Modeling and Simulation 建模与仿真, 2025, 14(9), 30-44
Published Online September 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.149582

Hans汉斯

基于改进USAD模型的用电数据时序无监督异常检测

孙　慧，智路平

上海理工大学管理学院，上海

收稿日期：2025年8月4日；录用日期：2025年8月28日；发布日期：2025年9月5日

摘　要

异常检测在电力行业具有重要意义。然而，在实际应用中往往会出现正常或异常样本标签不足的情况，导致检测能力下降。因此，本文针对用电数据无监督异常检测中缺乏时间上下文信息及异常样本稀缺的问题，提出了一种基于改进USAD模型的用电数据时序无监督异常检测。在模型训练阶段，首先提取多维时序特征包括小时、星期、月份、年内日序和年内周序，结合正余弦函数进行周期性编码，并构造关键滞后特征lag1、lag24、lag168，捕捉用电序列的短期波动和长期变化。其次，设计五种类型的合成异常样本包括尖峰、趋势、模式断裂、水平偏移、方差变点注入用电序列数据中，使模型学习各类异常样本的特征。在模型检测阶段，基于USAD的检测架构引入了一种EMA平滑下的异常分数组合与自适应阈值设定机制，缓解异常分数中的噪声波动，从而提升模型的泛化性和鲁棒性。实验结果显示该模型AUC达到了83.91%，准确率达到98.14%，Recall值为43.34%，F1值为60.13%，继而在十个数据集上进行泛化性能测试，结果表明该检测方法具有较好的检测异常样本的能力。

关键词

异常检测，无监督，AUC，尖峰，特征

Unsupervised Anomaly Detection in Electricity Consumption Time Series Based on an Improved USAD Model

Hui Sun, Luping Zhi

Business School, University of Shanghai for Science and Technology, Shanghai

Received: Aug. 4th, 2025; accepted: Aug. 28th, 2025; published: Sep. 5th, 2025

文章引用: 孙慧, 智路平. 基于改进 USAD 模型的用电数据时序无监督异常检测[J]. 建模与仿真, 2025, 14(9): 30-44.
DOI: 10.12677/mos.2025.149582

Abstract

Anomaly detection plays a vital role in the power industry. However, labeled normal or anomalous samples are often scarce in practice, which degrades detection performance. To address the lack of temporal context information and the scarcity of anomalous samples in unsupervised anomaly detection of electricity consumption data, the author proposed an improved USAD-based unsupervised anomaly detection method for electricity consumption time series. In the training stage, the author first extracted multi-dimensional temporal features—including hour, weekday, month, day of year and week of year—encoded periodicity via sine and cosine transformations, and constructed key lag features (lag 1, lag 24 and lag 168) to capture short-term fluctuations and long-term variations of the consumption series. Next, five types of synthetic anomalies—spike, trend, pattern break, level shift and variance change—were injected into the consumption time series, enabling the model to learn the characteristics of various anomaly types. In the detection stage, the author introduced an EMA smoothing approach for combining anomaly scores and an adaptive threshold-setting mechanism into the USAD detection framework to mitigate noise fluctuations in anomaly scores, thereby improving the model's generalization ability and robustness. Experimental results show that the model achieves an AUC of 83.91%, an accuracy of 98.14%, a recall of 43.34% and an F1 score of 60.13%. Furthermore, generalization performance tests on ten additional datasets demonstrate that this detection method exhibits strong capability in identifying anomalous samples.

Keywords

Anomaly Detection, Unsupervised, AUC, Spikes, Features

1. 引言

传统的电力异常检测方法主要依靠人力实现，通过人工检查和对比异常电表读数，检查被绕行的输电线路[1]。然而这种方法不仅费时费力，还会导致过高的人工成本和较低的效率。随着智能电网中大量智能硬件设备的铺设，由数据驱动主导的电力异常检测方法得到广泛应用。这类检测方法通过对海量数据进行分析和挖掘，从而学习数据中的异常模式。从数据挖掘的角度来看，电力异常检测方法可分为有监督学习和无监督学习两类[2]。有监督学习依赖于标注了正常用电行为与异常用电行为的样本，但这类带标签的数据在实际场景中往往难以获得。因而，本文聚焦于无监督的用电数据异常检测，这种方法可以在数据标签是未知的情况下，通过挖掘该用户自身的周期性用电模式来识别出疑似窃电的异常用电数据。

早期无监督用电数据异常检测研究主要依赖于单一的无监督学习框架。庄池杰等[3]提出了通过主成分分析便于计算离群因子，网格处理筛选出低密度区域的数据点，最后基于无监督学习的异常用电检测模型进行异常检测。Jesmeen 等[4]使用无监督异常检测模型基于高斯混合模型(GMM)和隔离森林(iForest)，将这些模型与局部异常值因子(LOF)和单类支持向量机(OCSVM)进行了比较，发现 iForest 作为检测模型表现出色。Kaleta 等[5]使用了五种无监督异常检测方法：K 最近邻、局部异常值因子、主成分分析、隔离森林和基于直方图的异常值分数，并应用了两种时间序列相似度量，来检测电力异常。随着算法的多样化发展，无监督用电数据异常检测研究走向多模型融合。王建元和张少锋[6]提出一种基于线性判别分析和密度峰值聚类的双判据无监督异常用电检测模型，即遵循“特征构造－维度规约－聚类－异常检测”的流程，借助

聚类算法将进行异常检测。傅世元[7]等针对基于动态选择集成的无监督异常检测方法存在衡量基检测器性能的评价基准不准确且只使用单一性能指标的问题，提出混合选择集成思想和基检测器动态选择集成策略进行检测。华回春等[8]提出了结合发电企业的串谋预警指标体系和无监督的变分自编码高斯混合模型(VAEGMM)，进行准确的密度估计，从而可以准确地预警串谋风险。近些年无监督用电数据异常检测研究开始聚焦于无监督技术的改进和半监督机制。雷磊等[9]提出了一种建筑能耗数据动态异常检测算法，实现了点异常和集体异常的动态检测。该算法将无监督聚类算法与监督算法相结合，建立了半监督匹配机制，避免了误差标签的影响，提高了异常检测效率。Almalki [10]提出了一种独特的识别窃电的混合技术，即通过将监督学习模型(随机森林)与无监督学习算法(隔离森林、单类支持向量机(SVM)、局部异常值因子(LOF))和基于密度的噪声应用空间聚类(DBSCAN)融合在一起进行电力异常检测。Mishra 等[11]通过采用一种称为基于噪声应用的基于分层密度的空间聚类(HDBSCAN)的新型无监督数据驱动聚类技术来检测窃电事件。

目前国内外对无监督用电数据异常检测的研究存在以下不足。首先，缺乏时间上下文信息以及对时间依赖性建模不足，难以捕捉用电序列的短期波动和长期变化，这会导致用电相关方无法对突发用电峰谷做出及时响应，影响电网调度的精准性和风险预警的有效性。其次，真实场景中异常往往很少，模型难以学习到多样化的故障模式，这会使运维方在面对新型或少见故障时检测效率下降，从而增加巡检成本。针对上述问题，本文提出基于改进 USAD 模型的用电数据时序无监督异常检测，旨在以下三个方面做出突破：1) 提取多维时序特征包括小时、星期、月份、年内日序和年内周序，结合正余弦函数进行周期性编码，并构造关键滞后特征 lag1、lag24、lag168，从而捕捉用电序列的短期波动和长期变化；2) 设计五种类型的合成异常样本包括尖峰、趋势、模式断裂、水平偏移、方差变点注入用电序列中，使模型学习各类异常信号的特征，从而在实际检测中能够利用这些特征准确识别多样化的异常模式；3) 为抑制异常分数中的噪声波动，在 USAD 模型检测中引入了一种 EMA 平滑下的异常分数组合与自适应阈值设定机制，从而提升模型的泛化性和鲁棒性。

2. 基于改进 USAD 模型的用电数据时序无监督异常检测

本文用电数据异常检测的框架如图 1 所示。

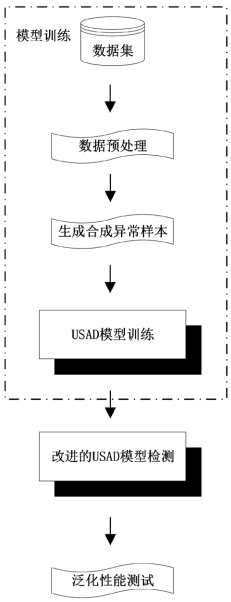

Figure 1. Abnormal electricity consumption detection framework
图 1. 用电数据异常检测框架

2.1. 改进的 USAD 模型训练

2.1.1. 数据预处理

本文使用的窃电数据集是低碳伦敦(Low Carbon London, LCL)数据集，该数据集记录了某用户 738 天的连续用电记录，采样间隔为 30 min，单位 KWh，默认该用户为诚实用户，则该用户的用电序列为正常用电序列。

首先从数据集中提取多维特征包括小时(*Hour*) (即一天中不同时间段的波动)、星期(*DayOfWeek*) (即周一至周日的用电差异)、月份(*Month*) (即年内季节性变化)、年内日序(*DayOfYear*) (即一年中第几天，细化年内逐天变化)和年内周序(*WeekOfYear*) (即一年中第几周，捕捉跨月周的周期模式)，为模型提供时间上下文信息。其次，由于小时和星期这两个短周期特征极易发生端点不连续性问题，影响模型捕捉真实的连续用电模式，对 *Hour* 和 *DayOfWeek* 这两种具有短而强信号的周期采用正余弦函数[12]生成 $Hour\_\sin$、$Hour\_\cos$、$DayOfWeek\_\sin$ 和 $DayOfWeek\_\cos$ 四个周期性特征，使模型能够捕捉用电模式的昼夜模式和周末模式，正余弦函数的数学表达式如下：

$$Hour\_\sin = \sin\left(2\pi \cdot Hour/24\right) \tag{1}$$

$$Hour\_\cos = \cos\left(2\pi \cdot Hour/24\right) \tag{2}$$

$$DayOfWeek\_\sin = \sin\left(2\pi \cdot DayOfWeek/7\right) \tag{3}$$

$$DayOfWeek\_\cos = \cos\left(2\pi \cdot DayOfWeek/7\right) \tag{4}$$

为了更好地建立时间依赖性，定义用电量特征 *KWHhh* (单位：KWh，表示每小时的用电量)，基于此，构造关键滞后特征 $KWHhh\_lag1$ (前一小时用电量)捕捉短期依赖，$KWHhh\_lag24$ (前一天相同时刻用电量)捕捉日周期模式，$KWHhh\_lag168$ (前一周相同时刻用电量)捕捉周周期模式。缺失值采用前向填充(*ffill*)与后向填充(*bfill*)结合的双向填充进行处理，缺失值定义如下。

1) 定义原始序列数据，即 $x_1, x_2, \cdots, x_n$ (含缺失值 *NaN*)；

2) 进行前向填充，即对于每一个缺失点 t，找到离它最近且在它之前非缺失的时间点，记为：

$$x_i^{ffill} = \begin{cases} x_i, x_i \notin NaN \\ x_j, \text{else}, j = \max\left(k \middle| k < i, x_k \notin NaN\right) \end{cases} \tag{5}$$

3) 进行后向填充，即对于每一个缺失点 t，找到离它最近且在它之后非缺失的时间点，记为：

$$x_i^{bfill} = \begin{cases} x_i, x_i \notin NaN \\ x_j, \text{else}, j = \min\left(k \middle| k > i, x_k \notin NaN\right) \end{cases} \tag{6}$$

考虑到实际用电数据可能存在异常值，对所有数值特征(除三角函数特征以外)进行四分位距(*IQR*)异常值修正，计算每个特征的 *IQR* 并将超出 $\left[Q_1 - 1.5IQR, Q_3 + 1.5IQR\right]$ 范围的值裁剪至边界，消除极端值对后续分析的干扰，四分位距异常值修正的定义如下：

1) 计算四分位数：

$$Q_1 = Quantile_{0.25}\left(X\right), Q_3 = Quantile_{0.75}\left(X\right) \tag{7}$$

其中，$Quantile_p\left(X\right)$ 表示 p 分位数。

2) 计算四分位距：

$$IQR = Q_3 - Q_1 \tag{8}$$

3) 确定上下界：

$$L = Q_1 - 1.5IQR, U = Q_3 + 1.5IQR \tag{9}$$

其中，L 和 U 分别是对偏小和偏大极端值的控制临界点。

4) 对每个样本值 x_i，得到裁剪后值 x_i^*：

$$x_i^* = \begin{cases} L, x_i < L \\ x_i, L \le x_i \le U \\ U, x_i > U \end{cases} \tag{10}$$

最后对特征使用标准化，使所有特征均值为 0，方差为 1，标准化的公式如下：

$$z_i = \frac{x_i^* - \mu}{\sigma}, \mu = \frac{\sum_{i=1}^{n} x_i^*}{n}, \sigma = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i^* - \mu\right)^2}{n}} \tag{11}$$

在用电时间序列处理阶段，将处理后的用电时间序列数据重构为 24 小时(1 天)长度的连续序列样本，即每个样本构成一个 264 维的特征向量，作为模型的输入，重构样本的定义如下。

设原始时间序列为 $x = x_1, x_2, \cdots, x_T$，$x_t \in \mathbb{R}^{11}$，即设定每个时间步 x_t 是一个 11 维特征向量(包含 $KWHhh$、$KWHhh\_lag1$、$KWHhh\_lag24$、$KWHhh\_lag168$、$Hour\_\sin$、$Hour\_\cos$、$DayOfWeek\_\sin$、$DayOfWeek\_\cos$、$Month$、$DayOfYear$、$WeekOfYear$)。

通过滑动窗口提取连续 24 个时间步的特征向量构成样本矩阵：

$$S_i = \begin{bmatrix} x_i \\ x_{i+1} \\ \vdots \\ x_{i+23} \end{bmatrix} \tag{12}$$

该矩阵维度为 24×11，表示从第 i 小时起始的连续 24 小时数据。为形成模型输入，将此矩阵按行展平为特征向量 v_i：

$$v_i = vec\left(S_i\right) = \left[x_{i,1}, x_{i,2}, \cdots, x_{i,11}, x_{i+1,1}, \cdots, x_{i+23,11}\right]^{\mathrm{T}} \tag{13}$$

其中 $x_{t,k}$ 是时间步 t 的第 k 个特征值。进行展平使 24 个时间步的 11 维特征首尾拼接，生成一个固定的 264 维输入向量($24 \times 11 = 264$)，即：

$$v_i \in \mathbb{R}^{264} \tag{14}$$

通过这一系列操作，原始的时间序列数据被转化为具有固定长度的输入样本，为后续的模型训练做准备。

2.1.2. 合成异常样本生成

为了模拟不同的电力异常场景，本文设计五种类型的合成异常样本包括尖峰、趋势、模式断裂、水平偏移、方差变点注入用电序列样本中，使模型学习各类异常信号的特征，具体内容如下：

1) 尖峰异常(Spike)，该方法通过突发峰值或电流尖峰来模拟电力系统的异常，尖峰异常的定义如下：使用均匀分布，随机选取 k 个索引集合 $S \subset \{1, \cdots, d\}$，$|S| = k$，对每个 $j \in S$ 施加：

$$\tilde{x}_j = x_j\left(1 + \delta_j\right), \delta_j \sim Uniform\left(a, b\right), a = 3, b = 5 \tag{15}$$

2) 趋势异常(Trend)，该方法模拟设备老化或外部环境变化导致的异常，即通过时间递增的方式，模

拟电力使用趋势的变化，趋势异常的定义如下：

在整个向量长度上按线性趋势叠加：

$$\tilde{x}_j(t)=x_j(t)\left[1+\mathrm{sgn}(\tau)\frac{|\tau|}{T}t\right],\tau\sim Uniform(-c,c),c\in[2,3] \tag{16}$$

其中，$t=1,\cdots,T$ 为时刻索引，τ 代表设备老化的速率或外部环境波动的强度。

3) 模式断裂(Pattern Break)，该方法模拟电力操作过程中的意外中断，模式断裂的定义如下：

选定某一通道 j，以及起始位置 t_0 和长度 L，在该区间用噪声覆盖：

$$\tilde{x}_j(t)=\begin{cases}\mu_j+\epsilon_j, t_0\le t<t_0+L\\ x_j(t), \text{otherwise}\end{cases},\epsilon_j\sim Uniform(-2\sigma_j,2\sigma_j) \tag{17}$$

其中，μ_j 表示故障前的正常电力水平，ϵ_j 表示故障引起的偏差。

4) 水平偏移(Level Shift)，该方法通过给电力使用值添加一个随机的偏移量来模拟电力使用的水平变化，水平偏移的定义如下：

对整条序列加上常数偏移：

$$\tilde{x}_j(t)=x_j(t)+\Delta,\Delta=s\sigma_j,s\sim Uniform([-3,-2]\cup[2,3]) \tag{18}$$

5) 方差变点(Variance Change)，该方法模拟电力数据中由于故障或突发事件导致的波动增大，方差变点的定义如下：

在原始样本基础上加入放大噪声：

$$\tilde{x}_j(t)=x_j(t)+\eta_t,\eta_t\sim N\left(0,(m\sigma_j)^2\right),m\sim Uniform(3,5) \tag{19}$$

2.1.3. USAD 模型训练

USAD (Unsupervised Anomaly Detection)是一种基于对抗训练的双自编码器结构，由三个元件组成：编码器网络 E 和两个解码器网络 D_1 和 D_2，这三个元件形成两个自编码器 AE_1 和 AE_2，它们共享相同的编码器 E，其中，

$$AE_1(X)=D_1(E(X)),AE_2(X)=D_2(E(X)) \tag{20}$$

USAD 模型的训练过程分为两个阶段进行，见图 2 所示，具体过程如下。

第一阶段：自编码器训练。此阶段目的在于训练两个自编码器学习正常输入窗口 X 的重构能力。输入数据 X 经编码器 E 压缩至隐空间 Z，随后由两个解码器分别重构。训练目标通过最小化重构误差实现，其损失函数定义为：

$$\mathcal{L}_{AE_1}=\|X-AE_1(X)\|_2,\mathcal{L}_{AE_2}=\|X-AE_2(X)\|_2 \tag{21}$$

第二阶段：对抗训练。此阶段采用对抗机制训练两个自编码器，即 AE_1 试图欺骗 AE_2，而 AE_2 学习区分原始数据 X 与 AE_1 的重构输出 $AE_1(X)$。具体而言，AE_1 的生成数据被编码器 E 再次压缩至 Z，再由 AE_2 重构。训练目标定义为：

$$\min_{AE_1}\max_{AE_2}\|X-AE_2(AE_1(X))\|_2 \tag{22}$$

该目标分解为两个损失函数：

$$\mathcal{L}_{AE_1}=+\|X-AE_2(AE_1(X))\|_2,\mathcal{L}_{AE_2}=-\|X-AE_2(AE_1(X))\|_2 \tag{23}$$

自编码器具有双重目标，即 AE_1 需最小化对 X 的重构误差，同时最小化 AE_2 对其输出的重构误差；AE_2 需最小化对 X 的重构误差，但最大化对 AE_1 输出的重构误差。这两个目标通过权重调度进行动态调整，其中目标的权重会随着训练轮次的增加而动态调整，进一步优化的损失函数公式如下：

$$\begin{aligned}\mathcal{L}_{AE_1} &= \frac{1}{n}\left\|X - AE_1(X)\right\|_2 + \left(1-\frac{1}{n}\right)\left\|X - AE_2\left(AE_1(X)\right)\right\|_2,\\ \mathcal{L}_{AE_2} &= \frac{1}{n}\left\|X - AE_2(X)\right\|_2 - \left(1-\frac{1}{n}\right)\left\|X - AE_2\left(AE_1(X)\right)\right\|_2\end{aligned} \tag{24}$$

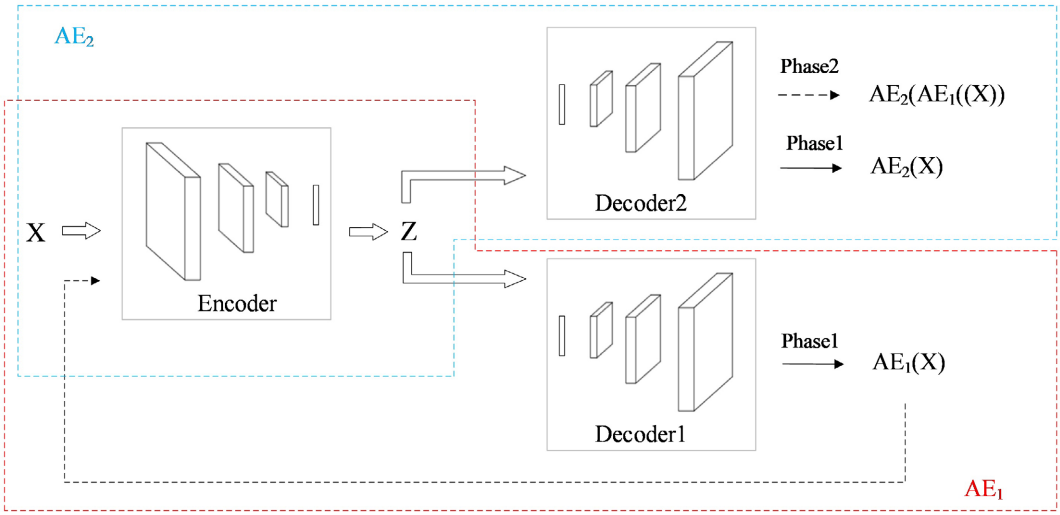

Figure 2. USAD model training
图 2. USAD 模型训练

2.2. 改进的 USAD 模型检测

针对基础 USAD 模型检测仅依赖解码器重构误差、忽视潜在表示的问题，本文在其基础上引入了一种更全面且稳定的异常分数组合与自适应阈值设定方法。

首先通过编码器将输入数据映射为潜在表示 z_t，同时计算潜在空间误差项：

$$\mathcal{L}_{latent} = \left\|z_t\right\|_2 \tag{25}$$

其次，重构过程由两个自编码器 AE_1 与 AE_2 完成，生成的重构误差包含 MSE 和 $L1$ 损失，重构误差的定义如下：

$$MSE\left(x_t,\hat{x}_t\right) = \frac{1}{d}\sum_{j=1}^{d}\left(x_{j,t}-\hat{x}_{j,t}\right)^2, L1 = \sum_{j=1}^{d}\left|x_{j,t}-\hat{x}_{j,t}\right| \tag{26}$$

其中，d 为通道数。

最终异常分数 s_t 在时间步 t 由加权组合生成：

$$s_t = \alpha MSE\left(x_t,\hat{x}_t\right) + \beta\left\|x_t-\hat{x}_t\right\|_1 + \gamma\left\|z_t\right\|_2 \tag{27}$$

其中，α、β、γ 是三项误差的权重系数，最佳权重组合由网格搜索确定，x_t 为原始输入，$\hat{x}_t$ 为重构输出，z_t 为潜在编码。

为了消除异常分数中由短期波动引起的显著变化，采用指数移动平均(EMA)对原始分数序列进行平

滑处理，从而有效减少噪声的干扰。平滑后的分数 $\tilde{s}_t$ 定义为：

$$\tilde{s}_t = \alpha_{EMA} \cdot s_t + (1-\alpha) \cdot \tilde{s}_{t-1} \tag{28}$$

其中 $\alpha_{EMA} \in (0,1)$ 为平滑因子，用于控制当前值和历史值的权重。

为提升模型对噪声与极端值的鲁棒性，本文提出一种基于滑动窗口的分位数自适应阈值机制。该机制设定时间窗口长度 W，基于平滑后的分数计算窗口内分数的分位数作为动态基准，进而构建自适应阈值，阈值的公式如下：

$$\tau_t = Q_p\left(\{\tilde{s}_{t-w+1}, \tilde{s}_{t-w+2}, \cdots, \tilde{s}_t\}\right) \tag{29}$$

其中，Q_p 表示 p^- 分位数，文中设定 $p = 0.95$。当分数分布偏移时，τ_t 自动调整以避免固定阈值导致的误报或漏报，若平滑分数超过阈值($\tilde{s}_t > \tau_t$)，则判定当前时间步为异常，否则为正常。通过这种方法，模型能够动态适应电力系统的变化，提高异常检测的准确性和鲁棒性。改进的 USAD 模型检测过程如图 3 所示。

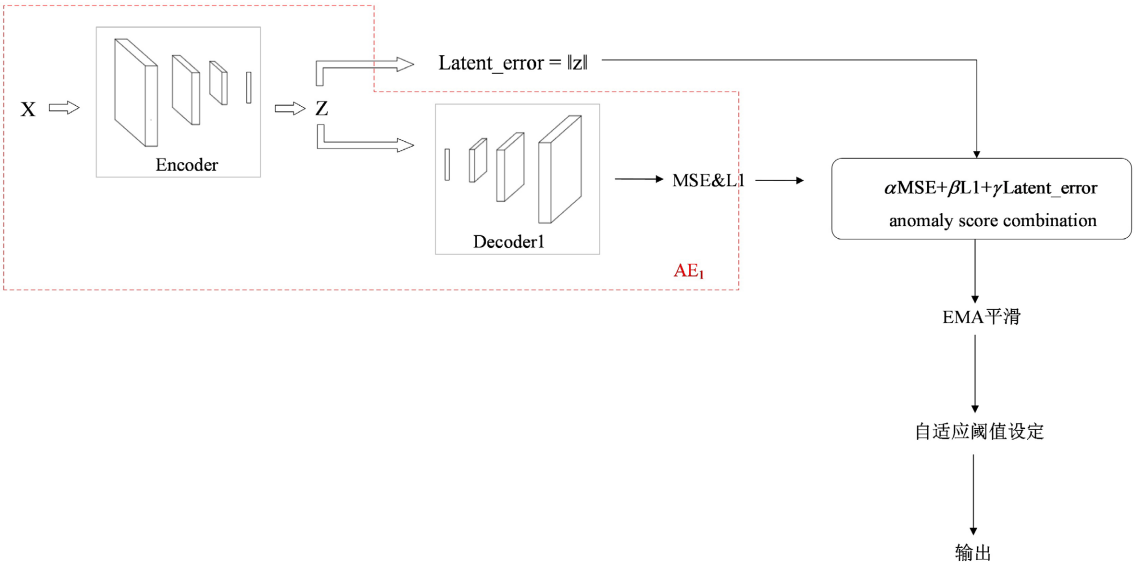

Figure 3. Improved USAD model detection process
图 3. 改进的 USAD 模型检测过程

3. 实验

3.1. 实验设置

实验环境如表 1 所示。

Table 1. Experimental environment
表 1. 实验环境

	环境设置
操作系统	Ubuntu 20.04
RAM	184 GB
GPU	NVIDIA A100-PCIE-40 GB
CPU	Intel (R) Xeon (R) Gold 6248R CPU @ 3.00 GHz
软件	Python 3.10、PyTorch 2.7.1

3.2. 评估指标

在电力数据异常检测场景下，由于异常样本在数据集中占比较低，模型容易产生误报。误报不仅会引发频繁的错误警告，还会导致不必要的人工巡检，从而增加运维成本。因此，针对这一问题，本文选用以下几个评价指标：AUC、Precision (精确率)、Recall (召回率)、F1 值，指标的计算公式如下：

$$\text{Precision} = \frac{TP}{TP + FP} \tag{30}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{31}$$

$$F_1 = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{32}$$

式中，TP 为实际异常用电用户被检测为异常用户数量，FP 为正常用户被误识别为异常用户数量，TN 为实际正常用户被检测为正常用户数量，FN 为异常用户被误识别为正常用户数量。

3.3. 数据预处理

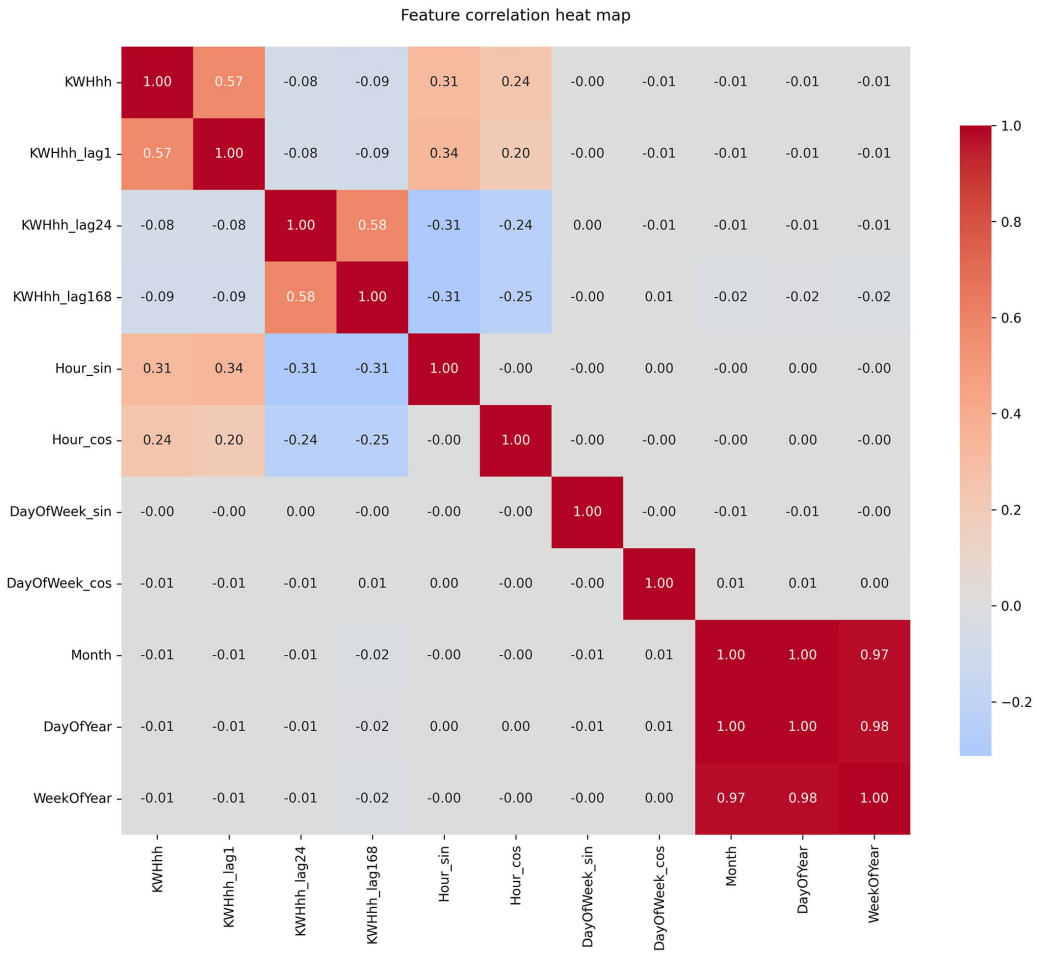

Figure 4. Feature heatmap
图 4. 特征热力图

为了验证特征间的相关性，本文绘制了热力图对其进行分析，如图 4 可见，$KWHhh\_lag1$ 与当前用

电量 *KWh* 的相关系数为 0.57，这表明前一小时用电量对当前预测具有显著影响。同时， *KWHhh _ lag*24 与 *KWHhh _ lag*168 之间存在强相关性($r=0.58$)，反映了日周期与周周期的共同协作，进一步证实了用电行为的强周期性特征，这为后续模型构建提供理论依据。 *Month* 与 *DayOfYear* 存在完全共线性($r=1.0$)，而 *DayOfYear* 与 *WeekOfYear* 高度相关($r=0.98$)，因此，需剔除冗余特征 *Month* 或 *DayOfYear* 。*Hour* _ sin ($r=0.31$)和 *Hour* _ cos ($r=0.24$)与当前用电量呈中等相关，保留其为关键特征，而特征 *DayOfWeek* _ sin 和 *DayOfWeek* _ cos 与当前用电量的相关性极低($|r|<0.01$)，这反映出该用户的用电行为无显著的周模式差异，即周一至周日的用电模式高度相似，导致特征失效，因此需剔除。

3.4. 合成异常样本生成

在训练阶段，通过对比合成异常样本与正常样本的可视化结果，可以清晰地看出本文添加的五种不同类型的异常模式，以此验证了本文提出的合成异常样本方法的有效性。如图 5 可见，在时间步 10~12 小时区间，红色曲线出现显著尖峰，波动超出正常范围 3 倍以上，这种瞬时突变的形态可能模拟了真实场景中空调压缩机突然故障或大功率电器误启动导致的瞬时过载；在时间步 0~2 小时区间，红色曲线呈现持续单向趋势偏离的走势，趋势不仅与蓝色曲线相反，其波动幅度也明显增大，这种模式可能模拟了真实场景中设备的持续性故障；在时间步 15~17 小时区，红色曲线的均值整体上抬并持续偏高，明显高于蓝色曲线，显示出持续的上移偏移状态，这种偏移可能模拟了电表校准错误引发的测量偏差或电压波动导致所有电器功耗同步增加等情况；在时间步 0~4 小时区间内，红色曲线的标准差区间显著宽于正常样本，波动幅度显著增大，此时红线峰值达到 4，高谷差极大，这种模式可能模拟了线路接触不良引发的随机功耗抖动或恶劣天气导致的间歇性负载波动；在时间步 15 小时，红色曲线意外出现低谷，破坏正常的日周期模式，这种断裂可能模拟了用户的行为突变或定时设备的程序错乱。

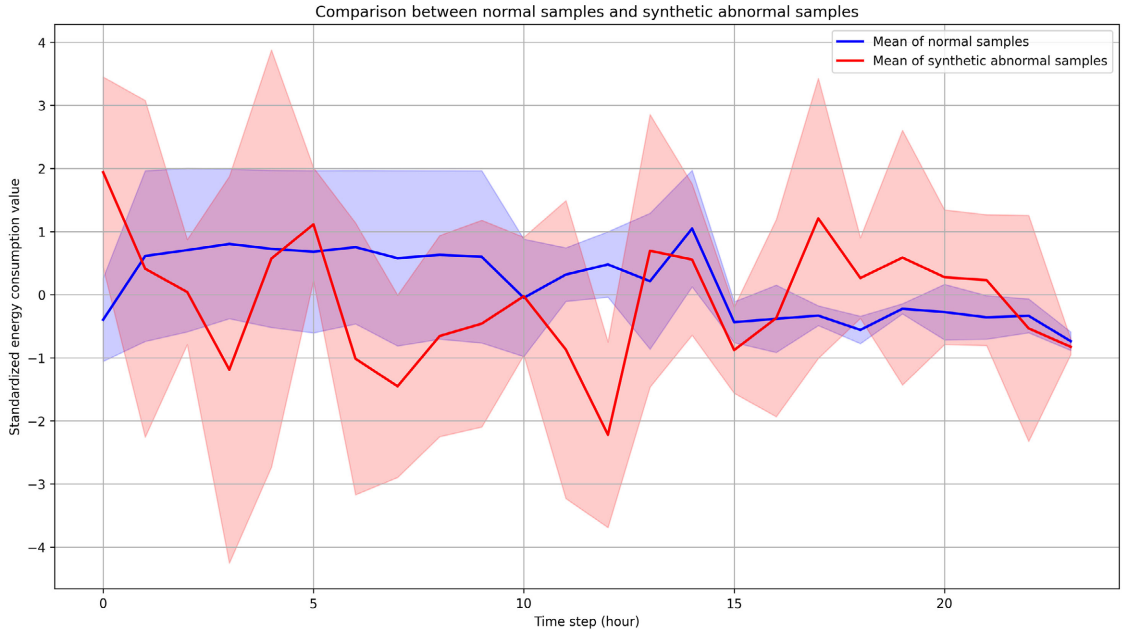

Figure 5. Mean comparison chart of normal samples and synthetic abnormal samples
图 5. 正常样本与合成异常样本的均值对比图

3.5. 自适应阈值设定

在检测阶段，如图 6 所示，模型对正常样本的评分集中于 0.006~0.010 区间，该区间内极少出现极端

值，评分越高，样本被判为异常的可能性越大。异常样本的评分则主要分布于 0.009~0.020 区间，且样本密度随评分升高呈递减趋势，但仍显著高于正常样本的密度分布。基于此，一方面，为了确保评分超过此阈值的样本几乎全部为异常样本，保证模型对显著异常样本的高召回率，另一方面，正常样本的评分极少超过该阈值，从而将误报率控制在极低水平，本文将判别阈值设定为 0.0155。

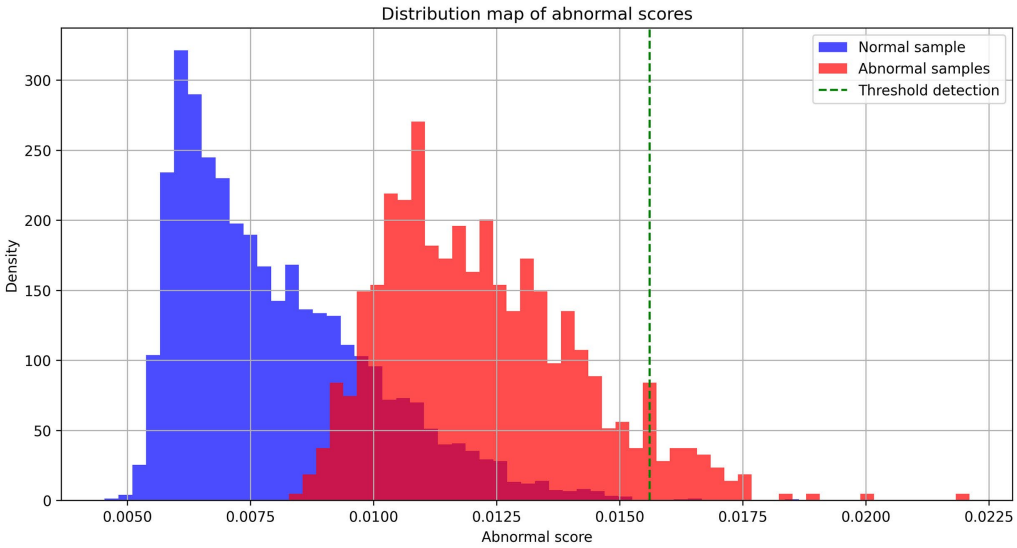

Figure 6. Anomaly score distribution
图 6. 异常分数分布图

3.6. 实验结果

本文针对某用户 738 天内采样间隔为 30 min 的连续用电记录进行无监督用电数据异常检测，为充分捕捉时序数据的长期依赖性，同时兼顾模型轻量化、降低计算资源消耗与过拟合风险，本文选用 USAD 模型进行改进训练。该模型通过双自编码器的对抗训练，克服了单一重构网络易陷入局部最优的缺陷，在保持数十万参数规模的小型网络下，显著提升了对尖峰、漂移等多种异常形态的判别能力和鲁棒性。在已定阈值下，为探究本文模型的检测效果，将本文模型与 AE、VAE、LSTM_AE、LSTM_VAE、IF、One-Class-SVM、PCA 和 Transformer 进行对比实验。实验结果如图 7 和表 2 所示。

Table 2. Model performance metrics
表 2. 模型性能指标

模型	Precision	Recall	F1
AE	0.9038	0.3213	0.4740
VAE	0.9354	0.4074	0.5677
LTSM_AE	0.9292	0.2255	0.3628
LTSM_VAE	0.9306	0.3953	0.5367
IF	0.8925	0.3079	0.4579
One-Class SVM	0.9185	0.3564	0.5135
PCA	0.6931	0.3146	0.4327
Transformer	0.9364	0.4219	0.5853
Ours	0.9814	0.4334	0.6013

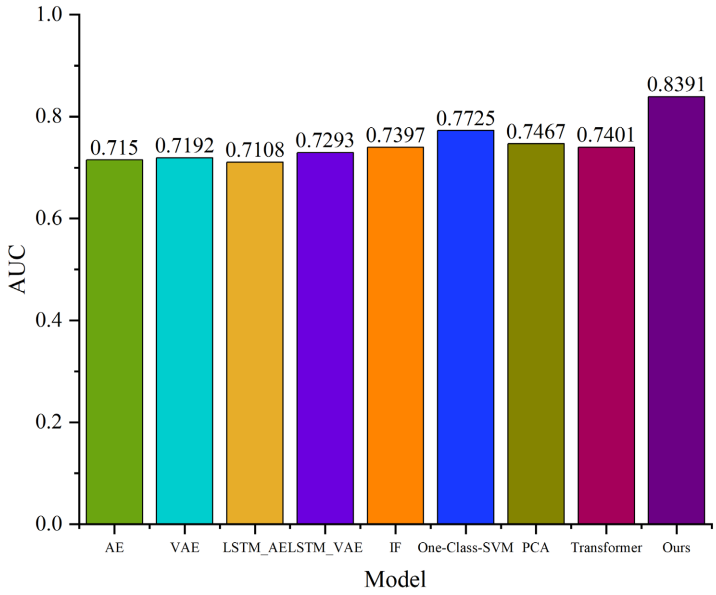

Figure 7. Comparison of model AUC performance
图 7. 模型 AUC 性能对比

结果发现，本文模型的 AUC 值较 AE、VAE、LSTM_AE、LTSM_VAE、IF、One-Class-SVM、PCA 和 Transformer 分别提升了 12.41%、11.99%、12.83%、10.98%、9.94%、6.66%、9.24%、9.9%，Precision 值分别提升了 7.76%、4.6%、5.22%、5.08%、8.89%、6.29%、28.83%、4.5%，Recall 值分别提升了 11.21%、2.6%、20.79%、3.81%、12.55%、7.7%、11.88%、1.15%，F1 值分别提升了 12.73%、3.36%、23.85%、6.46%、14.34%、8.78%、16.86%、1.6%。由此验证，本文模型在电力数据异常检测方面具有一定的优势。

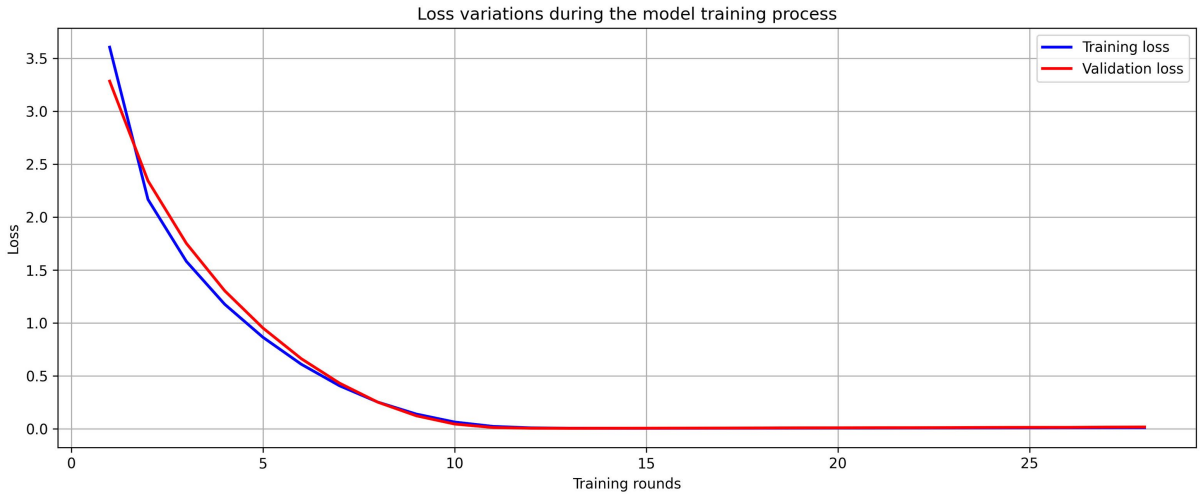

Figure 8. Loss plot
图 8. 损失曲线

图 8 展示了本文所提出的模型的损失变化曲线，从图中可以看出，模型的训练损失与验证损失在初始阶段迅速下降，在训练了 10 轮之后，下降速度明显减缓，并逐步趋于平稳，且两条曲线彼此靠拢，这表明模型已在训练集上充分收敛，同时在未见过的数据上的表现相对一致，体现出良好的泛化能力，能够有效地识别电力数据中的异常模式。

4. 消融实验

为验证本文模型的有效性，本文设计了消融实验，以此评估不同模块的性能差异，消融实验如下所示：

1) 消融 1：仅使用重建误差，无异常分数组合、EMA 平滑与自适应阈值设定。
2) 消融 2：使用异常分数组合，无 EMA 平滑与自适应阈值设定。
3) 消融 3：使用异常分数组合与自适应阈值设定，无 EMA 平滑。
4) 消融 4：使用重建误差、EMA 平滑与自适应阈值设定，无异常分数组合。
5) 本文模型：EMA 平滑下的异常分数组合与自适应阈值设定。

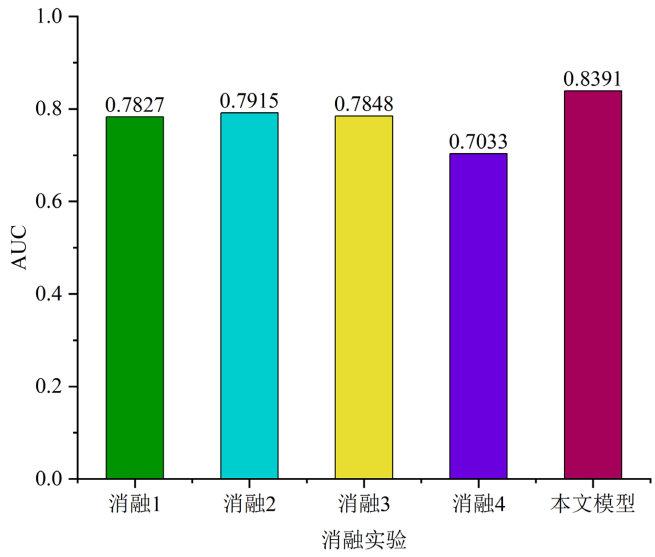

Figure 9. Comparison of model AUC performance
图 9. 模型 AUC 性能对比

Table 3. Model performance metrics
表 3. 模型性能指标

实验	Precision	Recall	F1
消融 1	0.9528	0.4124	0.5757
消融 2	0.9663	0.3408	0.5039
消融 3	0.9621	0.3621	0.5915
消融 4	0.8438	0.2581	0.3563
本文模型	0.9814	0.4334	0.6013

根据图 9 和表 3 的结果所示，本文模型的 AUC 值较消融 1、消融 2、消融 3 和消融 4 分别提升了 5.64%、4.76%、5.43%、13.58%，Precision 值分别提升了 2.86%、1.51%、1.93%、13.76%，Recall 值分别提升了 2.1%、9.26%、7.13%、17.53%，F1 值分别提升了 2.56%、9.74%、9.8%、24.5%，进一步验证了 EMA 平滑下的异常分数组合与自适应阈值设定机制在提高异常检测性能方面的效果。

5. 泛化实验

本节使用 LCL 中的其他十位用户的数据集进行泛化对比实验。实验结果如图 10 和表 4 所示，表中

每个用户代表一个独立的测试集。

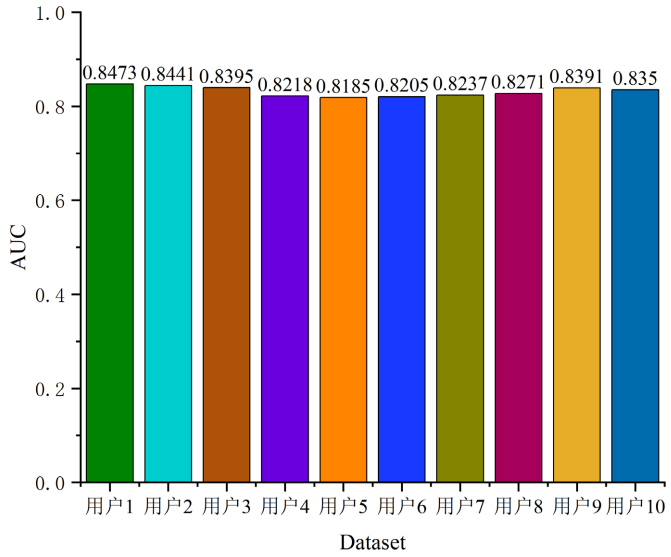

Figure 10. Comparison of AUC performance of different users
图 10. 不同用户的 AUC 性能对比

Table 4. Performance metrics for different users
表 4. 不同用户的性能指标

不同用户	Precision	Recall	F1
用户 1	0.9783	0.4436	0.6102
用户 2	0.9962	0.4542	0.6238
用户 3	0.9968	0.4547	0.6246
用户 4	0.9861	0.4157	0.5850
用户 5	0.9867	0.4087	0.5778
用户 6	0.9817	0.4136	0.5822
用户 7	0.9954	0.4196	0.5902
用户 8	0.9951	0.4267	0.5970
用户 9	0.9938	0.4431	0.6128
用户 10	0.9952	0.4339	0.6046

结果发现，本文模型在所有用户中都表现出高精确率，这说明当模型将某个样本判定为异常时，其误报率极低，能够有效地避免对正常用电波动的误判。同时，模型在召回率和 F1 值上波动较小，这表明模型在所有用户中的性能表现稳定，具有较好的鲁棒性。因此，本文模型在保证高精确度的前提下，兼具良好的综合性能，为电力系统的实时监控与故障预警提供了有力支持。

6. 结论

为有效应对用电数据异常检测中无标签样本的挑战及捕捉用电数据时间序列的短期波动与长期依赖，本研究在 USAD 模型上进行了优化。在 USAD 模型训练的基础上，首先添加提取多维时序特征包括小时、星期、月份、年内日序和年内周序，结合正余弦函数进行周期性编码，并构造关键滞后特征 lag1、

lag24、lag168；其次，设计五种类型的合成异常样本包括尖峰、趋势、模式断裂、水平偏移、方差变点注入用电序列数据。在 UASD 模型检测的基础上引入指数移动平均(EMA)平滑组合异常分数，并基于此设定自适应阈值判定最终异常数据。

通过在 LCL 中的某用户数据集上进行实验分析，结果显示该方法 AUC 达到了 83.91%，准确率达到 98.14%，Recall 值为 43.34%，F1 值为 60.13%，并将该方法与多种异常检测方法进行对比，验证了本文方法的有效性和优越性。最后，将本文方法用于 LCL 中的其他十位用户的数据集进行泛化对比实验，结果表明该方法实现了均衡的性能表现，为电网的在线监测与故障预警提供了坚实可靠的技术保障。

未来可进一步探索在少量可用标签下，如何有效引入这些监督信息来提升模型的性能。

基金项目

上海市哲学社会科学规划资助项目(2024BGL001)。

参考文献

[1] Hu, T., Guo, Q., Shen, X., Sun, H., Wu, R. and Xi, H. (2019) Utilizing Unlabeled Data to Detect Electricity Fraud in AMI: A Semi-Supervised Deep Learning Approach. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 3287-3299. https://doi.org/10.1109/tnnls.2018.2890663

[2] 刘钊瑞, 高云鹏, 郭建波, 等. 基于深度自编码器高斯混合模型的窃电行为检测[J]. 电力系统保护与控制, 2022, 50(18): 83-91.

[3] 庄池杰, 张斌, 胡军, 等. 基于无监督学习的电力用户异常用电模式检测[J]. 中国电机工程学报, 2016, 36(2): 379-387.

[4] Jesmeen, M.Z.H., Murthy, G.R., Hossen, J., *et al*. (2021) Detecting Abnormal Electricity Usage Using Unsupervised Learning Model in Unlabeled Data. *International Journal of Advanced and Applied Sciences*, **8**, 102-111. https://doi.org/10.21833/ijaas.2021.09.014

[5] Kaleta, J., Dubiński, J., Wojdan, K., *et al*. (2021) Detection of Anomalous Consumers Based on Smart Meter Data. *Journal of Power Technologies*, **101**, 202-212.

[6] 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.

[7] 傅世元, 高欣, 张浩, 等. 基于元学习动态选择集成的电力调度数据异常检测方法[J]. 电网技术, 2022, 46(8): 3248-3261.

[8] 华回春, 邓彬, 刘哲, 等. 基于变分自编码高斯混合模型的发电企业串谋智能预警[J]. 电力系统自动化, 2022, 46(4): 188-196.

[9] Lei, L., Wu, B., Fang, X., Chen, L., Wu, H. and Liu, W. (2023) A Dynamic Anomaly Detection Method of Building Energy Consumption Based on Data Mining Technology. *Energy*, **263**, Article ID: 125575. https://doi.org/10.1016/j.energy.2022.125575

[10] Almalki, A.J. (2024) Unsupervised Learning with Hybrid Models for Detecting Electricity Theft in Smart Grids. *IEEE Access*, **12**, 187027-187040. https://doi.org/10.1109/access.2024.3498733

[11] Mishra, A.K. and Das, B. (2025) A Novel Density Based Clustering Approach for Electricity Theft Detection. *IEEE Transactions on Industry Applications*, **61**, 5537-5548. https://doi.org/10.1109/tia.2025.3544167

[12] Vaswani, A., Shazeer, N., Parmar, N., *et al*. (2017) Attention Is All You Need. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.