

DOI: 10.19783/j.cnki.pspc.201534

# 基于概率预测的用电采集终端电量异常在线实时识别方法

刘宣<sup>1</sup>, 唐悦<sup>1</sup>, 卢继哲<sup>1</sup>, 阿辽沙·叶<sup>1</sup>, 侯帅<sup>1</sup>, 叶方彬<sup>2</sup>

(1. 中国电力科学研究院有限公司, 北京 100192; 2. 国网浙江省电力有限公司营销服务中心, 浙江 杭州 311121)

**摘要:** 电力市场环境下用电信息采集系统采集的用电量成为市场结算的重要依据。实时识别用电采集终端上送的异常电量, 不但可以提升数据质量, 也可以为发现采集终端的故障、识别异常用电行为提供参考。针对现有异常数据识别方法识别性能和实时性不高的问题, 提出基于概率预测的电量异常在线实时识别方法。首先, 在分析电量异常类型和特点的基础上, 提出离线训练概率预测模型、在线实时识别异常数据的检测方法。其次, 提出了基于状态空间模型的结构化用电量模型对用户用电规律进行建模, 并采用变分贝叶斯推断训练模型, 以实现用电量的概率分布预测。最后, 利用预测标准分数衡量电量实测数据与电量概率预测结果之间的差异, 从而实现异常数据的在线识别。采用实际电量数据进行验证, 并与其他方法进行对比, 验证了该方法的实用性和有效性。

**关键词:** 用电信息采集终端; 异常识别; 概率预测; 结构化电量模型; 变分贝叶斯推断

## Online real time anomaly recognition method for power consumption of electric energy data acquisition terminal based on probability prediction

LIU Xuan<sup>1</sup>, TANG Yue<sup>1</sup>, LU Jizhe<sup>1</sup>, YE Aliaosha<sup>1</sup>, HOU Shuai<sup>1</sup>, YE Fangbin<sup>2</sup>

(1. China Electric Power Research Institute, Beijing 100192, China; 2. Marketing Service Center of State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 311121, China)

**Abstract:** The electricity consumption collected by a power user electric energy data acquisition system in the electricity market environment becomes an important basis for market settlement. Real-time identification of abnormal power data can not only improve the quality of power data, but also provide a reference for detection of faults in the collection terminal and identification of abnormal power consumption behavior. In order to solve the problems of existing anomaly recognition methods, a method based on probability prediction is proposed. First, on the basis of analyzing the types and characteristics of electric power anomalies, a detection method for offline training probability prediction models and online identification of abnormal data is proposed. Then, a structured power consumption model based on the state space model is proposed to model the user power consumption rules. The Variational Bayesian Inference is used to train the model in order to realize the probability prediction of power consumption. Finally, it uses the prediction standard score to measure the difference between the measured electricity data and the electricity probability prediction result, so as to realize the online identification of abnormal data. Real electricity data are used for verification and comparison with other methods to verify the practicability and effectiveness of this method.

This work is supported by the Science and Technology Project of State Grid Corporation of China (No. 1100-201919158A-0-0-00).

**Key words:** electric energy data acquire terminal; abnormal recognition; probabilistic prediction; structured electricity model; variational Bayesian inference

## 0 引言

用电采集终端是智能电网的重要基础设施<sup>[1]</sup>。随

着越来越多的负荷侧资源参与到电力系统实时调度运行中<sup>[2-4]</sup>, 该终端采集的负荷侧电力用户电量将成为电力市场结算的主要依据, 其数据的一致性、完整性直接影响到电力市场的公平性。

用电采集终端在实际运行过程中经常存在异常电量数据<sup>[5]</sup>。实时识别异常电量数据, 可以提高数

基金项目: 国家电网公司科技项目资助(1100-201919158A-0-0-00)

据质量, 评估采集终端的健康状态<sup>[6]</sup>、辨识用户的异常用电行为<sup>[7-8]</sup>。早期通常采用人工设定规则的方式识别异常电量数据, 需要消耗大量的人力和物力资源。为了解决上述问题, 国内外学者针对数据异常检测提出了许多解决方案<sup>[9]</sup>。其中, 基于机器学习技术的异常数据检测方法应用较多, 并且取得较好的效果, 主要包括基于有监督学习的分类方法和基于无监督学习的聚类方法。

有监督学习的分类方法主要有神经网络<sup>[10]</sup>、随机森林<sup>[11]</sup>和孤立森林<sup>[12]</sup>等异常数据识别方法。有监督异常数据识别方法基于已标注异常的数据训练分类模型, 需要人工事先完成异常电量数据的标注, 并且异常数据的样本较少, 存在样本不平衡问题, 成为制约此类方法性能的主要瓶颈。为了解决该问题, 许多无监督机器学习方法被应用于异常数据识别领域, 主要包括基于聚类<sup>[13]</sup>或 DBSCAN 聚类的方法<sup>[14]</sup>、电量时序建模与聚类相结合的方法<sup>[15-16]</sup>等。目前, 越来越多的基于深度学习的无监督学习方法也被应用于电力数据异常识别<sup>[17-19]</sup>。但是, 将该类方法应用在用电采集终端异常数据识别中存在两方面的问题。

(1) 用户用电量受人们生产生活规律的影响, 有其内在的时序变化规律, 其异常数据的识别需要考虑用电量的时序变化规律, 从而识别不同类型的异常电量。

(2) 基于机器学习的方法需要利用大量的历史数据训练识别模型, 训练时间较长, 无法满足实时识别异常电量的要求。

为了解决以上问题, 本文分析了采集终端中异常电量的类型和特点, 提出了采用离线训练、在线识别的异常电量识别方法。离线训练阶段利用结构化电量时序模型和变分贝叶斯推断 (Variational Bayes Inference, VBI) 建立能够反映用电量变化规律的概率预测模型; 在线识别阶段利用模型预测未来电量概率分布, 并与实测电量进行对比实现电量异常数据的在线异常识别。采用实际的用电量数据验证了该方法的识别性能。

### 1 终端电量异常数据的类型和特点

图 1 是采集终端采集的某市一个企业园区 2019 年 6 月 13 日至 20 日的用电量数据。从图 1 可以看出, 由于不同的原因导致其存在 3 种典型的异常电量数据。

(1) 点异常又称为全局异常, 如果采集的电量值远不同于正常的电量值, 该电量值为点异常值。该类异常通常是由采集装置故障、通信通道中断等原

因导致。

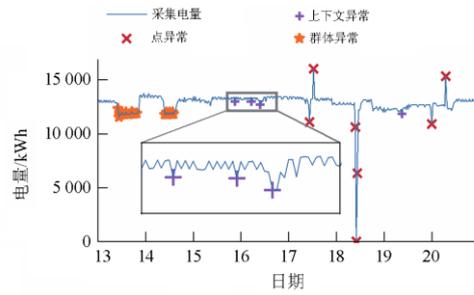


图 1 3 种电量异常数据类型

Fig. 1 Three types of abnormal electrical energy data

(2) 上下文异常又称为条件异常, 如果采集的电量值与其上下文环境的其他电量值有明显的不同, 则该电量值称为上下文异常值。在时间序列数据中, 其上下文表示包含该异常值的某一时间窗口。

(3) 群体异常也可称为集体异常, 当某一个时间窗口内的电能量数据集与整个数据集不一致时, 该时间窗口内的数据点称为群体异常数据点。群体异常中的每个数据点既不是点异常也不是上下文异常。

图 2 为采用 Z-score 法<sup>[20]</sup>、孤立森林<sup>[12]</sup>(Isolation Forest, IF)、基于密度的空间聚类<sup>[14]</sup>(Density-Based Spatial Clustering of Applications with Noise, DBSCAN)、单类别支持向量机<sup>[21]</sup>(One-class SVM) 4 种算法, 对该企业园区 2019 年 6 月 13 日至 20 日的用电量数据进行异常识别的结果。

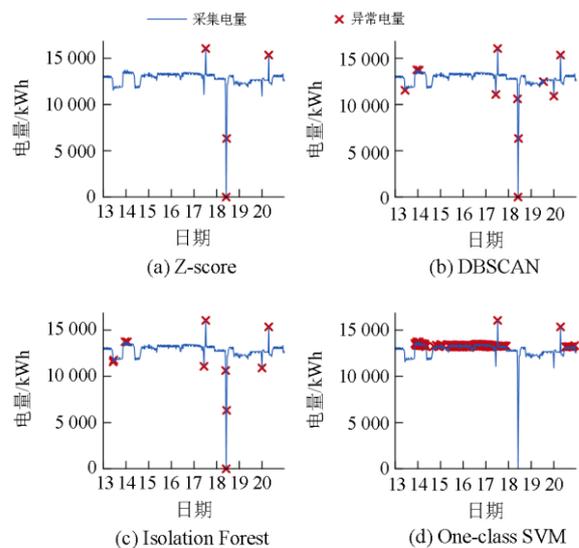


图 2 不同异常识别方法的电量异常识别结果

Fig. 2 Electric energy anomaly identification results of different anomaly identification methods

由图 2(a)一(c)可以看出, 针对点异常 3 种算法都能够进行有效的识别, 但是无法识别出上下文异

常和群体异常。图 2(d)虽然能够识别上下文异常和群体异常,但是误检率较高,即很多正常的电量数据被误判为异常数据。

## 2 基于概率预测的电量异常识别

为了能够在用电终端中实时识别电量数据的上下文异常和群体异常,提升异常识别的效率和准确率,将电量异常识别过程分为离线训练和在线识别 2 个阶段,如图 3 所示。

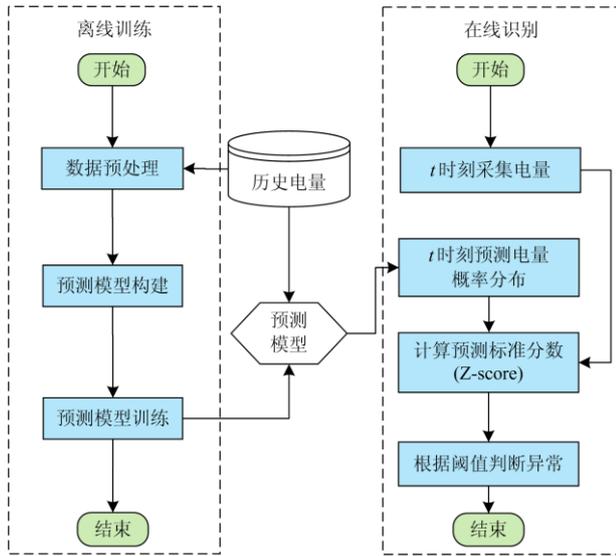


图 3 基于概率预测的电量异常数据识别流程

Fig. 3 Electric energy abnormal data recognition process based on probability prediction

离线训练阶段,基于历史电量数据训练电量概率预测模型,该阶段主要包括 3 步:

(1) 数据预处理。采用 Z-score 法<sup>[20]</sup>对历史电量数据进行点异常初步判断,利用异常数据点前后时刻的电量数据的平均值进行初步校正,形成质量较高的训练数据集。

(2) 预测模型构建。综合考虑用电量的趋势成分、季节成分及外部影响因素对用户用电行为的影响,利用结构化时间序列模型构建电量概率预测模型。

(3) 预测模型训练。基于预处理后历史电量数据,采用 VBI 算法训练电量概率预测模型。

在线识别阶段,将预测模型线上部署,实时预测未来时刻电量数据的概率分布,并与实际电量进行对比实现异常电量的实时在线识别,主要包括 2 步:

(1) 基于  $t$  时刻之前的实际电量,利用部署的电量概率预测模型对  $t$  时刻电量的概率分布进行预测。

(2) 将  $t$  时刻采集的电量与预测电量概率分布进行对比,采用预测标准分数(Z-score)实时计算实际采集电量偏离预测电量概率分布均值的程度,当偏差过大时,判断为异常电量。预测标准分数定义为式(1)。

$$\hat{Z}_t^{\text{score}} = \frac{|y_t - E(\hat{y}_t)|}{\hat{s}} \quad (1)$$

式中:  $y_t$  为  $t$  时刻采集的实际电量;  $E(\hat{y}_t)$  表示  $t$  时刻预测电量概率分布的期望,即均值;  $\hat{s}$  为方差。

预测标准分数定义了以预测电量的标准差为单位度量的实际电量偏离预测电量均值的程度,当其大于某一设定阈值  $\lambda$  时,可以判定该电量为异常电量。从另一个角度讲,根据阈值  $\lambda$  生成电量预测区间,区间上下限的定义如式(2)所示。

$$\begin{cases} L = y_t - \lambda E(\hat{y}_t) \\ U = y_t + \lambda E(\hat{y}_t) \end{cases} \quad (2)$$

式中:  $L$  为区间的下限;  $U$  为区间上限。当实际电量的预测标准分数  $\hat{Z}_t^{\text{score}}$  大于阈值  $\lambda$  时,说明实际电量在预测区间之外,可以判定为异常电量。

## 3 电量概率预测模型的构建和训练

准确预测未来电量的概率分布是提升电量异常识别准确率的关键。用户的用电量是典型的时间序列数据,受人们生产生活规律的影响,具备趋势性、季节性等时间序列的基本特征。采用状态空间模型对用电量的不同成分进行建模,实现电量概率分布预测。

### 3.1 用电量状态空间模型

用电量是人们生产活动在用电行为上的体现。基于状态空间模型的基本思想,可以将用电行为抽象为状态随时间不断变化的系统,用户用电量就是系统状态的外在体现。用电量是可观测的时间序列,系统状态是不可观测的。状态空间模型的一般表达式见式(3)<sup>[22]</sup>。

$$\begin{cases} y_t = \mathbf{Z}_t^T \boldsymbol{\alpha}_t + \varepsilon_t \\ \boldsymbol{\alpha}_{t+1} = \mathbf{T}_t \boldsymbol{\alpha}_t + \mathbf{R}_t \boldsymbol{\eta}_t \\ \varepsilon_t \sim N(0, \sigma_t^2), \boldsymbol{\eta}_t \sim N(0, \mathbf{Q}_t) \end{cases} \quad (3)$$

式中:  $y_t$  为  $t$  时刻的用户采集电量;  $\boldsymbol{\alpha}_t$  为  $m$  维的系统状态向量;  $\mathbf{Z}_t$  为  $m$  维的输出向量,表示系统状态与用电量之间的转换;  $\mathbf{T}_t$  为  $m \times m$  的状态转移矩阵;  $\mathbf{R}_t$  为  $m \times q$  的控制矩阵;  $\varepsilon_t$  为观察误差;  $\boldsymbol{\eta}_t$  为  $q$  维的系统状态误差;  $\mathbf{Q}_t$  为  $q \times q$  的状态扩散矩阵。其中  $q > m$ 。式(3)中,第一部分为量测方程,描述观

测值用电量和系统的状态之间的变化关系, 第二部分为状态方程描述动态系统的状态从前一时刻到当前时刻的演化规律。

受人们生产生活规律的影响, 其用电量通常由趋势成分、季节成分构成, 并受外部影响因素的影响。基于状态空间模型的基本思想可以构建不同用电量成分的变化模型。

### 3.2 用电量趋势模型

在某一个时间范围内, 电力用户用电量的变化呈现趋势性, 趋势模型定义为式(4)。

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ \mu_{t+1} = \mu_t + \delta_t + \eta_{u,t} \\ \delta_{t+1} = \delta_t + \eta_{\delta,t} \\ \eta_{u,t} \sim N(0, \sigma_{u,t}^2), \eta_{\delta,t} \sim N(0, \sigma_{\delta,t}^2) \end{cases} \quad (4)$$

式中:  $\mu_t$  为用电量  $t$  时刻的趋势(level);  $\delta_t$  为用电量在  $t$  和  $t+1$  时刻之间趋势的变化量, 即  $t$  时刻的斜率(slope);  $\eta_{u,t}$ 、 $\eta_{\delta,t}$  分别为趋势和变化量的误差, 分别服从标准差为  $\sigma_{u,t}$  和  $\sigma_{\delta,t}$  的标准正态分布。用电量趋势模型假设用电量的趋势  $\mu_t$  和变化  $\delta_t$  都是随机的。趋势模型可用状态空间模型的一般形式表示, 如式(5)所示。

$$\begin{cases} y_t = [1 \ 0] \begin{bmatrix} \mu_t \\ \delta_t \end{bmatrix} + \varepsilon_t \\ \begin{bmatrix} \mu_{t+1} \\ \delta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \delta_t \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{u,t} \\ \eta_{\delta,t} \end{bmatrix} \end{cases} \quad (5)$$

此时, 式(3)中定义的状态空间模型中各变量如式(6)所示。

$$\begin{cases} \mathbf{Z}_t^T = [1 \ 0]^T \\ \boldsymbol{\alpha}_t = [\mu_t \ \delta_t]^T \\ \mathbf{T}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\ \mathbf{R}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \boldsymbol{\eta}_t = [\eta_{u,t} \ \eta_{\delta,t}]^T \end{cases} \quad (6)$$

### 3.3 用电量季节模型

受生活规律的影响, 电力用户用电量具备以日、周、月度为变化周期的季节性, 季节模型可以反映用电量周期性的变化规律, 定义为式(7)。

$$\begin{cases} y_t = \gamma_t + \varepsilon_t \\ \gamma_{t+d} = \sum_{i=0}^{s-2} \gamma_{t-ixd} + \eta_{\gamma,t} \end{cases} \quad (7)$$

式中:  $\gamma_t$  为季节因素对用电量的影响程度;  $s$  为季节的数量;  $d$  为每个季节的持续时间。该模型可以当作用电量在  $s$  个虚拟变量上的回归。例如, 在以小时为时间粒度的用电量数据中, 以天为单位的季节模型中  $s=24$ 、 $d=1$ ; 以周为单位的季节模型中  $s=7$ 、 $d=24$ 。

季节模型采用状态空间模型的一般形式表示, 如式(8)。

$$\begin{cases} y_t = [1 \ 0 \ \dots \ 0] \begin{bmatrix} \gamma_t \\ \gamma_{t-d} \\ \vdots \\ \gamma_{t-(s-2)d} \end{bmatrix} + \varepsilon_t \\ \begin{bmatrix} \gamma_{t+d} \\ \gamma_t \\ \gamma_{t-d} \\ \vdots \\ \gamma_{t-(s-4)d} \\ \gamma_{t-(s-3)d} \end{bmatrix} = \begin{bmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma_t \\ \gamma_{t-d} \\ \gamma_{t-2d} \\ \vdots \\ \gamma_{t-(s-3)d} \\ \gamma_{t-(s-2)d} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \eta_{\gamma,t} \end{cases} \quad (8)$$

### 3.4 影响用电量的回归模型

电力用户的用电量受气温、节假日等外部因素的影响, 基于状态空间模型, 构建外部影响因素与用电量之间的回归模型, 如式(9)。

$$\begin{cases} y_t = \mathbf{x}_t \boldsymbol{\beta}_t = \sum_{j=1}^J x_{j,t} \beta_{j,t} \\ \beta_{j,t+1} = \beta_{j,t} + \eta_{\beta,j,t} \end{cases} \quad (9)$$

式中:  $\mathbf{x}_t$  为影响负荷各类因素在  $t$  时刻的值构成的向量;  $\boldsymbol{\beta}_t$  为各类影响因素在  $t$  时刻的回归系数;  $J$  为影响因素的数量。

此时, 式(3)中定义的状态空间模型中各变量如式(10)所示。

$$\begin{cases} \mathbf{Z}_t^T = \mathbf{x}_t \\ \boldsymbol{\alpha}_t = \boldsymbol{\beta}_t \\ \mathbf{T}_t = \mathbf{R}_t = \mathbf{I}_{J \times J} \end{cases} \quad (10)$$

式中,  $\mathbf{I}_{J \times J}$  为  $J \times J$  的单位矩阵。

### 3.5 结构化用电量模型

将每个成分的状态转化向量  $\mathbf{Z}_t$  进行拼接, 并把每个成分的状态转移矩阵  $\mathbf{T}_t$  和控制矩阵  $\mathbf{R}_t$  分别组合成分块对角矩阵, 形成统一的考虑各类用电量成分的电力用户用电量状态空间模型-结构化用

电量模型。以趋势模型和季节模型为例, 构建完成后, 式(3)中的  $Z_t$ 、 $T_t$  和  $R_t$  如式(11)所示。

$$Z_t = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, T_t = \begin{bmatrix} 1 & 1 & & & & \\ 0 & 1 & & & & \\ & & -1 & -1 & \dots & -1 & -1 \\ & & 1 & 0 & \dots & 0 & 0 \\ & & 0 & 1 & \dots & 0 & 0 \\ & & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 0 & 0 & \dots & 1 & 0 \\ & & 0 & 0 & \dots & 0 & 0 \end{bmatrix}, R_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ & & 1 \\ & & 0 \\ & & \vdots \\ & & 1 \end{bmatrix} \quad (11)$$

### 3.6 结构化用电量模型的训练

如果利用结构化用电量模型进行用电量的概率分布预测, 需要基于历史数据训练模型中每个成分的参数的后验分布。需要训练的参数如表 1 所示。

表 1 结构化用电量模型的需要训练的参数

Table 1 Parameters to be trained for the structured

power load model		
负荷模型	参数	说明
趋势模型	$\mu_t$	趋势水平
趋势模型	$\delta_t$	趋势变化斜率
季节模型	$\gamma_t$	季节因素对负荷影响系数
回归模型	$\beta_t$	外部因素对负荷影响系数

假设  $\theta = (\mu_t, \delta_t, \gamma_t, \beta_t)$ , 模型训练的主要目的是基于历史用电量数据  $y$ , 构建参数  $\theta$  的后验分布的  $p(\theta|y)$ 。训练完成后可以从模型参数的后验分布进行采样, 实现未来电量的概率分布预测。

根据贝叶斯定理, 参数  $\theta$  的后验概率分布计算如式(12)。

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y,\theta)d\theta} \quad (12)$$

式中:  $p(\theta)$  为参数  $\theta$  的先验分布;  $p(y|\theta)$  为似然分布;  $\int p(y,\theta)d\theta$  为证据。由于  $\int p(y,\theta)d\theta$  的积分难以计算, 可以假设一个代理分布  $q(\theta)$ , 利用 VBI 算法来近似逼近参数的后验概率分布。

VBI 算法的主要思想是使用优化方法来最小化代理分布  $q(\theta)$  和后验分布  $p(\theta|y)$  之间以负证据因子下界(Evidence Lower Bound, ELBO)衡量的变分损失<sup>[23]</sup>, 从而近似逼近参数  $\theta$  的后验概率分布。ELBO 定义如式(13)。

$$ELBO(q) = E[\log p(y|\theta)p(\theta)] - E[\log q(\theta)] \quad (13)$$

式中:  $p(\theta)$  为参数的先验分布;  $p(y|\theta)$  为似然分布;  $q(\theta)$  为代理分布。

## 4 电量异常识别最优阈值确定和性能评估

电量异常识别是典型的二分类问题。因此, 采用准确率和召回率评估异常识别的性能, 从而确定最优的算法参数或者判定阈值<sup>[24]</sup>。准确率是判定为指标异常的数据点中正确判定的比例, 如式(14)。

$$P_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \times 100\% \quad (14)$$

召回率是判定为异常的数据点占有异常数据点的比例, 如式(15)。

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \times 100\% \quad (15)$$

式(14)和式(15)中,  $N_{\text{TP}}$ 、 $N_{\text{FP}}$  和  $N_{\text{FN}}$  定义如表 2。

表 2 准确率和召回率的定义

Table 2 Definition of precision and recall

类别	异常电量数据点 (实际)	正常电量数据点 (实际)
异常电量数据点 (预测)	$N_{\text{TP}}$ : 实际为异常的数据点被判定为异常点	$N_{\text{FP}}$ : 实际为正常的数据点被判定为异常点
正常电量数据点 (预测)	$N_{\text{FN}}$ : 实际为异常的数据点被判定为正常点	$N_{\text{TN}}$ : 实际为正常的数据点被判定为正常点

由准确率和召回率定义可知, 当两者都达到 100% 时, 模型的性能将达到最优。然而, 两个指标都受异常识别算法的关键参数设置的影响。以本文方法为例, 两个指标受阈值  $\lambda$  的影响, 准确率随着阈值  $\lambda$  的增加而上升, 而召回率随着阈值  $\lambda$  的增加而下降。随着  $\lambda$  的变化可构成准确率和召回率 2 条曲线, 2 条曲线的交点处, 作为最优阈值  $\lambda$  的位置。

异常判定阈值  $\lambda$  取不同的值时, 得到相应的准确率和召回率。随着  $\lambda$  的变化, “召回率-准确率”构成的曲线可以反映模型的识别性能。该曲线越靠近 (100%, 100%) 的位置 (准确率和召回率都为 100%), 异常识别性能越好<sup>[19]</sup>。

通常也采用 F-score 衡量异常识别模型的整体识别性能<sup>[24]</sup>, 定义如式(16)。

$$F_{\text{score}} = \frac{2 \times P_{\text{precision}} \times R_{\text{recall}}}{P_{\text{precision}} + R_{\text{recall}}} \quad (16)$$

## 5 算例分析

### 5.1 基于实际电量数据的概率预测模型的训练

以某市一个企业园区 2019 年 1 月至 6 月的用电

量数据对本文方法进行验证,数据的时间分辨率为 30 min。其中,1-5 月份数据作为训练数据集,6 月份数据作为测试集,对 6 月份电量数据中存在的异常进行人工标注,共标注了 185 个实际存在异常的电量数据点。实验采用 Python 语言作为编程语言,其他异常算法采用机器学习算法库 Scikit-learn 中的算法,文中结构化用电量模型的构建和训练采用 TensorFlow Probability 软件库实现。硬件配置方面,处理器为 Intel Core i7-6700, RAM 为 8 GB。

在构建预测模型时,利用趋势模型对该企业园区用电量趋势进行建模;利用季节性模型分别对以日和周为周期的用电量季节性进行建模;利用回归模型对气温、是否节假日两个外部影响因素对用电量的影响进行建模。模型构建完成后,采用变分推理算法训练模型参数的后验概率分布。训练过程中 ELBO 损失的变化曲线如图 4 所示。

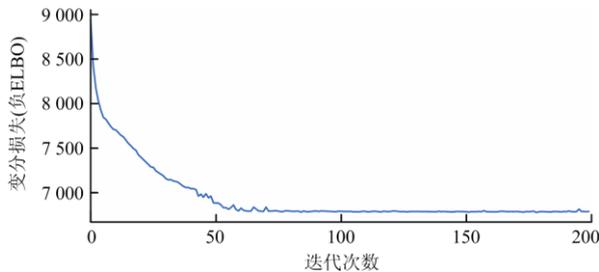


图 4 ELBO 损失的变化曲线  
Fig. 4 Change curve of ELBO loss

由图 4 可知,随着迭代次数的增加,变分损失逐渐下降。迭代次数在 100 次以后,损失不再下降,训练基本收敛。训练完成后,各参数后验分布的均值和标准差如表 3 所示。

表 3 训练完成后参数后验分布的均值和标准差

Table 3 Mean and standard deviation of the posterior distribution of the parameters after the training is complete

负荷模型	参数	均值	标准差
趋势模型	趋势水平( $\mu$ )	0.07	0.55
	趋势斜率( $\delta$ )	0.05	0.03
季节模型	日( $\gamma_{,1}$ )	21.34	1.61
	周( $\gamma_{,2}$ )	18.78	16.12
回归模型	气温( $\beta_{,1}$ )	-1.64	1.32
	是否节假日( $\beta_{,2}$ )	0.14	4.31

### 5.2 最优阈值确定

基于电量概率预测模型和已标注异常的测试数据,利用第 4 节的方法确定最优阈值。“阈值-准确率”和“阈值-召回率”曲线如图 5 所示。

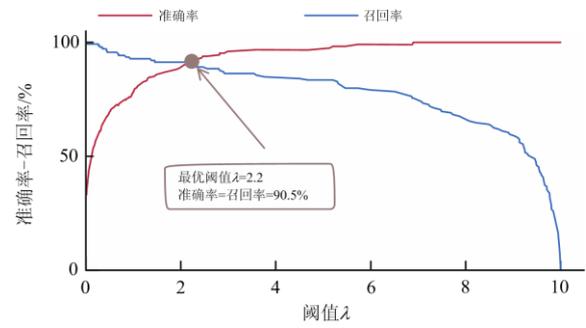


图 5 准确率和召回率随阈值的变化曲线

Fig. 5 Change curves of accuracy rate and recall rate vs. threshold

由图 5 可知,2 条曲线相交的位置为最优阈值的位置,此时  $\lambda = 2.2$ 。

### 5.3 电量概率预测及异常识别

基于训练完成的模型,在模型参数后验分布中进行采样,采样次数设置为 1 000,实现对该企业园区 2019 年 6 月份每 30 min 的电量概率预测。根据概率分布的均值、标准差和最优阈值  $\lambda = 2.2$ ,采用式(2)计算电量预测区间。其中 6 月 1 日电量预测区间和异常识别结果如图 6 所示。

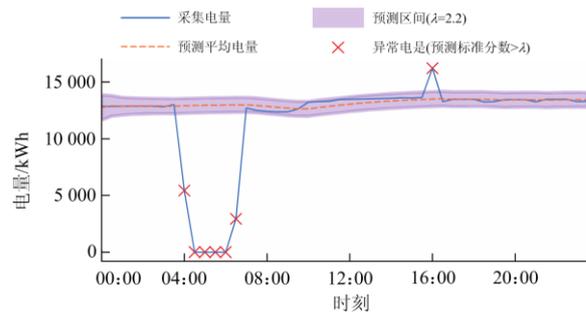


图 6 2019 年 6 月 1 日电量预测区间及异常识别结果  
Fig. 6 Electric energy probability interval prediction and abnormal identification results on June 1, 2019

由图 6 可知,基于概率预测的电量识别方法不仅能够识别点异常,也可以对电量数据中存在的上下文异常和群体异常进行有效的识别。

图 6 中 00:00、05:00、12:00、16:00 的预测电量概率分布如图 7 所示。

由图 7 可知,每个时刻的预测电量概率分布基本呈正态分布。05:00 和 16:00 的实际电量在由  $\lambda = 2.2$  计算的预测电量区间之外,可以判断 2 个时刻电量为异常电量。

### 5.4 电量异常识别性能比较

基于 6 月份已标注异常的电量数据,采用第 4 节所述的“召回率-准确率”曲线和 F-score 评估电量异常识别的性能。将本文方法与 Z-score<sup>[20]</sup>、

Isolation Forest<sup>[12]</sup>、DBSCAN<sup>[14]</sup>、One-class SVM<sup>[21]</sup>4种方法进行对比, 结果如图 8 所示。

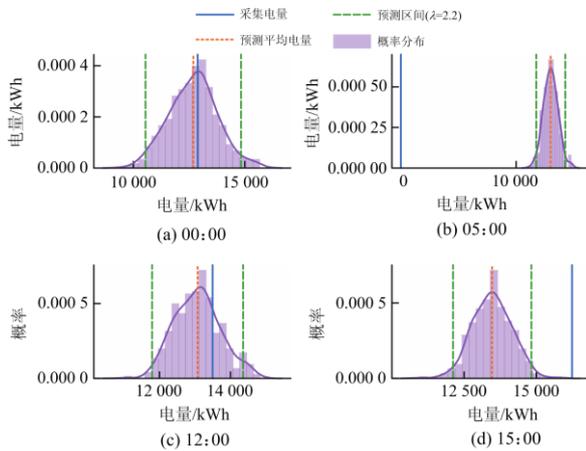


图 7 2019 年 6 月 1 日 4 个时刻预测电量概率分布

Fig. 7 Probability distribution of electric energy predicted at 4 times on June 1, 2019

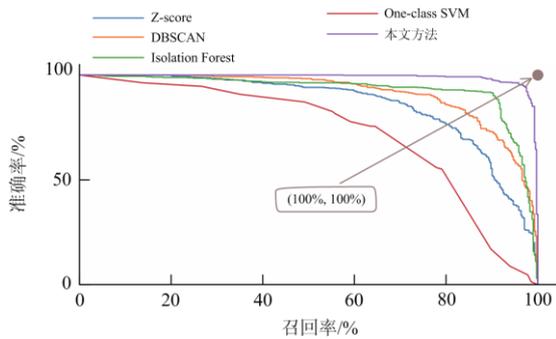


图 8 不同异常识别算法的“召回率-准确率”曲线

Fig. 8 The "recall-accuracy" curve of different anomaly recognition algorithms

由图 8 可以看出, 本文方法的“召回率-准确率”曲线更靠近(100%, 100%)的位置, 识别性能比其他几种方法更好。

利用第 4 节的方法确定每种方法参数的最优设置(参数的含义见文献[12,14,21]), 然后进行异常电量识别, 每种方法的参数设置、F-score 和异常识别时间如表 4 所示。

表 4 不同算法的异常识别性能和识别时间对比

Table 4 Comparison of anomaly recognition performance and recognition time of different algorithms

算法名称	参数的最优设置	F-score	异常识别时间/s
本文方法	$\lambda = 2.2$	0.87	1.2
Z-score	$\lambda = 2.9$	0.64	3.4
DBSCAN	$\text{eps} = 0.04, \text{min\_samples} = 20$	0.71	96
Isolation Forest	$\text{contamination} = 0.14$	0.79	54
One-class SVM	$\text{kernel} = \text{sigmoid}, \text{nu} = 0.03$	0.62	78

由表 4 可知, 本文方法终端电量异常识别性能(F-score)高于其他几种算法, 同时异常识别所需时间较短, 能够满足在线实时识别电量异常的要求。

## 6 结论

本文提出了基于电量概率预测的分阶段电量异常在线实时识别方法。基于实际的电量数据进行了验证, 实验结果表明:

(1) 该方法能够有效识别电量数据中的上下文异常和群体异常, 识别性能与其他异常识别算法相比更优。

(2) 将异常识别模型的训练和异常的识别分阶段进行, 电量异常识别的时间大幅度缩短, 满足实时识别的要求。

构建准确的用户用电量概率预测模型是实现准确电量异常识别的关键, 采用更多的深度学习技术对用户用电量进行建模, 提升概率预测性能以提高异常识别的效果是未来的研究方向之一。

## 参考文献

- [1] 胡江溢, 祝恩国, 杜新纲, 等. 用电信息采集系统应用现状及发展趋势[J]. 电力系统自动化, 2014, 38(2): 131-135. HU Jiangyi, ZHU Enguo, DU Xingang, et al. Application status and development trend of power consumption information collection system[J]. Automation of Electric Power Systems, 2014, 38(2): 131-135.
- [2] 康守亚, 李嘉龙, 李燕珊, 等. 考虑峰谷分时电价策略的源荷协调多目标发电调度模型[J]. 电力系统保护与控制, 2016, 44(11): 83-89. KANG Shouya, LI Jialong, LI Yanshan, et al. Multi-objective generation scheduling model of source and load considering the strategy of TOU price[J]. Power System Protection and Control, 2016, 44(11): 83-89.
- [3] 张宁, 代红才, 胡兆光, 等. 考虑系统灵活性约束与需求响应的源网荷协调规划模型[J]. 中国电力, 2019, 52(2): 61-69. ZHANG Ning, DAI Hongcai, HU Zhaoguang, et al. A source-grid-load coordinated planning model considering system flexibility constraints and demand response[J]. Electric Power, 2019, 52(2): 61-69.
- [4] 袁晓冬, 费骏韬, 胡波, 等. 资源聚合商模式下的分布式电源、储能与柔性负荷联合调度模型[J]. 电力系统保护与控制, 2019, 47(22): 17-26. YUAN Xiaodong, FEI Juntao, HU Bo, et al. Joint scheduling model of distributed generation, energy storage and flexible load under resource aggregator mode[J]. Power System Protection and Control, 2019, 47(22): 17-26.
- [5] 田力, 向敏. 基于密度聚类技术的电力系统用电量异常分析算法[J]. 电力系统自动化, 2017, 41(5): 64-70. TIAN Li, XIANG Min. Abnormal power consumption analysis based on density-based spatial clustering of applications with noise in power systems[J]. Automation of Electric Power Systems, 2017, 41(5): 64-70.
- [6] 陆春艳. 基于 GPU 的电表健康状态评估与预测[J].

- 电测与仪表, 2020, 57(11): 135-140.
- LU Chunyan. Assessment and prediction of the health status of electric energy meters based on GPU[J]. *Electrical Measurement & Instrumentation*, 2020, 57(11): 135-140.
- [7] 李洪乾, 韩松, 周忠强. 利用 Rayleigh 熵和并行计算的大规模电网异常负荷快速识别[J]. *电力系统保护与控制*, 2019, 47(23): 37-43.
- LI Hongqian, HAN Song, ZHOU Zhongqiang. Efficient abnormal load identification in large-scale power system employing Rayleigh quotient and parallel computing technology[J]. *Power System Protection and Control*, 2019, 47(23): 37-43.
- [8] 郑思达, 梁琪琳, 彭鑫霞, 等. 基于模糊聚类的异常用电行为识别研究[J]. *电测与仪表*, 2020, 57(19): 40-44.
- ZHENG Sida, LIANG Qilin, PENG Xinxia, et al. Fuzzy clustering based abnormal power consumption behavior detection on resident level[J]. *Electrical Measurement & Instrumentation*, 2020, 57(19): 40-44.
- [9] 周忠强, 韩松. 基于样本协方差矩阵最大特征值的低信噪比环境电网异常状态检测[J]. *电力系统保护与控制*, 2019, 47(8): 113-119.
- ZHOU Zhongqiang, HAN Song. MESCM based abnormal state detection of power system in low SNR environment[J]. *Power System Protection and Control*, 2019, 47(8): 113-119.
- [10] 张承智, 肖先勇, 郑子萱. 基于实值深度置信网络的窃电行为检测[J]. *电网技术*, 2019, 43(3): 1083-1091.
- ZHANG Chengzhi, XIAO Xianyong, ZHENG Zixuan. Electricity theft detection for customers in power utility based on real-valued deep belief network[J]. *Power System Technology*, 2019, 43(3): 1083-1091.
- [11] LIANG X, ZHAO B, MA Q, et al. Terminal access data anomaly detection based on random forest for power user electric energy data acquisition system[C] // *International Conference on Advanced Information Networking and Applications*, 2019, Cham: 166-175.
- [12] 黄福兴, 周广山, 丁宏, 等. 基于孤立森林算法的电能异常数据检测[J]. *华东师范大学学报: 自然科学版*, 2019(5): 123-132.
- HUANG Fuxing, ZHOU Guangshan, DING Hong, et al. Electric energy abnormal data detection based on isolation forests[J]. *Journal of East China Normal University: Natural Science*, 2019(5): 123-132.
- [13] 高金兰, 康迪, 雷星宇, 等. 基于改进模糊聚类分析的电力系统不良数据辨识[J]. *电气自动化*, 2018, 40(5): 30-33.
- GAO Jinlan, KANG Di, LEI Xingyu, et al. Identification of bad data in power system based on improved fuzzy clustering analysis[J]. *Electrical Automation*, 2018, 40(5): 30-33.
- [14] 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 Yong, ZHENG Kaihong, YU Zhongzhong, et al. Power data anomaly detection based on holt-winters model and DBSCAN clustering[J]. *Power System Technology*, 2020, 44(3): 1099-1104.
- [15] YANG Z, LIU H, BI T, et al. Bad data detection algorithm for PMU based on spectral clustering[J]. *Journal of Modern Power Systems and Clean Energy*, 2020, 8(3): 473-483.
- [16] 王方雨, 刘文颖, 陈鑫鑫, 等. 基于“秩和”近似相等特性的同期线损异常数据辨识方法[J]. *电工技术学报*, 2020, 35(22): 4771-4783.
- WANG Fangyu, LIU Wenying, CHEN Xinxin, et al. Abnormal data identification of synchronous line loss based on the approximate equality of rank sum[J]. *Transactions of China Electrotechnical Society*, 2020, 35(22): 4771-4783.
- [17] CHAHLA C, SNOUSSI H, MERGHEM L. A novel approach for anomaly detection in power consumption data[C] // *8th International Conference on Pattern Recognition Applications and Methods*, 2019, Prague, Czech Republic.
- [18] JIANG W, HONG Y, ZHOU B, et al. A GAN-based anomaly detection approach for imbalanced industrial time series[J]. *IEEE Access*, 2019, 7: 143608-143619.
- [19] 庞传军, 牟佳男, 余建明, 等. 基于运行关键指标和 Seq2Seq 的大电网运行异常识别[J]. *电力建设*, 2020, 41(7): 17-24.
- PANG Chuanjun, MOU Jianan, YU Jianming, et al. Identification of abnormal operation of large power grids according to key operating indicators and Seq2Seq[J]. *Electric Power Construction*, 2020, 41(7): 17-24.
- [20] ROUSSEEUW P J, HUBERT M. Anomaly detection by robust statistics[J]. *Data Mining and Knowledge Discovery*, 2018, 8(2): 1236-1250.
- [21] MA Bin, YUAN Long, XU Shaozhe, et al. Positive active power outlier detection based on one-class SVM[C] // *12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, September 20-23, 2020, Nanjing, China: 1-4.
- [22] BRODERSE N, GALLUSSE R, KOEHLE R, et al. Inferring causal impact using Bayesian structural time-series models[J]. *The Annals of Applied Statistics*, 2015, 9(1): 247-274.
- [23] BLEI D M, KUCUKELBIR A, MCAULIFFE J D. Variational inference: a review for statisticians[J]. *Journal of the American Statistical Association*, 2017, 112: 859-877.
- [24] 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 Qixin, ZHENG Kedi, KANG Chongqing, et al. Detection methods of abnormal electricity consumption behaviors: review and prospect[J]. *Automation of Electric Power Systems*, 2018, 42(17): 189-199.

收稿日期: 2020-12-10; 修回日期: 2021-02-24

作者简介:

刘 宣(1979—), 男, 硕士, 高级工程师, 研究方向为用电信息采集技术; E-mail: liuxuan@epri.sgcc.com.cn

唐 悦(1975—), 男, 本科, 高级工程师, 主要研究方向为用电信息采集技术; E-mail: tangyue@epri.sgcc.com.cn

卢继哲(1992—), 男, 通信作者, 硕士, 中级工程师, 主要研究方向为用电信息采集技术。E-mail: lujizhe@epri.sgcc.com.cn

(编辑 魏小丽)