

doi: 10.7690/bgzd.2025.12.009

一种双阶段电力数据异常分析模型

刘 浩, 高明海, 俞文帅, 季溢贤, 刘立运, 黄金贤, 耿 斌, 王 彬
(国网镇江供电公司, 江苏 镇江 212001)

摘要: 为改善现有电力异常检测时效率、精度低的问题, 提出一种双阶段电力数据异常识别模型。模型由一步超前负荷预测器(one-step ahead load predictor, OSALP)和基于规则引擎的负荷异常检测器组成。OSALP结合线性自回归综合移动平均(autoregressive moving average, ARMA)和非线性人工神经网络(artificial neural networks, ANN)的优点, 可有效提高模型负荷预测精度。基于规则引擎的负荷异常检测器集成支持向量机(support vector machines, SVM)、 k 邻近(k -nearest neighbor, k NN)方法和交叉熵损失函数等模型优点, 可在不受预测结果影响下完成电力异常检测。实时检测结果表明: 模型 R^2 、均方根误差(root mean square error, RMSE)和 F_1 分数指标分别为 69.67%, 5.03%和 99.53%, 明显优于随机森林(random forest, RF)、SVM、ANN 以及长短期记忆(long short term memory, LSTM)等模型。

关键词: 电力系统; 异常检测; 负荷预测; SVM

中图分类号: TP393 **文献标志码:** A

A Two-stage Model For Analyzing Anomalies in Power Data

Liu Hao, Gao Minghai, Yu Wenshuai, Ji Yixian, Liu Liyun, Huang Jinxian, Geng Bin, Wang Bin
(State Grid Zhenjiang Power Supply Company, Zhenjiang 212001, China)

Abstract: To solve the problem of low efficiency and accuracy in existing power anomaly detection, a two-stage power data anomaly recognition model is proposed. The model consists of a one-step ahead load predictor (OSALP) and a rule-engine-based load anomaly detector. The one-step ahead load predictor combines the advantages of autoregressive moving average (ARMA) and artificial neural networks (ANN), effectively improving the accuracy of load prediction. The rule-engine-based load anomaly detector integrates the strengths of models such as support vector machines (SVM), k -nearest neighbor (k NN), and the cross-entropy loss function, enabling power anomaly detection independent of prediction results. Real-time detection results show that the model achieves R^2 , RMSE, and F_1 score metrics of 69.67%, 5.03%, and 99.53%, respectively, significantly outperforming models such as random forest (RF), SVM, ANN, and long short term memory (LSTM).

Keywords: power system; abnormal detection; load forecasting; SVM

0 引言

随着网络、大数据、物联网、通信技术^[1-3]的不断发展, 智能电网基于智能电表可收集大量用电数据。进一步, 可对这些海量数据进行分析, 从而推动电力系统服务质量提升; 然而, 在电力供应过程中存在有意或无意攻击引起的异常用电行为^[4-5]。这种行为不仅严重扰乱了电力的正常使用, 而且给电力系统造成了巨大的经济损失。同时, 未经授权修改线路或仪表容易导致停电和火灾等事故, 并对相关电力系统的安全构成严重威胁。

大量学者对电力系统异常检测方法进行研究, 并取得了丰硕成果。文献[6]提出了基于 GA-BP 神经网络窃漏电用户的识别方法, 可实现电力系统实时运行数据和窃漏电用户识别。文献[7]提出了一种

基于熵随机森林的电网用户窃电检测方法。文献[8]提出一种基于改进 K -medoids 聚类和支持向量机(SVM)的用电异常行为在线检测方法。文献[9]基于差分整合移动平均自回归模型和递归贝叶斯算法, 构建了一种针对配电网低压窃电行为的识别方法。文献[10]提出了一种基于双向长短期记忆神经网络的用户异常用电行为检测方法。人工神经网络^[6]对输入数据的形式很敏感, 存在过度拟合的问题。随机森林(RF)、SVM^[7-9]等模型调整参数困难。整体来说, 这些传统的机器学习方法易于实现, 适用于小样本的用电异常检测; 然而, 这些模型存在特征提取能力低和检测精度有限的问题。相对而言, 深度神经网络^[10]不仅具有强大的特征提取能力, 而且可以映射复杂的非线性关系, 使其比传统机器方法

收稿日期: 2024-10-26; 修回日期: 2024-11-23

第一作者: 刘 浩(1970—), 男, 江苏人。

具有更高的检测精度。深度学习方法需要大量样本数据集和消耗大量计算时间，检测效率及精度限制了其广泛应用。

为改善上述问题，笔者提出了一种双阶段电力数据异常识别模型。该模型由一步超前负荷预测器和基于规则引擎的负荷异常检测器组成，可有效提高模型检测精度并提升模型运算效率。

1 双阶段电力数据异常识别模型

双阶段电力数据异常识别模型旨在实现实时住宅电力负荷预测和异常数据检测，模型结构如图 1 所示。模型的输入为过去 24 h 的实时负荷数据和历史负荷数据。为了避免实时预测中的过拟合或欠拟合问题，同时平衡预测精度和计算速度的要求，该研究将滑动窗口设计为在每步中更新预测模型。在每步执行过程中，接收到实时数据之后，滑动窗口选择实时数据和过去 24 h 的历史数据作为输入，从而更新一步超前负荷预测器 (OSALP)。在更新 OSALP 并获得预测结果之后，使用基于规则引擎的负荷异常检测器确定实时数据的正确性。电力数据异常识别模型的输出由下一步的预测结果和实时数据的检测结果组成。

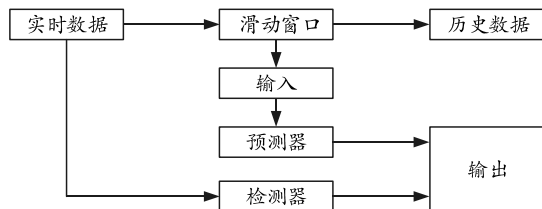


图 1 双阶段电力数据异常识别模型结构

1.1 住宅用户用电行为分析

一般情况下，家庭用电量数据受居民的生活习惯和随机事件 2 方面的影响。机器学习模型可以通过良好训练的数据集学习居民的生活习惯。然而，随机事件是不可预测的，这使得学习过程变得非常困难。一些随机活动，如突然打开浴室灯或煮咖啡，会随机、不规则、不受控制地发生。这些随机行为一方面反映了现代人生活的多样性，另一方面也增加了负荷预测和检测的难度。更具挑战性的是，住宅小区的总负载是每个家庭中所有此类随机事件的总和，从而使学习过程更加复杂。

在住宅家庭用户日负荷变化中也可以总结出一些规律，这些规律的某些部分有显著差异，例如，白天中居民负荷数据显示出很高的随机性；因此，可基于实施振动分析和自聚类来捕获住宅负载的固

有特征，并设计一种新的检测方法来降低异常检测的难度。

1.2 振动分析

振动分析通常指从时间序列的角度挖掘负荷序列特征。由于负载序列通常以强非线性方式随时间动态变化，因此首要任务是将非平稳序列转换为平稳序列。笔者使用的方法是计算负荷序列的 x 阶差分，并使用增广迪基-福勒 (augmented dickey-fuller, ADF) 检验来检测时间序列数据是否平稳。实际情况中，从晚上 7 点到次日早上 7 点，居民的用电量趋于稳定。然而，在其他时期，人类生活中随机事件对电力消耗产生剧烈的振动影响；因此，振动分析确定了振动和非振动周期。振动分析的结果为根据不同模型的不同周期和特征选择合适的载荷预测模型提供了指导。

1.3 自聚类

考虑到电力负荷的独立性、多样性和随机性不可避免地影响异常检测结果，笔者将负荷数据分类到不同的聚类中，从而降低随机事件引起的异常检测的难度。与一般的基于监督学习的分类技术相比，笔者提出的自聚类不依赖于训练数据集，并提供了一种处理复杂数据的简单数据探索方法。 K -means 聚类因其快速且简单等优点，已被广泛应用在众多数据分析领域，故使用具有 K -means 和轮廓系数的自动聚类方法。对于所有聚类模型，当使用 K -means 将 N 个数据分配给 k 个不相交的聚类时，必须预先设置 k 的大小 (簇的数量)。然而， k 对于新数据是未知的，因此，采用轮廓系数来衡量一个物体与自身簇群中其他物体相比与其他簇群中的物体相似性大小。轮廓系数定义如下：

$$S_c(k) = \frac{b(k) - a(k)}{\max(a(k), b(k))} \quad (1)$$

式中： $a(k)$ 为平均簇内距离； $b(k)$ 为 k 的平均最近簇距离。轮廓系数范围为 $[-1, +1]$ ，其中 $+1$ 为最佳值。

假设有 m 个聚类场景。对于每个场景，使用 K -means 对数据进行聚类，并使用轮廓系数来评估聚类结果。在 m 个聚类时期之后，将获得 m 个轮廓系数。最终选择轮廓系数最大值的聚类过程为最佳聚类结果。

自聚集后，将数据设置为 5 种负荷模式 (标记为 0~4)。模式 3 和模式 4 发生在振动期间，这种模式在白天是随机的。通过自动聚类，将不规则负

荷数据转换为不同的功耗模式。在异常检测阶段,检测器只需要检测实时负荷数据是否属于正常学习模式。

2 一步负荷预测

一般情况下,电力负荷时间序列数据包含线性特征和非线性特征 2 种;因此,应用于时间序列数据的综合预测因子应同时具有线性回归和非线性补偿:线性回归用于拟合数据,非线性补偿用于减少回归误差。

2.1 线性回归

在各种线性回归模型中,自回归移动平均 (ARMA) 模型因其准确性和灵活性的优点,已被广泛应用于许多时间序列数据预测研究中。ARMA 模型可以表示如下:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + n_t + b_1 n_{t-1} + \cdots + b_q n_{t-q} \quad (2)$$

式中: y_t 为 t 时的预测值,且与前 p 时的值相关,即 $y_{t-1}, y_{t-2}, \cdots, y_{t-p}$; n_t 为 t 时的误差; $a_i (i=1, 2, \cdots, p)$ 为自相关系数, p 为自回归项的数量; $b_i (i=1, 2, \cdots, q)$ 为移动平均系数, q 为滞后预测误差的数量。

同时,考虑到电力负荷数据具有典型的非平稳性,直接使用 ARMA 拟合数据会存在发散问题。笔者将基础 ARMA 模型进行推广,具体思路为:在拟合序列数据之前,所提改进的 ARMA 首先将非平稳时间序列转换为平稳的 d 阶差分序列数据,其中 d 为差分的阶数。

当应用于实时负荷预测时,扩展 ARMA 的一个限制是必须根据目标数据序列的特征预设 p 和 q 的值。通常,为了避免由设置不当的 p 和 q 值引起的过拟合或欠拟合问题,可使用自相关函数 (autocorrelation function, ACF) 和偏自相关函数 (partial autocorrelation function, PACF) 来分析目标数据序列。根据 ACF 和 PACF 的结果选择和预设 p 和 q 的最佳值。这种方法忽略了非平稳负荷序列数据随时间发生显著变化的问题,即用于先验分析和实时监控的负荷序列数据的特征可能会发生巨大变化。固定的参数可能不足以反映实时数据的特性,最终导致模型收敛性极差。

为改善上述问题,笔者引入了贝叶斯信息准则 (Bayesian information criterion, BIC) 从一组有限模型中选择最优模型的标准。BIC 的一般思想是使用信息先验将无限制模型缩小到精确基准,从而减少参数不确定性并提高预测精度。在对每个模型的拟

合结果进行评分之后,选择最低 BIC 分数作为最佳拟合模型的指标。对于自聚类过程,该研究假设 p 和 q 的最大值都为 n ,且产生编号从 1 到 n^2 个回归场景。故该过程可获得 n^2 个 BIC 分数,然后选择具有最低 BIC 的回归作为最终结果。BIC 定义如下:

$$\text{BIC} = k \ln m - 2 \ln \hat{L} \quad (3)$$

式中: $\hat{L}$ 为模型的最大似然函数值; m 为样本大小; k 为模型参数的数量。

2.2 非线性补偿

非线性补偿是预测模型的重要组成部分。尽管基于 BIC 可有效改进扩展 ARMA,然而现有模型精度仍然不够,无法适应复杂变化的电力负荷预测。为此,笔者选择非线性人工神经网络 (ANN) 作为非线性补偿模型,以补偿非线性回归中扩展的 ARMA 的拟合误差。ANN 是一种模拟大脑思考的非线性建模技术。ANN 训练数据集的输入是扩展 ARMA 的回归序列数据和相关的预测时间。输出为拟合误差,即回归结果和实际负载值之间的差异。同时,ANN 滑动窗口设计为记录训练数据集的输入和输出,并在每个步骤中更新 ANN 模型。

此外,考虑到居民用电数据的非线性特征,该研究选择双曲函数作为神经网络激活函数。同时,为了平衡计算速度和精度,网络的学习率设置为 0.01。

3 基于规则引擎的负荷异常检测器

基于规则引擎的负荷异常检测器中规则引擎用来实时检测电力异常。在振动期间,随机事件和预测误差频繁发生;因此,为了准确识别异常数据并减少随机事件对检测的影响,规则引擎按以下顺序运行:首先,执行结果分类;其次,执行分类结果比较;接着,进行分类错误测试;最后,对分类结果进行评估。

3.1 结果分类

为了避免预测误差对异常检测的影响,不能简单地比较预测结果和实际负载的差异。使用 SVM 来探索结果是否表现出相同的用电模式。

SVM 是一种广泛使用的分类器,通过引入核函数,SVM 可从正常数据中分离异常值。由于核函数是非参数的,并且是局部操作的,因此不需要对所有数据具有相同的函数形式。在该研究中,SVM 模

型采用高斯核函数。通过在高维空间中定位解空间，基于高斯核的 SVM 可以有效地解决由非线性决策函数引起的精度低的问题，从而提高模型的非线性分类性能。SVM 的训练数据集由正常用电数据集和异常用电数据集组成。该研究使用的正常用电数据集主要为自聚类结果。异常用电数据集包含窃电和漏电 2 种异常事件。需要注意的是，由于窃电和漏电数据量较少，在历史数据集中，与正常数据的数量相比，异常的数量所占比重很低。为了防止数据不平衡导致的分类错误，笔者基于数据增强操作丰富样本数据。具体步骤如下：

1) 随机选择来自正常数据集的 n 个数据以形成新的正常数据集 x ，即 $x \in \{x_1, x_2, \dots, x_n\}$ 。

2) 考虑到异常用电时电力消耗的结果比正常使用时要低，则异常数据集 y 可以表示为：

$$y = \alpha x. \quad (4)$$

式中 α 为超参数，通常 $\alpha \in [0, 1]$ 。

3) 基于 SVM 的泛化特性来生成异常数据集。

3.2 对比分类结果

分类结果可分为 3 种情况：

1) 如果预测结果和实际负荷呈现相同的模式，并且为正常模式，则当前负荷值是正常的；

2) 如果预测结果和实际负荷呈现相同的模式，且为异常模式，则负荷异常，并执行告警；

3) 如果预测结果和实际负荷呈现不同的模式，则有 2 个潜在的原因：随机事件影响负荷并使真实数据发生很大变化，从而导致不同的模式；存在分类错误。此时，所提算法执行基于最邻近 (k NN) 的分类错误检测，以重新评估分类结果。如果没有观察到错误，则需要确定分类结果是否可接受。

3.3 分类错误检测

预测和实际数据呈现出不同的模式时，则基于 k NN 执行分类错误检测。该过程具体步骤如下：

1) 使用 k NN 方法查找最接近实际负荷的训练数据集模式，并将其标记为 P_{kNN} ；

2) 如果实际负荷与 P_{kNN} 相同，则无分类错误。居民用电的日常随机事件是预测和实际结果呈现不同模式的原因；反之如果实际结果是正常的，则跳转至 3)。

3) 如果实际值和 P_{kNN} 表现出不同的模式，则存在分类错误。

需注意，当 k NN 结果与 SVM 结果不同时，规

定 k NN 结果为正确。分析原因，当训练数据质量不足或正则化参数不合适时，SVM 可能会产生错误分类。相比之下， k NN 方法依赖于有限数量的邻居；因此，当出现分类错误时， k NN 可用于校正分类结果。

3.4 分类结果评估

在执行完分类后，该研究基于交叉熵损失函数，从全局优化的角度评估分类结果是否可接受。交叉熵损失函数是一种主流的分类/聚类结果评估方法。交叉熵损失函数定义如下：

$$L(y, \hat{y}) = \sum_{i=1}^T y_i \log(\hat{y}_i). \quad (5)$$

式中： y 为训练数据集中的真实标签； $\hat{y}$ 为 SVM 的分类结果； T 为训练数据集的大小。此外，交叉熵损失越低，表征分类效果越好。

4 仿真与分析

4.1 数据集

为验证所提模型有效性，基于中国某电网公司智能电表数据集进行模拟和分析。该数据集统计了中国某市 2018 年 3 月至 2020 年 12 月住宅及部分商业用户用电情况，用电电力的采样周期设置为 30 min。

4.2 仿真环境与参数设置

仿真软件环境为 Pycharm 搭建算法框架，并由 Python 搭建基础算法。同时，算法运行硬件环境为酷睿 i7 CPU，内存为 128 G ARM 的联想服务器，操作系统为 Ubuntu 18.04 64 位，显卡为 NVIDIA RTX2080Ti 11G。

SVM 模型训练时采用 SGD 优化器训练模型，实验时部分参数定义如下：批量大小设置为 16，初始学习率为 10^{-2} ，学习率衰减率为 10^{-1} ，学习率衰减周期设置为 10，最大迭代次数设置为 15 000。

4.3 性能分析

4.3.1 模型综合性能

对比了所提模型与基础模型：RF、support SVM、人工神经网络 ANN、长短时记忆 (LSTM) 的性能。同时，将基础模型加入预测模型，从而构成双阶段模型：ARIMA/ANN-AD 和 LSTM-AD 从而验证所提模型的优势。实验时选取拟合优度 (R^2)、均方根误差 (RMSE) 以及 F_1 分数等指标来衡量模型

性能。其中, R2 回归度量用于根据结果的总变化比例提供每个模型复制观察结果的直观度量。R2 范围从 0 到 1, 且分数越高, 拟合越好。RMSE 是模型预测值与观测值之间差异的另一种常用度量, 计算如下:

$$\text{RMSE} = \frac{1}{N} \sqrt{\sum_i^N (P_i - O_i)^2} \quad (6)$$

式中: N 为样本个数; P_i 为第 i 个样本的真实值; O_i 为第 i 个样本的估计值。需注意, RMSE 较小表示模型估计能力较高, 较大表示模型估计能力较低。

F_1 分数是一种提供分类器整体性能的平衡评估度量。 F_1 分数为精度 Precision 和召回率 Recall 的调和平均值, 定义如下:

$$\left. \begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{FP} + \text{TP}} \\ \text{Recall} &= \frac{\text{TP}}{\text{FN} + \text{TP}} \\ F_1 &= \frac{2\text{PrecisionRecall}}{\text{Precision} + \text{Recall}} \end{aligned} \right\} \quad (7)$$

式中: TP 为正常数据被正确分类的数量; FP 为异常数据被分类为正常的数量; TN 为异常数据被分类为异常的数量; FN 为正常数据被分类为异常的数量。

不同模型异常检测性能比较结果如表 1 所示。不同模型中, 基础 RF、SVM、ANN 以及 LSTM 模型无负荷预测模型, 直接进行电力数据异常检测; ARIMA/ANN-AD、LSTM-AD 和所提模型包含预测模型来预测当前或未来负荷, 并基于预测结果执行电力异常检测。

表 1 不同模型异常检测性能比较结果

模型	R2	RMSE	F_1
RF	N/A	0.139 1	0.763 5
SVM	N/A	0.120 2	0.795 6
ANN	N/A	0.120 6	0.806 9
LSTM	N/A	0.108 9	0.835 8
ARIMA/ANN-AD	0.592 8	0.098 0	0.842 7
LSTM-AD	0.363 1	0.050 6	0.973 7
所提模型	0.696 7	0.050 3	0.995 3

从表 1 可以看出: 双阶段模型的 R2 和 RMSE 指标明显优于基础模型(无预测)。这表明预测模型的引入增强了异常检测模型性能, 可以有效地减少异常检测模型的过拟合或欠拟合问题的影响, 从而提高了预测精度。同时, 所提模型 R2、RMSE 和 F_1 分数指标最优, 分别为 0.696 7、0.050 3 和 0.995 3。分析原因, 所提模型包含预测器和异常检测器双阶段, 并且滑动窗口在每个步骤中更新模型。同时,

考虑到住宅负荷中人为引起的随机事件对预测结果的影响, 每个模型的预测误差是不可避免的。在这种情况下, 确定如何实现自动预测误差补偿成为提高实时预测精度的关键。对于所提模型, 即使出现预测误差, 仍可通过 k NN 对模型进行修正, 从而实时拟合滑动窗口选择的最新历史数据。该设计确保了观测值与模型下预期值之间的差异较小。相比之下, 尽管 LSTM-AD 使用了更复杂的深度神经网络, 但该模型仍然缺乏减少预测误差的反馈机制。为了更直观地展示每个模型的预测结果, 图 2 为不同双阶段模型实时预测结果。表 2 为不同模型预测性能指标。

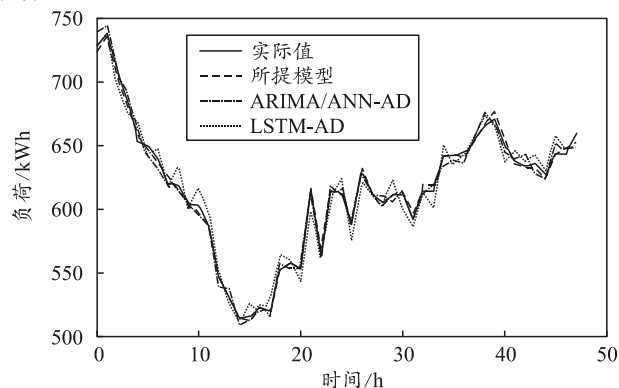


图 2 不同双阶段模型实时预测结果

表 2 不同模型预测性能指标

模型	R2	RMSE
ARIMA/ANN-AD	0.609 9	0.128 0
LSTM-AD	0.738 4	0.040 6
所提模型	0.771 1	0.031 8

可以看出, 所提模型预测性能均优于其余 2 种双阶段方法。

此外, 在所有双阶段异常检测结果中, ARIMA/ANN-AD 表现出最低的 F_1 分数。分析原因, 用于识别异常的基于一般回归模型的异常检测方法过度依赖于预测模型的结果。所提模型的结果表明: 为了提高基于回归模型的异常检测精度, 需要独立的异常检测过程。根据 LSTM-AD 和 ARIMA/ANN-AD 的 F_1 分数对比结果, LSTM-AD 具有较高的检测准确率, 主要原因: 1) LSTM 网络的预测精度高于 ARIMA/ANN; 2) LSTM-AD 的检测过程没有比较特定时间的预测负载和实际负载之间的误差, 但在分析趋势时, 其证明了检测的合理性。同时, 与其他方法相比, 基础模型(RF、SVM、ANN 以及 LSTM 模型)无法提供有关未来负荷的更多信息, 这限制了其推广应用。

综上, 实验结果验证了所提模型在所有异常检

测器中最佳的预测和异常检测性能。

4.3.2 运行效率分析

对 ARIMA/ANN-AD、SVM-AD 和所提模型的运行速度进行对比。

经过 100 次实验验证，并选取平均值为平均计算时间指标值，不同模型运行速度对比结果如表 3 所示。

表 3 不同模型运行速度对比结果

模型	预测速度	检测速度	总时间/s
ARIMA/ANN-AD	378	468	0.846
LSTM-AD	1 876	1 144	3.020
所提模型	403	500	0.903

所提模型的平均计算时间约为 400 ms，预测和检测的总计算时间约为 0.9 s，接近实时检测。然而，LSTM-AD 的运行速度比其他方法慢，约为 3 s。因此，在实际应用中，计算速度太慢，导致 LSTM-AD 模型无法进行实时检测。同时，LSTM-AD 的局限性在于，该方法为基于深度神经网络的模型，其实用性受到硬件配置的限制。相比之下，所提模型以简单的硬件配置在计算机上实现了实时负荷预测和异常检测，实用性较高。

5 结束语

笔者对电力数据进行研究与分析，建立了一种双阶段电力异常检测模型：首先，基于实时负荷数据和历史负荷数据进行一步超前负荷预测；其次，在获得预测结果之后，使用基于规则引擎的负荷异常检测器确定实时数据的正确性，从而输出电力数

据异常识别结果。实时检测结果表明：该模型为电力数据分析及安全故障隐患的发现提供了一定借鉴作用。

参考文献：

[1] 毛龙灿, 杨南. 基于大数据背景的皮革人才培养优化研究[J]. 中国皮革, 2021, 50(9): 38-41.

[2] 杨涛. 互联网时代下皮革行业电商物流体系研究分析[J]. 中国皮革, 2021, 50(8): 82-85.

[3] 钟建棚, 余少锋, 廖崇阳, 等. 基于云计算的电力设备智能监测系统[J]. 云南师范大学学报(自然科学版), 2022, 42(3): 37-41.

[4] 严莉, 张凯, 徐浩, 等. 基于缺失值补全与 BiLSTM 的用电行为异常检测[J]. 软件导刊, 2022, 21(10): 136-141.

[5] 苏鹏涛, 孙晨辉. 基于 LOF 分析算法的异常用电行为研究[J]. 信息与电脑(理论版), 2022, 34(15): 55-57.

[6] 韩建富, 肖春, 宋小兵, 等. 基于 GA-BP 神经网络的能源互联网窃电行为识别方法[J]. 电气传动, 2022, 52(14): 38-44.

[7] 阙华坤, 冯小峰, 刘盼龙, 等. Grassberger 熵随机森林在窃电行为检测的应用[J]. 计算机科学, 2022, 49(S1): 790-794.

[8] 胡聪, 徐敏, 洪德华, 等. 基于改进 K-medoids 聚类 and SVM 的异常用电模式在线检测方法[J]. 国外电子测量技术, 2022, 41(2): 53-59.

[9] 胡一伟, 刘珊, 黄浩. 基于 ARIMA 和递归贝叶斯的窃电用户识别算法[J]. 电测与仪表, 2022, 59(6): 196-200.

[10] 谢禄江, 段立, 谭刚, 等. 基于双向 LSTM 的用户用电行为识别[J]. 微电子学与计算机, 2021, 38(5): 36-41.