H}$,如式(7)所示。

$$\mathbf{H} = \text{Concat}(\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_i) \mathbf{W}^o \quad (7)$$

式中: $\text{Concat}(\cdot)$ 为拼接函数, $\mathbf{W}^o$ 为权重矩阵。

2 用电异常识别方法

2.1 变分自动编码器

在用户负荷数据中,大多数为正常负荷样本,异常样本仅占少数。基于自动编码器 (autoencoder, AE) 的异常用电检测,以大量的正常负荷序列进行训练。完成训练后,对于测试集中的正常负荷序列,AE 能将其有效重构。对于异常负荷序列,由于其负荷特征与正常负荷序列特征偏离,重构时会产生较大的误差。因此通过设定重构误差的阈值,即可对异常负荷序列进行检测。

变分自动编码器 (variational autoencoder, VAE)^[24-25]作为扩展 AE 的一种,与传统 AE 存在明显的差别,其网络结构如图 2 所示。VAE 并非像其他编码器一样通过确定的映射将原始负荷序列映射至隐藏层中,而是在隐藏层中设计了一种服从独立多元高斯分布的潜变量 $\mathbf{z}$ 。以

高斯分布实现高维原始负荷序列到潜变量 z 的随机映射,不断迭代训练优化参数,使得潜变量 z 能够有效表示原始负荷序列的特征。

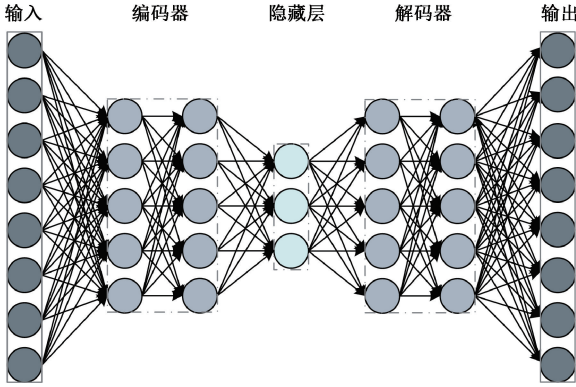


图2 变分自动编码器网络结构

Fig.2 Network architecture of VAE

对于用户融合行业关联度的负荷特征序列数据 $r = (r_1, r_2, \dots, r_{144})$, VAE 编码网络将原始特征序列采样至潜变量 z 。 z 的真实后验分布符合下式:

$$P_\theta(z|r) = P_\theta(r|z)P_\theta(z)/P_\theta(r) \quad (8)$$

式中: $P_\theta(z|r)$ 是 z 的真实后验分布; $P_\theta(r|z)$ 是给定潜变量 z 时数据 r 的分布; $P_\theta(z)$ 是 z 的先验分布; $P_\theta(r)$ 是 r 的先验分布; θ 是隐含层网络的真实后验分布参数。

由于分布 $P_\theta(r)$ 难以计算, VAE 依托于编码器网络拟合近似后验分布 $U_\phi(z|r)$ 来逐渐逼近真实后验分布 $P_\theta(z|r)$ 。此时特征序列 r 的边缘似然函数为:

$$\log P_\theta(r) = D_{KL}(U_\phi(z|r) \| P_\theta(z)) + L(\theta, \phi; r) \quad (9)$$

式中: ϕ 为编码器的网络参数; $D_{KL}(\cdot)$ 为近似后验分布与先验分布之间的 KL 散度, 用于表示两者的相似程度; $L(\theta, \phi; r)$ 是负荷数据点的边际似然函数的变分下界。由于 KL 散度总是大于 0, 所以式(9)可改写为:

$$\log P_\theta(r) \geq L(\theta, \phi; r) = -D_{KL}(U_\phi(z|r) \| P_\theta(z)) + E_{U_\phi(z|r)}[\log P_\theta(r|z)] \quad (10)$$

式中: 等号右侧第 1 项为潜变量 z 的近似后验分布和先验分布的 KL 散度; 第 2 项为特征序列 r 在经过近似后验分布函数 $U_\phi(z|r)$ 和似然分布函数 $P_\theta(r|z)$ 后的重构误差。

VAE 综合考虑重构序列与原始序列之间的差异及重构过程的可变性。其通过概率编码器拟合原始序列在隐藏层中潜变量的分布, 运用随机采样的方式考虑到潜变量的可变性。这使得正常数据和异常数据即使在隐藏层中拥有相同均值, 由于方差不同, 仍可正确做出判断。

2.2 考虑行业关联度的异常用电检测模型

本文构建的考虑行业关联度的 MHA-VAE 工业电力用户用电异常检测模型总体流程如图 3 所示, 其主要包

含如下 6 个步骤。

1) 数据清洗: 剔除用户缺失数据过多的日负荷曲线, 采用近邻插补与近邻平均值插补法修补异常点负荷数据。比对校验日负荷数据均值与智能电表记录的日平均负荷的偏差, 将偏差大于 2% 的序列剔除。

2) 数据集划分: 采用最大最小归一化方法将经过数据清洗的数据归一化处理, 通过权重公式计算得到每个工业用户的典型负荷曲线, 构建正常负荷数据集。按照平均随机原则, 数据集被划分为数据集 A 和 B。

3) 行业关联特征提取: 采用 K-means 算法将数据集 A 按照所属的 48 个行业类别聚类, 生成行业典型特征曲线。计算数据集 B 中所有工业用户与 48 个行业的行业关联度, 生成行业关联序列。

4) 用户负荷特征提取: 将数据集 B 中的负荷序列通过多头注意力机制处理获得每个负荷点的注意力值, 将原始负荷值乘注意力值获得用户负荷特征序列。

5) 模型训练与检测: 合并用户负荷特征序列与行业关联序列。处理完成后获得训练集与测试集。训练集用以构建 VAE 异常检测模型, 测试集用以评估模型检测效果。训练集包含 80% 正常负荷曲线, 测试集中包含 10% 的正常负荷曲线与 10% 异常负荷曲线。

6) 模型性能评估: 通过模型输出的重构误差数值判断序列状况, 当序列重构误差到达阈值时被归类为异常。最后通过测试集负荷数据标签, 评估模型性能。

2.3 评价指标

如表 1 所示, 本文通过混淆矩阵中的真正例 (true positive, TP)、真负例 (true negative, TN)、假正例 (false positive, FP) 和假负例 (false negative, FN) 4 个参数, 分别定义准确率 δ_{ACC} 、检出率 δ_{DR} 和误检率 δ_{FPR} 这 3 个性能指标。各指标计算公式如式(11)所示。

表 1 混淆矩阵

Table 1 Confusion matrix

测试集	检测正常用户	检测异常用户
实际正常用户	TP	FN
实际异常用户	FP	TN

$$\begin{cases} \delta_{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\ \delta_{DR} = \frac{TP}{TP + FN} \times 100\% \\ \delta_{FPR} = \frac{FP}{FP + TN} \times 100\% \end{cases} \quad (11)$$

式中: δ_{ACC} 、 δ_{DR} 与 δ_{FPR} 的输出值范围为 $[0, 1]$, δ_{ACC} 与 δ_{DR} 的输出结果与 1 越靠拢, δ_{FPR} 的输出结果与 0 越靠拢, 则检测效果越好, 即模型算法性能越好。

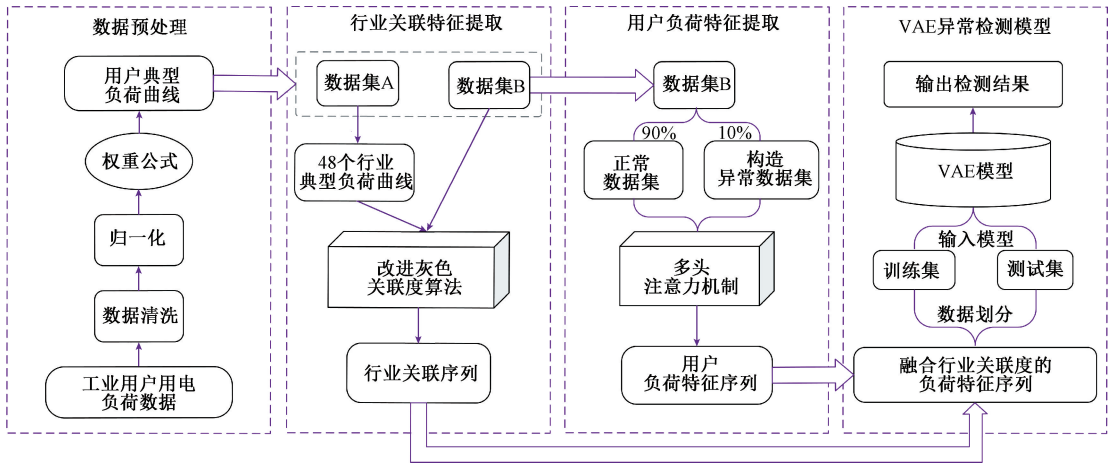


图3 考虑行业关联度的深度异常用电检测模型

Fig. 3 Deep abnormal power consumption detection model considering industry relevance

3 算例分析

本文采用中国南方某省由智能设备采集的2018年初~2020年末的实际用电负荷数据作为研究对象,分析考虑行业关联度的MHA-VAE模型的异常检测效果。负荷数据时间跨度达987天,采样频率为15 min每次,每日采集96点负荷数据,共包括5 918家工业电力用户。每家电力用户均有档案记录的四级行业类别标签。本文在其基础上,按照行业数量分布,提炼出48个行业类别。

3.1 异常负荷曲线构建

从数据集B中随机挑选10%正常负荷曲线,参考文献[26]的异常样本生成方法,根据五类工业用户典型异常负荷特征对其进行变换,构建异常负荷测试集。

五类异常特征如图4所示,A类异常模拟用户在较长的一段时间内,产业升级或者转型导致用电需求发生变换的异常情况;B类异常模拟用户在某些时间点因生产需求等因素产生的特殊用电情况;C类异常模拟用户在总体负荷水平未产生重大变化时,用电特征发生改变的异常情况;D类异常模拟用户改变峰谷用电的窃电行为;E类异常模拟负荷采集设备故障造成的异常。

异常负荷序列变换公式如表2所示。表中, $G_a(\cdot)$ 、 $G_b(\cdot)$ 、 $G_c(\cdot)$ 、 $G_d(\cdot)$ 、 $G_e(\cdot)$ 分别为A、B、C、D、E类异常负荷变换函数; $X=(x_1, x_2, \dots, x_{96})$ 为电力用户某日典型负荷数据; x_t 是电力用户在第 t 个时间节点的负荷值, $t=1, 2, \dots, 96$; $\alpha_t, \beta_t, \xi_t, \eta_t$ 均为第 t 个时间节点的变换系数, τ_t 为 $[0, 1]$ 之间的随机取值,用以确定第 t 个时间节点的负荷值是否变换; $mean(\cdot)$ 为取均值函数, $random(w_1, w_2)$ 为 $[w_1, w_2]$ 范围内的随机数生成函数, $reverse(\cdot)$ 为翻转峰谷电函数; t_{ST} 为E类异常负荷变换

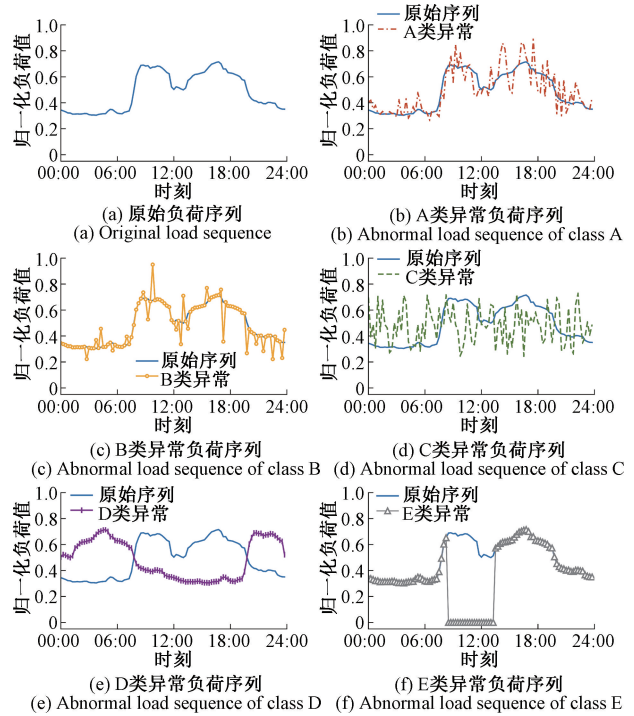


图4 原始序列及其对应的5类异常负荷序列

Fig. 4 Original sequence and its corresponding five types of abnormal load sequences

中变换起始时间节点, t_{END} 为变换结束时间节点, s 为变换区间内包含的负荷点数量。

3.2 行业关联度分析

取某一行业典型负荷序列作为待测序列,由关联度算法得到此行业与剩余47个行业的负荷关联系数。此行业与其某一高、中、低关联行业负荷曲线如图5所示。曲线特征相似性与以改进灰色关联度算法计算出的行业关联度相契合。行业负荷关联度真实有效反映了行业用户用电特征的关联情况。

表 2 异常负荷序列变换公式

Table 2 Transformation formula of abnormal load sequences

异常类别	变换公式
A	$G_a(x_t) = (1 + \alpha_t)x_t; \alpha_t = \text{random}(-1, 1)$
B	$G_b(x_t) = \begin{cases} (1 + \beta_t)x_t, & 0 \leq \tau_t < 0.3 \\ x_t, & 0.3 \leq \tau_t < 1 \end{cases}; \begin{cases} \beta_t = \text{random}(-1, 1) \\ \tau_t = \text{random}(0, 1) \end{cases}$
C	$G_c(x_t) = (1 + \xi_t) \cdot \text{mean}(X); \xi_t = \text{random}(-1, 1)$
D	$G_d(x_t) = \text{reverse}(x_t)$
E	$G_e(x_t) = \eta_t x_t; \eta_t = \begin{cases} 0, & t_{ST} < t < t_{END} \\ 1, & \text{其他} \end{cases}; \begin{cases} t_{END} = t_{ST} + s \\ t_{ST} = \text{random}(0, 39) \\ s = \text{random}(8, 48) \end{cases}$

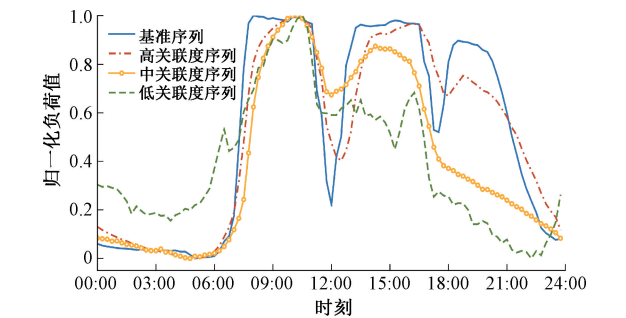


图 5 行业负荷曲线关联图

Fig. 5 Relevance diagram of industry load curve

48 个行业典型负荷序列关联度如图 6 所示,用户行业以数字 1~48 进行编号。图中仅有 5 个行业与其他行业关联度小于 0.7,由此可见行业用电负荷特征存在明显关联性,行业关联性越高,则互相影响负荷特征的概率越高。

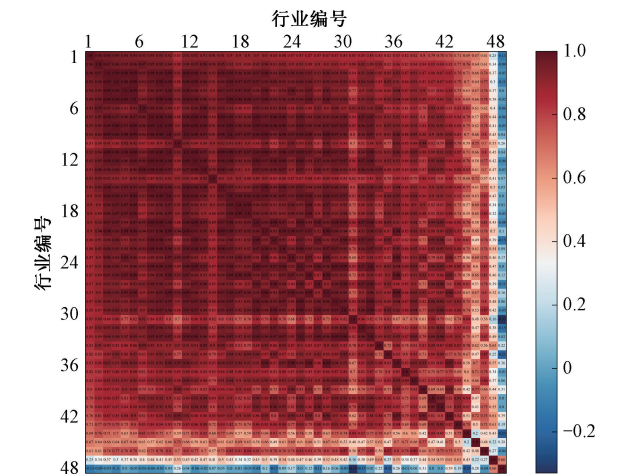


图 6 行业负荷特征关联性

Fig. 6 Relevance of industry load characteristics

3.3 多头注意力机制有效性分析

为验证多头注意力机制对负荷序列异常检测结果的积极影响,分别采用原始负荷序列、注意力机制处理后的序列和多头注意力机制处理后的序列进行异常检测。结果如表 3 所示。由检测结果指标可见,多头注意力机制处理后的负荷序列在异常检测算法中的整体检测效果更加优秀。

表 3 不同负荷序列处理方式异常检测效果对比

Table 3 Comparison of anomaly detection effects of different load sequences processing methods %

处理方式	δ_{ACC}	δ_{DR}	δ_{FPR}
未处理	93.67	95.95	8.61
注意力机制	94.68	96.79	7.43
多头注意力机制	95.78	98.14	6.59

图 7 为原始负荷序列经过变换后的曲线。由识别结果对比图 7 可知,经典注意力机制对于重要的负荷特征有一定的刻画能力,这也解释了其为何能在一定程度上提升异常检测算法的效果。但多头注意力机制对于一些表征不明显却十分重要的特征描述能力更强,例如在 6~7 时、17~18 时这两个时刻,由于原始曲线上特征不明显,注意力机制选择弱化其特征表达,而多头注意力机制对此却有较强的关注度。得益于多头注意力机制更加优异的负荷特征刻画表达能力,异常检测算法拥有更加准确的检测效果。

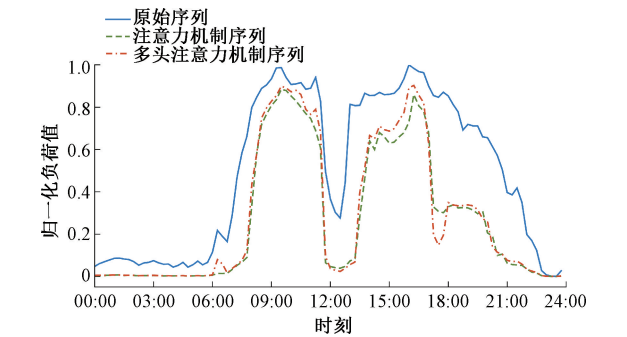


图 7 注意力机制特征提取效果对比

Fig. 7 Comparison of feature extraction results between attention and MHA

3.4 异常检测算法效果对比

为验证本文 MHA-VAE 异常检测算法的性能,与常规异常检测算法线性主成分分析 (principal component analysis, PCA)、机器学习异常检测算法支持向量机 (support vector machine, SVM)、深度学习异常检测算法 AE 和堆叠自动编码器 (stacked autoencoder, SAE) 进行对比实验。实验结果如表 4 所示。

表4 不同算法异常检测效果对比

Table 4 Comparison of anomaly detection effects of different algorithms			%
算法	δ_{ACC}	δ_{DR}	δ_{FPR}
PCA	77.87	78.21	22.47
SVM	82.01	81.76	17.74
AE	85.73	88.01	16.55
SAE	89.93	92.74	14.19
VAE	93.67	95.95	8.61
MHA-PCA	79.48	81.25	22.30
MHA-SVM	80.91	82.94	21.11
MHA-AE	88.26	91.05	14.53
MHA-SAE	92.15	94.59	10.30
MHA-VAE	95.78	98.14	6.59

由上述检测结果可知,除了 SVM 异常检测算法,其余算法通过 MHA 处理负荷序列后,检测效果都得到了提升。SVM 算法通过 MHA 处理后 δ_{ACC} 指标有所下降。究其原因,PCA 与基于自动编码器的异常检测算法依托于负荷特征表达来检测异常情况,通过 MHA 处理后的负荷序列特征更加明显,因此在 PCA 与自动编码器异常检测算法中,MHA 对于检测结果都有一定的提升。SVM 关注的是负荷序列在高维特征的表达,贸然改变负荷序列特征表示会对其在高维特征中的表达造成影响,从而影响检测结果。由此可见,通过 MHA 处理负荷序列并非适用于所有的异常检测算法。

而通过对比基于自动编码器的异常检测算法,VAE 异常检测算法的检测效果最好,这是由于 VAE 学习正常负荷序列在隐空间中的概率分布,相较于 AE 与 SAE 这类通过编码器和解码器对正常负荷序列模式进行强行记忆的检测算法,具有更强的灵活性和检测能力。

3.5 异常检测算法误检分析

针对 MHA-VAE 模型存在的误检情况,对误检负荷序列进行提取,探求其误检原因。

从检出率来看,MHA-VAE 模型达到 98.14%,即算法将 11 条正常负荷序列检测为异常序列。探究原因,由于 VAE 异常检测算法的本质是通过负荷序列重构误差的高低判定序列是否异常,为尽量避免将异常负荷序列识别为正常负荷序列,异常判定的阈值设置较低,11 条曲线的重构误差均超过了阈值因此被判定为异常。从误检率来看,MHA-VAE 模型误检率为 6.59%,即算法将 39 条异常负荷序列判定为正常序列。其中,五类异常负荷序列分别占 12 条、16 条、11 条、0 条、0 条,分别对五种序列误检情况进行分析。

由图 8 可知,A 类误检负荷序列变换前后负荷特征差异较大,但在算法解码时重构误差未超过阈值,因此被判定为正常负荷序列。对此误检序列与训练集中所有正常负荷序列进行误差对比,采用欧式距离作为相似度判定依据。发现此序列与训练集中另一企业正常负荷序列特征相似,因此被模型认定为另一企业的正常负荷序列,导致误判。

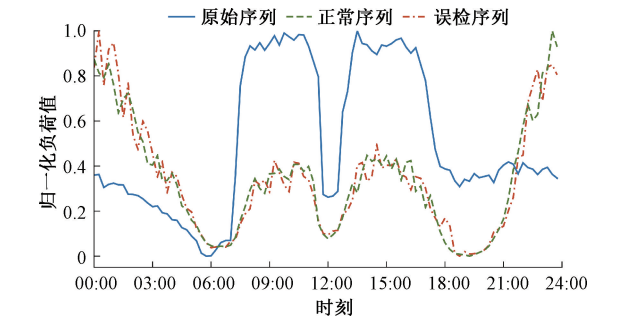


图8 A类异常负荷误检典型序列对比
Fig. 8 Comparison of typical sequences for misdetection of class A abnormal load

由图 9 可知,B 类误检负荷序列变换前后,仅存在有较少的异常离群负荷点,曲线特征仍与原负荷序列相似,因此在重构时模型忽略少数离群点产生的异常,认定异常负荷序列为正常。

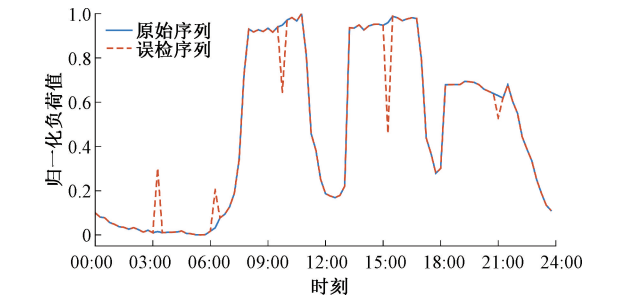


图9 B类异常负荷误检典型序列对比
Fig. 9 Comparison of typical sequences for misdetection of class B abnormal load

由图 10 可知,C 类误检负荷与 A 类负荷类似。原始负荷序列经过 MHA 变换后,序列特征产生差异,但变化后的序列特征与其他企业正常负荷特征恰好类似,导致模型判定其为正常序列。

MHA-VAE 算法对 D、E 这两类异常负荷拥有 100% 的检出率,主要得益于此两类负荷异常特征较为明显,且正常负荷序列中不可能存在有类似的负荷特征,因此检测效果较好。

综上所述,B 类异常负荷序列当且仅当存在少数离群点时,异常检测算法难以检测。但用户负荷产生少数离群异常数据几乎无法产生危害性行为,而当离群点较

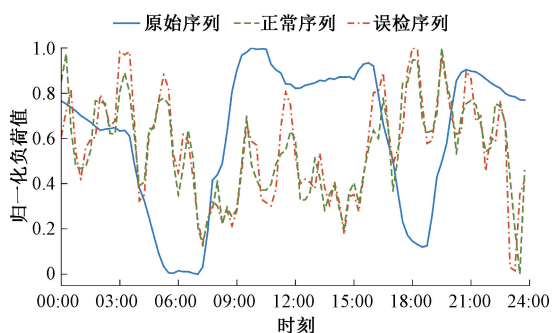


图 10 C类异常负荷误检典型序列对比

Fig. 10 Comparison of typical sequences for misdetection of class C abnormal load

多时则会被算法检测出来,因此可短期忽略此类异常负荷被误检为正常负荷的情况。A、C类异常负荷序列都存在与其他用户正常负荷序列特征契合的情况,导致检测失误,这是以负荷特征为导向的异常检测算法的局限性所致,即算法本身只以负荷序列特征是否吻合作为评判标准。

3.6 考虑行业关联度的用电异常检测

将用户典型负荷曲线与48个行业典型负荷特征曲线通过改进灰色关联度算法计算其关联度值,即增加48个维度的关联度值作为关联参考因素。某一行业用户典型负荷曲线增加行业关联度后曲线如图11所示。图中实线部分为行业用户归一化后的负荷序列,虚线部分为与其对应的48个行业典型负荷的关联度。

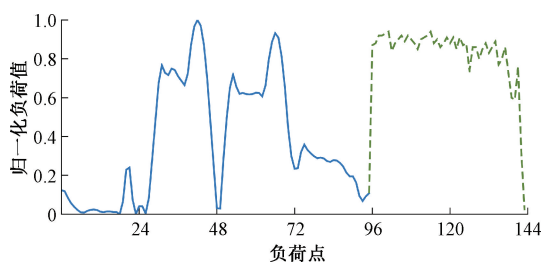


图 11 融合行业关联度的负荷曲线

Fig. 11 Load curve combining industry relevance

引入行业关联特征后,MHA-VAE模型的异常检测结果为 $\delta_{ACC}=96.84\%$ 、 $\delta_{DR}=98.02\%$ 、 $\delta_{FPR}=4.35\%$ 。如图12所示, δ_{DR} 指标与原先近乎持平, δ_{ACC} 提高1.06%, δ_{FPR} 降低2.24%。考虑行业关联度后在5种异常类型的检测中,对于D类、E类异常负荷序列仍然达到100%的检出率。对于B类异常负荷仅存在少数异常离群点的负荷序列难以检测,此类异常负荷共有8条被误检。而对于A、C两类异常负荷序列的检测效果大大提升,因为在考虑行业关联度后,即使用户特征序列中的负荷特征与正常

负荷部分相似,但在特征序列中的行业关联部分中,其行业信息关联特征存在大量偏差,导致重构误差超过阈值,被判定为异常。

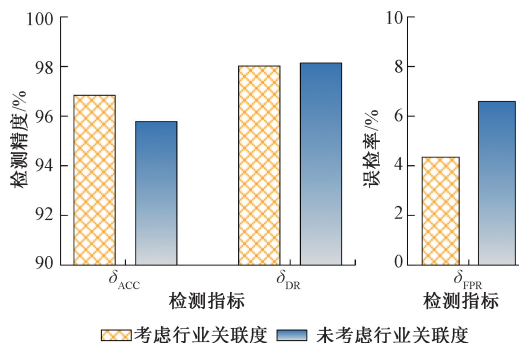


图 12 考虑行业关联度前后异常检测效果对比

Fig. 12 Comparison of anomaly detection effect before and after considering industry relevance

MHA-VAE异常检测模型仅依靠负荷特征难以检测出异常序列的局限性被行业关联参考因素的引入所改善,使算法检测效果进一步提升。

4 结 论

本文针对工业电力用户用电异常识别设计了基于深度学习的异常检测算法,并引入行业关联参考因素对异常用电识别进行联合优化,提高深度异常检测模型的识别精度。针对电力用户异常负荷模式多样,难以通过人工大量标注负荷数据的情况,提出MHA-VAE用电异常检测模型。相较于现有用电异常检测算法,其对于多种异常负荷能够更加有效甄别,模型效果更佳。针对影响负荷特征的关联因素,提出行业关联度这一负荷关联因素。整理现有电力用户行业数据,生成行业关联度矩阵,分析行业典型用电特征关联性。证明其对于异常用电识别具有积极的作用。

参考文献

- [1] 陈启鑫,郑可迪,康重庆,等.异常用电的检测方法:评述与展望[J].电力系统自动化,2018,42(17):189-199.
CHEN Q X, ZHENG K D, KANG CH Q, et al. Detection methods of abnormal electricity consumption behaviors: Review and prospect [J]. Automation of Electric Power Systems, 2018, 42(17): 189-199.
- [2] 林振智,崔雪原,金伟超,等.用户侧窃电检测关键技术[J].电力系统自动化,2022,46(5):188-199.
LIN ZH ZH, CUI X Y, JIN W CH, et al. Key

- technologies of electricity theft detection at consumer side[J]. Automation of Electric Power Systems, 2022, 46(5): 188-199.
- [3] 赵文清, 沈哲吉, 李刚. 基于深度学习的用户异常用电模式检测[J]. 电力自动化设备, 2018, 38(9): 34-38.
- ZHAO W Q, SHEN ZH J, LI G. Anomaly detection for power consumption pattern based on deep learning[J]. Electric Power Automation Equipment, 2018, 38(9): 34-38.
- [4] 赵天辉, 王建学, 马龙涛, 等. 基于非参数回归分析的工业负荷异常值识别与修正方法[J]. 电力系统自动化, 2017, 41(18): 53-59.
- ZHAO T H, WANG J X, MA L T, et al. Outlier detection and correction method for industrial loads based on nonparametric regression analysis[J]. Automation of Electric Power Systems, 2017, 41(18): 53-59.
- [5] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile edge computing: A survey[J]. IEEE Internet of Things Journal, 2018, 5(1): 450-465.
- [6] ELAHE M F, JIN M, ZENG P. Review of load data analytics using deep learning in smart grids: Open load datasets, methodologies, and application challenges[J]. International Journal of Energy Research, 2021, 45(10): 14274-14305.
- [7] 刘慧自, 汪颖, 胡文曦, 等. 考虑信息动态表达的异常用电模式识别云边协同方法[J]. 电力自动化设备, 2022, 42(7): 59-67.
- LIU H Z, WANG Y, HU W X, et al. Cloud-edge collaboration method for abnormal power consumption pattern recognition considering dynamic expression of information[J]. Electric Power Automation Equipment, 2022, 42(7): 59-67.
- [8] 朱天怡, 艾芊, 贺兴, 等. 基于数据驱动的用电行为分析方法及应用综述[J]. 电网技术, 2020, 44(9): 3497-3507.
- ZHU T Y, AI Q, HE X, et al. An overview of data-driven electricity consumption behavior analysis method and application[J]. Power System Technology, 2020, 44(9): 3497-3507.
- [9] WANG X L, AHN S H. Real-time prediction and anomaly detection of electrical load in a residential community[J]. Applied Energy, 2020, 259: 114145.
- [10] 李想, 王鹏, 刘洋, 等. 考虑类别不平衡的海量负荷用电模式辨识方法[J]. 中国电机工程学报, 2020, 40(1): 128-137.
- LI X, WANG P, LIU Y, et al. Massive load pattern identification method considering class imbalance[J]. Proceedings of the CSEE, 2020, 40(1): 128-137.
- [11] ZANETTI M, JAMHOUR E, PELLENNI M, et al. A tunable fraud detection system for advanced metering infrastructure using short-lived patterns[J]. IEEE Transactions on Smart Grid, 2019, 10(1): 830-840.
- [12] 肖勇, 郑楷洪, 余忠忠, 等. 基于三次指数平滑模型与DBSCAN聚类的电量数据异常检测[J]. 电网技术, 2020, 44(3): 1099-1104.
- XIAO Y, ZHENG K H, YU ZH ZH, et al. Power data anomaly detection based on Holt-winters model and DBSCAN clustering[J]. Power System Technology, 2020, 44(3): 1099-1104.
- [13] 赵天辉, 张耀, 王建学. 基于空间密度聚类 and 异常数据域的负荷异常值识别方法[J]. 电力系统自动化, 2021, 45(10): 97-105.
- ZHAO T H, ZHANG Y, WANG J X. Identification method of load outlier based on density-based spatial clustering and outlier boundaries[J]. Automation of Electric Power Systems, 2021, 45(10): 97-105.
- [14] 金伟超, 张旭, 刘晟源, 等. 基于剪枝策略和密度峰值聚类的行业典型负荷曲线辨识[J]. 电力系统自动化, 2021, 45(4): 20-28.
- JIN W CH, ZHANG X, LIU SH Y, et al. Identification of typical industrial power load curves based on pruning strategy and density peak clustering[J]. Automation of Electric Power Systems, 2021, 45(4): 20-28.
- [15] 杜章华, 苏盛, 刘正谊, 等. 基于生产经营状态识别的低误报率窃电检测二次筛查方法[J]. 电力系统自动化, 2021, 45(2): 97-104.
- DU ZH H, SU SH, LIU ZH Y, et al. Second inspection method for electricity theft detection with low false alarm rate based on identification of production and operation status[J]. Automation of Electric Power Systems, 2021, 45(2): 97-104.
- [16] LONG H, CHEN C, GU W, et al. A data-driven combined algorithm for abnormal power loss detection in the distribution network[J]. IEEE Access, 2020, 8: 24675-24686.

- [17] 严强, 李扬, 樊友杰, 等. 基于加权表决集成聚类的居民用电行为回归分析[J]. 电网技术, 2021, 45(11): 4435-4446.
YAN Q, LI Y, FAN Y J, et al. Regression analysis of residential electricity consumption behavior based on weighted voting ensemble clustering[J]. Power System Technology, 2021, 45(11): 4435-4446.
- [18] 王建元, 张少锋. 基于线性判别分析和密度峰值聚类的异常用电模式检测[J]. 电力系统自动化, 2022, 46(5): 87-98.
WANG J Y, ZHANG SH F. Anomaly detection for power consumption patterns based on linear discriminant analysis and density peak clustering[J]. Automation of Electric Power Systems, 2022, 46(5): 87-98.
- [19] SHEHZAD F, JAVAID N, ASLAM S, et al. Electricity theft detection using big data and genetic algorithm in electric power systems[J]. Electric Power Systems Research, 2022, 209: 107975.
- [20] 刘小峰, 叶榕婷, 柏林, 等. 基于灰色理论的航空发动机剩余寿命预测[J]. 电子测量与仪器学报, 2021, 35(1): 74-81.
LIU X F, YE R T, BO L, et al. Remaining useful life estimation for aero-engines based on grey theory[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(1): 74-81.
- [21] SHI J, DING Z, LEE W J, et al. Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features[J]. IEEE Transactions on Smart Grid, 2014, 5(1): 521-526.
- [22] 徐健, 刘长良, 王梓齐, 等. 基于缺失数据填补的风电齿轮箱状态监测研究[J]. 仪器仪表学报, 2022, 43(9): 88-97.
XU J, LIU CH L, WANG Z Q, et al. Research on condition monitoring of wind turbine gearbox based on missing data imputation[J]. Chinese Journal of Scientific Instrument, 2022, 43(9): 88-97.
- [23] ASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems 30-Proceedings of the 2017 Conference, California, Neural Information Processing Systems Foundation, 2017: 5999-6009.
- [24] 于晓升, 许茗, 王莹, 等. 基于卷积变分自编码器的异常事件检测方法[J]. 仪器仪表学报, 2021, 42(5): 151-158.
YU X SH, XU M, WANG Y, et al. Anomaly detection method based on convolutional variational auto-encoder[J]. Chinese Journal of Scientific Instrument, 2021, 42(5): 151-158.
- [25] AN J, CHO S. Variational autoencoder based anomaly detection using reconstruction probability[J]. Special Lecture on IE, 2015, 2(1): 1-18.
- [26] JOKAR P, ARIANPOO N, LEUNG C M. Electricity theft detection in AMI using customers' consumption patterns[J]. IEEE Transactions on Smart Grid, 2016, 7(1): 216-226.

作者简介



陈静, 分别于 2013 年和 2016 年于厦门大学获得硕士和博士学位, 现为福州大学副教授, 主要研究方向为电力大数据和智能电网。

E-mail: chenjing@fzu.edu.cn

Chen Jing received her M. Sc. degree and Ph. D. degree both from Xiamen University in 2013 and 2016, respectively. She is currently an associate professor at Fuzhou University. Her main research interests include power big data and smart grid.



江灏(通信作者), 分别于 2011 年和 2013 年于厦门大学获得硕士和博士学位, 现为福州大学副教授, 主要研究方向为电力故障辨识和智能感知系统。

E-mail: jiangh@fzu.edu.cn

Jiang Hao (Corresponding author) received his M. Sc. degree and Ph. D. degree both from Xiamen University in 2011 and 2013, respectively. He is currently an associate professor at Fuzhou University. His main research interests include power fault identification and intelligent sensing systems.